Engineering nitrilases for enhanced thermostability

Lenye Sebenzile Dlamini

Thesis Presented for the Degree of

DOCTOR OF PHILOSOPHY in Medical Biochemistry

in the Department of Integrative Biomedical Sciences, Faculty of Health Sciences

UNIVERSITY OF CAPE TOWN

April 2024

Supervisors: Prof. BT Sewell

Dr. JD Woodward

Prof. E Sturrock

Declaration

I, Lenye Sebenzile Dlamini, hereby declare that the work on which this thesis is based is my original work (except where acknowledgements indicate otherwise) and that neither the whole work nor any part of it has been, is being, or is to be submitted for another degree in this or any other university. I authorise the University to reproduce for the purpose of research either the whole or any portion of the contents in any manner whatsoever.

Signature Date:21 April 2024

Acknowledgements

First, I'd like to extend my gratitude to the Organisation for Women in Science from Developing countries (OWSD), the L'Oreal for Women in Science Foundation (FWIS) and The International Union of Crystallography (IUCr) for their support, financial and otherwise. Their backing has been crucial in enabling me to pursue this degree. I am grateful to OWSD and FWIS for advocating for and supporting women in science. Their contribution to my life and career will last long after this degree is complete.

To Professor Bryan Trevor Sewell, I extend my heartfelt gratitude for your unwavering belief in my abilities and for your kindness, patience, and encouragement throughout this journey. I am also appreciative of Jeremy Woodward for generously sharing his expertise in the CryoEM aspects of this project. I would like to express my gratitude to Prof. Edward Sturrock for his genuine care for my well-being and for his encouragement. It is often the seemingly small gestures that make the greatest impact, and the meetings and random check-ins were truly meaningful to me.

I extend my sincere appreciation to all the staff at the Electron Microscope Unit at UCT, with special gratitude to Mr. Mohammed Jaffer. His expertise in microscopy has been instrumental in shaping my understanding and skill in the technique, and his consistent support as my "second pair of hands" during the CryoEM sample screens has been invaluable. I am grateful for his constant kindness, encouragement, and funny "dad jokes" which lightened even the hardest of days. A heartfelt thank you goes to Mr. Andrew "Andy" Rabagliati for always being "a text away", assisting with all my IT troubles and for tirelessly installing various programs,

most of which did not even make it into this thesis. I also want to acknowledge Mr. Sean Karriem for his consistent support in the lab, from checking up on me daily to assisting with repairs and shipments. His kindness and encouragement have meant a great deal to me, and I am truly grateful. To Michael Woodward, a special thank you for all the random and seemingly never-ending equipment repairs. To the entire staff at the Electron Microscope Unit, thank you for including me into your activities and making me feel at home among you.

I extend my thanks to the lab managers at the Centre for Bioprocess Engineering (CeBER), Tichaone Simkange, Sarah Fernandez, and Sharon Rademeyer, for allowing me to use their equipment and for never tiring of my constant presence in their lab. I also want to express my gratitude to the staff and students at the Membrane Protein Lab for their warm hospitality, especially James Birch, who taught me quite a lot during the visit. I am deeply thankful to Sue Huddy, Michael Bester, Taariq Firfirey, and all the staff in the research group at Roche Diagnostics for allowing me to use their Prometheus, their generous support, kindness, and hospitality during my time at their company. I'm very grateful.

To my colleagues who I've worked with at different stages—Stanley Makumire, Portia Makhubu, Phillip Venter, Zanele Makhado, Lizelle Lubbe, Melissa Marx, Jeremy Burgess, Kaylene Baron, Hendrina Shipanga and Veneshley Samuels—I am grateful for the camaraderie, shared laughter, and mutual support despite the challenges of their own lives and work. A special acknowledgment goes to Andani Mulelu for his assistance in reviewing and correcting portions of this thesis, as well as for the many random and insightful discussions about nitrilases.

I extend my deepest gratitude to my family for their unwavering support, encouragement, and love albeit from afar. To my mom, and I am profoundly grateful. Lastly, I offer my gratitude to God, whose presence has provided me with strength, resilience, and comfort throughout this journey.

Funding acknowledgements

The following organisations are acknowledged and thanked for funding this project:

The National Research Foundation of South Africa (NRF)

Synchrotron Techniques for African Research Technologies (START)

Instruct-ERIC

For my mom, Sauline Lomhlangano Matsenjwa-Dlamini

Abstract

Cyanide dihydratase from *Bacillus pumilus* C1 ($CynD_{pum}$) catalyses the hydrolysis of cyanide to formic acid and ammonia. $CynD_{pum}$ has the potential to remediate cyanide-containing wastewater. However, two obstacles hinder the commercial use of recombinantly expressed $CynD_{pum}$ as a biocatalyst: reduced activity at pH>8 which is typical of cyanide-rich environments, and inactivation at temperatures above 42 °C. Several variants with either enhanced thermostability or activity at alkaline pH have been discovered by random mutagenesis and directed evolution; however, these methods are slow and limited in their explorable sequence space. The aim of this project is to investigate the structural and chemical determinants of thermostability in cyanide dihydratases. This was achieved through rational site-saturation mutagenesis, followed by experimental validation using nanoscale differential fluorimetry, negative stain and cryogenic electron microscopy, and single-particle analysis.

In this study, atomic resolution structures of wild-type $CynD_{pum}$ and its variant $CynD_{pum}$ (Q86R/H305K/H308K/H323K) were employed to predict, using empirical force fields, the change in Gibb's free energy resulting from mutating each amino acid in the atomic resolution structure of the variant (amino acids 3-319) to each of the remaining 19 residues. Favourable Gibb's free energy changes were used as indicators of the thermostability of $CynD_{pum}$ variants

and validated experimentally using chemical denaturation monitored by nanoscale differential scanning fluorimetry. The thermodynamic favourability of six variants (S29A, T217I, T260I, Q86M, N119R, and E155R) was successfully validated. Negative stain electron microscope micrographs of these variants revealed that these variants formed helical fibres with increased length relative to wild-type $CynD_{pum}$, indicating a positive correlation between a favourable change in Gibb's free energy of protein unfolding and fibre length.

The observed favourable Gibb's free energy change of the surface variants (S29A, T217I, and T260I) is attributed to the reduced number of solvent-accessible amino acid sidechains, promoting protein folding and oligomerisation. Variants N119R and E155R, located along the groove of the helical assembly, form electrostatic interactions across the grooves, thereby enhancing the structural rigidity of the quaternary structure and leading to a favourable Gibb's free energy change. These interactions were visualised in the 2.78 Å resolution atomic model of the E155R variant of $CynD_{pum}$ and the 3.26 Å resolution structure of $CynD_{stu}$ solved by cryo-electron microscopy. Analyses of the conventional interfaces (named A-, C-, D-, and E-) through which the $CynD_{pum}$ helix and cyanide dihydratase from *Pseudomonas stutzeri* AK61 ($CynD_{stu}$) spiral are formed, revealed that the helical and spiral assemblies are predominantly stabilised through hydrophobic and electrostatic interactions occurring at the A and C interfaces, respectively. Additionally, examination of the C-terminal tail regions of the

atomic-resolution structures of wild-type $CynD_{pum}$, its variants (Q86R/H305K/ H308K/H323K, and E155R), and $CynD_{stu}$ revealed that the C-terminal tail stabilises all interfacial regions through specific interactions, demonstrating its critical structural and functional role in assembly formation and thermostability.

Abbreviations

Å	Angstrom (s)
℃	Degrees Celsius
2D	Two-dimensional
3D	Three-dimensional
ΔG	Change in Gibb's free energy of protein unfolding
ΔΔG	Difference in the free energy of protein unfolding between a native protein and its variant
µL	Microlitre
µm	Micron
λ_{ex}	Excitation wavelength
λ_{ex}	Tryptophan fluorescence emission wavelength
ASP	Ammonium sulphate precipitation
C50	Concentration of denaturant at which half the protein in solution is folded and the other half unfolded
CTF	Contrast transfer function
CynD	Cyanide dihydratase
CynD 8A3	Cyanide dihydratase from strain 8A3
$CynD_{pum}$	Cyanide dihydratase from *Bacillus pumilus* C1
$CynD_{stu}$	Cyanide dihydratase from *Pseudomonas stutzeri* AK61
CHT	Cyanide hydratase
Cryo-EM	Cryogenic-electron microscopy
CTF	Contrast transfer function
CV	Column volume
Da	Dalton
DMSO	Dimethylsulfoxide
DNA	Deoxyribonucleic acid
DSF	Differential scanning fluorimetry
e^-	Electron
EM	Electron microscopy

eV	Electron volt
FI	Fluorescence intensity
FSC	Fourier-shell correlation
Fwd	Forward primer
G	Gibb's free energy
HCN	Hydrogen cyanide
His	Histidine
IEX	Ion-exchange chromatography
IMAC	Immobilised metal affinity chromatography
IPTG	Isopropyl β-D-1 thiogalactopyranoside
J	Joule
kcal	Kilocalories
KCN	Potassium cyanide
K_{eq}	Equilibrium constant of protein unfolding
kJ	Kilojoule
kDa	Kilodalton
L	Litres
LB	Luria-Bertani broth
M	Molar (mole/litre)
mAu	Milli Absorbance units
mA	Milli Amperes
mL	Millilitres
mL/min	Millilitres per minute
min	Minute
mM	Millimolar
MPL	Membrane Protein Lab
MW	Molecular weight
MWCO	Molecular weight cut-off

nm	Nanometre
OD_{600}	Optical density at a wavelength of 600 nm
PCR	Polymerase chain reaction
px	Pixel (s)
Rev primer	Reverse primer
s/sec	Second (s)
SDS-PAGE	Sodium-dodecyl sulphate polyacrylamide gel electrophoresis
SEC	Size exclusion chromatography
SEC-MALS	Size exclusion chromatography multi-angle light scattering
SSM	Site-saturation mutagenesis
SPA	Single-particle analysis
T_a	Annealing temperature
TEM	Transmission electron microscope/ microscopy
TEMED	N, N, N', N'-tetramethyl ethylenediamine
T_m	Melting temperature
Tris	Tris (hydroxymethyl) aminomethane
U	Units
UA	Uranyl acetate
UV	Ultra-violet
v/v	Volume per volume
w/v	Weight per volume
xg	Multiple of gravitational force

List of figures

List of tables

Table of contents

Chapter I

Literature review

1.1 The nitrilase superfamily

Enzymes of the nitrilase superfamily are thiol enzymes that catalyse the hydrolysis of nonpeptide nitrile bonds and amides using nonpeptide nitriles as substrates. Enzymes and proteins of the superfamily are diverse and play a critical role in various biological and industrial processes. They were first identified as the enzymes catalysing the synthesis of the plant growth hormone indole acetic acid (Thimann and Mahadevan 1964). Nitrilases occur widely in nature and are found in animals, fungi, plants, archaea, and bacteria, where they are involved in product synthesis, detoxification, and post-translational modification (Thimann and Mahadevan 1964, Pace and Brenner 2001, Gong, Lu et al. 2012).

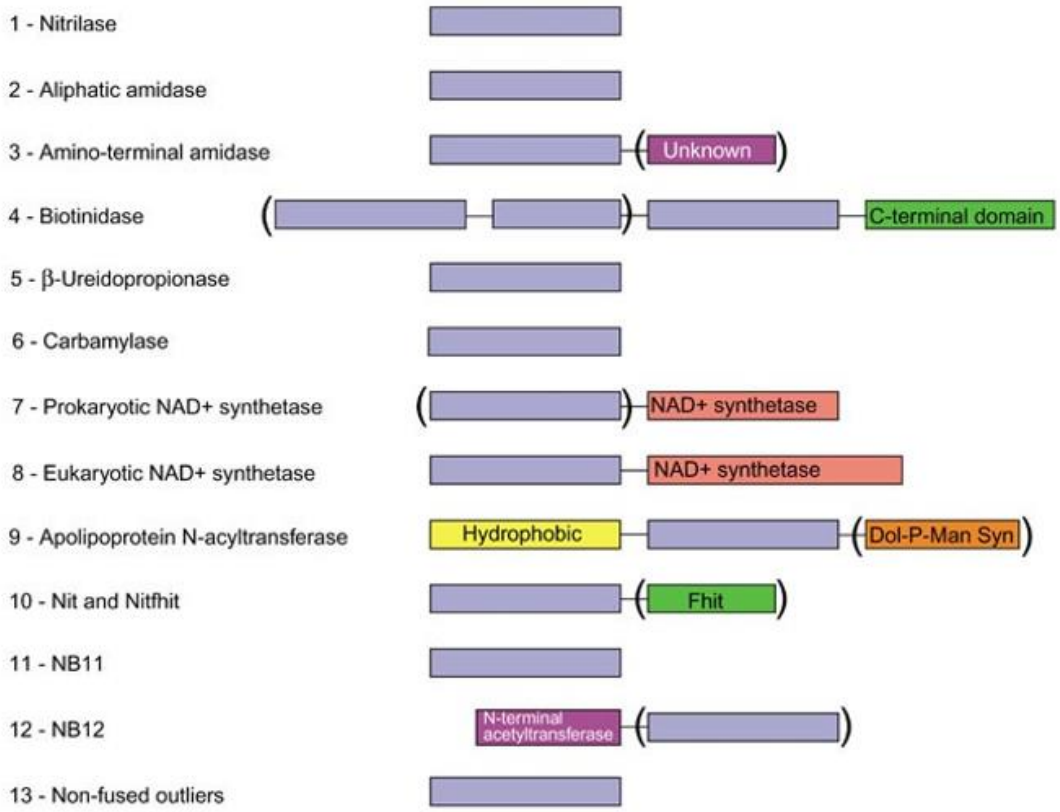
Interest in enzymes of the nitrilase superfamily has emerged and persists because of their potential applications as industrial biocatalysts in product synthesis owing to their regio-, enantio-, and stereoselectivity in catalysing nonpeptide nitrile bonds (Kiziak, Conradt et al. 2005, Zhang, Yin et al. 2014). Furthermore some members of the nitrilase superfamily such as glutamine synthetase and apolipoproteins from infectious organisms such as *Mycobacterium tuberculosis* are studied for their potential role as drug targets (Pace and Brenner 2001, Legood, Boneca et al. 2021). Thus, in-depth study into their natural reaction mechanisms, substrates, biochemical pathways are a constant area of investigation

1.2 Classification of the nitrilase superfamily

The nitrilase superfamily is currently classified into 13 branches, distinguished by global sequence, structure-based similarity, and additional domains (Figure1.1) (Pace and Brenner 2001). Despite the shared structural homology and conserved catalytic amino acids within the superfamily, members of this superfamily catalyse the conversion of a broad range of substrates. Pace and Brenner (2001) described four primary reactions catalysed by superfamily members: nitrilases, amidase, carbamylase, and the reverse amidase reaction illustrated in Figure1.1 (b) (Pace and Brenner 2001). Among these, amidase reactions are the most observed and extensively studied reactions.

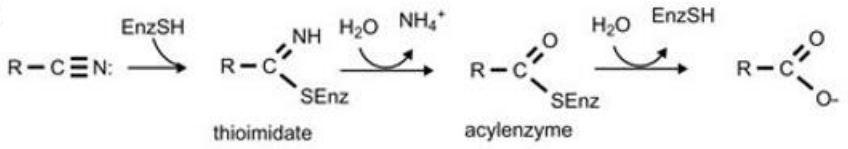
Notably, among the 13 branches described by Pace and Brenner (2001), only the first branch comprises enzymes with authentic nitrilase activity, facilitating the conversion of nitrile bonds (Brenner 2002)

(a)

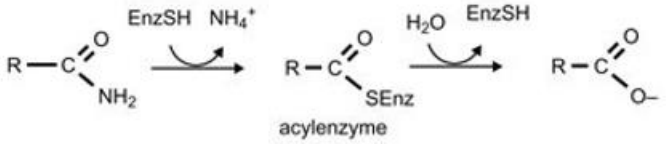

1 - Nitrilase
2 - Aliphatic amidase
3 - Amino-terminal amidase
Unknown
4 - Biotinidase
C-terminal domain
5 - β-Ureidopropionase
6 - Carbamylase
7 - Prokaryotic NAD+ synthetase
NAD+ synthetase
8 - Eukaryotic NAD+ synthetase
NAD+ synthetase
9 - Apolipoprotein N-acyltransferase
Hydrophobic
Dol-P-Man Syn
10 - Nit and Nitfhit
Fhit
11 - NB11
12 - NB12
N-terminal acetyltransferase
13 - Non-fused outliers

(b)

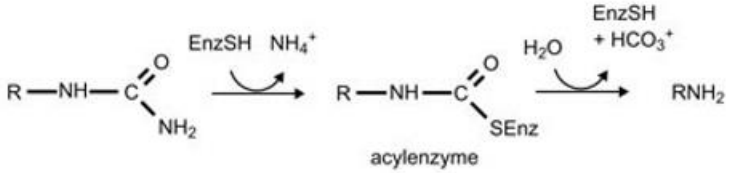
i


EnzSH
NH
SEnz
thioimidate
H_2O
NH_4^+
acylenzyme
H_2O
EnzSH

ii


EnzSH
NH_4^+
NH_2
SEnz
acylenzyme
H_2O
EnzSH

iii


EnzSH
NH_4^+
NH_2
SEnz
acylenzyme
H_2O
EnzSH
+ HCO_3^+
RNH_2

iv

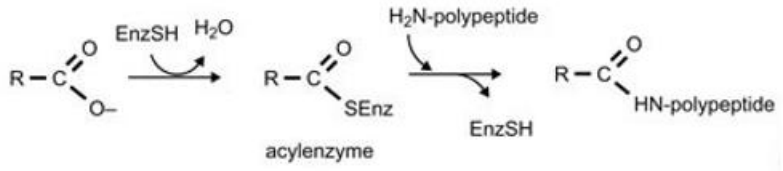

EnzSH
H_2O
SEnz
acylenzyme
H_2N-polypeptide
EnzSH
HN-polypeptide

Figure 1.1: Classification of the nitrilase superfamily. (a) The nitrilase superfamily is categorised into 13 branches based on global sequence similarity and conserved domains. The conserved the conserved superfamily domains are coloured in blue; additional domains are coloured differently for each branch. The first branch of the superfamily, wherein the cyanide-degrading enzymes cyanide dihydratases and cyanide hydratases belong, is broadly further categorised based on substrate specificity. (b) Enzymes of the nitrilase superfamily catalyse four main reactions, all of which are postulated to proceed through a tetrahedral acyl-enzyme intermediate. (I) Hydrolysis of nitrile functional groups is mediated by enzymes belonging to the first branch of the superfamily. (II) Amidase activity is the most frequently observed reaction among enzymes of the nitrilase superfamily. Members of branches 2-4 catalyse the conversion of different types of amide substrates. (III) Enzymes responsible for the addition of carbamyl functional groups to molecules and other proteins are categorised into branches 5 and 6. (IV) The reverse amidase reaction is carried out by branch-9 N-acyltransferases, which involve the addition of an acyl functional group to the N-terminus of polypeptides (Pace and Brenner 2001).

1.3 Applications of nitrilase superfamily enzymes

Nitrilases have potential use as industrial biocatalysts and are used in various industrial, pharmaceutical, and environmental applications.

1.3.1 Commercial chemical synthesis

Nitrilases are enantio-, regio-, and stereoselective, catalysing the conversion and formation of specific enantiomers (Gong, Lu et al. 2012). This property is particularly advantageous in drug development, precursor synthesis, and fine chemical production, where the properties of one enantiomer may be more desirable over the other, unlike conventional chemical methods which often result in the formation of racemic mixtures and lower yields of the desired enantiomer. Nitrilases have been employed and are being developed to produce various high-value chemicals and therapeutic drugs; some examples are described here.

The R-stereoisomer of mandelic acid is an important intermediate in the synthesis of important pharmaceutical products such as semi-synthetic penicillin and cephalosporins (Tang, Yi et al. 2009, Zhang, Wang et al. 2020), and is used to prevent urinary tract infections (van Putten 1979, Kumar 2017). The nitrilase-mediated conversion of the racemic mixture (R, S) mandelonitrile using the nitrilase from the WT10 strain of *Alcaligenes faecalis* results in the formation of the pure *R*-stereoisomer of mandelic acid, making it a potential candidate for the synthesis of mandelic acid (Xue, Xu et al. 2011).

Nitrilases have also been employed in the commercial production of various acids, such as nicotinic acid (vitamin B3), an important additive in the food industry, from 3-cyanopyridine instead of nicotinic acid, using a nitrilase from *Rhodococcus* sp. NDB 1165 (Almatawah and Cowan 1999, Prasad, Misra et al. 2007, Sharma, Sharma et al. 2011). Additionally, glycolic acid, which is widely used in the skincare industry, is synthesised from glycolonitrile using a nitrilase enzyme, whereas acrylic acid synthesised from acrylonitrile is used in the synthesis of various polymers, including adhesives, paints, polishes, and textiles (Nagasawa, Wieser et al. 2000, Gong, Lu et al. 2012, Sun, Yamada et al. 2017).

Nitrilases have proven to be effective in the synthesis of intermediates crucial to the production of important therapeutic drugs. For example, the nitrilase enzyme from *Acidovorax facilis* ZJB09122 has been used to catalyse the formation of an acid intermediate

in the synthesis of gabapectin, a structural analogue of an inhibitory neurotransmitter used for treating neuropathic pain and epilepsy (Xue, Wang et al. 2015). Similarly, atorvastatin, a medication which decreases low-density lipoproteins thus preventing coronary heart disease, involves the contribution of nitrilases. The nitrilase enzymes catalyses the hydrolysis of 3-hydroxyglutaronitrile to the corresponding acid, (R)-4-cyano-3-hydroxybutyric acid. This acid is subsequently esterified and serves as a crucial intermediate in the synthesis of atorvastatin (Patel 2009).

1.3.2 Waste management and bioremediation

Compounds containing the nitrile functional group occur naturally and are one of the most commonly synthesised functional groups in organic chemical synthesis (Hinzmann, Yavuzer et al. 2023). Nitrile-containing compounds are found in nature as cyanogenic glycosides and lipids, ricinine, and phenyacetonitriles where they are involved in various processes, including plant defence mechanisms and natural product biosynthesis (Gong, Lu et al. 2012). Synthetic nitriles are used in the chemical, pharmaceutical, and agricultural industries. For example, the production of herbicides and their intermediates, such as bromoxynil (Cutulle, Armel et al. 2014) and iminodiacetic acid, are used in the manufacture of chelators, herbicides, and surfactants (Liu, Lu et al. 2018), respectively. Nitriles are also used in the textile,

petrochemical, coal coking, and mining industries for the production of clothing dyes, nitrile-based rubber, and nylon (Gong, Lu et al. 2012, Peter, Singh et al. 2023).

Despite their usefulness, nitriles are known for their carcinogenicity, mutagenicity, and toxicity, with hydrogen cyanide as a notable example (Gong, Lu et al. 2012, Peter, Singh et al. 2023). Therefore, nitrile waste from these industries poses a health risk to workers in those industries and communities of people living near those areas. Nitrile waste also contributes significantly to environmental pollution, damaging ecosystems, and contaminating freshwater sources. Although various chemical processes are employed to recycle and mitigate the toxicity of nitrile wastes, residual amounts often persist. These chemical methods often convert some nitriles into less toxic forms using conventional chemical methods which require large inputs of energy and resources and often result in the production of additional undesirable compounds as reaction by-products.

Nitrilases present an opportunity to convert toxic nitrile waste compounds used in these economically important industries into useful products of economic value, without the production of undesirable reaction by-products, owing to their enantioselectivity and ability to synthesise target compounds without the formation of undesirable by-products. Moreover, nitrilases present an environmentally benign alternative to conventional chemical degradation processes, posing minimal risk to the environment (Gong, Shi et al. 2017). They are nontoxic

and nonlethal to both animals and humans. In addition, nitrilases are easily inactivated and are completely degradable. Furthermore, studies have demonstrated retained nitrilase activity in the presence of some common co-contaminants of cyanide wastewater such as silver or copper salts, mercury and nickel (Vargas-Serna, Carmona-Orozco et al. 2020, Sedova, Rucká et al. 2021, Martínková, Kulik et al. 2023). This suggests that they are not inactivated by most of the co-contaminants likely to be found in cyanide wastewater and thus potentially do not require extensive pre-treatment of the wastewater before being used as biocatalysts.

However, nitrilases being protein in nature are susceptible to denaturation and inactivation under industrial conditions. They may not naturally catalyse the reactions for which they are desirable or have turnover rates lower than the demands of industrial production, necessitating modification. Numerous nitrilases have been modified to enhance key attributes such as substrate specificity (Mulelu, Kirykowicz et al. 2019), enantioselectivity (Patel 2009) and catalytic efficiency (Crum, Sewell et al. 2016, Yu, Yao et al. 2019) and stability (Crum, Park et al. 2015).

1.4 Nitrilase structure and function

1.4.1 Oligomerisation

Early crystal structures of members of the superfamily revealed multimeric protein structures with a conserved αββα sandwich monomer fold. Generally, monomers self-associate, forming inactive dimers (αββα- αββα) (Park, Mulelu et al. 2016). During protein assembly, the dimers become repeating units along the length of larger, active homo-oligomeric helices and spirals of variable length (Figure 1.2) (Thuku, Brady et al. 2009). This structural characteristic is exhibited by the microbial nitrilases cyanide dihydratases (CynD) from *Pseudomonas stutzeri* AK61 ($CynD_{stu}$) (recently reclassified to *Stutzerimonas stutzeri*), *Bacillus pumilus* C1 ($CynD_{pum}$), *and Bacillus pumilus* 8A3 which exist as a spiral and helix of 14, 18, and 22 subunits, respectively, at optimum pH (Figure 1.2). (Jandhyala, Berman et al. 2003, Sewell, Berman et al. 2003, Eicher 2007). Helical assemblies are characterised by subunits arranged along a central axis with a constant radius along the height of the helix, whereas in spiral assemblies, subunits wind around a central point with varying radii. Nitrilases of fungal and eukaryotic origin also adopt similar quaternary structure (Lundgren, Lohkamp et al. 2008, Woodward, Weber et al. 2008, Dent, Weber et al. 2009).

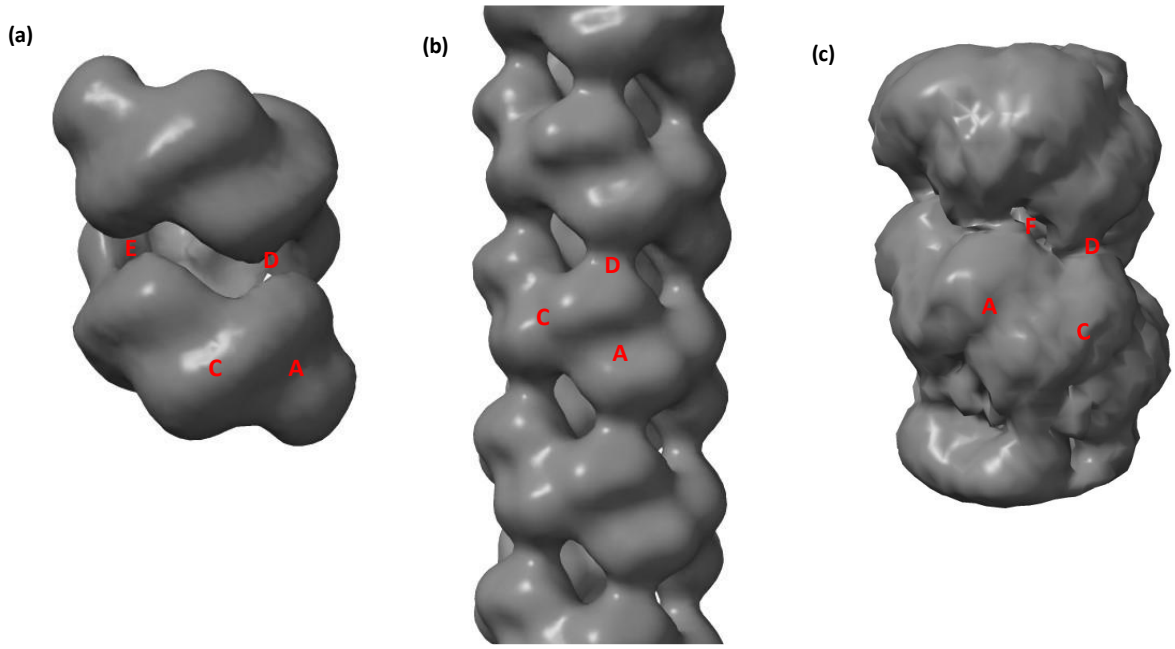


Figure 1.2: Nitrilases occur as spirals (a) and helices (b and c) of various lengths that are stabilised through interfacial regions A, C, D, E, and F. (a) Cyanide dihydratase from *Pseudomonas stutzeri* AK61 is a 14mer self-terminating spiral with C2 symmetry (EMD-1050), (b) The nitrilase from *Rhodococcus rhodochrous* J1 (EMD-1313), and (c) Cyanide hydratase from *Neurospora crassa* (Makumire *et al* unpublished).

The biochemical and biophysical basis of oligomerisation is an area of active study. Oligomerisation appears to be influenced by a variety of chemical and physical changes, including the presence of specific substrates, exposure to elevated temperatures (Nagasawa, Wieser et al. 2000), specific solvent conditions such as changes in pH (Jandhyala, Willson et al. 2005), and post-translational modifications (Thuku, Weber et al. 2007). Furthermore, these factors seem to be protein-specific; certain solvent conditions may lead to the formation of specific quaternary structures in one protein but not in its homologue. For instance, the CynD enzyme from *Bacillus pumilus* C1 forms long fibres of indeterminate length at pH 5.4, whereas

its homologue from strain 8A3 (CynD 8A3) does not exhibit this behaviour (Eicher 2007). However, alignment of the amino acid sequences of nitrilases capable of forming larger homo-oligomers revealed two unique insertion regions conserved in their amino acid sequences and an extended C-terminal region, which are attributed to their ability to polymerise (Thuku, Brady et al. 2009). In certain instances, the substitution of amino acids in these regions for cysteine in $CynD_{pum}$ has been shown to modify the oligomeric structure and, in some instances, decrease enzymatic activity (Park, Mulelu et al. 2016).

The quaternary structure of nitrilases is directly linked to their substrate specificity and activity. Even slight structural changes can lead to altered substrate specificity, making this an important area of study for reengineering of nitrilases for specific uses. For instance, plant nitrilase 4 from *Arabidopsis thaliana* undergoes changes in the angle of rotation along the helical axis, resulting in a reduction in the size of the substrate-binding site, thus allowing for the catalysis of smaller substrates that are not typical of plant nitrilases (Mulelu, Kirykowicz et al. 2019).

1.4.2 Interfacial regions

The interfaces that facilitate interactions between individual monomers and subunits to form dimers and multimers within the nitrilase superfamily are described in Sewell, Thuku et al. (2005) and illustrated and shown in Figures 1.2 and 1.3, respectively. These interfaces are

commonly observed in both spiral and helical nitrilases. Extensive directed evolution and site-directed mutagenesis studies have increased our understanding of the intermolecular interactions occurring between monomers and subunit dimers to maintain the quaternary structure of nitrilases. Furthermore, as more atomic-resolution structures of nitrilases are solved, the chemical interactions maintaining these interfaces are better understood, thus increasing our understanding of the fundamental chemical and structural principles that lead to substrate specificity, oligomerisation, and tolerance to unfavourable environmental conditions. These interfaces have been termed A, C, D, E, and F interfaces (Figure 1.2).

Individual monomers interact across interfaces A and C. The A-interface serves as the dimerisation interface, where two monomers interact to form a dimer, which subsequently becomes a repeating subunit along the length of the spiral or helix, (Figure 1.2 (a)). This interface, present in all known structures of the nitrilase superfamily, is vital for maintaining enzyme activity (Sewell, Thuku et al. 2005).

The C-interface or oligomerisation interface is essential for oligomerisation and occurs between pairs of dimers at approximate right angles to the A-interface of each dimer (Figure 1.2 (a)). Mutations at this interface alter the oligomeric form of the enzyme leading to different-sized homo-oligomers and in some cases loss of activity due to changes in quaternary structure (Sewell, Thuku et al. 2005, Park, Mulelu et al. 2016, Mulelu 2017). This interface,

closely related to the A-interface directly influences substrate specificity. For example, the angle of rotation about the helical axis which is determined by specific chemical interactions, dictates the 3-dimensional arrangement of catalytic amino acids in space, thus altering the substrate binding site and substrate specificity of the enzyme (Woodward, Trompetter et al. 2018, Mulelu, Kirykowicz et al. 2019). This explains the observed loss or reduced wildtype activity, without loss of global quaternary structure, when certain amino acids on this interface are substituted, as observed with the cyanide dihydratase from *Bacillus pumilus* C1 (Park, Mulelu et al. 2016). This discovery is significant for the use of nitrilases as industrial biocatalysts. An in-depth understanding of amino acid interactions across the interfaces can enable the engineering of nitrilases capable of catalysing various nitrile substrates. This has implications for altering substrate specificities, ultimately allowing the structure-based engineering of custom enzymes tailored for the conversion of any desired substrate.

The D-interface occurs across the major groove when the spiral or helix completes a turn (Figure 1.2 (a) and 1.3). According to studies based on homology models, this interface results from strong electrostatic interactions between helices α1 and α3 of each subunit. Notably, these helices have significant sequence variability between nitrilases (Sewell, Thuku et al. 2005, Thuku, Brady et al. 2009). The E-interface is present only in self-terminating spirals and helices, as observed in the cyanide dihydratase from *Pseudomonas stutzeri* AK61 which forms

discrete 14-subunit spirals. Interactions in this region are asymmetric, arising from electrostatic interactions from non-conserved amino acids. These interactions are thought to contribute to an angular tilt towards the core of the spiral, reducing the space required for the addition of another subunit, thus leading to termination of the spiral (Sewell, Berman et al. 2003, Thuku, Brady et al. 2009). The F-interface, which is less common among nitrilases, also occurs across the groove, similarly to the D-interface (Figure 1.2 (c)). This interface is thought to provide additional stabilisation, with nitrilases possessing it showing increased helical twist (Nagasawa, Wieser et al. 2000, Thuku, Weber et al. 2007, Thuku, Brady et al. 2009), which, as mentioned previously, has implications for substrate specificity. The C-terminal tail region (Figure 1.2 (b)) is formed by four β-strands, forming an antiparallel β-sheet along the central core of the enzyme (Mulelu, Kirykowicz et al. 2019).

1.4.3 Nitrile catabolism and the nitrilase reaction mechanism

Nitrile bioconversion proceeds through two distinct reactions. The first occurs through the direct conversion of a nitrile to ammonia and the corresponding carboxylic acid (Figure 1.3 (a)), catalysed by a nitrilase enzyme. The second reaction is catalysed by nitrile hydratases (EC 4.2.1.84), metalloenzymes which convert nitriles to amides. An amidase enzyme (EC 3.5.1.4) is required to complete the conversion, forming ammonia and the corresponding carboxylic acid (Figure 1.3 (b)). Although nitrile hydratases convert chemically similar substrates to those

of nitrilases, they are structurally and mechanistically unrelated to nitrilases and are therefore not currently categorised as part of the nitrilase superfamily (Pace and Brenner 2001, Brenner 2002).

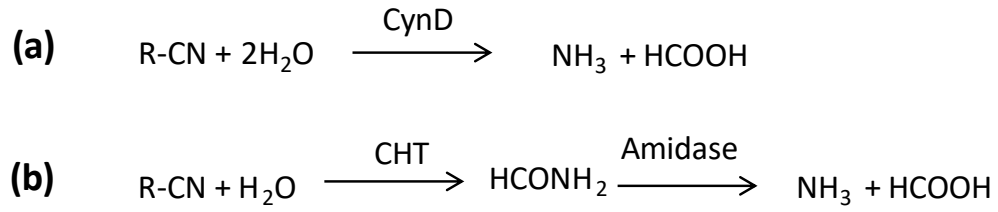


Figure 1.3: Cyanide catabolism in the nitrilase superfamily. (a) The CynD-catalysed reaction is a single step hydrolysis reaction utilising two water molecules, directly converting cyanide-containing compounds, such as HCN, into ammonia and the corresponding acid. (b) The CHT-catalysed reaction, however, proceeds through an amide intermediate which is converted to the final reaction products by an amidase.

Nitrilase enzymes catalyse the complete and irreversible hydrolysis of nitrile substrates into ammonia and carboxylic acids. Multiple studies have demonstrated that catalysis in the nitrilase superfamily is typically mediated by a conserved tetrad of amino acids (CEEK) (Pace and Brenner 2001, Kimani, Agarkar et al. 2007, Weber, Kimani et al. 2013).

Despite the diversity of this superfamily, with various branches and sub-branches each exhibiting different substrate specificities and reaction mechanisms, these amino acids are consistently conserved. Generally, nitrilases attack the amide carbon of nitrile substrates and proceed through the formation of two potential tetrahedral intermediates. The hydrolysis reaction involves the addition of two water molecules, resulting in the formation of ammonia

and a carboxylic acid. The cysteine and one of the glutamic acid residues were first identified by multiple sequence alignment of nitrile hydrolases (Bork and Koonin 1994).

The catalytic cysteine was the first to be unequivocally shown to play a critical role in catalysis, with its substitution resulting in inactivity (Kobayashi, Komeda et al. 1992, Grifantini, Pratesi et al. 1996). The catalytic triad was first observed at atomic resolution in the crystal 2.8 Å structure of NitFhit, a tumour suppressor protein that occurs as a fusion protein in some invertebrates, and N-carbamyl-D-amino acid amidohydrolase from *Agrobacterium* (Nakai, Hasegawa et al. 2000, Pace, Hodawadekar et al. 2000). Mutagenesis studies involving the substitution of glutamic acid and lysine in *Pseudomonas aeruginosa* amidase revealed their catalytic essentiality, in addition to the cysteine (Novo, Farnaud et al. 2002), with subsequent studies and atomic resolution structures of members of the superfamily confirming their essentiality to catalysis in the superfamily and providing insights into the potential mechanism by which bioconversion mediated by these enzymes occurs.

The nitrilase reaction mechanism is currently an area of active research. It is proposed to proceed through a four-step process. Initially, the side-chain carboxyl group of glutamic acid acts as a general base catalyst, removing a proton from the thiol group of the cysteine, thereby activating it. Subsequently, the active nucleophilic cysteine attacks the amide carbon of the nitrile substrate, forming a covalent tetrahedral intermediate which is stabilised by the side-

chain amino group of a lysine residue, which could potentially be a thioimidate (Stevenson, Feng et al. 1992, Nakai, Hasegawa et al. 2000, Fernandes, Mateo et al. 2006). The addition of a water molecule leads to the release of ammonia and a thioester intermediate. Following this, glutamic acid acts as a general base catalyst again, abstracting a proton from a water molecule, thus activating it. The activated water molecule attacks the thioester, releasing the corresponding carboxylic acid and recycling the catalytic amino acids (Stevenson, Feng et al. 1992, Nakai, Hasegawa et al. 2000, Brenner 2002, Jandhyala, Willson et al. 2005, Raczynska, Vorgias et al. 2011). The second glutamic acid residue was later identified in the *Geobacillus pallidus* amidase crystal structure, and its catalytic essentiality revealed through mutagenesis studies. The second glutamic acid, discovered last, plays an indispensable role in correctly positioning the substrate, enabling the initial nucleophilic attack by the cysteine (Kimani, Agarkar et al. 2007, Weber, Kimani et al. 2013).

1.5 Protein stability

Proteins are crucial to almost all processes that are essential to life, where they are involved in multiple processes including catalysis, transport, cellular signalling, immune response, and defence mechanisms (Lodish 2008). As biocatalysts, they offer environmentally benign, renewable, less toxic, and highly selective single-molecule catalysts, serving as alternatives to

conventional chemical catalysts (Sheldon and Woodley 2018, Bell, Finnigan et al. 2021). Consequently, they are sought after for use in various industrial processes and in the synthesis of pharmaceuticals and biologics. However, their utility in any industry is determined mainly by their ability to function as required under conditions that may inactivate and denature them or lead to suboptimal performance.

Protein stability is crucial for proper and efficient functioning of proteins. The correct 3-dimensional (3D) structure ensures the correct conformation of ligand-binding sites on enzymes and proteins, enabling the binding of specific molecules. In addition, changes in ligand-binding sites may alter ligand specificity, thereby changing the expected or desired function of the protein. Therefore, one of the most critical requirements for protein function is the maintenance of a stable 3D structure. Protein stability refers to the retention of a protein's 3D fold under specific physical and chemical conditions. It is characterised by a net balance of intra- and intermolecular forces that contribute to the formation and maintenance of the 3D structure (Gromiha 2010, Goldenzweig and Fleishman 2018, Kazlauskas 2018). These forces typically include non-covalent bonds such as hydrogen bonds, salt bridges, disulphide bridges, hydrophobic and van der Waals interactions (Ahmad 2022).

Various types of protein stability and the forces that maintain them have been extensively described in the contexts in which they apply (Lai and Topp 1999, Chi, Krishnan et al. 2003,

Costanzo, O'Brien et al. 2014, Randolph, Schiltz et al. 2015, Deller, Kong et al. 2016, Manning, Liu et al. 2018, Ahmad 2022). However, relevant to this study is thermostability. Thermostability refers to resistance to chemical and physical changes in intra- and intermolecular interactions that maintain the integrity of a protein's 3D fold at non-optimal temperatures (Kumwenda, Litthauer et al. 2013). Thermostability is an essential property of an efficient biocatalyst, allowing for increased operational stability and enabling enzymes to tolerate processes and reaction conditions that would normally inactivate or denature them. The benefits of using thermostable biocatalysts include increased substrate solubility, increased reaction rates, and consequently, higher conversion rates, leading to improved efficiency and productivity in industrial processes, thus ensuring that the biocatalyst is economically viable and decreasing potential contamination from biological agents (Turner, Mamo et al. 2007). However, most enzymes, including nitrilases, are mesophilic and therefore optimally active at ambient temperatures. This has limitations for turnover rates, process efficiency and possible storage conditions (Nigam, Arfi et al. 2017). Multiple studies on organisms with nitrilase activity and isolated thermophilic nitrilases have been conducted to understand and exploit the properties leading to their increased thermostability for the synthesis of commercially valuable chemicals (Almatawah and Cowan 1999, Almatawah, Cramp et al. 1999). Others have sought to engineer properties that improve thermostability

in mesophilic nitrilases by introducing changes to their amino acid sequences and thus protein structure and physical modifications (Xu, Cai et al. 2018, Khatik, Jain et al. 2022, Xiong, Lv et al. 2022, Lu, Tang et al. 2023).

Properties conferring thermostability are introduced using various protein engineering methods. The choice of technique often depends on the available information about the protein's primary, tertiary, and quaternary structures. Common strategies include targeted site-directed mutagenesis, directed evolution and consensus sequence analysis (Kazlauskas and Bornscheuer 2009). These approaches involve substituting or deleting specific amino acids and validating their effects. Additionally, structural data or homology models can guide chemical modifications such as introducing salt bridges and disulfide linkages. In some cases, the protein's hydrophobic core is reinforced by increasing its hydrophobicity, which reduces the degrees of conformational freedom and increases structural rigidity, leading to enhanced thermostability (Kazlauskas 2018). Another approach involves promoting interactions between protein domains to encourage oligomerisation or the formation of protein-protein complexes (Álvarez-Cao, González et al. 2018). Lastly, *de novo* protein design where proteins are created with specific structural and/ or functional properties based a fundamental understanding of protein chemistry (Watson, Juergens et al. 2023) and rational protein engineering methods, utilising thermodynamic simulations and computations, are

increasingly used to create proteins with desired structural and functional properties for improved thermostability.

Most nitrilases are naturally mesophilic (Gong, Lu et al. 2012). The few species that have been identified as thermostable have been isolated from thermophilic organisms (Almatawah and Cowan 1999). The chemical and structural basis of the observed thermostability is poorly understood, and the observed thermostability varies widely among these enzymes (Table 1). However, several studies have been conducted to increase the thermostability of mesophilic nitrilases (Wang, Watermeyer et al. 2012, Crum, Park et al. 2015, Xu, Cai et al. 2018, Xiong, Lv et al. 2022, Lu, Tang et al. 2023) using various techniques such as directed evolution and consensus sequence analysis of thermostable nitrilases. From these studies, mutations at the D-interface and C-terminal tail regions were found to enhance thermostability. However, due to lack of structural data the chemical basis of thermostability at the C-terminal tail is poorly understood and an Arg – Arg stacking interaction believed to result in increased thermostability of the Q86R variant of $CynD_{pum}$ (Wang, Watermeyer et al. 2012, Mulelu 2017)

Table 1.5.1: Reported thermostable nitrilases.

Organism	Molecular weight (kDa)	Optimum temperature (℃)	Substrate specificity	Reference

Pseudomonas sp. 13	1050	55	Aliphatic nitriles	(Yanase, Sakai et al. 1983)
Acinetobacter sp. AK226	580	50	Wide range	(Yamamoto and Komatsu 1991)
Bacillus pallidus Dac521	41	65	Aromatic nitriles	(Almatawah and Cowan 1999)
Pseudomonas fluorescens DSM 7155	40 and 38	55	Arylacetonitrile	(Layh, Parratt et al. 1998)
Acidovorax facilis 72W	570	65	Aliphatic dinitriles	(Chauhan, Wu et al. 2003)
Thermatoga maritima MSB8	30.07	45	Aliphatic nitriles	(Chen, Chen et al. 2015)
Parabulkhoderia phymatum	~666	45-55	Aromatic nitriles	(Bessonnet, Mariage et al. 2021)
Staphylococcus KP789138.1	-	50	Heterocyclic nitriles	(Arfi and Nigam 2020)
Rhodococcus rhodochrous K22	650	50	Aliphatic nitriles	(Kobayashi, Komeda et al. 1992)
Alcaligenes faecalis JM3	275	45	Arylacetonitrile	(Nagasawa, Mauger et al. 1990)

1.6 Cyanide

The cyano- functional group occurs in organic compounds, wherein the compound is often referred to as a nitrile. The functional group also occurs in inorganic compounds known as cyanides and is often found as simple solid cyanide salts. Cyanide, also known as prussic acid, is a weak, water-soluble acid. At alkaline pH, cyanide ions mostly remain in solution as salts but form volatile hydrogen cyanide (HCN) at physiological pH (Broderius 1981). The cyano functional group has a high affinity for metal ions and from this property stems its utility in

mining (discussed in Section 1.5), and toxicity. Cyanides, such as hydrogen cyanide, are synthesised through various industrial processes, such as the Andrussow process, but also occur in nature, where they are synthesised as metabolites in bacteria, fungi, and plants. In plants, for instance, where it is used as a natural defence mechanism, cyanide is synthesised through the breakdown of cyanogenic glycosides (Engler, Spencer et al. 2000, Gail, Gos et al. 2000).

The toxicity of cyanide stems from its ability to arrest cellular respiration. Cyanide ions bind to the trivalent iron molecule on cytochrome c oxidase, the final enzyme involved in aerobic respiration, inhibiting the binding of oxygen and consequently arresting cellular respiration, leading to death within minutes or hours depending on the dose (Way, Leung et al. 2007, Marziaz, Frazier et al. 2013). Mitochondrial rhodanese enzymes detoxify cyanide in both prokaryotic and eukaryotic cells by converting it to thiocyanate which is relatively less toxic (Chaudhary and Gupta 2012). Despite this detoxification mechanism, cyanide poisoning often leads to death owing to its high absorption rate and low lethal dose (Henschler 2003, Simeonova, Fishbein et al. 2004).

Although, cyanide is toxic and lethal, it is still used in large amounts in multiple industries, primarily because of its ability to bind metals. In a process called cyanide leaching or cyanidation, cyanide is used to extract precious metals, such as gold and iron from their ores

(Azizi and Ghaedrahmati 2015). The resulting waste exists as cyanide tailings i.e., cyanide remaining in complex with metals after extraction of the ores from the cyanide-ore slurry and in solution as cyanide wastewater. Cyanide tailings are stored as solid waste in tailing storage facilities or tailing dams. Some methods are actively being developed to recover precious metals from tailings (Hewitt, Breuer et al. 2012, Anning, Wang et al. 2019, Zhao, Zhao et al. 2023), however, the cyanide by-product still remains. Cyanide wastewater, which often includes other toxic co-contaminants, such as arsenic, is maintained at an alkaline pH in wastewater ponds, preventing the formation and volatility of hydrogen cyanide (Gail, Gos et al. 2000). These wastewater ponds pose a threat to all forms of life near them, with the risk of spillage and pollution of nearby ecosystems a constant reality.

Several chemical and physical methods and combinations thereof have been used to degrade cyanide. Chemical methods include alkaline chlorination, sulphite processes such as the use of oxone monopersulphate, and hydrogen peroxide oxidation (Hewitt, Breuer et al. 2012, Roldán, Olaya-Abril et al. 2021). Other combinations of chemical and physical properties include oxidation-reduction roasting which is used to degrade cyanide and recover precious metals remaining in the cyanide tailings (Zhao, Zhao et al. 2023). These methods do not result in complete degradation of the cyanide but reduce it to acceptable levels for storage in

wastewater effluents. Nitrilase enzymes have potential to completely convert and degrade free cyanide in wastewater effluents.

1.7 Cyanide degrading enzymes

Cyanide dihydratases (CynD) and cyanide hydratases (CHT) are nitrilases of microbial and fungal origin, respectively (Martínková, Kulik et al. 2023). They catalyse the hydrolysis of cyanide salts into useful chemicals and reaction intermediates. The CHT catalyses the conversion of cyanide salts to formamide, which requires an amidase enzyme to catalyse the final conversion of formamide to ammonia and formic acid. (Figure 1.3 (b)) (Dent, Weber et al. 2009). CynD catalyses the direct conversion of cyanide forming formic acid and ammonia without the amidase intermediate. This makes CynD the more economically viable enzyme, as it catalyses the reaction without the need for a second enzyme to bring the reaction to completion. However, both enzymes catalyse the hydrolysis of cyanide with high efficiency, without the need for co-factors or secondary substrates. This study is concerned with the investigation of CynD homologues from *Bacillus pumilus* C1 and *Pseudomonas stutzeri* AK61.

1.7.1 Cyanide dihydratase from *Bacillus pumilus* C1 ($CynD_{pum}$)

The *Bacillus pumilus* C1 ($CynD_{pum}$) strain was first isolated from mining wastewater, because of its ability to convert cyanide. The enzyme responsible for the conversion was later isolated,

characterised, and identified as a cyanide dihydratase (Meyers, Gokool et al. 1991, Meyers, Rawlings et al. 1993, Jandhyala, Berman et al. 2003). Based on residual activity analysis, $CynD_{pum}$ was found to be optimally active between pH 7-7.8 and temperatures 37–42°C, with less than 25% activity at 55°C.

$CynD_{pum}$ shares the characteristic monomer αββα fold found in enzymes of the nitrilase superfamily. The ~37 kDa monomers associate forming short left-handed spirals with C2 symmetry at optimum pH (Meyers, Rawlings et al. 1993, Jandhyala, Berman et al. 2003, Mulelu 2017, Sewell, Frangakis et al. 2017). Under specific solvent conditions (pH 5.4), these spirals self-associate forming inactive and long fibres of indeterminate length (Wang, Watermeyer et al. 2012).

However, the structural integrity and activity of wildtype $CynD_{pum}$ are abolished at pH higher than 8, rendering it inactive for the conversion of cyanide at the alkaline pH, at which cyanide wastewater is kept (Jandhyala, Willson et al. 2005, Wang, Watermeyer et al. 2012). As such, multiple studies have been conducted to understand the fundamental basis of formation of the structure of $CynD_{pum}$ under different solvent conditions (Mulelu 2017) and to engineer variants of the enzyme that are stable and active at alkaline pH. These studies have included the cysteine scanning of C-interface amino acids (Park, Mulelu et al. 2016), directed evolution of amino acids at the C-terminal tail (Sewell, Thuku et al. 2005, Crum, Park et al. 2015, Crum,

Park et al. 2015) and error prone polymerase chain reaction (PCR) generating libraries for the prediction of structurally stable variants (Wang, Watermeyer et al. 2012). Valuable insights have been gained from these studies such as the importance of the C-terminal region in maintaining the structural integrity of the enzyme, the identification of amino acid substitutions that contribute to increased stability and studies into the reaction mechanism of the nitrilases. Findings from these studies based on homology models have been hampered, in some instances by the lack of an atomic resolution structure of $CynD_{pum}$ resulting in incorrect assignment of amino acids to specific interfaces.

1.7.2 Cyanide dihydratase from *Pseudomonas stutzeri* AK61 ($CynD_{stu}$)

A cyanide degrading strain identified as *Pseudomonas stutzeri* AK61 was isolated from the wastewater of a metal-plating plant in Japan, where whole cells were able to degrade cyanide without the addition of any additives (Watanabe, Yano et al. 1998). $CynD_{stu}$ is optimally active between a pH range of 7 and 7.8 with a ~50% loss of activity at pH 8 and temperatures between 35- 37°C (Watanabe, Yano et al. 1998, Jandhyala, Willson et al. 2005). $CynD_{stu}$ also shares sequence and structural homology with members of the nitrilase superfamily, existing as a short left-handed spiral with C2 symmetry (Figure 1.2 (a)). $CynD_{stu}$ is unique among the nitrilases, having been identified as self-terminating, i.e. possessing specific interactions at the E-interface that prevent the addition of monomers to the spiral thus forming discrete 14mer

spirals (Sewell, Berman et al. 2003). The catalytic triad of amino acids (CEK) was identified in $CynD_{stu}$, prior to identification of the fourth catalytic amino acid, with C163 identified as the catalytic nucleophile cysteine based on loss of activity upon substitution with a serine (Watanabe, Yano et al. 1998, Watanabe, Yano et al. 1998).

1.8 Rationale

Cyanide dihydratases (CynD) are potential biocatalysts for the bioconversion of cyanide waste to useful products with economic value, and bioremediation of sites contaminated with cyanide waste. They catalyse the direct conversion of cyanide without the formation of reaction intermediates or the requirement for additional co-factors, co-enzymes, or additives.

One of the desirable properties for an industrial biocatalyst is tolerance to non-optimal reaction conditions, including high temperatures which have the benefit of allowing the reaction to proceed more efficiently by accelerating substrate turnover rates, thus meeting production demands. However, nitrilases are protein in nature and are therefore, susceptible to degradation and inactivation at the alkaline pH at which cyanide waste is kept. Furthermore nitrilases are mesophilic, exhibiting optimal activity at a maximum temperature of 37 ℃ (Jandhyala, Willson et al. 2005). Consequently, these native properties necessitate modification to adapt them for industrial use. Important to the engineering of a recombinantly

expressed thermostable biocatalyst is an in-depth understanding of the chemical and structural determinants that lead to variants with increased thermostability. Thus, this study is concerned with the chemical and structural determinants of enhanced thermostability in $CynD_{pum}$.

Among known CynD enzymes, $CynD_{pum}$ is the most studied (Meyers, Gokool et al. 1991, Jandhyala, Berman et al. 2003, Sewell, Thuku et al. 2005, Wang, Watermeyer et al. 2012, Crum, Park et al. 2015, Crum, Sewell et al. 2016, Park, Mulelu et al. 2016, Mulelu 2017). $CynD_{pum}$ undergoes changes in its oligomeric state, forming large homo-oligomers (fibres) at pH 5.4, a phenomenon attributed to changes to the ionisation state of histidine amino acids at the C-terminal tail (Mulelu 2017). This pH-dependent change has not been observed in its homologues, CynD 8A3 and $CynD_{stu}$ (Eicher 2007, Mulelu 2017). Additionally, some directed evolution studies have produced mutants with enhanced thermostability. These include the Q86R variant which retains activity at pH 9.0, identified using error-prone PCR (Wang, Watermeyer et al. 2012). The CynDpum- CynDstu chimera created by replacing the C-terminal region of $CynD_{pum}$ with that of $CynD_{stu}$, that retains approximately 40% of its activity after incubation at 42°C for 20 hours (Sewell, Thuku et al. 2005, Crum, Park et al. 2015).

Although various studies on directed evolution have identified thermostable variants of $CynD_{pum}$, with thermostability measured using methods such as melting temperature, residual

activity after incubation at a specific temperature and ΔΔG, the underlying chemical basis for these observations remains poorly understood. Consequently, it remains challenging to reproduce more thermostable variants systematically and logically. Additionally, previous directed evolutionary studies have often relied on structural data from homology models from distant homologs, which were sometimes inaccurate. As a result, the exact locations of some regions have been misinterpreted, leading to inaccuracies in the interpretation of mutations and their effects. Lastly, random mutagenesis experiments often yield a large proportion of inactive variants due to the limited amount of available amino acid sequence space that can be sampled without abolishing enzyme activity.

1.9 Aim and objectives

The overarching aim of this study is to elucidate the determinants of enhanced thermostability in $CynD_{pum}$.

This aim will be achieved by addressing the following objectives:

1. Identify and investigate substitution mutations that potentially lead to enhanced thermostability in $CynD_{pum}$.

Our approach involves a comprehensive *in silico* exploration of the entire amino acid sequence space of the $CynD_{pum}$ homo-oligomer, utilising the atomic resolution structures of $CynD_{pum}$

and its Q86R/H305K/H308K/H323K variant. This methodology allows for a thorough assessment of the impact of substituting each amino acid at every position on the overall thermostability of the enzyme. By integrating computational predictions with experimental validation, we aim to elucidate the chemical and structural determinants of thermostability for $CynD_{pum.}$

2. **Experimental validation of selected predicted mutations**

Following the identification of potential mutagenesis targets, experimental validation will be conducted to verify the predicted improvements in thermostability. This entails generating variants of $CynD_{pum}$ through site-directed mutagenesis, activity analysis using the picric acid calorimetric assay, structural analysis by transmission electron microscopy, and biochemical and biophysical analysis by nanoscale differential scanning fluorimetry.

3. **Determine the chemical basis of enhanced thermostability in variants with enhanced thermostability.**

To determine the chemical basis of enhanced thermostability in variants with improved thermostability using structural analysis.

4. Solve and interpret the near atomic resolution structure of $CynD_{stu}$ and investigate the biochemical contributions of the C-terminal region of $CynD_{stu}$ leading to enhanced thermostability in its homologue $CynD_{pum}$.

The enhanced thermostability of the $CynD_{pum}$-$CynD_{stu}$ chimeric protein is attributed to new interactions brought about by the substitution of the C-terminal tail region of $CynD_{pum}$ with that of $CynD_{stu}$. To investigate the chemical basis of these observations, we will solve the structure of $CynD_{stu}$ at near-atomic resolution using cryogenic electron microscopy and single particle analysis. Furthermore, the chemical basis for the observed self-termination of the $CynD_{stu}$ spiral will be elucidated to re-engineer the interface contributing to the self-termination in its homologue, $CynD_{pum}$.

Chapter II

Investigating the Determinants of Thermostability in Cyanide Dihydratase from *Bacillus pumilus* C1 (A Structural Analysis)

2.1 Introduction

As discussed in Chapter I, cyanide dihydratases (CynD) are promising biocatalysts for cyanide conversion in wastewater treatment and remediation of cyanide-contaminated sites (Gong, Lu et al. 2012). Essential to the viability of a biocatalyst is its thermostability, which influences process efficiency, storage duration and conditions, and operational versatility. Moreover, within the nitrilase family, thermostability often correlates with tolerance to alkaline pH, to the extent that the ionisation states of catalytic amino acids, CEEK, will allow. As outlined in Chapter I, cyanide-containing wastewater is maintained at alkaline pH to prevent the release of volatile HCN. Therefore, in addition to withstanding non-optimal temperatures, thermostable nitrilases may also exhibit prolonged operational stability under alkaline conditions. Currently, wild-type $CynD_{pum}$ is optimally active at 37 °C and pH 7.4 (Jandhyala, Willson et al. 2005). Although whole cells have been shown to catalyse the conversion of cyanide more efficiently at elevated temperatures, recombinantly expressed $CynD_{pum}$ has been shown to be more catalytically efficient in the degradation of cyanide at alkaline pH (Vargas-Serna, Carmona-Orozco et al. 2020). Previous work has demonstrated potential correlation between thermostability and tolaerance to alkaline pH (Wang, Watermeyer et al. 2012, Crum, Park et al. 2015).

2.2 Rationale

Several studies have focused on specific sites on $CynD_{pum}$ to gain insights into its structure and improve its tolerance to alkaline pH and non-optimal temperatures (Sewell, Thuku et al. 2005, Wang, Watermeyer et al. 2012, Crum, Park et al. 2015). Due to the absence of an atomic-resolution structure, many of these studies relied on homology models. While confirming the αββα monomer fold and αββα-αββα dimer subunit, these models were prone to inaccuracies in predicting the locations of specific secondary structure elements, leading to off-target substitution mutations.

Several protein engineering strategies have been developed to improve the thermostability of $CynD_{pum}$ such as directed evolution using various forms of the polymerase chain reaction (PCR) (Wang, Watermeyer et al. 2012, Crum, Park et al. 2015, Crum, Sewell et al. 2016, Xu, Cai et al. 2018) and consensus sequence analysis (Xiong, Lv et al. 2022). However, these methods only explore a fraction of the amino acid sequence space, and in the case of consensus sequence analysis, may face challenges due to insufficient homology between the thermostable and the target mesophilic protein. To address these limitations, our study aimed to investigate the structural and chemical factors influencing thermostability in $CynD_{pum}$ through *in silico* substitution of all amino acids within the homo-oligomer of the solved structure.

The primary objective of this chapter is to investigate a rational and targeted approach to enhance the thermostability of $CynD_{pum}$ through a comprehensive analysis of its atomic resolution structure solved by cryo-electron microscopy (Mulelu 2017) and identification of potential structural and chemical determinants of its thermostability. This chapter addresses objectives 1,2 and 3.

2.3 Materials and methods

2.3.1 Refinement and interpretation of the reference structure $CynD_{pum}$ (Q86R/H305K/H308K/H323K)

Obtaining an accurate spatial configuration of the protein's constituent atoms, which is determined by the charge density map's resolution in atomic resolution structures, is essential for conducting reliable *in silico* site-saturation mutagenesis (SSM) experiments (Buß, Rudat et al. 2018). As a result, refining and interpreting the current model are the initial steps in this process, thereby enabling the precise analysis of subsequent computations. We utilised the 3.15 Å resolution structure of the $CynD_{pum}$ variant Q86R/H305K/ H308K/H323K (Mulelu 2017, Sewell, Frangakis et al. 2017), as the reference model for this study.

The reference model was refined using the phenix.real_space_refine program of the PHENIX suite version 1.20.1-4487-000 (Afonine, Poon *et al.* 2018) and flexible molecular dynamics fitting using Interactive Structure Optimisation by Local Direct Exploration (ISOLDE) plugin in ChimeraX (Croll 2018, Pettersen, Goddard *et al*. 2021). The deposited co-ordinates are assigned the PDB ID: 8c5i. Interfacial regions were identified and analysed by computing the interaction energy and binding affinity between monomers using the AnalyseComplex program in the FoldX5 suite 5.0 (Schymkowitz, Borg *et al.* 2005) and the PRODIGY online

server, respectively (Schymkowitz, Borg *et al.* 2005, Xue, Rodrigues *et al.* 2016). Additionally, specific amino acids involved in interactions between the protein molecules and the interactions they participate in were identified. Both programs were run at 37 ˚C, the AnalyseComplex program was run with the additional parameter, pH 7.4. Computation of the binding free energy (ΔG_{bind}) between two monomers in FoldX is given by ΔG_{bind} defined by Equation 3 (Table 2.3.1).

2.3.2 Structure guided *in silico* site saturation mutagenesis (SSM)

SSM was used to conduct an exhaustive and targeted analysis of the effect of mutagenesis on the Gibb's free energy change of protein unfolding (ΔG) on the entire available sequence space of $CynD_{pum}$ (Q86R/H305K/H308K/H323K). The refined model was truncated by deleting specific chains, including only a single turn of the helix, consisting of all known interfaces in the final model used for SSM, to reduce computing time and resources. SSM was performed using the BuildModel program of the FoldX suite (Schymkowitz, Borg et al. 2005), where each of the resolved amino acids (3-319) was substituted with the remaining 19 standard amino acids. The mutation_list file included the 20 standard amino acids, and the BuildModel program was run at 37 °C and pH 7.4, for 10 iterations. To facilitate data representation and visualisation, the FoldX BuildModel job was run from within MutateX (Tiberti, Terkelsen et al. 2022).

The approach used in this study to quantify the thermostability of $CynD_{pum}$ variants is based on the Gibb's free energy change of protein unfolding (ΔG^u). Proteins are stabilised by the cumulative and synergistic effect of interactions between constituent amino acids and with the environment. The most stable conformation corresponds to the lowest energy state or minimum. ΔG^u is a measure of all the energetic contributions of the system i.e. all the atoms constituting the amino acid of the protein and their interactions with the environment under specific conditions such as temperature and pH (Table 2.3.1) (Lutz and Bornscheuer 2012, Sanavia, Birolo et al. 2020). The difference between the ΔG^u of the wildtype protein (ΔG^u_{wt}) and the variant with the substitution mutation(s) (ΔG^u_{var}), $\Delta\Delta G^u$ is used to measure the thermodynamic favourability of the introduced mutation. Negative values typically indicate the thermodynamic favourability of the introduced mutation while positive values indicate that the system (i.e. protein and environment) occupies a thermodynamic minimum that is higher than that of the wildtype and is therefore unfavourable. For simplicity, we refer to ΔG^u as ΔG from here onwards. In relation to protein thermostability, variants with a low energy minimum than the wildtype i.e. proteins with $\Delta\Delta G<0$ have been shown to generally denature at higher temperatures (Rees and Robertson 2001).

Table 2.3.1: Equations for computing the Gibb's free energy change of protein unfolding and protein-protein binding. Equation 1 is the general equation that outlines the relationship between ΔG and the equilibrium constant of protein unfolding (K_{eq}). Equations 1-3 are employed by the FoldX algorithm to compute the thermodynamic favourability between interacting protein molecules. Equation 2 represents the general equation for ΔG computed for each molecule. Equation 3 is used to compute the interaction energy and thermodynamic favourability between molecules. Equations 4-5 are used to compute the thermodynamic favourability of amino acid changes. The individual thermodynamic contributions to Gibb's free energy for each amino acid are outlined in Equation 4, these parameters are described in detail in Table 2.3.2. Equation 5 is used to determine the thermodynamic favourability of the mutation. Equations 6-12 are used in the derivation of ΔG from experimental chemical denaturation curves. C50 is the concentration of denaturant at the folded and unfolded states exist in equal proportions.

Equation 1	$\Delta G = -RT \ln K_{eq}$
Equation 2	$K_{eq} = [Unfolded]/[Folded]$
Equation 3	$\Delta G_{bind} = -RT \ln K_d$
Equation 4	$\Delta G_{bind} = \Delta G_{AB} - (\Delta G_A + \Delta G_B)$
Equation 5	$\Delta G = a\Delta G_{vdw} + b\Delta G_{solvH} + c\Delta G_{solvP} + d\Delta G_{wb} + e\Delta G_{hbond} + f\Delta G_{el} + g\Delta G_{kon} + hT\Delta S_{mc} + kT\Delta S_{sc} + l\Delta G_{clash}$
Equation 6	Folded $\leftrightarrow$ Unfolded
Equation 7	$\Delta G(C) = -RT \ln K_{eq}(C)$ (where C is the specific denaturant concentration)
Equation 8	$\Delta G (H_2O) = \Delta G (C)$ limC $\rightarrow 0$
Equation 9	$\Delta G(C) = \Delta G (H_2O) - m (C)$
Equation 10	$C50 = \Delta G (H_2O)/m$
Equation 11	$K_{eq}(C) = F(unfolded)/F(folded)(C) = y(folded)(C) - FI(C)/ FI(C) - y(unfolded)(C)$
Equation 12	$\Delta\Delta G = \Delta G_{wt} - \Delta G_{var}$

Table 2.3.2: Thermodynamic properties used in FoldX to compute the Gibb's free energy of protein unfolding and definition of parameters.

ΔG	Gibb's free energy change of protein unfolding
ΔG_{vdw}	van der Waals term
ΔG_{solvH}	desolvation term for hydrophobic groups
ΔG_{solvP}	desolvation term for electrostatic/polar groups
ΔG_{wb}	water molecules persistently hydrogen bonded with the protein

ΔG_{hbond}	hydrogen bonds
ΔG_{el}	electrostatic contribution
ΔG_{kon}	electrostatic contribution between atoms of different chains for oligomeric proteins
$T\Delta S_{mc}$	entropic penalty for fixing incorrect angles along the main chain
$T\Delta S_{sc}$	entropic penalty for fixing incorrect side chain angles
ΔG_{clash}	measure of steric overlap between atoms in the structure
ΔG_{wt}	Gibb's free energy change for the wildtype protein
ΔG_{var}	Gibb's free energy change for the variant protein
ΔΔG	Difference in the change in Gibb's free energy between wildtype and variant proteins

2.3.3 Site-directed mutagenesis (SDM)

In this study, we used the thermodynamic favourability of the Gibb's free energy of protein unfolding as an indicator of thermostability.

Mutations predicted through FoldX computations to be thermodynamically favourable were selected based on insights gained from the analysis and interpretation of the *in silico* SSM reference structure; however, attention was given to including substitutions representing a wide range of predicted ΔΔG values and different locations on the protein to prevent bias. Mutations occurring at least 10 Å away from the substrate binding site were selected to prevent mutagenesis of catalytic amino acids or residues that facilitate substrate binding in order to minimize the risk of deactivating the enzyme.

Site-directed mutagenesis was used to introduce target substitution mutations into the *cyndpum* gene. Suitable primers with the desired substitution mutations were designed based on the template sequence of wild-type $CynD_{pum}$ (GenBank: AF492815.1). The primers were designed to be no more than 36 base pairs in length with a low GC content, where possible, and the annealing temperature (T_a) ranged between 55 and 65 °C. Primer design, calculation of the GC content and T_a were conducted using SnapGene.

Oligonucleotides were synthesised at Source Bioscience, UK (MPL, Membrane Protein Lab) and Inqaba Biotech, SA. All primers were diluted to 10 µM working stock solutions with deionised water. Template DNA was extracted from NEBα cells (New England Biolabs, USA) transformed with the pET26b-*cyndpum* expression construct using the QIAprep Spin Miniprep Kit (Qiagen) (MPL) and the ThermoScientific Miniprep kit (ThermoScientific, USA). Substitution mutations were introduced into the template DNA using PCR, with the reaction components outlined in Table 2.3.5. Phusion Flash DNA polymerase was selected for its high fidelity, and 4% (v/v) dimethyl sulfoxide (DMSO) was added to the reaction mixture to improve amplification owing to the high GC content of the template DNA.

The reaction was run on a Veriti Cycler (Applied Biosystems, USA) (MPL) and a T100 thermal cycler (Bio-Rad, USA) using the cycling parameters outlined in Table 2.3.6. The PCR products were analysed on precast 48 well 1% agarose e-gels (Thermo Fisher Scientific, USA) using an

e-gel® system (Invitrogen, USA) (MPL). The unmutated methylated template DNA was degraded using DpnI restriction enzyme digestion. DpnI (5U) (NEB) was added to each 25 µL PCR reaction and incubated at 37 °C for 30 min. The enzyme was inactivated by incubating at 80 °C for 20 min in a thermal cycler. Each DpnI-digested reaction (3 µL) was used to transform 20 µL of chemically competent Top10 cells (MPL) or NEB-alpha high-efficiency cells (NEB).

When creating a multi-site substituted variant, mutations were introduced sequentially following the protocol outlined above. Sanger sequencing was performed after each iteration to verify the presence of introduced mutations. The mutated plasmid served as the template for subsequent iterations, thereby introducing the second mutation for creation of the multi-site substituted enzyme.

Table 2.3.3: Primer sequences used for site-directed mutagenesis of expressed mutants. The mutated codons are underlined and shown in the 5' – 3' direction.

Primer	Sequence
I4M	ATGACAAGTATGTACCCAAAGTTTCGAGCAGCTGC
S25V	CTACTTAAATTTGGAAGCAGTCGTTGAGAAATCATGTG
S29A	AAGCAAGCGTTGAGAAAGCATGTGAACTGATCGAC
C30V	CAAGCGTTGAGAAATCAGTTGAACTGATCGACGAGGC
D34L	ATCATGTGAACTGATCCTCGAGGCAGCTTCAAATGG
D34R	GAAATCATGTGAACTGATCCGCGAGGCAGCTTCAAATG
E35A	AACTGATCGACGCCGCAGCTTCAAATGGTGCAAAGCTT
H62Y	TTTGCTTTTATTGGTTATCCAGAATATACGAGA
T66Q	CATCCAGAATATCAGAGAAAGTTCTATCATGAATT

H71R	CGAGAAAGTTCTATCGTGAATTATATAAAAATGCCG
L83W	GTTGAAATCCCTAGCTGGGCCATTCAAAAAATAAGTG
Q86M	CCCTAGCTTAGCCATTATGAAAATAAGTGAGGCAGCAAAG
Q86K	CGTTGAAATCCCTAGCTTAGCCATTAAAAAAATAAG
E90K	GCTTAGCCATTCAAAAAATAAGTAAGGCAGCAAAG
K93R	CAAAAAATAAGTGAGGCAGCAAGGAGAAATGAAACGTACGTTTG
E96G	GAGAAATGGAACGTACGTTTGTATATCATGC
E105V	TGTATATCATGCAGTGTAAAAGATGGCGGTTCTCTC
N119S	CTCTCTATTTAGCTCAGCTTTGGTTTTCTCCTAATGGGGA
N119R	CTATTTAGCTCAGCTTTGGTTTCGTCCTAATGGGGATTTAATAGG
D123E	GTTTAATCCTAATGGGGAGTTAATAGGAAAACATCGG
S147I	GGGATGGAAGTGGAATTATGATGCCGGTGTTTC
E155R	GTATGATGCCGGTGTTTCAAACTCGGATTGGAAACCTTGGCG
D172N	GCATCAAGTCCCACTTAATCTTATGGCGATG
N177L	CTTGATCTTATGGCGATGCTTGCCCAAAATGAGC
E181A	GCGATGAATGCCCAAAATGCACAGGTACATGTAGCCTCTTG
D193V	GCCAGGTTATTTTGTGGATGAAATATCAAG
S197L	GATGAAATATTGAGTAGATATTATGCTATCGCG
T217I	GTGCTGATGACATCATCTATATATATTGAAGAAATGAAAGAGAT
D232L	GATTTGTTTAACGCAGGAGCAAAGAATTACTTTGAAACATTTAAG
S239L	CTTTGAAACATTTAAGATCGGGCATACGTGCATT
E250R	ATTTACGGGCCGGACGGGAGACCGATCAGTGATATGG
T260I	GATATGGTTCCAGCTGAAATCGAGGGAATTGCTTACGC
D275E	AGAAAGAGTCATTGAATACAAGTATTATATTGATC
H285M	TATTGATCCGGCTGGAATGTACTCCAATCAAAGTTTG
N288R	CTGGACATTACTCCGCTCAAAGTTTGAGTATG
Q289R	CCGGCTGGACATTACTCCGATCAAAGTTTGAGTATG
S292R	CAATCAAAGTTTGAGGATGAATTTTAATCAGC
N294C	CAATCAAAGTTTGAGTATGTGTTTTAATCAGCAGCCCAC

T300Y	GAATTTTAATCAGCAGCCCTATCCTGTTGTGAAACAT
N311R	CATTTAAATCATCAAAAAGCTGAAGTATTCACATATGAGG
E312F	AAATCATCAAAAAAATTTGTATTCACATATGAGGACA
T315I	CAAAAAAATGAAGTATTCATATATGAGGACATTCAATATC
D318Y	GTATTCACATATGAGTACATTCAATATCAAC

Table 2.3.4: Primer sequences used for Sanger sequencing are shown in the 5' – 3' direction.

Primer	Sequence	Source
T7 promoter	TAA TAC GAC TCA CTA TAG GG	DNA Sequencing Unit (Stellenbosch University) and Source Bioscience, UK
T7 terminator	GCT AGT TAT TGC TCA GCG G	DNA Sequencing Unit (Stellenbosch University), Source Bioscience, UK

Table 2.3.5: PCR master mix used for construction of single-site substitution mutants.

For a 25 µL reaction	
Reagent	**Amount**
2X Phusion Flash MasterMix (Thermo Fisher Scientific, USA)	12.5 µL
Template plasmid DNA	0.4 ng
Dimethyl sulphoxide (DMSO)	4 % (v/v)
Deionised water	Up to 25 µL
Fwd primer	0.1 µM
Rev primer	0.1 µM

Table 2.3.6: Cycling parameters used for PCR.

Step	Temperature (℃)	Duration
1. Initial denaturation	98	10 sec
2. Denaturation	98	5 sec
3. Annealing	Primer dependent	Primer dependent

4. Elongation	72	1 min
5. Final elongation	72	2 min
6. Final hold	4	
30 cycles completed for each reaction		

2.3.4 Transformation of chemically competent bacterial cells

All the recombinant expression vectors used in this study were transformed into chemically competent NEBα cells (New England Biolabs) for plasmid propagation. The plasmid was extracted and transformed into BL21 (DE3) cells (New England Biolabs) for overexpression of the target protein.

Cells (20 µL) were thawed on ice and plasmid DNA (3 µL) was added to the cells. The mixture was incubated on ice for 30 min, and the cells were heat-shocked at 42 ℃ for 30-60 sec in a water bath and cooled on ice for 2 min. Super Optimal broth with Catabolite repression (SOC) media (150 µL) (NEB) was added to the cells and incubated at 37 °C for 1 h with shaking at 600 rpm. The transformed cells were centrifuged for 1 min at 14 000 xg and resuspended in 80 µL of the cell culture medium. A 50 µL sample of the cells was plated onto Luria Bertani (LB) agar media plates (Table 2.3.7) supplemented with 50 µg/mL kanamycin (Sigma Aldrich) and incubated at 37 °C overnight. Cell colonies growing on the plates were picked and grown in liquid media to confirm the presence of the correct substitution mutation and prepare glycerol stock solutions for target protein expression.

A single colony was inoculated into terrific broth (TB) medium (5 mL) and incubated overnight at 37 °C. The plasmid was extracted from the cells using the Thermo Scientific Miniprep kit (Thermo Scientific, USA), and the nucleotide sequence was determined by Sanger sequencing to confirm the presence of target mutants. Sanger sequencing was done at the DNA Sequencing Unit at Stellenbosch University and Source Bioscience (UK, MPL) using standard T7 promoter and terminator sequencing primers (Table 2.3.4).

Preparation of glycerol for protein expression was performed by inoculating a single colony of transformed cells into 5 mL LB medium supplemented with 50 µg/mL kanamycin and incubating the cells at 37 °C for 6 h. Glycerol (30%, v/v) was added to the cells to a final volume of 1.5 mL. The glycerol stock solutions were stored at -80 °C.

2.3.5 Recombinant protein expression and purification of cyanide dihydratase from *Bacillus pumilus* C1 ($CynD_{pum}$) and its variants

2.3.5.1 Recombinant protein expression

The pET26b- *cyndpum* expression vector (Jandhyala, Berman et al. 2003) was transformed as described in section 3.2.4. The starter culture was prepared by inoculating 10 mL LB medium supplemented with 50 µg/mL kanamycin with a scraping of cells transformed with the $CynD_{pum}$ expression vector stored in glycerol stock. The cells were grown overnight at 37 °C

with shaking in a Gerhardt Thermoshake incubator (Gerhardt Analytical Systems, ID) at 200 rpm.

LB media (1 L) was inoculated with the 10 mL starter culture and incubated at 37 °C until an OD_{600} (optical density at 600 nm) of between 0.4 and 0.6 was reached. Thereafter, target protein overexpression was induced by activating the lac operon by adding 100 mM isopropyl β-D-1 thiogalactopyranoside (IPTG), and the culture was incubated for an additional 18 h at 25 °C. The cell culture was centrifuged at 4500 × g for 15 min to collect the cell pellet which was resuspended in 5 mL (for 1 L culture) of lysis buffer (350 mM NaCl, 50 mM Tris-HCl pH 7.4, protease inhibitor (cOmplete EDTA-free protease inhibitor cocktail, Roche).

Uninduced and induced samples were collected before and after IPTG addition for the analysis of target protein overexpression by sodium dodecyl sulphate polyacrylamide gel electrophoresis (SDS-PAGE).

The expression construct expressing $CynD_{pum}$ (Q86R/H305K/H308K/H323K) peT26b-*cyndpum* was sourced from Mulelu (2017) and expressed following the protocol outlined for the wild-type enzyme. All $CynD_{pum}$ variants generated in this study were similarly expressed, according to the procedures detailed in this section.

Table 2.3.7: Cell culture media and additives

Name	Description	Use
Media		

Luria-Bertani (LB) media	1% (w/v) tryptone, 0.5% (w/v) yeast extract, 0.05% NaCl	Bacterial protein expression
Terrific broth (TB)	2.4% yeast extract(w/v), 1.2% tryptone (w/v), 0.94% dipotassium phosphate, 0.22% monopotassium phosphate, 0.2% glycerol (Scharlau)	Bacterial cell growth
LB agar solid media	0.5 % NaCl, 0.5 % yeast extract powder, 1 % agar	Solid transformation selection media
Additives		
Kanamycin	50 µg/mL (Sigma Aldrich)	Bacterial transformation selection
IPTG (isothiogalactopyranoside)	100 mM (Melford)	*lac* operon activation

Cell lysis

The cells were thawed at room temperature and disrupted by sonication on a Sonicator® 3000 (Misonix, USA) with a blunt-tip horn (Sigma Aldrich). Cell disruption was conducted with temperature monitoring and control for 4 min, interspersed with 15 s of cooling and 10 s of lysis. The cells were kept in an ice slurry during sonication to prevent overheating and potential protein denaturation. The crude cell lysate was then centrifuged at 29 000 xg at 4 °C for 30 min. Samples from the soluble and insoluble fractions were collected to confirm the solubility of the target protein by SDS-PAGE.

For small-scale expression of $CynD_{pum}$ variants, cells resuspended in lysis buffer were sonicated using a microtip attachment (Sigma Aldrich) instead of a blunt tip horn. The total sonication time was 2 min, with intervals of 15 s for cooling, and 10 s for lysis.

2.3.5.2 Protein purification

Protein purification was conducted following a three-step process involving ammonium sulphate precipitation (ASP), ion-exchange chromatography (IEX), and size exclusion chromatography (SEC), which was used as the final polishing step.

Ammonium sulphate precipitation

Because the $CynD_{pum}$ expression construct is untagged, ammonium sulphate precipitation (ASP) was used as the initial purification step to isolate the target protein from the crude cell lysate.

To determine the percentage of ammonium sulphate at which $CynD_{pum}$ became insoluble, ammonium sulphate was added to the cell lysate in 10% (w/v) saturation increments from 10% to 50%, and the solution was mixed on a roller mixer at 4 °C for 30 min. The mixture was centrifuged at 29 000 xg for 10 min at 4 °C, and the soluble and insoluble fractions of each percentage saturation were collected and analysed by SDS-PAGE to confirm the amount of ammonium sulphate at which $CynD_{pum,}$ and its mutants were eluted. All $CynD_{pum}$ was in the soluble fraction at 40% ammonium sulphate saturation; thus, for purification of large-scale cell cultures, cells were disrupted as described previously and incubated in 40% ammonium sulphate overnight at 4 °C on a roller mixer. The soluble and insoluble fractions were separated by centrifugation at 29 000 xg for 15 min at 4 °C.

Ion exchange chromatography

Ion exchange chromatography (IEX) was used as the intermediate purification step following the ASP. Prior to IEX, buffer exchange was conducted using the HiPrep™ 26/10 desalting column (Cytiva) equilibrated with two column volumes (CV) of low-salt buffer (150 mM NaCl, 50 mM Tris-HCl pH 7.4).

IEX was performed in two steps with columns of different resin sizes owing to inadequate protein purification in a single step. The first step of IEX was performed on a HiTrap 16/10 IEX column (Cytiva) with a 90 µm bead size equilibrated with 2CV of low-salt buffer. The target protein was eluted using high-salt buffer (50 mM Tris-HCl, 1 M NaCl, pH 7.4). The second step was performed on a CaptoQ ImpRes column (Cytiva) with a 34 µm bead size using the same procedure outlined in the first step of the IEX. IEX steps were performed using a step elution gradient. Fractions were collected and analysed using SDS-PAGE, as described in Section 2.3.6. Fractions with bands corresponding to the expected monomeric size of $CynD_{pum}$ (~ 37 kDa) were pooled together and concentrated in an Amicon centrifugal filter (Merck, USA) with a 100 kDa membrane filter limit until a volume of 5 mL was reached.

Size exclusion chromatography (SEC)

The column was calibrated prior to running the sample using lyophilised protein molecular weight (M_r) standards (Bio-Rad, USA) reconstituted in deionised water.

Protein samples (5 mL) were filtered using a 0.2 µm filter and loaded onto a HiPrep™ 16/10 Sephacryl™ S300 HR column (Cytiva, USA) pre-equilibrated with 240 mL (2CV) of SEC buffer (50 mM Tris-HCl, 50 mM NaCl, pH 7.4). SEC buffer was used as the mobile phase at a flowrate of 0.5 mL/min. The sample was run through the column at 4 °C. Fractions corresponding to an increase in absorbance at 280 nm were collected and analysed by SDS-PAGE and negative-stain EM.

Buffer exchange was performed to investigate the changes in protein oligomerisation at different pH (pH 5.4, 9, and 10). Purified samples were concentrated to 1 mL using an Amicon centrifugal filter (Merck, USA) with a 100 kDa membrane filter limit. Subsequently, 14 mL of SEC buffer (pH 9 and 10) and citrate buffer (pH 5.4 (10 mM citrate, 50 mM NaCl) were added to the Amicon centrifugal filter, and the samples were concentrated to 1 mL. The process was repeated by adding another 14 mL of buffer, and the samples were then kept at 4 °C for approximately 12 h. Following this incubation period, the diluted samples were concentrated again to at least 1 mg/mL concentration and characterised.

2.3.6 Protein analysis

2.3.6.1 Sodium dodecyl sulphate -polyacrylamide gel electrophoresis (SDS-PAGE)

Protein expression and purification were analysed using sodium dodecyl sulphate - polyacrylamide gel electrophoresis (SDS-PAGE) as described by Laemli (Laemmli 1970). All analyses were performed on 10% polyacrylamide gels, and 10 µL of sample application buffer was added to 25 µL of the sample. However, the sample volumes were adjusted, and dilutions were made to obtain the best separation of the protein bands. The sample was incubated at 95 °C for 5 min, centrifuged for 2 min at 14 000xg on a Tabletop minispin centrifuge (Eppendorf, Germany), and loaded onto the gel. Electrophoresis was performed at a constant current of 20 mA on a BIORAD MiniPROTEAN gel apparatus (Bio-Rad, USA).

SDS-PAGE gels were stained with Aqua stain (Vacutec, SA) and visualised to assess the presence of target proteins. The expected monomeric size of the target proteins was determined using the Color Prestained Protein Standard Broad Range (NEB) protein molecular weight marker, unless otherwise indicated. The components of the buffers used for SDS-PAGE are listed in Table 2.3.8.

Table 2.3.8: Buffers used for SDS-PAGE.

Name	Description

Stacking buffer (4%, 5mL)	30% (v/v) acrylamide/bis solution (Biorad, USA) 0.67 mL, 0.5 M Tris-HCl pH 6.8 (1.25 mL), 10% (w/v) ammonium persulphate, 0.1 % (v/v) TEMED (N, N, N', N'-Tetramethyl ethylenediamine) (5µL), dH_2O
Resolving buffer (8%)	30% (v/v) acrylamide/bis solution (Biorad, USA), 1.5M Tris-HCl pH 8.8, 1% (w/v) ammonium persulphate (APS), 0.1% TEMED
(5X) Tris-Glycine SDS (TGS) running buffer	1.51% (w/v) Tris, Glycine, 0.005% (w/v) SDS
4X sample application buffer	200 mM Tris-HCl (pH 6.8), 8% (w/v) SDS, 0.4 % (w/v) bromophenol blue, 40% (v/v) glycerol

2.3.6.2 Measurement of protein concentration

Protein concentration was calculated using the Beer-Lambert equation. The absorbance was determined spectrophotometrically using a MultiSkan®Go spectrophotometer (Thermo Fisher Scientific, USA) at a wavelength of 280 nm and the extinction coefficient (ε) of the specific protein sample. The final concentration was calculated by multiplying the resulting value by the M_r of the protein.

For experiments conducted at the MPL and Roche Diagnostics, a NanoDrop spectrophotometer (Thermo Fisher Scientific, USA) was used to quantify protein concentrations, with the extinction coefficient and molecular weight of each protein sample specified.

2.3.7 Screening predicted mutants for activity and enhanced thermostability

2.3.7.1 Enzyme activity analysis

Enzyme activity was used as a screening method to determine and select active variants of $CynD_{pum}$ predicted to be thermostable by FoldX and all protein preparations in general.

Enzyme activity was evaluated using the picric colorimetric assay following the procedure outlined by Fisher (Fisher and Brown 1952). The solution containing purified enzyme (40 µL) was mixed with 100 mM KCN to achieve a final concentration of 20 mM per 50 µL reaction, buffered in 50 mM Tris-HCl (pH 7.4). The reaction mixture was then incubated at 37 °C for 1 h. Subsequently, a solution containing 40 µL of picric acid-sodium carbonate (50% (w/v) picric acid (Sigma Aldrich) and 50% (w/v) sodium carbonate was added to the reaction mixture. To analyse residual KCN, the reaction solution was heated to 100 °C for 10 minutes, promoting the reaction of the alkaline picrate with any remaining cyanide in the solution, resulting in the formation of a brick-red coloured alkaline picrate complex (Figure 2.3.2). Absorbance was measured at 510 nm using a Thermo Fisher Scientific Multiskan spectrophotometer (Thermo Fisher Scientific, USA). This assay was used to assess the enzymatic activity of $CynD_{pum}$ and its variants.

For the screening and selection of active variants, KCN (20 mM) was added to the crude lysate (100 µL) and treated as described for the purified protein. Inactive variants were not purified and analysed further.

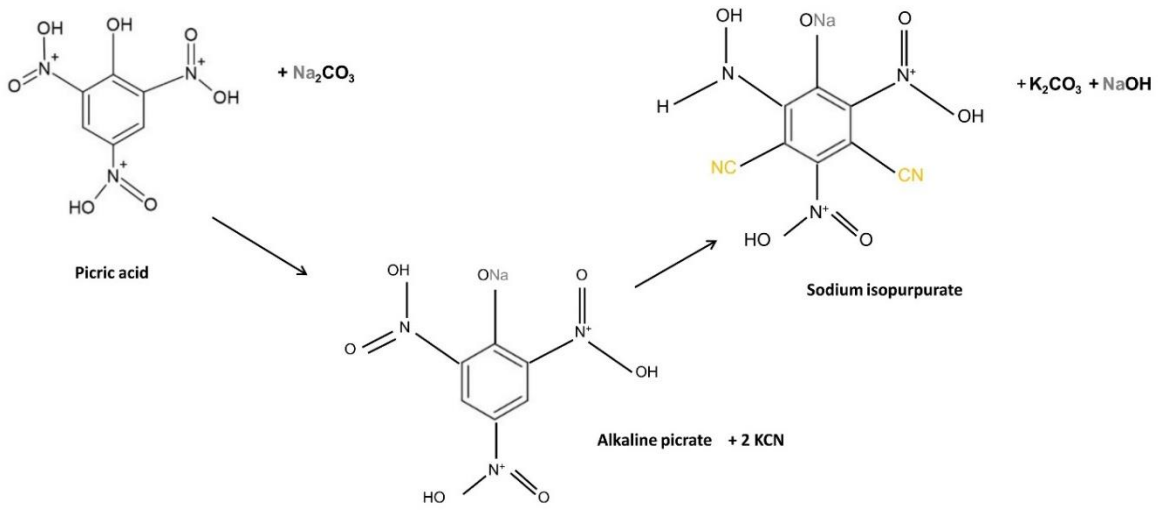


Figure 2.3.1: Colorimetric determination of cyanide ions. The reaction of alkaline picrate with cyanide ions produces sodium isopurpurate, which has a characteristic brick red-brown colour at 100 °C.

Determining enzyme activity at different pH

Purified protein samples were analysed by SDS-PAGE according to the procedure outlined in Section 2.3.6. Subsequently, these samples were subjected to buffer exchange into buffers at pH 5.4 (10 mM citrate, 50 mM NaCl), 8 and 9 (50 mM Tris-HCl, 50 mM NaCl), pH 10.0 (10 mM Na_2CO_3, 50 mM NaCl).

The activity assay, detailed in the preceding section, was conducted for each sample in the respective pH buffers. Residual cyanide levels were quantified spectrophotometrically at 510 nm, with measurements performed in triplicate.

2.3.7.2 Biophysical analysis by nanoscale-differential scanning fluorimetry (nanoDSF)

NanoDSF together with chemical denaturation was employed to compute the Gibb's free energy of protein unfolding and C50 and therefore, thermostability of $CynD_{stu}$, $CynD_{pum}$ and $CynD_{pum}$ variants. This technique allowed for high-throughput analysis of the protein samples with an experimentally determined Gibb's free energy calculation allowing for direct comparison with the *in silico* data. Thermal denaturation was used to determine if a correlation exists between thermostability as determined by chemical and thermal denaturation allowing for comparison with previous studies.

Chemical denaturation

Urea stock buffered in Tris-HCl (10 M urea, 50 mM NaCl, 50 mM Tris-HCl, pH 7.4) was used to prepare 24 dilutions of the denaturant ranging from 7 M to 2 M, with decrements of 0.217 M. Each protein sample (2 mg/mL) was combined with the denaturant at each concentration to achieve a final reaction volume of 30 µL. The individual reactions were incubated for 1 h at 20 °C. The PR ChemControl® software 1.4.1 (NanoTemper Technologies, Germany) was used to

measure the fluorescence signal emitted as a result of intrinsic Trp exposure resulting from the protein unfolding. The experiment was performed in triplicate.

The ΔG of protein unfolding was derived from graphs of the fluorescence intensity ratio FI350/FI330 at different concentrations of urea using the linear extrapolation method (LEM) (Makhatadze 1999, Pace and Shaw 2000). Generally, in this derivation, experimental data points corresponding to changes in the fluorescence signal as a result of chemical denaturation are used to fit a sigmoidal curve using the two-state model for protein unfolding which assumes that the protein transitions reversibly from its folded to unfolded state (Equation 6, Table 2.3.1). The Gibb's free energy of protein unfolding at each denaturant concentration is thus given by Equation 7 (Table 2.3.1). Of interest in the determination of the ΔG of protein unfolding from chemical denaturation data is the ΔG in buffer, i.e. in the absence of denaturant ((ΔG (H_2O) or $\Delta G_{b)}$) where b represents the buffer). This parameter is obtained by extrapolation of the ΔG values in the transition of the sigmoidal curve (red line in Figure 2.1 and Equation 8) to ΔG (H_2O). At this transition the ΔG has been shown to be related to the denaturant concentration by a linear function (Equation 9, Table 2.3.1) for most proteins. The Gibb's free energy of protein unfolding at any denaturant concentration is therefore, given by Equation 9 (Table 2.3.1). Chemical denaturation is monitored using the fluorescence intensity (FI) ratio (FI 350nm/330nm) emitted by intrinsic tryptophan amino acids, that are exposed

when the protein unfolds. This experimental observable (FI) is related to the Gibb's free energy of protein unfolding by Equation 11, where y(unfolded) (C) and y(folded) (C) are the concentration dependencies of the fluorescence intensity for the folded and unfolded states. The concentration at which half the protein particles in solution are folded and the other half unfolded is given by the C50 value which is calculated by Equation 10 (Table 2.3.1).

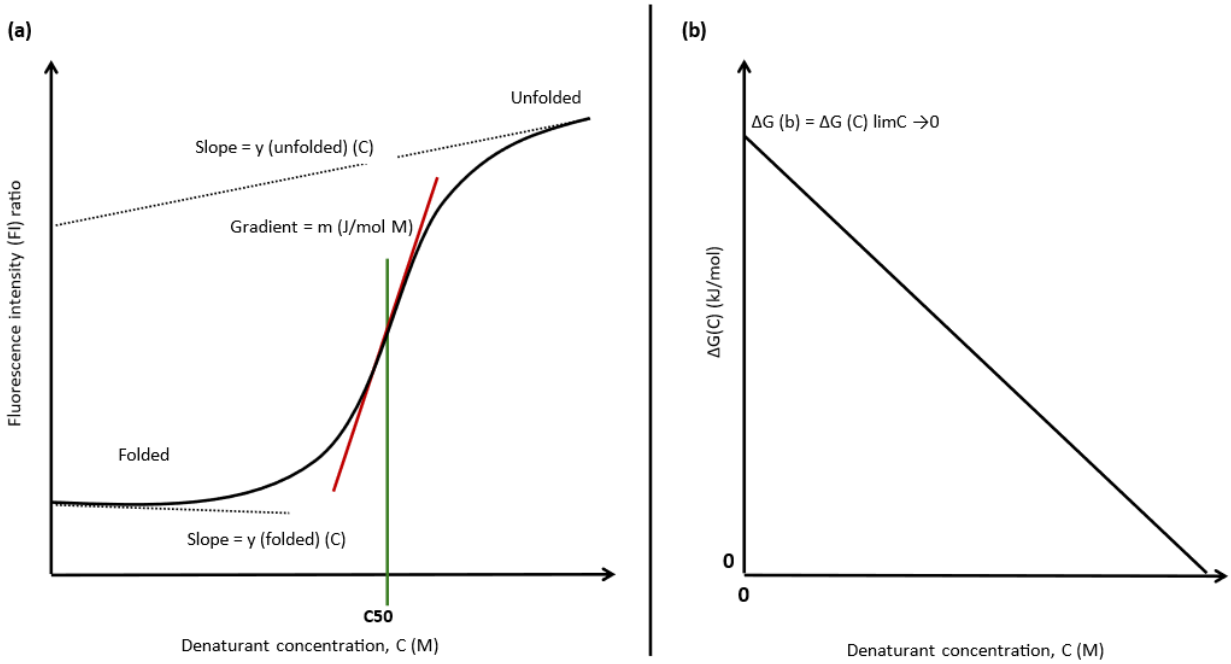


Figure 2.2.2: Illustration of a typical chemical denaturation reaction following the two-state model of protein unfolding. (a) The black sigmoidal curve represents the typical unfolding curve with two-state model fitting. The red line represents the linear line used to extrapolate the ΔG along the transition of the curve to 0 M denaturant concentration which corresponds to ΔG (H_2O) or ΔGb, where b denotes buffer (illustrated in (b)). The green line shows the midpoint of the linear portion of the sigmoidal curve corresponding to the C50 value. (b) The extrapolation to zero denaturant concentration is shown.

Computation of the thermodynamic parameters during experimental analysis was performed on the PR ChemControl® software (NanoTemper Technologies, Germany). The normalised

fluorescence ratio (FI 350nm/330nm) was used to account for fluorescence shifts resulting from Trp only, allowing for a more robust and accurate analysis. The mean and standard deviation of ΔΔG and C50 for each protein sample across three experiments were computed and plotted. The results of both the thermal and chemical denaturation experiments were analysed using the PR. Stability ® analysis software to compare data from the chemical and thermal denaturation experiments of the different protein samples and to generate the statistics for repeat experiments.

Thermal denaturation

Thermal denaturation was used to determine the melting temperature (T_m) of each protein sample and to elucidate the unfolding process induced by thermal denaturation.

An initial discovery scan was conducted at 100% excitation power using protein concentrations ranging from 0.5 to 2 mg/mL of purified protein to determine the optimal protein concentration emitting useable signals for the subsequent thermal and chemical denaturation experiments. The best signal was observed at a concentration of 2 mg/mL.

Purified protein samples were diluted to a concentration of 2 mg/mL in SEC buffer, and 9 µL of each sample was loaded onto standard glass capillaries (Prometheus®, NanoTemper Technologies, Germany). Protein denaturation was performed on the Prometheus® NT.48

instrument using PR.ThermControl® software 2.3.1 (NanoTemper Technologies, Germany), with temperatures ranging from 20 °C to 90 °C at a ramp rate of 2 °C/min and 100% excitation power. The emitted fluorescence signal was recorded at wavelengths of 350 nm and 330 nm, and the fluorescence intensity ratio, FI350 / FI330 ratio was calculated.

2.3.8 Negative stain transmission electron microscopy (NSTEM)

Negative stain TEM was used to assess protein stability in the buffer, the quality of the protein preparations and to visualise the oligomeric states of $CynD_{pum}$ and its variants.

Sample preparation

The purified protein samples were diluted to 0.1, 0.3, 0.5, and 1 mg/mL in SEC buffer to determine the protein concentration at which the best abundance and separation between protein particles could be achieved.

Carbon-coated grids (made at the Electron Microscope Unit (EMU), UCT) were glow-discharged at a current of 25 mA for 30 s on an EMS100 glow discharger (Electron Microscope Sciences, USA). Five microlitres of the protein sample (0.3 mg/mL) was applied onto a continuous carbon grid and incubated for 30 s. The sample was blotted on the edge of blotting paper (Whatman qualitative filter paper). The sample side of the grid was then stained in two drops of 15 µL of 2% uranyl acetate (UA), with blotting on filter paper between staining. The grids were air-dried and stored for imaging.

The grids were imaged using a Tecnai T20 transmission electron microscope (FEI, USA). Images were taken at 29 000x magnification using a Gatan US2000 CCD camera with sampling at 0.312 Å/pixel.

2.4 Results and discussion

2.4.1 Interface analysis and interpretation of the refined site saturation mutagenesis reference model: $CynD_{pum}$ Q86R/H203K/H203K/H223K

A truncated model of the 3.15 Å structure of $CynD_{pum}$ Q86R/H203K/H203K/H223K variant (PDB ID: 8c5i) (Mulelu 2017, Sewell, Frangakis et al. 2017) consisting of 12 of the possible 18 monomers was used as the reference model for site saturation mutagenesis (SSM) to approximate the effect of substituting amino acids on the stability of $CynD_{pum}$.

Before proceeding with the SSM experiment, we refined and analysed the model. The objective was to understand the specific interactions that stabilise the structure, thus ensuring accurate interpretation of the SSM experimental results. In addition, this provided insight into interactions stabilising the enzyme, potentially leading to increased thermostability and informed the selection of predicted stabilising mutations for experimental validation.

We begin by identifying secondary structure elements involved in interfacial contacts and analysing the chemical interactions stabilising them. The interfacial regions are potential sites for engineering thermostability in $CynD_{pum}$, since the enzyme is formed and stabilised by interactions occurring across them. We further provide an alternative interpretation of the variant's observed thermostability (Wang, Watermeyer et al. 2012, Mulelu 2017), aided by

the refined model of the solved atomic resolution structure. From this interpretation, we gained further insight into the thermostability of $CynD_{pum}$.

Table 2.4.1: Summary of interaction energies of monomers related by typical interfaces A-, C-, D- on the $CynD_{pum}$ Q86R/H305K/H308K/H323K atomic model. The C-interface is formed by four monomers related to one another along the dyad axis at the C-interface. Therefore, interactions at this interface are measured for both pairs of monomers across the axis. The electrostatic kon term refers to the electrostatic interactions between the two interacting molecules prior to complex formation.

Parameter	**A-interface (Group 1 K) (Group 2 -M)**	**C-interface (Group 1 I) (Group 2 -K)**	**C-interface (Group 1- I) (Group 2 -K)**	**(Group 1 – B) (Group 2- K)**
Interaction energy (kcal/mol)	**-42.12**	**-20.84**	**-18.72**	**0.16**
Backbone Hbond	-8.77	-7.09	-7.24	-2.84e-14
Sidechain Hbond	-7.48	-13.08	-12.92	-2.84e-14
Van der Waals	-37.19	-29.28	-26.90	-1.31
Electrostatics	-0.02	-3.088	-0.82	-0.22
Solvation Polar	33.79	38.55	35.33	2.19
Solvation hydrophobic	-54.90	-38.44	-35.47	-2.051
Helix dipole	-0.11	-0.57	-0.50	-0.04
Electrostatic kon	-0.12	-0.84	-0.69	-0.16
Energy Ionisation	0.0013	0.0048	0.006	8.89e-16
Interface Residues	88.0	71.0	71.0	9.0

The A-interface

The A-interface is formed through non-bonded interactions involving the side chains of amino acids on helices α6 and α7 from each monomer across the interface, along with the side chain and backbone atoms of loops L10 and L23 (Figure 2.4.1 (a and b)) of the C-terminal tail.

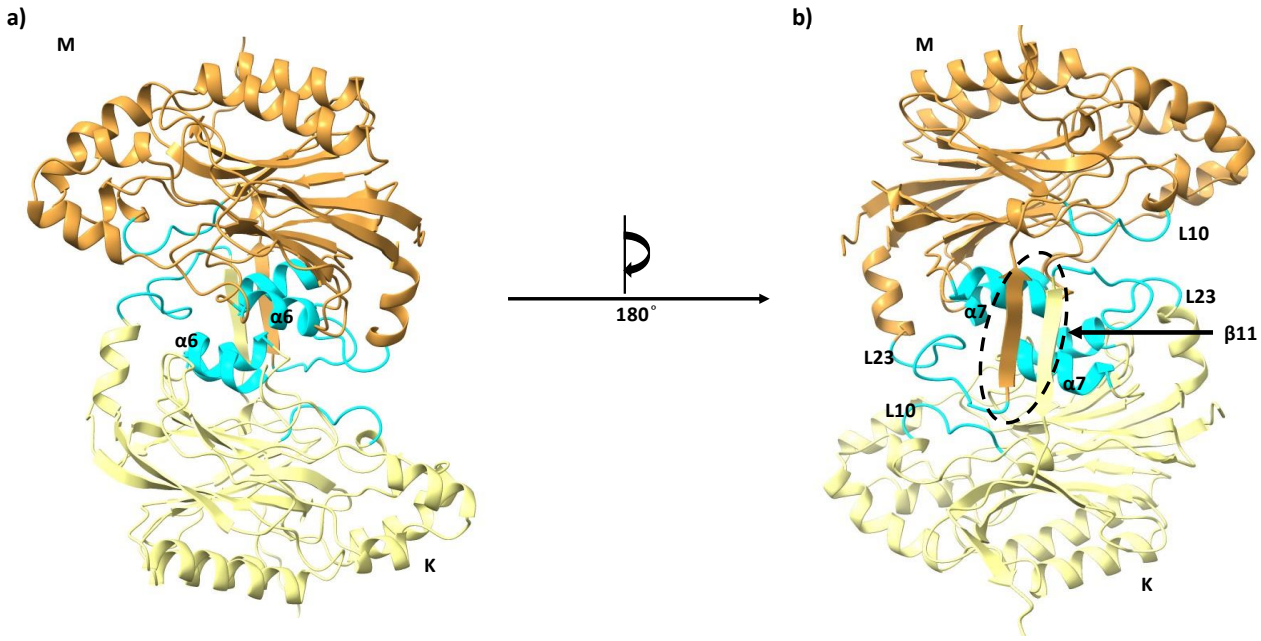


Figure 2.4.1: Secondary structure elements defining the A-interface. (a) Front-view of monomers related across the A-interface. Interactions across the A-interface occur between helices α6 and α7 and loops L10 and L23 of each chain (K and M) to form the dimer subunit. These secondary structure elements are coloured in cyan. (b) Interior -view of monomers related across the A-interface. Two β-strands of the C-terminal tail (dashed outline) stabilise the interface.

Analysis of the protein-protein binding energy at this interface suggests that the formation of this interface is highly thermodynamically favourable. The A-interface is formed primarily through hydrophobic and van der Waals interactions (Figure 2.4.2). Hydrogen bonding and electrostatic interactions play a lesser role in stabilising this interface. Interactions stabilising

the A-interface across the two α6 helices (SSRYYAIAT) are predominantly hydrophobic, with polar and charged residue side chains interacting across the interface via apparent van der

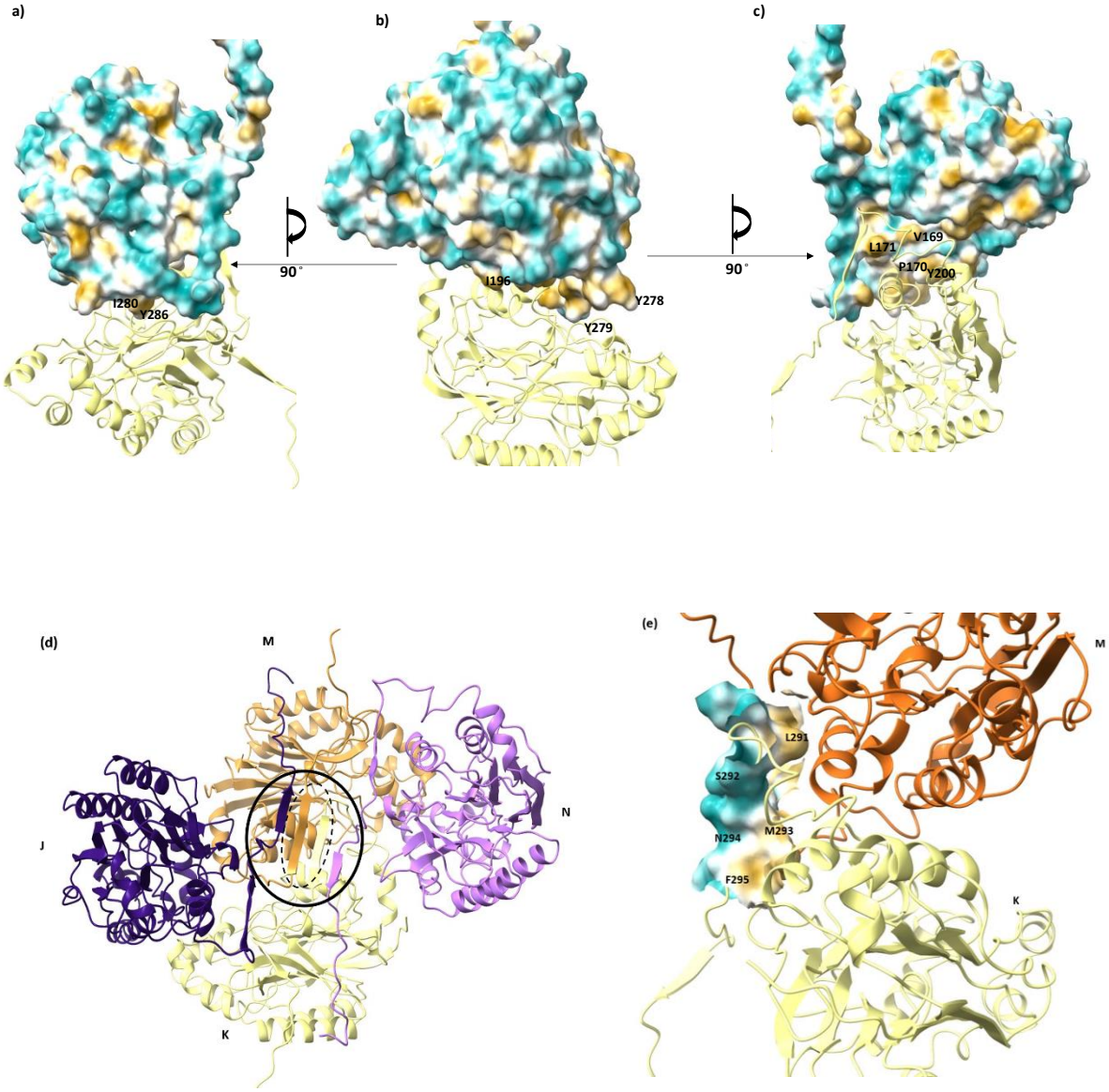


Figure 2.4.2: Interactions stabilising the A-interface. (a), (b), and (c): The A-interface is mainly stabilised by a network of hydrophobic interactions with aliphatic and aromatic amino acids (V169, P170, L171, Y200) located at the core where the two monomers meet (d) The C-terminal tail region contributes to the stabilisation of this interface primarily through interactions facilitated by amino acids on the β11 strands (oval with dashed outline) of the C-terminal β-sheet (circle with solid outline) formed by the two monomers (chains K and M) constituting the subunit and two other monomers from adjacent subunits (N and J).(e) The C-terminal tail stabilises the interface through hydrophobic interactions from the β11 strands of the two monomers related across the interface. (e) Their aliphatic

side chains project towards the A-interface while polar sidechains project towards the central hollow core of the enzyme assembly.

The longer α7 helices consist mostly of non-polar hydrophobic amino acids (170PLDLMAMNA178). The side chains of these amino acids project towards the central core of the enzyme, forming a hydrophobic region with the two β-strands contributed by the C-terminal tails and the α6 helix (Figure 2.4.2 (d and e)). This contributes significantly to increased hydrophobicity in this and therefore, core packing in this region. Additionally, electrostatic contributions to stability at the A-interface occur between the nucleophilic loop L10 (130KMRASVA136) and predominantly electrophilic loop L23 (278YYIDPAGHYSNQS290) from each monomer (Figure 2.4.3).

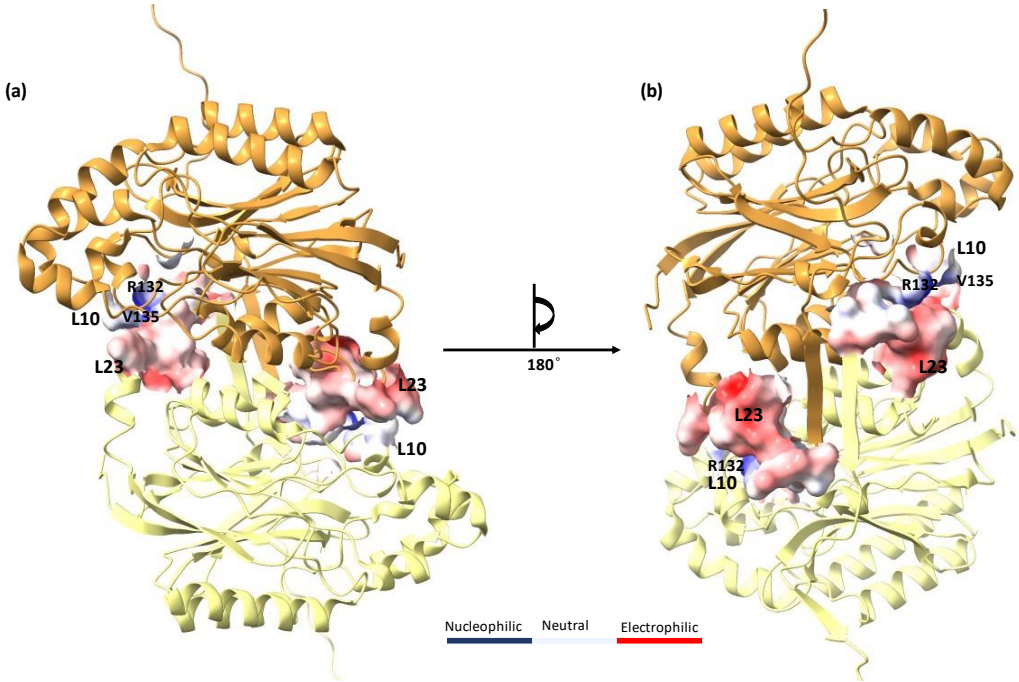


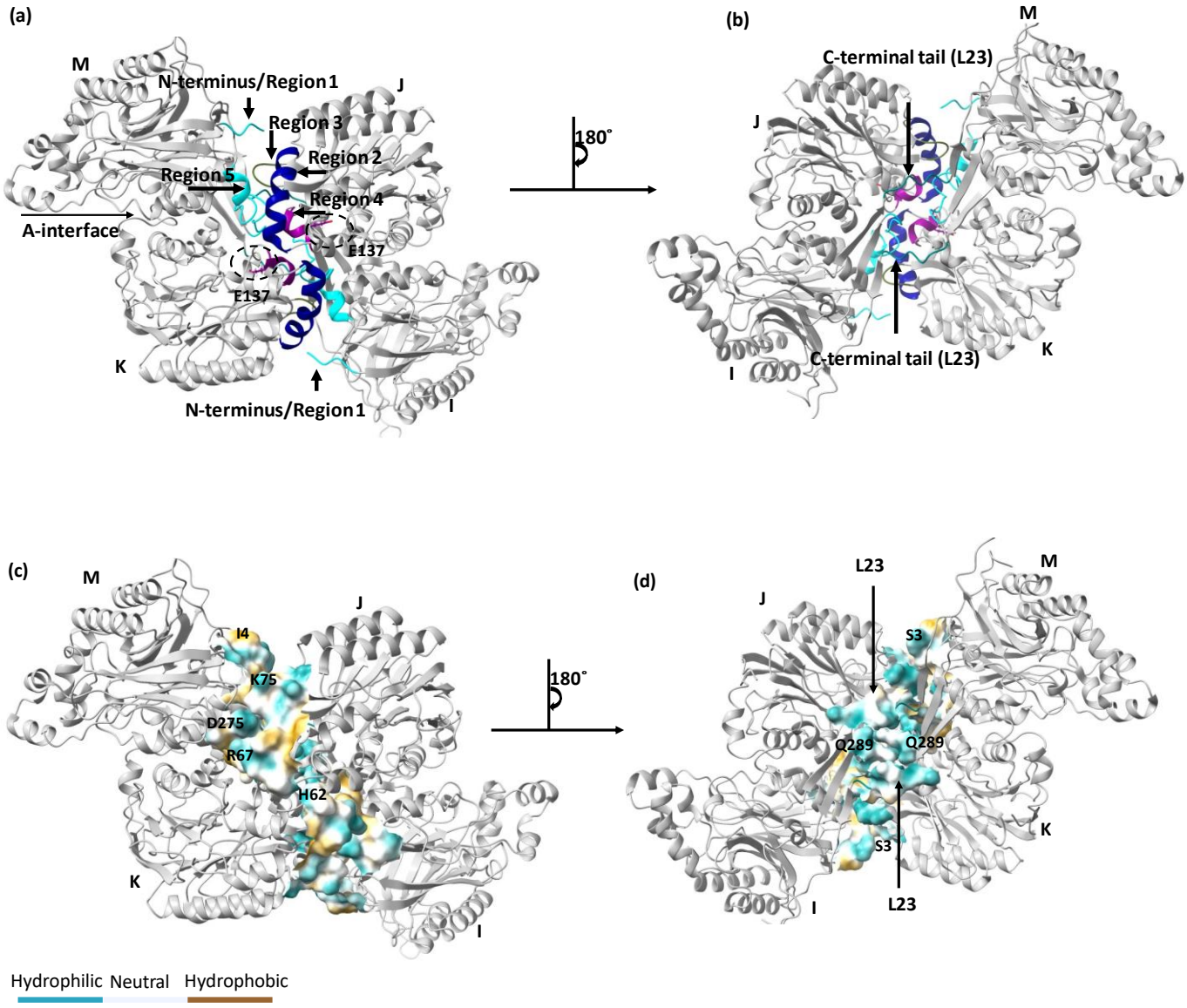
Figure 2.4.3: Interactions stabilising the A-interface across loops L10 and L23. (a) and (b) Exterior and interior view of the A-interface, respectively. Weak electrostatic interactions between main chain and sidechain amino acids amino acids of loops L10 and L23 contribute to stabilisation of monomers associated across the A-interface. The contribution of polar amino acids to the electrostatic nature of the loops is through the partial charges and resonance on mainchain atoms.

The C-interface

The C-interface is defined by five distinct regions (Figure 2.4.4 (a and b)). The first region is the N-terminal loop (amino acids 3-5). Region 2 is formed by helix α3 (amino acids 62-75) and region 3 by a three-amino-acid β-hairpin loop (amino acids 107-110) located between β-strands β4 and β5. The fourth region comprises amino acids 135-145 and is located on helix α5, region 5 (amino acids 271-290) is composed of helix α11 and an adjacent loop near the C-terminal tail.

The interface is stabilised by predominantly electrostatic and hydrophobic interactions (Figure 2.4.4 (c and d)). The interactions are increasingly hydrophobic where the C-interface meets the A-interface (e). The portion of the interface facing the hollow core of the enzyme is composed of mainly electrostatic amino acids (276YYIDPAGHYSNQS290) which interact with amino acids of the two C-terminal tail L23 loops, (Figure 2.4.4 (c) and (d)). The C-terminal loops further contribute weak hydrophobic interactions near the A-interface mainly occurring between V302, V303 and A283 and G284 of region 5 (not shown). The E137 (dashed outline (a)) of the catalytic tetrad is located on the helix defining region 4 of the interface. Because

the E137 side chain projects towards the core where the active site is located, it does not form any interactions across the interface. However, neighbouring amino acids A136 and R138 do.



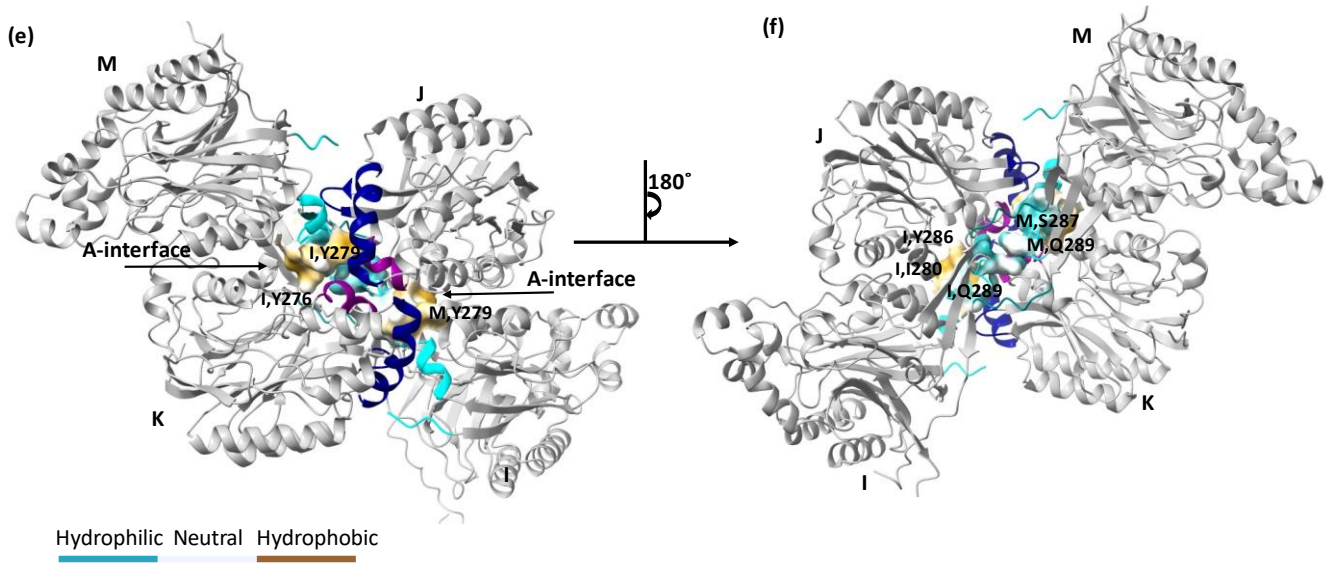


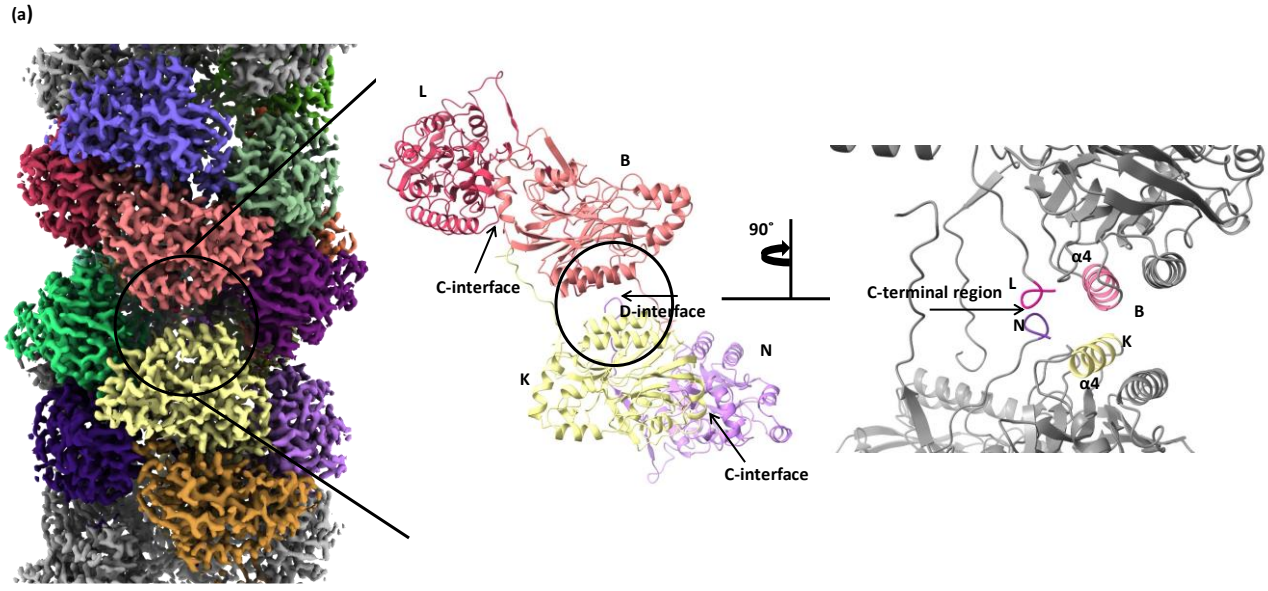
Figure 2.4.4: Regions defining the C-interface and the interactions stabilising it. (a), (c) and (e) show the exterior view, while (b), (d), and (f) display the interior view of the enzyme at the C-interface. (a) and (b) The C-interface is formed by interactions between two subunits (four monomers) positioned along the central axis. Five regions comprise amino acids involved in chemical interactions across the interface. Identical regions making up the C-interface and present on different monomers are depicted in the same colour. Regions 1 and 5= cyan, regions 2= blue, regions 3=olive green, and region 4=purple. These secondary structure elements are mirrored across the dyad axis at the A-interface, located on both subunits forming the interface. E137 of the catalytic tetrad is located on region 4 of the C-interface, projecting towards the core of the enzyme. (c) and (d) The general interactions (electrostatic and hydrophobic) stabilising the interface are shown along the length of the interface. Some amino acids contributing to stabilising the interface are labelled. (d) The loops (L23) of the C-terminal tails (teal (b)) of two monomers, positioned perpendicular to the C-interface, stabilise the interface mainly through electrostatic interactions. (e) and (f) Both loops stabilise the A-interface through hydrophobic interactions where the C-interface meets the A-interface.

The D-interface

Analysis of protein-protein binding between monomers related across to one another across the D-interface revealed a net positive Gibb's free energy value, indicating that interactions across the groove are thermodynamically unfavourable.

Our analysis of the Q86R mutation shown by Wang, Watermeyer et al. (2012) to result in the formation of long thermostable fibres reveals that new stabilising interactions at this

interface occur with amino acids of the C-terminal tail rather than across the groove. These interactions occur between a monomer associated across the C-interface with the one of the monomers associated across the groove and R86 on helix α4 of the opposite monomer (Figure 2.4.5 (a)). Although the guanidium side chain of arginine is generally basic, it possesses three methylene groups, that are in the current model positioned such that they create weak hydrophobic contacts with the aliphatic sidechain of I319 of the C-terminal tail (Figure 2.4.5 (b)). The guanidium side chain maintains a net positive charge at pH 5.4. Through resonance, the nitrogen atoms of the side chain of R86 form van der Waals and electrostatic interactions with the lone pair of electrons on the oxygen atoms of the carbonyl groups of amino acids Y316 and I319, and the side chain carboxyl group of E317 (Figure 2.4.5 (c)), thereby forming a network of new short-range interactions.



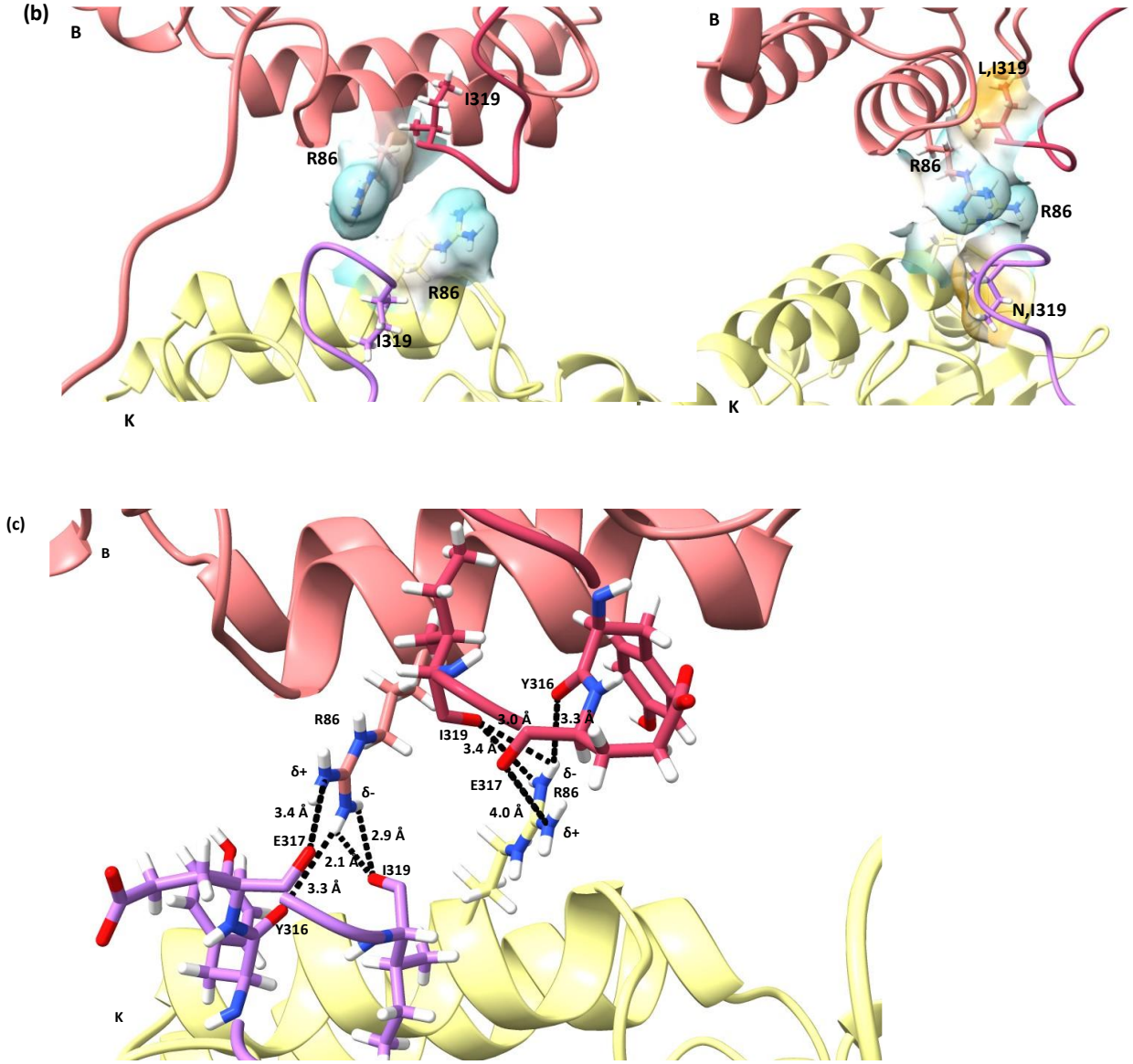


Figure 2.4.5: The D-interface and basis for observed thermostability in the Q86R/H305K/H308K/ H323K variant of CynD$_{pum}$. (a) Individual monomers are coloured differently to show the locations and contributions of amino acids participating in stabilising the interface. Helices α4 of each monomer across the interface (monomers B and K) interact with the C-terminal tail of monomers related to them across the C-interface (monomers L and N). ((b) and (c)) The three methylene groups on R86 are positioned such that R86 can form hydrophobic interactions with the sidechain of I319 and electrostatic interactions with the lone pair of electrons on the carbonyl oxygen of the same amino acid which is located at the C-terminal tail. (c) The R86 sidechain forms multiple interactions with the sidechains of E317 also located at the C-terminal tail.

The wild-type residue Q86 (PDB ID 8p4i) has a shorter side chain, which restricts its ability to reach into the groove and interact with amino acids at the C-terminal tail. Additionally, the Q86 side chain contains only one amino group, resulting in a lower net positive charge compared to R86. As a result, it establishes fewer ionic and hydrogen bonding interactions with the Q86 amino acid across the groove (Figure 2.4.6).

Conversely, the introduction of R86 facilitates multiple interactions with the C-terminal region, reinforcing the homo-oligomer and potentially contributing to the formation of the observed long, thermostable structures. Although the remaining 11 amino acids are not resolved in the current structure, we hypothesise that the C-terminal tail region traverses the α4 helices of the monomers related through the D-interface.

The C-terminal tail traverses all interfacial regions across the core of the enzyme, stabilising them through specific interactions at each interface as described above.

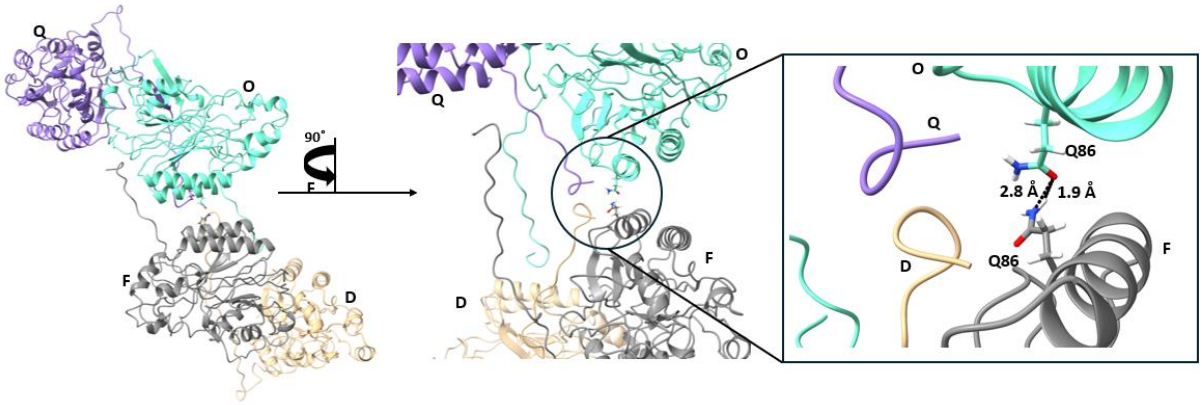


Figure 2.4.6: The Q86 amino acid at the D-interface. The Q86 at this region forms fewer interactions (potential ionic and hydrogen bonding interactions) with the Q86 amino acid across the groove.

2.4.2 Structure-guided site saturation mutagenesis

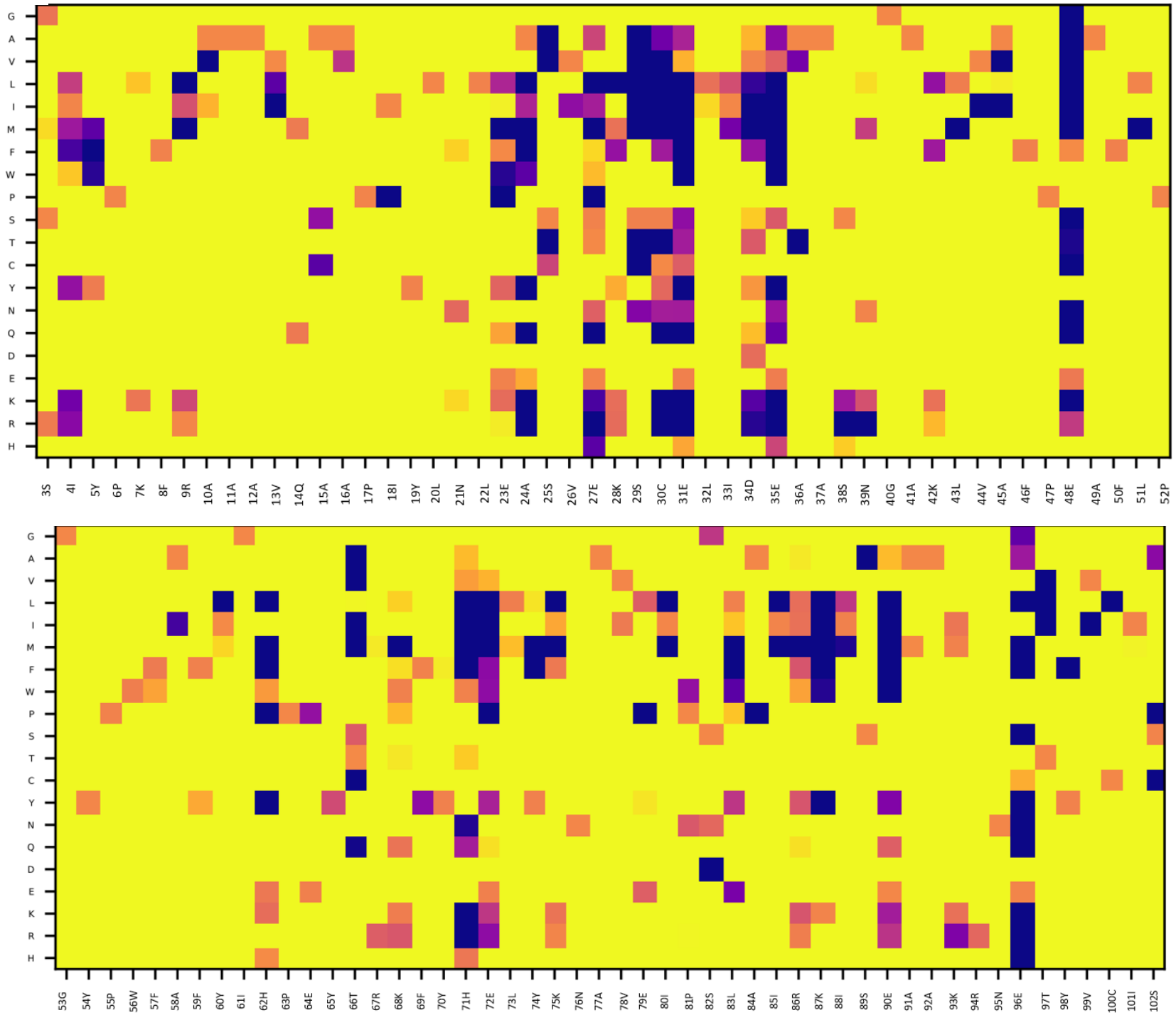
In this study, Gibb's free energy change (ΔG) associated with protein unfolding served as a metric for assessing protein thermostability. The quaternary structure of proteins is governed by precise interactions between amino acids and with the surrounding solvent, culminating in the formation of a stable three-dimensional structure. During protein folding, the most stable conformation corresponds to the state of the lowest energy, where ΔG between amino acids and the solvent is at its minimum. The ΔG of protein unfolding in the *in silico* SSM experiment includes entropic contributions brought about by amino acid main- and side chain clashes, and the contributions of interactions between amino acids, including van der Waals, electrostatic, and hydrophobic interactions. The parameters considered during the *in silico* computation of ΔG are listed in Table 2.3.2.

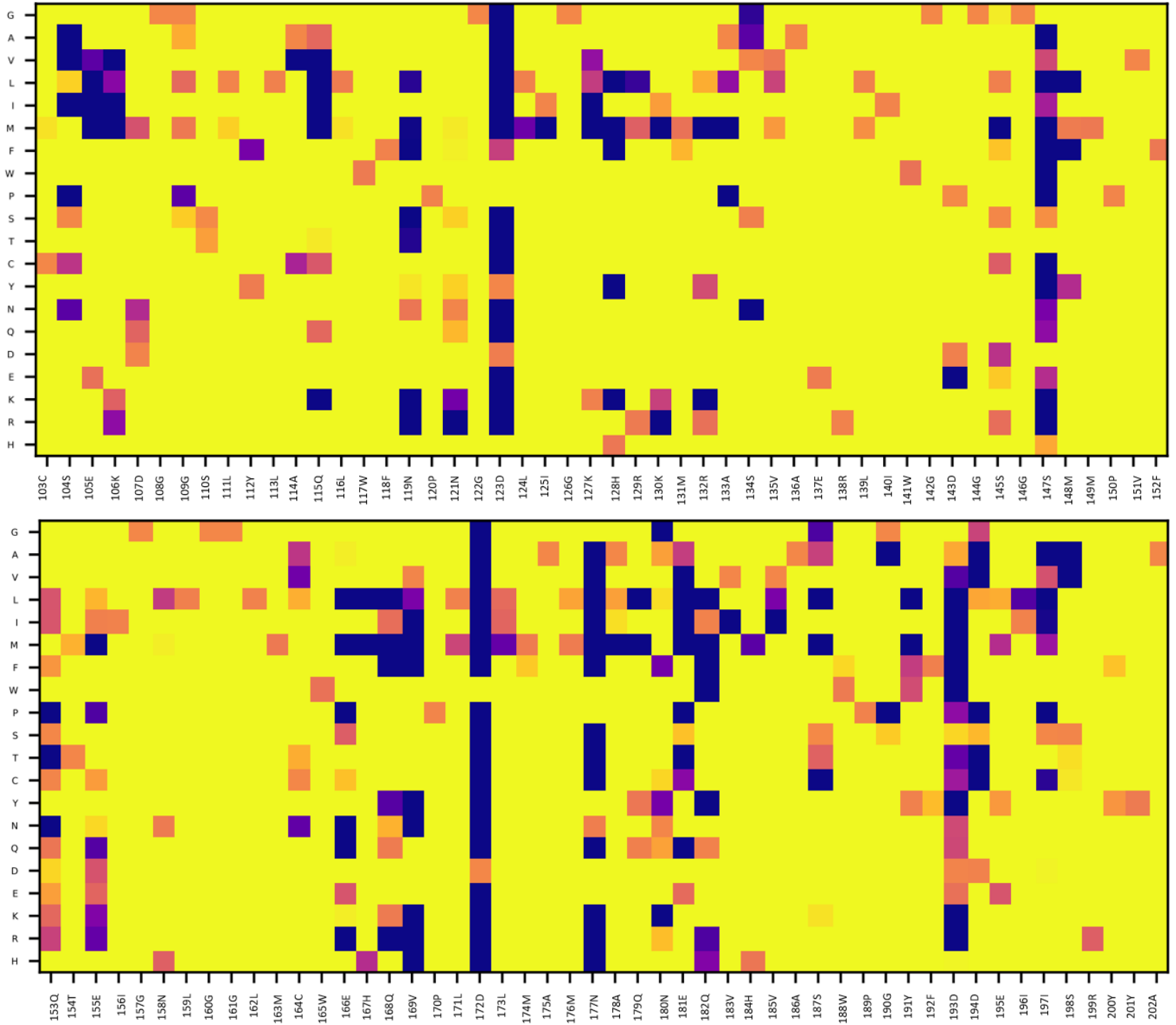
Variants with favourable (ΔG <0) would require more thermal energy to break the bonds and molecular interactions that form the overall structure of the enzyme and would therefore be more thermostable than the wild-type enzyme. The magnitude of ΔG observed between the wild-type and variant amino acids was used to determine the thermostability of $CynD_{pum}$ variants relative to wild-type $CynD_{pum}$. Negative values are indicative of thermodynamic favourability and, therefore, thermostability.

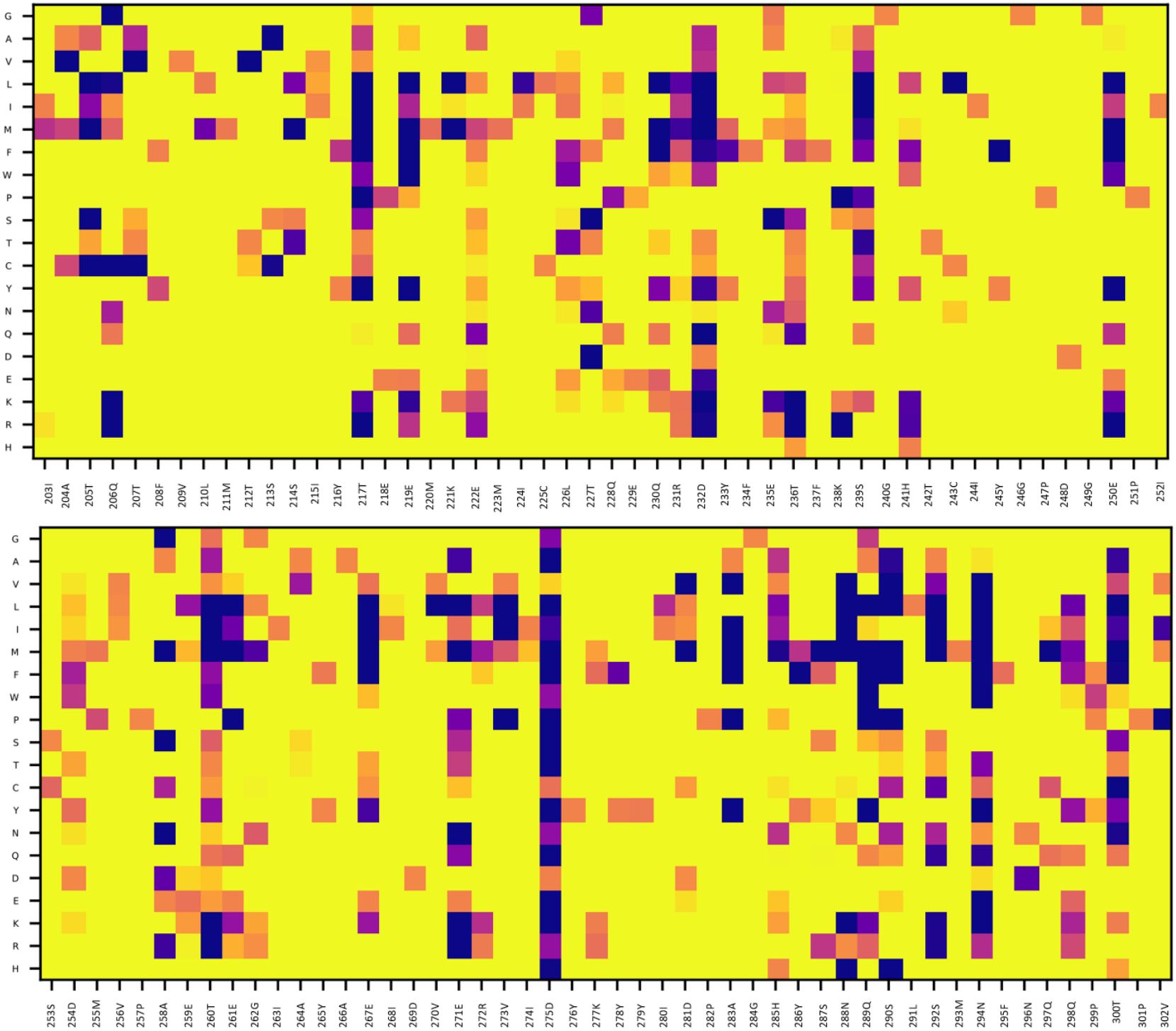
2.4.2.1 *In silico* site saturation mutagenesis and computation of Gibb's free energy change (ΔG) of protein unfolding

The *in silico* SSM data are illustrated using heat maps depicting the (ΔG) resulting from each amino acid substitution. The degree of thermostability is represented by a colour gradient ranging from blue to yellow, where blue corresponds to the most stable mutations and yellow indicates the least stable mutations (Figure 2.4.7 (a)). For ease of data presentation, data within the range of -3 and 3 are shown. A general overview of these data on the $CynD_{pum}$ model is illustrated on Figure 2.4.7 (b).

(a)


G
A
V
L
I
M
F
W
P
S
T
C
Y
N
Q
D
E
K
R
H
3S 4I 5Y 6P 7K 8F 9R 10A 11A 12A 13V 14Q 15A 16A 17P 18I 19Y 20L 21N 22L 23E 24A 25S 26V 27E 28K 29S 30C 31E 32L 33I 34D 35E 36A 37A 38S 39N 40G 41A 42K 43L 44V 45A 46F 47P 48E 49A 50F 51L 52P
53G 54Y 55P 56W 57F 58A 59F 60Y 61I 62H 63P 64E 65Y 66T 67R 68K 69F 70Y 71H 72E 73L 74Y 75K 76N 77A 78V 79E 80I 81P 82S 83L 84A 85I 86R 87K 88I 89S 90E 91A 92A 93K 94R 95N 96E 97T 98Y 99V 100C 101I 102S


G
A
V
L
I
M
F
W
P
S
T
C
Y
N
Q
D
E
K
R
H
103C 104S 105E 106K 107D 108G 109G 110S 111L 112Y 113L 114A 115Q 116L 117W 118F 119N 120P 121N 122G 123D 124L 125I 126G 127K 128H 129R 130K 131M 132R 133A 134S 135V 136A 137E 138R 139L 140I 141W 142G 143D 144G 145S 146G 147S 148M 149M 150P 151V 152F
153Q 154T 155E 156I 157G 158N 159L 160G 161G 162L 163M 164C 165W 166E 167H 168Q 169V 170P 171L 172D 173L 174M 175A 176M 177N 178A 179Q 180N 181E 182Q 183V 184H 185V 186A 187S 188W 189P 190G 191Y 192F 193D 194D 195E 196I 197I 198S 199R 200Y 201Y 202A


G
A
V
L
I
M
F
W
P
S
T
C
Y
N
Q
D
E
K
R
H
203I 204A 205T 206Q 207T 208F 209V 210L 211M 212T 213S 214S 215I 216Y 217T 218E 219E 220M 221K 222E 223M 224I 225C 226L 227T 228Q 229E 230Q 231R 232D 233Y 234F 235E 236T 237F 238K 239S 240G 241H 242T 243C 244I 245Y 246G 247P 248D 249G 250E 251P 252I
253S 254D 255M 256V 257P 258A 259E 260T 261E 262G 263I 264A 265Y 266A 267E 268I 269D 270V 271E 272R 273V 274I 275D 276Y 277K 278Y 279Y 280I 281D 282P 283A 284G 285H 286Y 287S 288N 289Q 290S 291L 292S 293M 294N 295F 296N 297Q 298Q 299P 300T 301P 302V

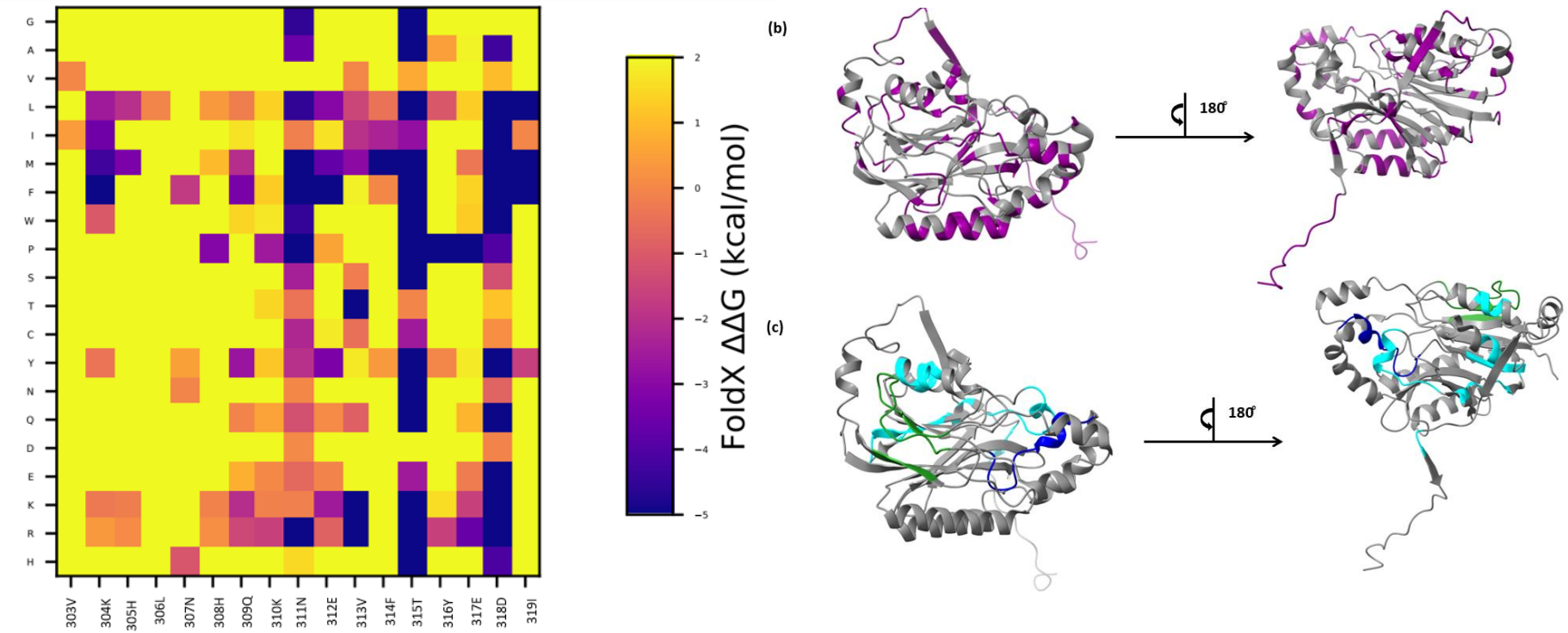


Figure 2.4.7: Final ΔΔG averages for substitution at each position on the available sequence space of the Q86R/H305K/H308K/H323K $CynD_{pum}$ variant of $CynD_{pum}$ (amino acids 3-319) with the remaining 19 standard amino acids. (a) Each residue (bottom panel) was substituted with each of the 20 standard amino acids including itself (left panel) and the ΔΔG of unfolding resulting from the substitution was computed. (b) Mutation 'hotspots' predicted by FoldX to be stabilising shown in purple occurring across all interfaces and the core but predominantly on the d-surface and c-terminal tail. (c) Regions with low numbers of predicted mutations are grouped into 3; regions involving the A-interface (cyan), located at the core (blue) and regions located at the surface (green).

Stabilising mutations are predicted to occur across the entire protein with more frequent mutations predicted on helices α1 (Q23-E35) and α4 (K88-I88) of the D-interface and the C-terminal tail region (β11 (S292-N294), loop L23 (A283-S290), loop L24 (Q298 and T300), and β12 (K304)). Fewer stabilising mutations were predicted for the C-interface, surface, and core of the protein (Figure 2.4.7(a)).

2.4.3 Experimental screening and validation of thermostability

To ascertain the dependability of the *in-silico* analyses, the catalytic activity of the variants was determined using the picric calorimetric assay, and their thermostability was assessed experimentally using nanoDSF together with chemical denaturation.

Ninety-seven substitution mutations anticipated to exhibit favourable outcomes in the in silico SSM experiment were selected based on structural data, with others chosen from the surface, core, and groove regions due to insufficient structural information regarding the effects of mutations at these regions. These mutations were subsequently introduced into the coding region of the expression plasmid encoding CynDpum using high-throughput site-directed mutagenesis.

Out of the 43 experimental samples, 24 were chosen based on their success in high-throughput PCR amplification. Poor PCRs resulting from the high GC content of the *cyndpum*

gene insert were optimised by adding 4% v/v DMSO to the reaction mixture and adjusting the annealing temperature. The samples that did not show improvement after this optimisation step were excluded from further analyses. The remaining 24 predicted mutations were expressed and assessed for enzyme activity. The active variants (20) were expressed recombinantly and their thermostability was evaluated through the measurement of experimentally determined ΔΔG. The quaternary structure of the active variants was also visualised to analyse any changes to the quaternary structure of the enzyme brought about by introduction of the mutation using negative-stain EM. In addition to the variants identified in this study, cyanide dihydratase from *Pseudomonas stutzeri* AK 61 ($CynD_{stu}$), the $CynD_{pum}$-$CynD_{stu}$ chimeric protein (both discussed in more detail in Chapter IV), and the $CynD_{pum}$ variant (Q86R/H305K/H308K/H323K) were also included in the analyses.

2.4.3.1 Screening

Recombinant protein expression, purification and characterisation by activity analysis and negative stain TEM

Eight of the 22 single mutation variants were inactive, as shown in Figure 2.4.8. The inactive variants are I4M, H71R, E75P, Q86K, E96G, N180K, D172N, and N177L. Mutations D172 and N177L are located at the A-interface, mutations H71R, E75P, and N180K are located at and

near the C-interface, and mutations E96G and Q86K at the D-interface, with I4M occurring at the N-terminus.

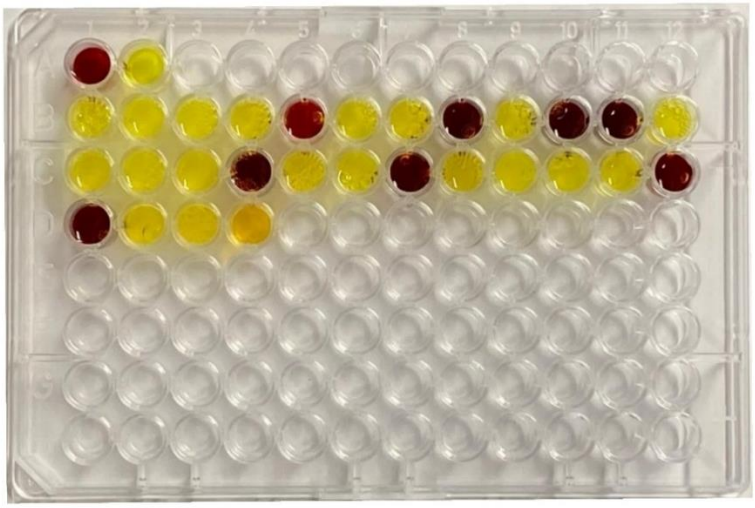


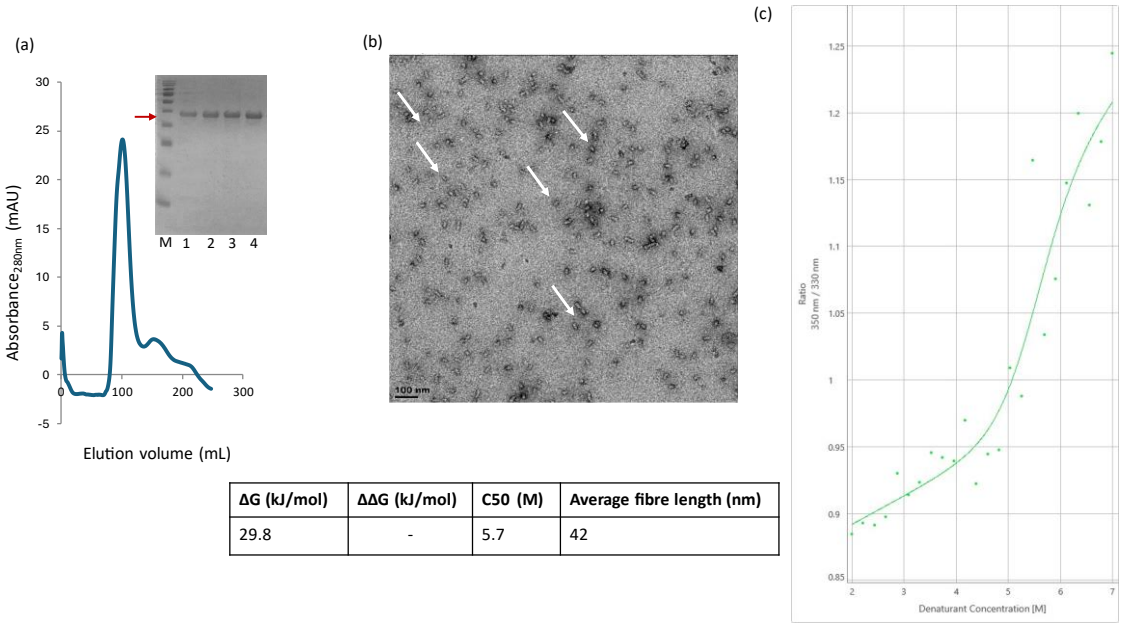
Figure 2.4.8: Screening plate for enzyme activity of CynDpum variants using the picric acid calorimetric assay. The presence of cyanide ions is indicated by brick red/brown wells, while yellow wells indicate absence of cyanide ions, indicating the conversion of cyanide to ammonia and formic acid. Wells A1 and A2 serve as the positive and negative control, respectively. Wells B1-B12 correspond to; wildtype $CynD_{pum}$, $CynD_{stu}$, $CynD_{pum}$ -$CynD_{stu}$ chimera, $CynD_{pum}$ (Q86R/H305K/H308K/ H323K), I4M, S29A, H62Y, H71R, K75L, E79P, Q86K and $CynD_{pum}$-$CynD_{stu}$ chimera+ Q86R. Wells C1-C12 correspond to $CynD_{pum}$ variants; S102C, Q115C, N119R, D172N, S147I, E155R, N177L, T260I, Q86M, E90K, E319L, E96G. Wells D1-D4 correspond to variants; N180K, T217I, $CynD_{pum}$ S29A/Q86R/N119R and I319L.

2.4.3.2 Biophysical analyses

The experimental ΔG was determined using chemical denaturation and nanoDSF. Protein unfolding curves were fitted using the two-state model of protein unfolding. The ΔG was derived from chemical denaturation curves using the linear extrapolation method (LEM) as described in section 2.3.7.2. The ΔΔG for each variant was calculated as the difference in ΔG

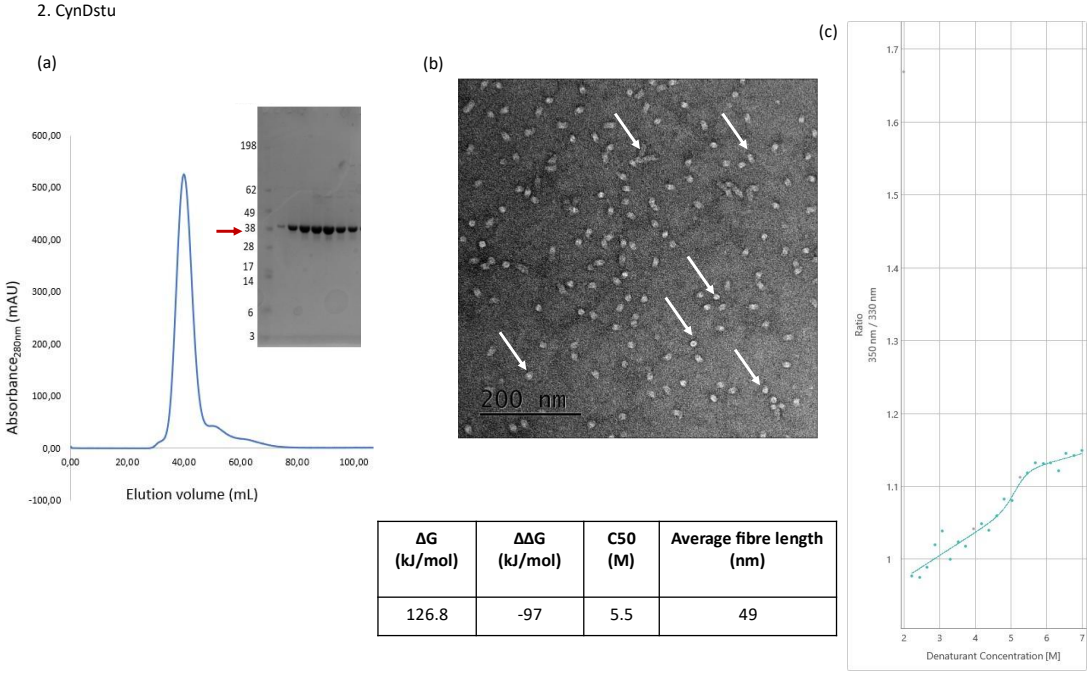
between the wildtype and each variant (Equation 12). These data are shown for each sample in Figure 2.4.9 and are summarised in Table 2.1.

1. CynDpum



ΔG (kJ/mol)	ΔΔG (kJ/mol)	C50 (M)	Average fibre length (nm)
29.8	-	5.7	42

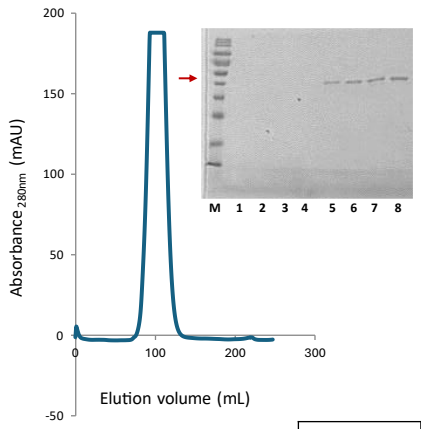
2. CynDstu



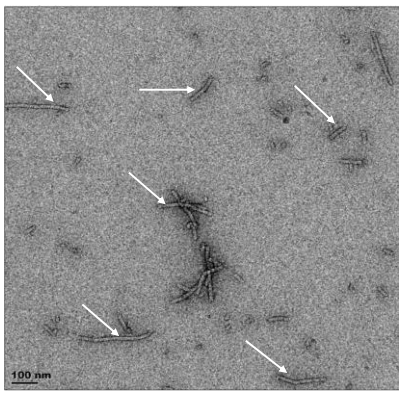
ΔG (kJ/mol)	ΔΔG (kJ/mol)	C50 (M)	Average fibre length (nm)
126.8	-97	5.5	49

3. CynDpum-CynDstu chimera

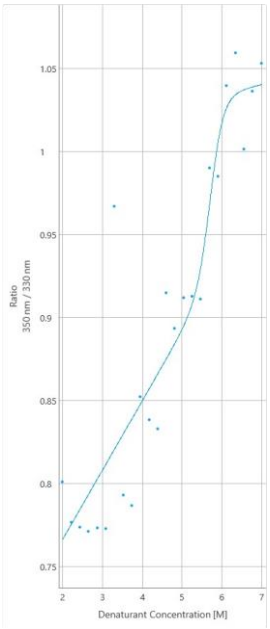
(a)



(b)



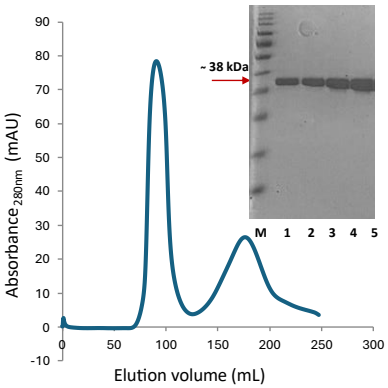
(c)



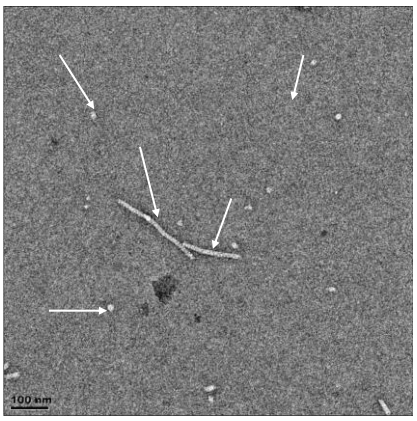
ΔG (kJ/mol)	ΔΔG (kJ/mol)	C50 (M)	Average length (nm)
46.6	-16.8	6.4	200

4. CynDpum Q86R/H305K/H308K/H323K variant

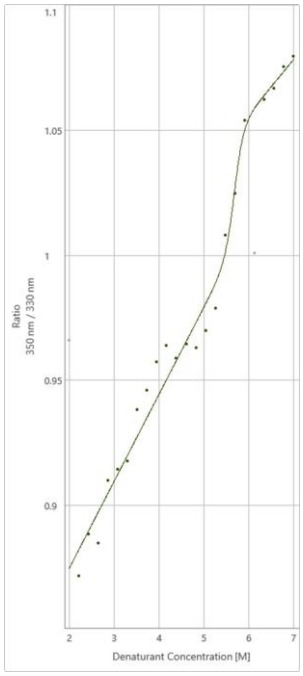
(a)



(b)



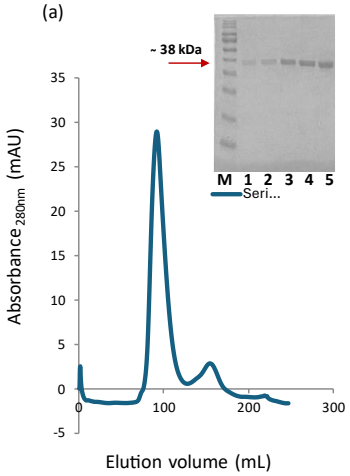
(c)



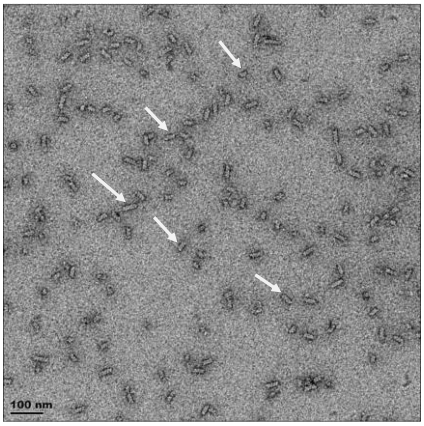
ΔG (kJ/mol)	ΔΔG (kJ/mol)	C50 (M)	Average length (nm)
137.7	- 104.9	5.7	178

5. S29A

(a)

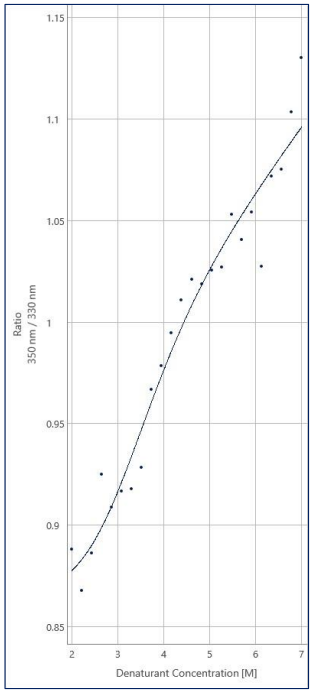


(b)



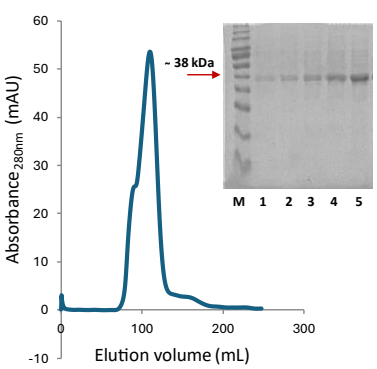
ΔG (kJ/mol)	ΔΔG (kJ/mol)	C50 (M)	Average length (nm)
35.1	-5.3	3.4	63

(c)

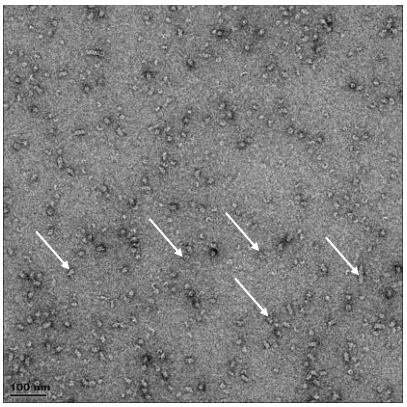


6. H62Y

(a)

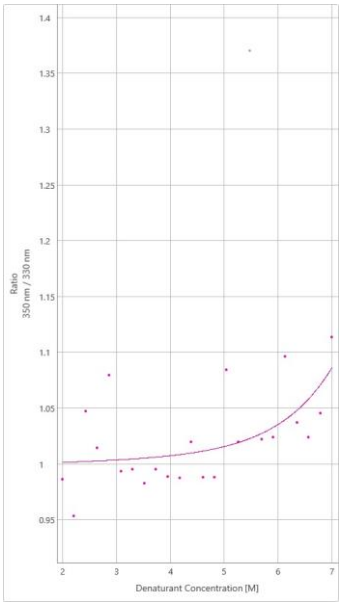


(b)



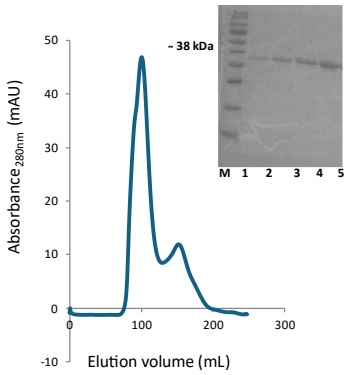
ΔG (kJ/mol)	ΔΔG (kJ/mol)	C50 (M)	Average length (nm)
-	-	-	41

(c)

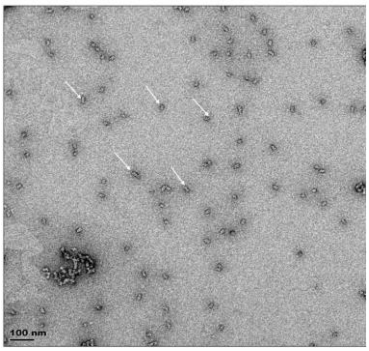


7. S102C

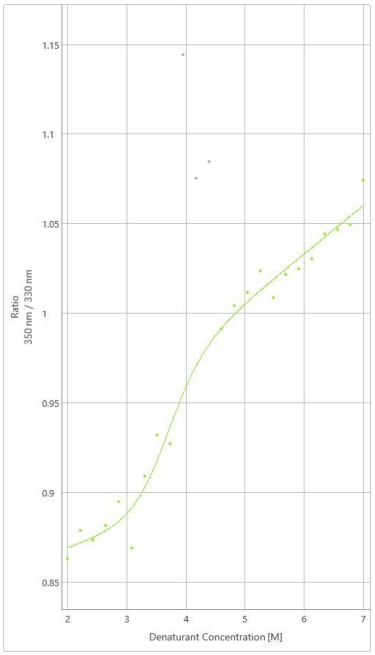
(a)



(b)



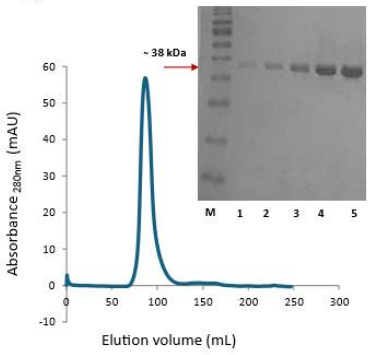
(c)



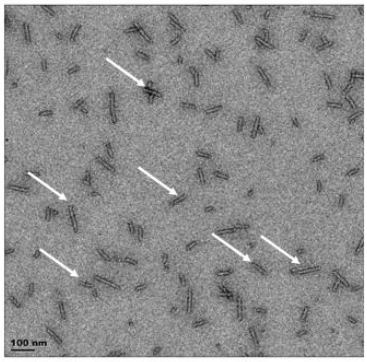
ΔG (kJ/mol)	ΔΔG (kJ/mol)	C50 (M)	Average length (nm)
29.6	0.2	3.6	34

8. E155R

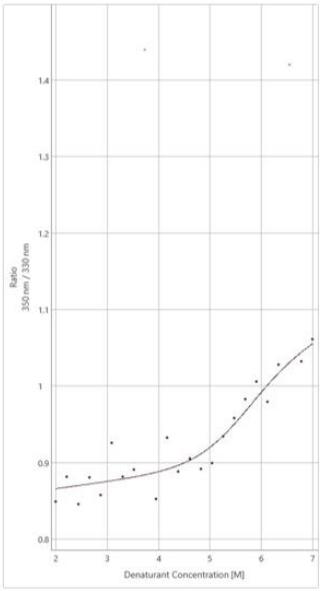
(a)



(b)



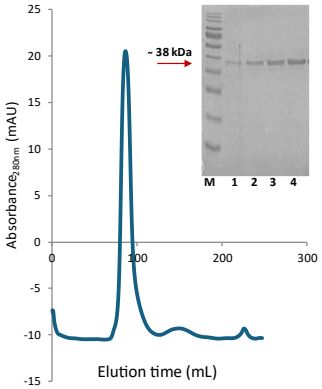
(c)



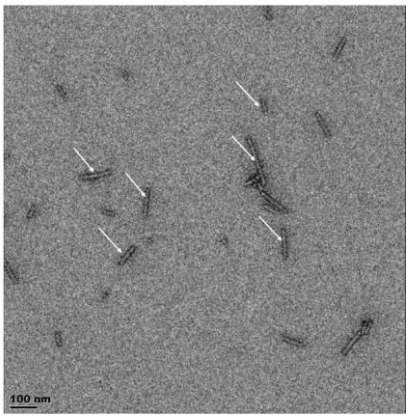
ΔΔG (kJ/mol)	ΔG (kJ/mol)	C50 (M)	Average length (nm)
44.8	-15	5.4	118

9. N119R

(a)

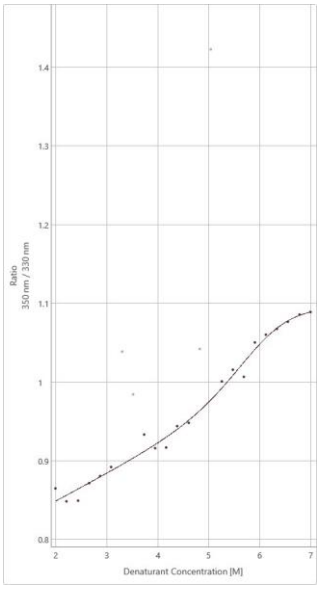


(b)



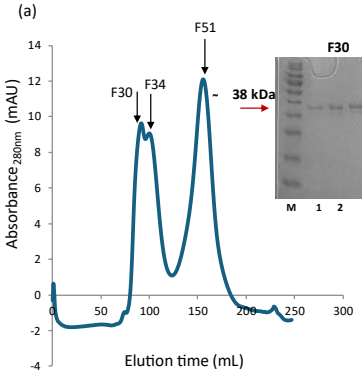
ΔG (kJ/mol)	ΔΔG(kJ/mol)	C50 (M)	Average length (nm)
35.6	-5.9	5.95	127

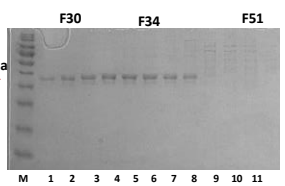
(c)



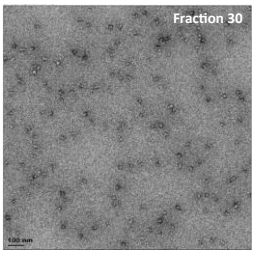
10. Q115C

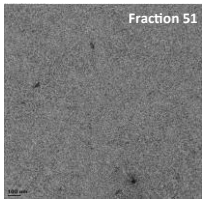
(a)

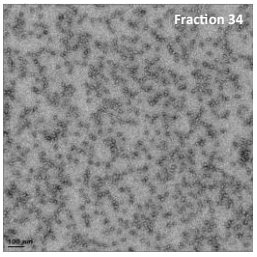




(b)

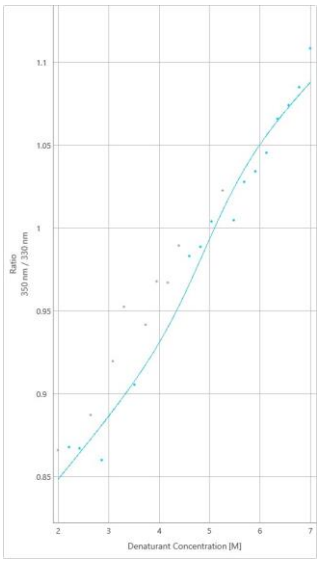






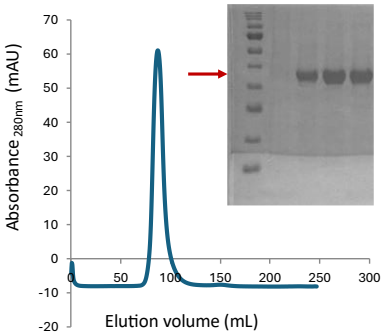
ΔG (kJ/mol)	ΔΔG (kJ/mol)	C50(M)	Average length (nm)
25.2	4.5	5.3	18

(c)

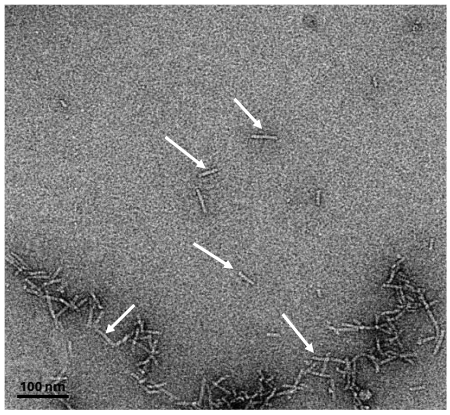


11. S29A/Q86R/N119R

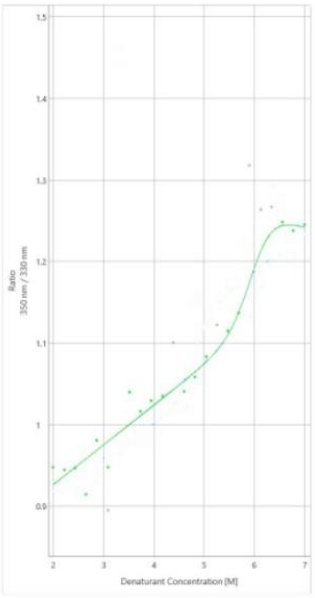
(a)



(b)



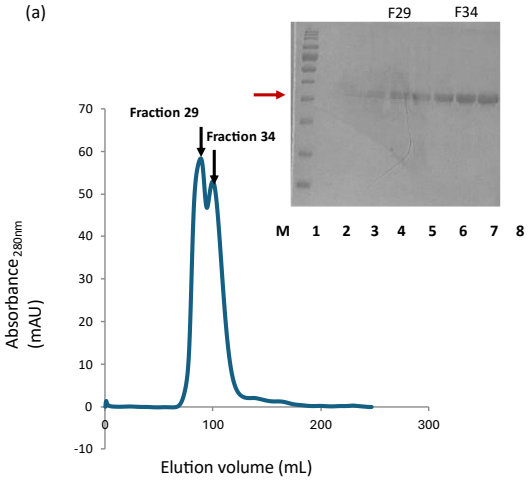
(c)



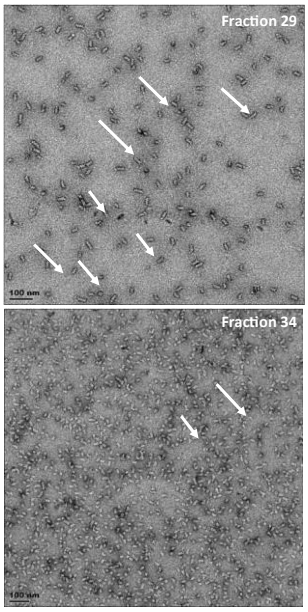
Average ΔG (kJ/mol)	ΔΔG (kJ/mol)	C50 (M)	Average length (nm)
62.3	-32.5	5.4	111

12.Q86M

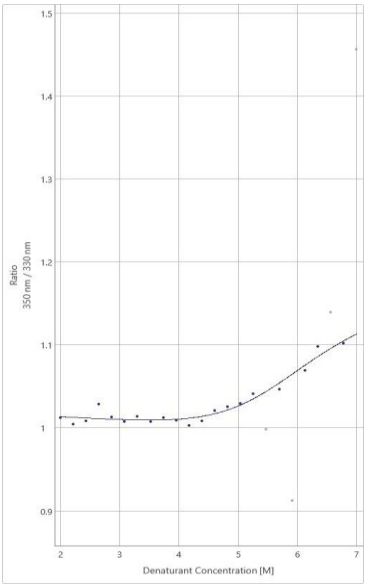
(a)



(b)



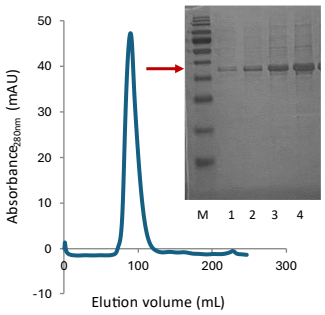
(c)



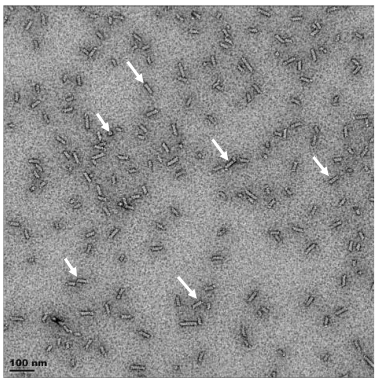
ΔG (kJ/mol)	ΔΔG (kJ/mol)	C50 (M)	Average fibre length (nm)
35.6	-5.9	5.3	51

13. T260I

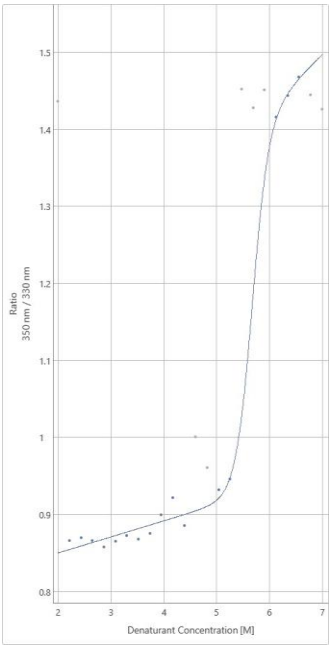
(a)



(b)



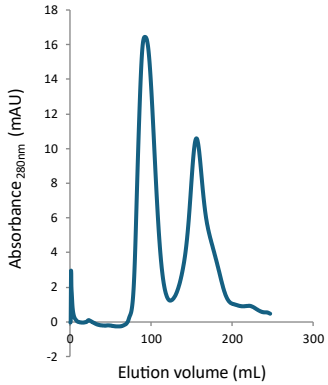
(c)



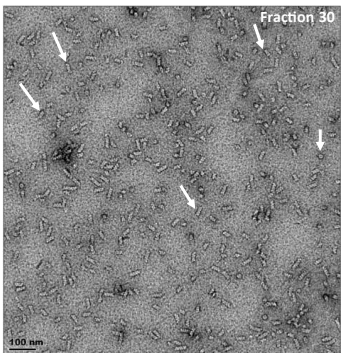
ΔG (kJ/mol)	ΔΔG (kJ/mol)	C50 (M)	Average length (nm)
30.1	-0.4	5.3	89

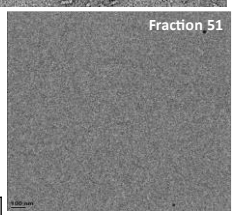
14. T217I

(a)

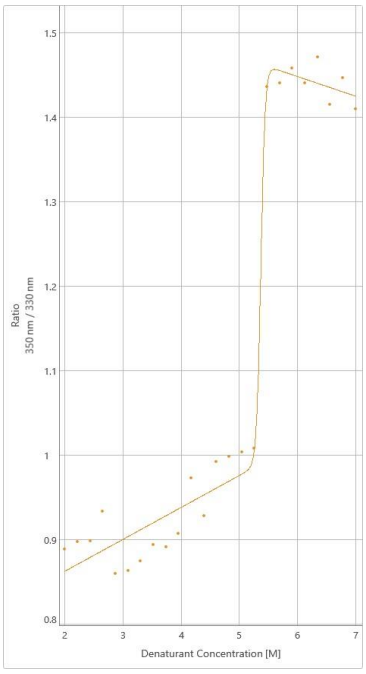


(b)





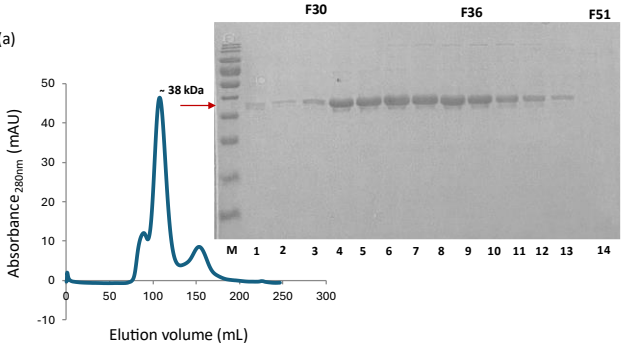
(c)



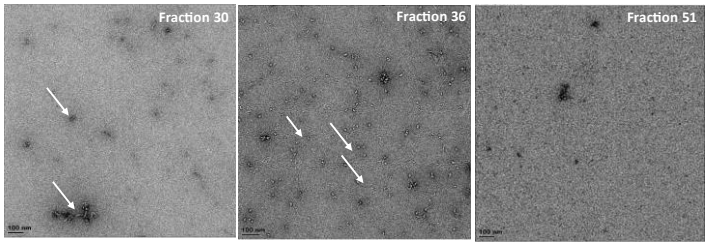
ΔG (kJ/mol)	ΔΔG (kJ/mol)	C50(M)	Average length (nm)
44.8	-15.1	5.3	62

15. S147I

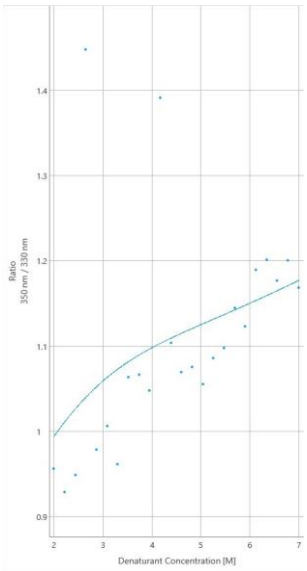
(a)



(b)



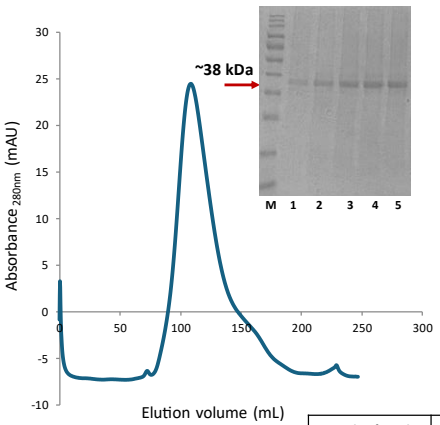
(c)



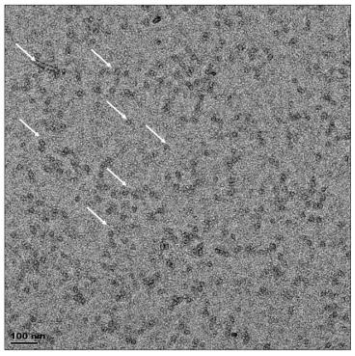
ΔG (kJ/mol)	ΔΔG (kJ/mol)	C50 (M)	Average length (nm)
-	-	-	35

16.E90K

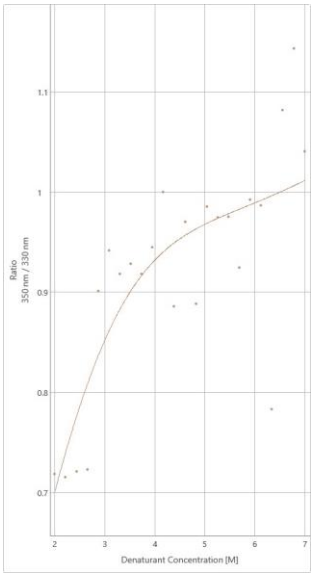
(a)



(b)

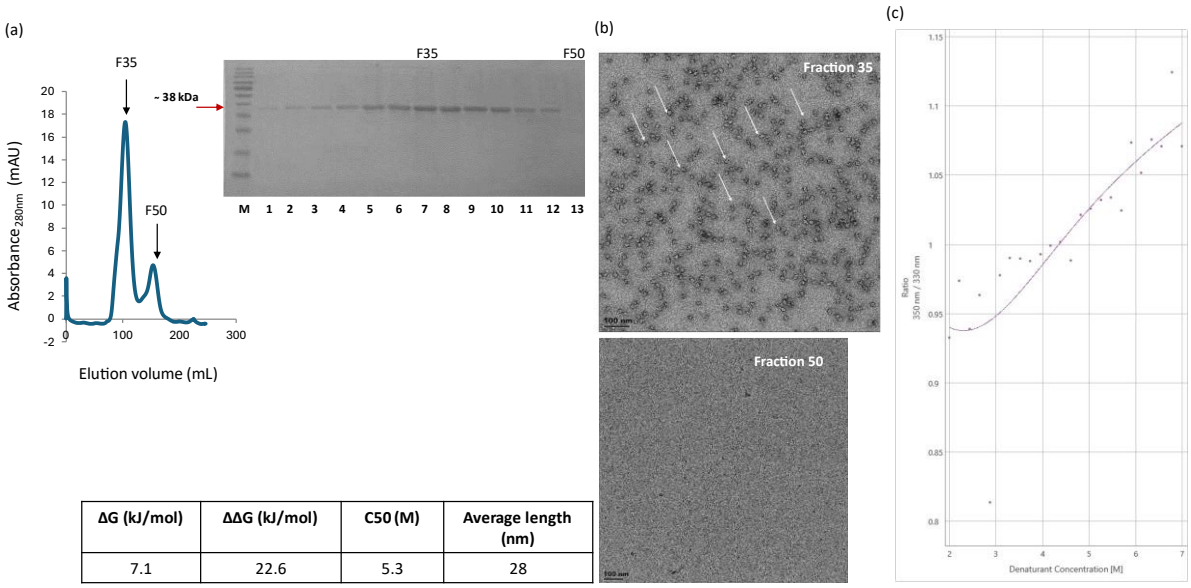


(c)



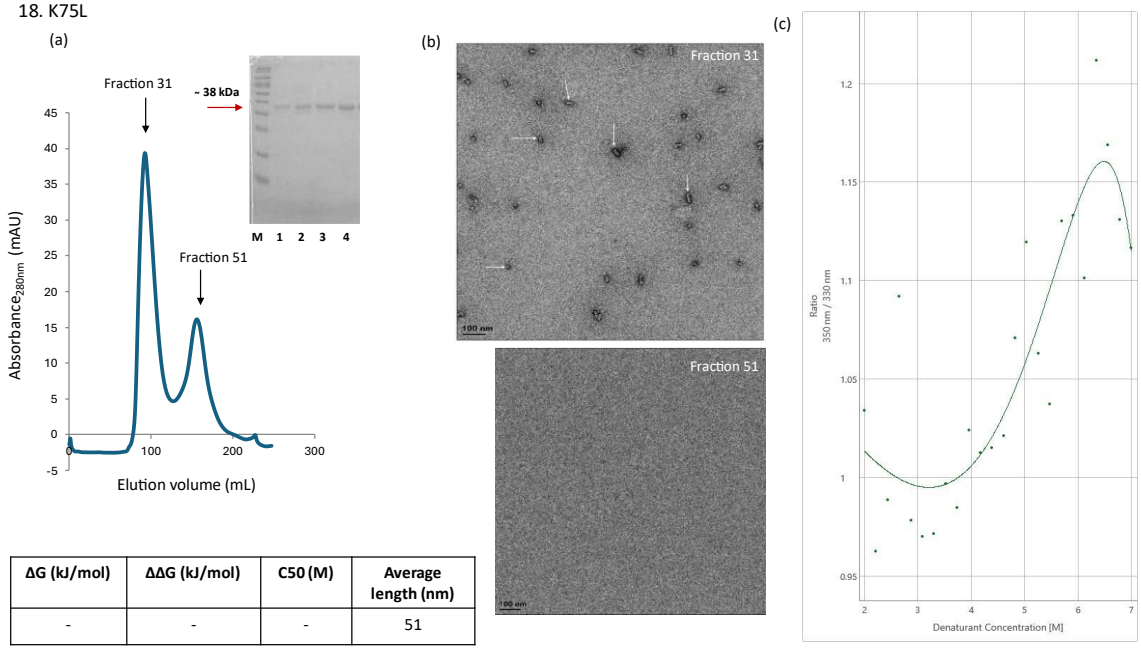
ΔG (kJ/mol)	ΔΔG (kJ/mol)	C50(M)	Average length (nm)
-	-	-	22

17. I319L



ΔG (kJ/mol)	ΔΔG (kJ/mol)	C50 (M)	Average length (nm)
7.1	22.6	5.3	28

18. K75L



ΔG (kJ/mol)	ΔΔG (kJ/mol)	C50 (M)	Average length (nm)
-	-	-	51

Figure 2.4.9: Purification analysis, structural characterisation by negative stain TEM and biophysical analysis of $CynD_{stu}$, $CynD_{pum}$ and $CynD_{pum}$ variants. Each substitution mutation is shown at the top of each image. Represented for each mutation are the following: (a) SEC elution profiles and corresponding 10 % SDS-PAGE gels. The red arrow highlights the expected monomeric size of each variant corresponding to that of wildtype $CynD_{pum}$ (~38kDa). (b) Negative-stain micrographs corresponding to the major peaks on the elution profile. The protein particles are highlighted with white arrows. (c) Unfolding profiles from chemical denaturation for each sample. The dots on each plot represent experimental data points, that are fitted using the two-state model for protein denaturation. The x-axis represents the denaturant (urea) concentration and the y, the normalised fluorescence ratio FI350nm/FI330nm. The average ΔG of three independent experiments, overall ΔΔG, average C50 and fibre length are shown in tables (inset). The average fibre length was calculated from five randomly selected protein particles for each sample.

Table 2.1: Summary of screening and characterisation results for $CynD_{pum}$ variants and $CynD_{stu}$. Protein samples and their corresponding computed and experimentally determined thermodynamic properties. ΔΔG from FoldX computations and experimentally determined ΔΔG and C50, are shown in detail in Figure 2.4.7. Included in the table are the observed oligomeric forms of the experimentally validated mutations shown in Figure 2.4.9.

Sample	Predicted average ΔΔG (kJ/mol)	Approximate experimental ΔΔG (kJ/mol)	C50 (M)	Location	Activity	Oligomeric description	Average length (nm)
$CynD_{pum}$	-	-	5.7	-	Active	Short helical fibres	42
$CynD_{stu}$	-	-97,1	5.4		Active	Short helical fibres	49
$CynD_{pum}$-$CynD_{stu}$ Chimera	-	-16.8	6.8	C-terminal tail	Active	Long fibres of variable length	200
$CynD_{pum}$ Q86R/ H305K/ H308K/ H323K	-	-104.9	5.7	-	Active	Long fibres of variable length	178
$CynD_{pum}$ N119R, S29A, Q86R	-	-34.5	5.4	C-interface, surface, D-interface	Active	Long fibres of variable length	111
I4M	-2.6	-	-	N-terminus	Inactive	-	-

S29A	- 10.2	-5.3	3.4	Core	Active	Short fibres longer than wildtype $CynD_{pum}$	63
H62Y	-6.7	-	-	Core near the C-interface	Active	Heterogenous helices shorter than the wildtype	41
H71R	-6.5	-	-	C-interface	Inactive	-	-
K75L	-6.4	20.1	3.6	C- interface	Active	Short helical fibres, slightly longer than the wildtype	51
E79P	-9.6	-	-	C-interface	Inactive	-	-
Q86K	7.8	-	-	D-interface	Inactive	-	-
S102C	-8.4	0.1	10.4	Core	Active	Same length as wild-type $CynD_{pum}$	34
Q115C	-1.1	4.1	5.3	Core	Active	Same length as wild-type $CynD_{pum}$	18
N119R	-6.3	-5.9	5.95	Groove	Active	Long homogenous fibres	127
D172N	-45.7	-	-	A-interface	Inactive	-	-

S147I	-2.5	27.9	1.1	Core/C-terminal tail	Active	Heterogenous sample, smaller helices than wildtype $CynD_{pum}$	35 22
E155R	-3.7	-15.1	4.6	Non-interfacial/groove	Active	Long homogenous fibres, longer than the wildtype	117
N177L	-38.9	-	-	A-interface	Inactive	-	-
T260I	-6.6	-0.4	2.7	Surface	Active	Homogenous fibres longer than wildtype $CynD_{pum}$	89
Q86M	-6.3	-5,9	5.3	D-interface	Active	Homogenous fibres of slightly longer than $CynD_{pum}$	51
E90K	-2.5	4,5	2.1	D-interface	Active	Smaller particles than the wildtype	22
I319L	-6.2	-	1.9	C-terminal tail	Active	Shorter helices than wildtype $CynD_{pum}$	28
E96G	-3.8	-	-	C-interface	Inactive	-	-
N180K	-6.1	-	-	C-interface	Inactive	-	-

T217I	-6.8	-15,0	4.8	Surface	Active	Homogenous fibres longer than wildtype $CynD_{pum}$	62

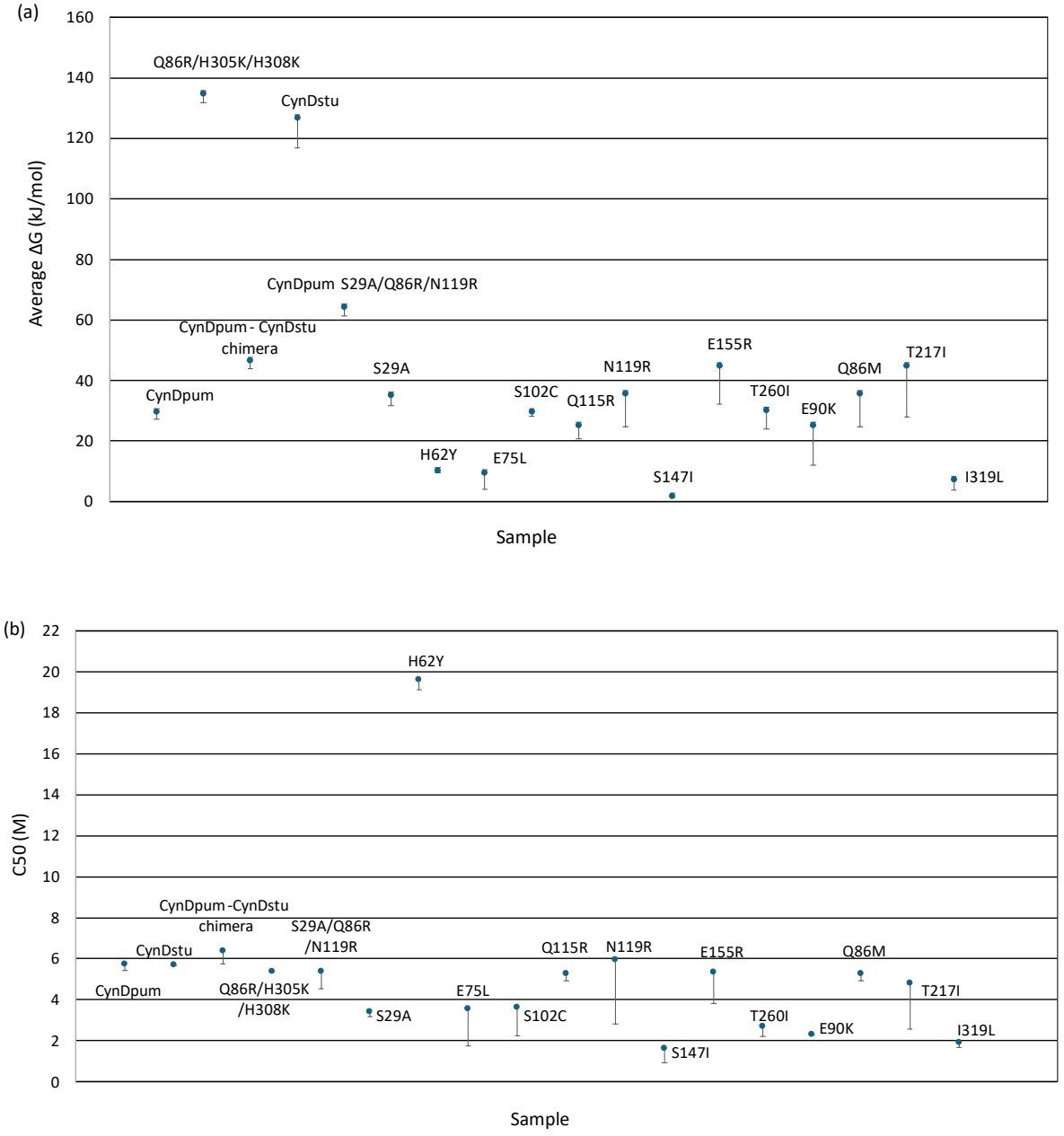


Figure 2.4.10: Comparison of ΔG and C50 of $CynD_{stu}$ and $CynD_{pum}$ variants. (a) Comparison of the relative ΔG of $CynD_{stu}$ and $CynD_{pum}$ variants relative to that of wildtype $CynD_{pum}$ (b) Comparison of the relative C50 of $CynD_{stu}$ and $CynD_{pum}$ variants relative to that of wildtype $CynD_{pum}$. Error bars represent the standard deviation of three independent experiments for each sample.

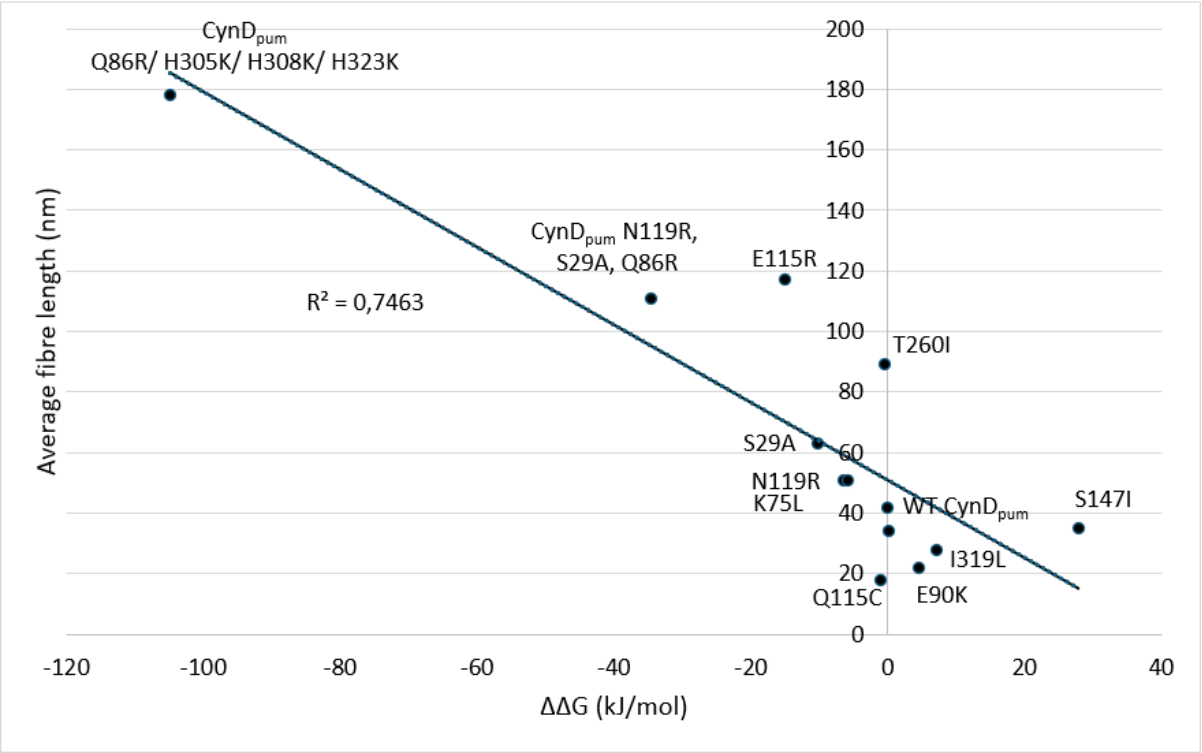


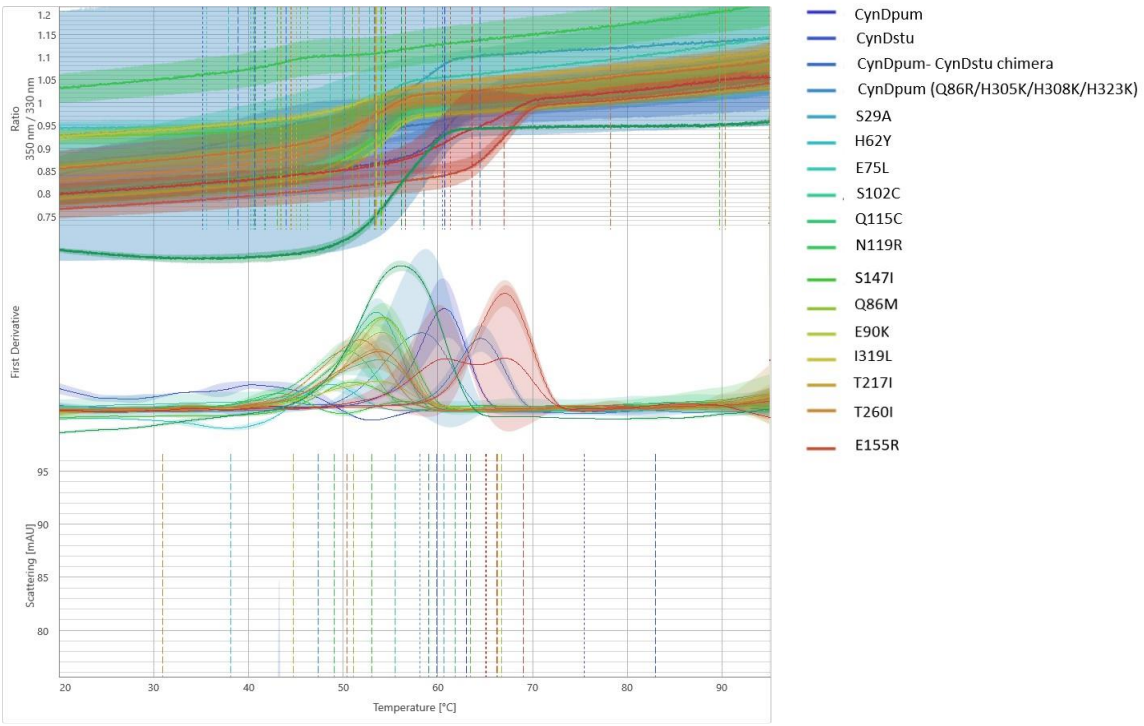
Figure 2.4.11: Correlation between the ΔΔG and the average length of the helical fibres formed by $CynD_{pum}$ variants in this study. The variants for which the ΔΔG could not be computed are excluded from the plot. The correlation plot including the $CynD_{pum}$ -$CynD_{stu}$ chimera, which was an outlier in this analysis is presented in the Appendix (Figure A2).

Of the 22 single substitution variants, six had a ΔG that was higher than that of wildtype $CynD_{pum}$. Of these, only N119R had an C50 value that was higher than that of the wildtype. The fitting of the experimental data points of the chemical denaturation for variants H62Y, E90K and S147I, resulted in curves for which the ΔG could not be computed, since the LEM derivation depends on the presence of a conformational transition between the folded and unfolded states. $CynD_{stu}$ and the multi-site substituted variants (Q86R/H305K/H308K/H323K, $CynD_{pum}$-$CynD_{stu}$ chimera, S29A/Q86R/N119R) had a higher ΔG than wildtype $CynD_{pum}$ with

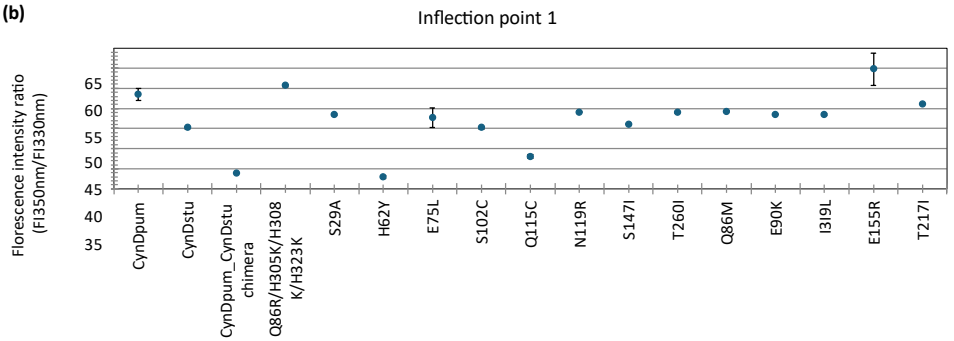
variants Q86R/H305K/H308K/H323K and the $CynD_{pum}$-$CynD_{stu\ chimera}$ having a higher C50 value. Variants with a favourable ΔΔG formed fibres with an average length longer than the wildtype. There was no correlation observed between the ΔΔG and C50. However, there was a strong linear correlation ($R^2 = 0.75$) (Figure 2.4.11) between the ΔΔG and the average length of the helical fibres formed by the variants.

Since ΔΔG and T_m are not comparable methods of determining thermostability, thermal denaturation of $CynD_{pum}$ and its variants was conducted to facilitate comparison with previous studies where thermostability was measured using T_m.

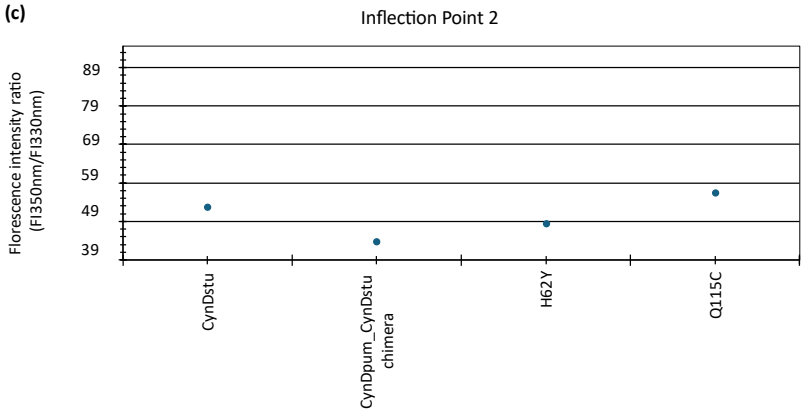
(a)


Ratio 350 nm / 330 nm
First Derivative
Scattering [mAU]
Temperature [°C]
CynDpum
CynDstu
CynDpum- CynDstu chimera
CynDpum (Q86R/H305K/H308K/H323K)
S29A
H62Y
E75L
S102C
Q115C
N119R
S147I
Q86M
E90K
I319L
T217I
T260I
E155R

(b)


Inflection point 1
Florescence intensity ratio (FI350nm/FI330nm)
CynDpum
CynDstu
CynDpum_CynDstu chimera
Q86R/H305K/H308K/H323K
S29A
H62Y
E75L
S102C
Q115C
N119R
S147I
T260I
Q86M
E90K
I319L
E155R
T217I

(c)


Inflection Point 2
Florescence intensity ratio (FI350nm/FI330nm)
CynDstu
CynDpum_CynDstu chimera
H62Y
Q115C

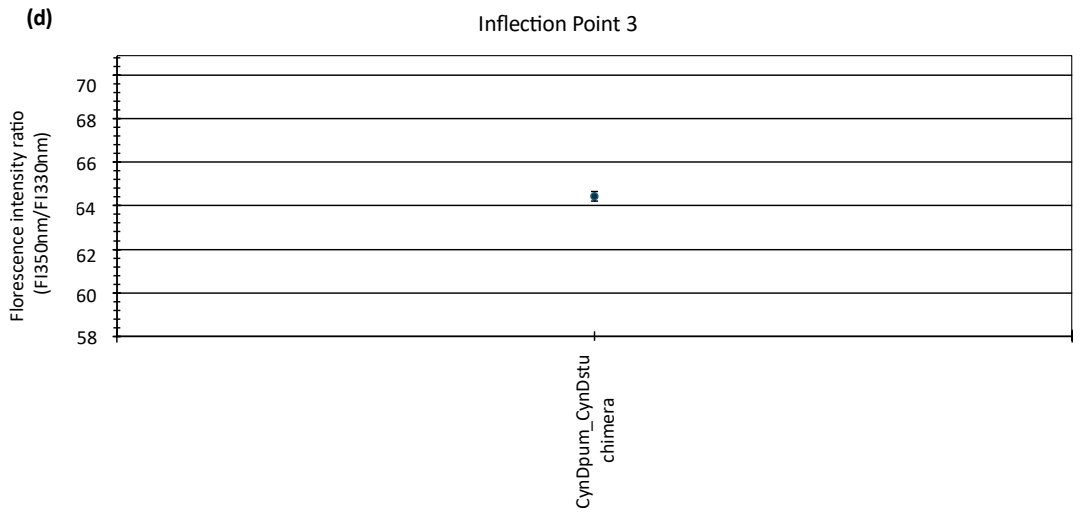


Figure 2.4.12: Thermal denaturation of $CynD_{stu}$, $CynD_{pum}$ and $CynD_{pum}$ variants predicted to be thermostable in FoldX. The top panel displays the fluorescence intensity resulting from the thermal unfolding of each sample (shown as a ratio of FI350nm/330nm). The middle panel presents the first derivative of these data, while the bottom panel shows the scatter plot corresponding to each fluorescence intensity peaks. Panels (b), (c), and (d) highlight the inflection points for each sample. The error bars in the figure indicate the standard deviation between four repeat experiments.

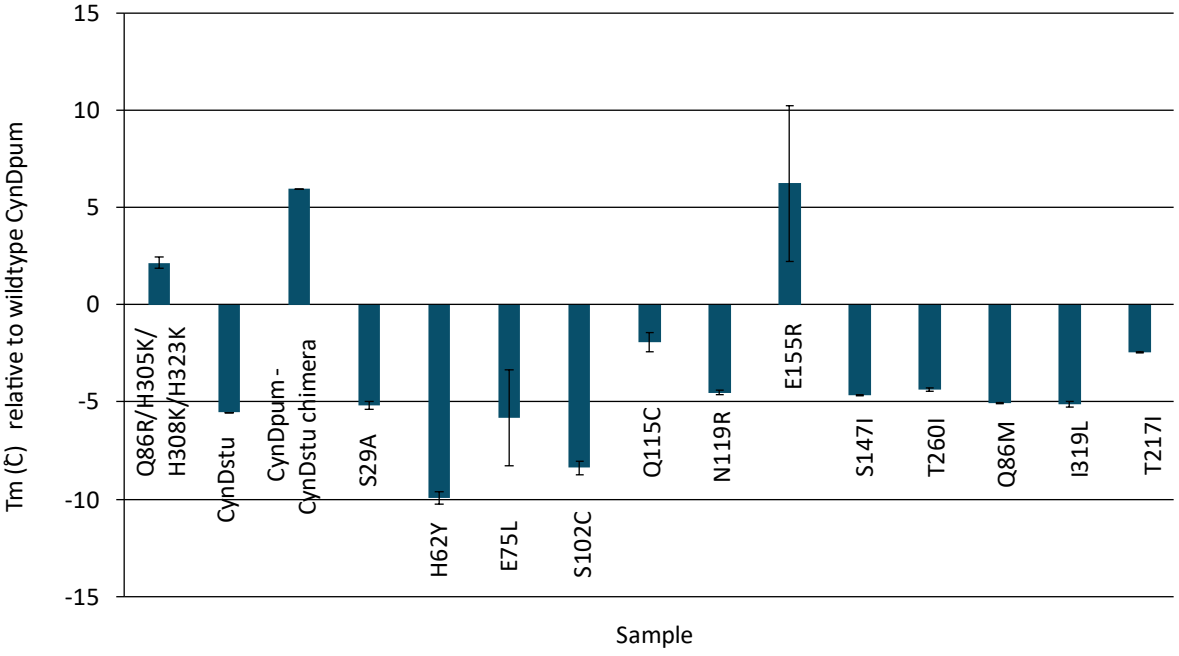


Figure 2.4.13: The melting temperature (T_m) for each protein sample relative to that of wildtype $CynD_{pum}$.

Variants Q68R/H305K/H308K/H323K and the chimera had a higher T_m than the wildtype. Of the single substitutions mutations predicted on FoldX, only the E155R variant of $CynD_{pum}$ had a higher T_m than wildtype $CynD_{pum}$. The T_m and average length of the helical fibres formed had a positive linear correlation of ($R^2 = 0.80$) (Figure 2.4.15).

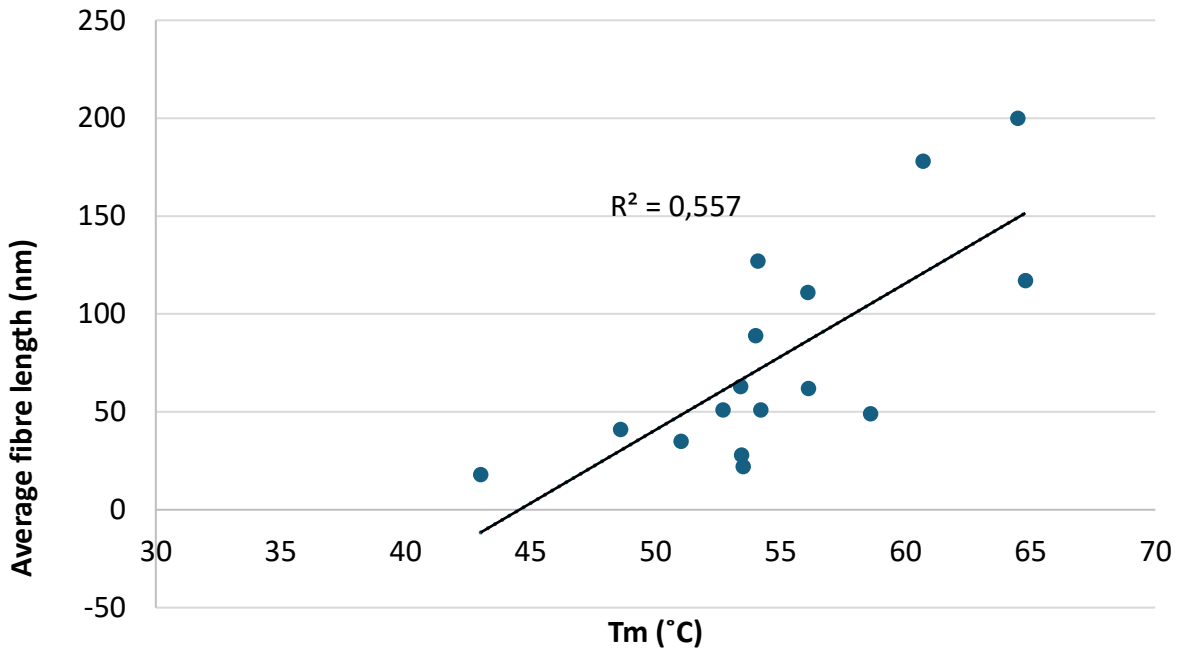


Figure 2.4.14: Correlation between T_m and the average length of the helical fibres formed by the variants.

2.4.3.3 Creating a multi-site substituted and thermostable variant of $CynD_{pum}$

Of the mutations occurring at the surface of the protein and exhibiting a more favourable ΔΔG, S29A was selected for creation of the multi-site substituted variant of $CynD_{pum}$. The introduction of the S29A mutation to the Q86R expression construct resulted in partial loss of activity (Lane A, Figure 2.4.12 (b)) the loss of activity was restored by the introduction of the N119R mutation (Lane B, Figure 2.4.12 (b)). However, addition of the E155R mutation to the S29A/Q86R/N119R expression construct resulted in loss of activity and insolubly expressed protein (Lane C, Figure 2.4.15 (b)). The S29A/Q86R/N119R/E155R variant was also inactive (Lane D, Figure 2.4.15 (b)). Attempts to optimise the expression of the S29A/Q86R/N119R/E155R variant by lowering the induction temperature to 20 °C and increasing the concentration of IPTG to 400 mM did not improve protein solubility. However, the resulting multi-site substituted variant with mutations S29A, N119R, and Q86R had increased thermostability (ΔΔG = -34.5 kJ/mol), retained activity at pH 9.0 and exists as fibres at pH 7.4.

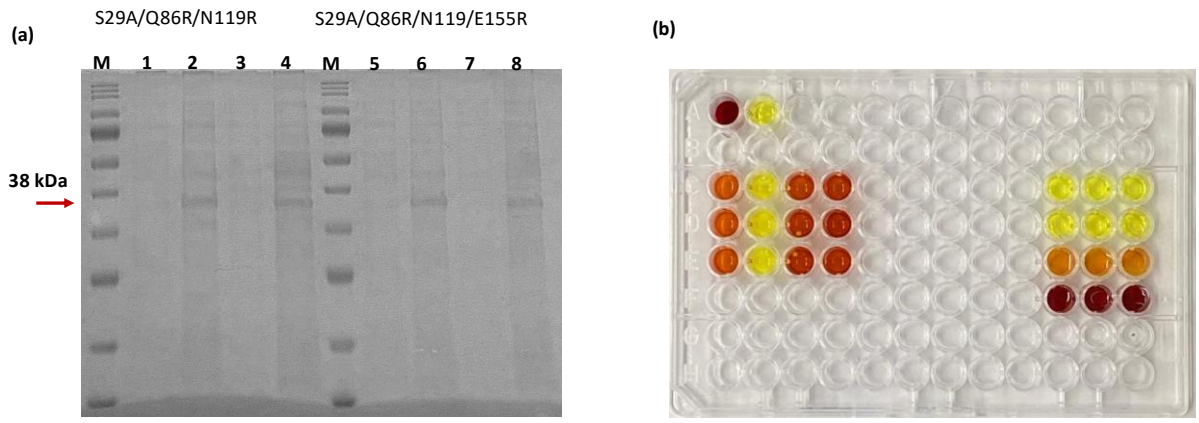


Figure 2.4.15: SDS-PAGE analysis of the recombinant expression of S29A/Q86R/N119R/E155R variant and enzyme activity screening plate for the sequential creation of a multi-site substituted thermostable variant of $CynD_{pum}$ and its screening for enzyme activity at different pH. (a) M denotes the molecular weight marker lanes. Lanes 1-4 correspond to soluble and insoluble fractions of the S29A/Q86R/N119R variant, lanes 5-8 correspond to the same for the S29A/Q86R/N119R/E155R variant. Lanes 1 and 2 and 5 and 6 correspond to the soluble and insoluble fractions for both variants expressed at 25 ℃ and 100 mM IPTG. Lanes 3 and 4 and 7 and 8 correspond to the soluble and insoluble fractions for both variants expressed at 20 ℃ and 400 mM IPTG. (b)Wells A1 and A2 correspond to the positive and negative controls. Wells C1, D1 and E1 correspond to three repeats of the activity of the S29A/Q86R variant. Wells C2, D2 and E2 correspond to three repeats of the activity of the S29A/Q86R/N119R variant. Wells C3, D3 and E3 and Wells C4, D4 and E4 correspond to three repeats of the activity analysis of variants S29A/Q86R/N119R/E155R and S29A/Q86R /E155R, respectively. Wells C10-C12 correspond to repeats for the activity analysis of the S29A/Q86R/N119R variant at pH 7.4. Wells D10-D12 correspond to repeats for the activity analysis of the S29A/Q86R/N119R variant at pH 8.0. Wells E10-E12 correspond to repeats for the activity analysis of the S29A/Q86R/N119R variant at pH 9.0. Wells F10-F12 correspond to repeats for the activity analysis of the S29A/Q86R/N119R variant at pH 10.0.

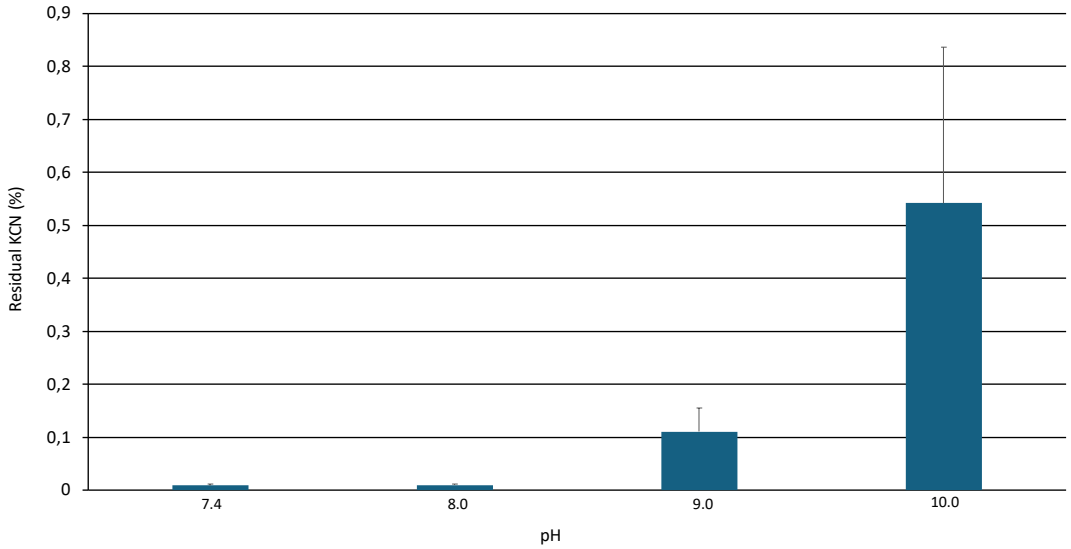


Figure 2.4.16: Percent residual KCN of $CynD_{pum}$ S29A/Q86R/N119R variant at different pH. The multi-site substituted variant retained approximately 90% and 50% of its activity at pH 9.0 and 10.0.

2.5 Discussion

2.5.1 Structure-guided *in silico* site saturation mutagenesis

Of the selected mutations, 46% had a more favourable ΔΔG than wild-type $CynD_{pum}$. Thus, the correlation between *in silico* predicted and experimentally determined ΔΔG was relatively low, with some mutations resulting in inactive- (I4M, H71R, E75P, Q86K, E96G, N180K, D172N, N177L) and others active but unstable variants (H62Y, K75L, E90K, S147I).

The quaternary structure of $CynD_{pum}$ is symmetric and the specific symmetry and oligomerisation state of the wildtype and variants are dictated by the amino acid sequence as observed in previous studies (Wang, Watermeyer et al. 2012, Crum, Park et al. 2015, Mulelu 2017, Woodward, Trompetter et al. 2018, Mulelu, Kirykowicz et al. 2019). Any changes in the helical assembly's symmetry can lead to inactivity by affecting the conformation of the substrate-binding site or causing improper protein folding. It is important to note that the initial calculation of the energy of the wild-type model in FoldX, considers the energetic contributions of both mainchain and sidechain amino acids. However, during *in silico* mutation, mainchain amino acids are fixed and only the sidechain contributions are considered, and their changes to the overall energy of the assembly are used to determine the thermodynamic favourability of a mutation (Sanavia, Birolo et al. 2020), without

considering potential conformational changes to the overall structure of the enzyme. Therefore, the conformational changes that arise from introduced substitution mutations in the helical assembly may have contributed to the inaccurate prediction of unstable and inactive variants as stable.

The Q86R/H305K/H308K variant of $CynD_{pum}$ was used as the reference model for *in silico* SSM instead of wildtype $CynD_{pum}$ because of the lower resolution (3.83 Å) of the wildtype structure compared to that of the variant (3.15 Å). This resulted in the overestimation of ΔΔG for each substitution by –0.522 kJ/mol for substitutions at all other positions except for those where the three mutations occur. Even so, the ΔΔG of all the variants, with the –0.522 kJ/mol overestimation considered, remains favourable (Table 2.1).

Despite the low correlation between *in silico* and experimentally determined ΔΔG, the degree of error in predicting inactive substitution mutations was significantly lower than that of random techniques used previously (Abou Nader 2012). Additionally, the multi-site substituted enzyme created in this study, based on predictions in FoldX, (S29A/Q86R/N119R) was thermostable forming long helical fibres at pH 7.4, and retained enzyme activity at pH 9.0.

2.5.2 Biophysical analyses

Variants H62Y, K75L, E90K, and S147I displayed significant heterogeneity and instability (Figure 2.4.9), which prevented the computation of reliable ΔΔG and C50 values. Therefore, fitting a two-state unfolding curve to the data points was not possible. The relationship between ΔΔG and C50 values could not be established for single substitution mutations. However, a positive linear correlation between the two parameters was identified in the $CynD_{pum}$-$CynD_{stu}$ chimera and the Q86R/H305K/H308K/H323K variant.

While a positive change in ΔΔG is typically associated with an increase in C50 for proteins following the two-state model of protein unfolding, the complex quaternary structure of $CynD_{pum}$ and its variants, consisting of multiple monomers and dimers connected through interfacial regions and C-terminal tail contributions, makes strict adherence to this model for all variants potentially unlikely. As a result, the chemical unfolding of some of the variants may produce multiple intermediates (resulting from potential dissociation) where the protein particles in solution do not all strictly occupy either the folded or unfolded state at any given point or unfold irreversibly. Therefore, the relationship between ΔΔG and C50 for $CynD_{pum}$ variants may not always be linear (Cellmer, Henry et al. 2007). The thermostability of variants

in this case may be computed as the difference in the energy of unfolding of the wildtype and that of the intermediate rather than the unfolded state (Moore, Kanu et al. 2011).

A positive linear relationship was observed between the average length of helical fibres formed by the $CynD_{pum}$ variants, and the ΔΔG (Figure 2.4.11). This suggests that the length of the helical fibres may influence the thermostability of $CynD_{pum}$. However, the basis for the increased length of the helical fibres remains unclear due to the varying nature of the substitution mutations which occur at all regions of the protein. There was no correlation observed between M_r and ΔΔG or T_m, despite the longer fibres of the variants resulting in a higher M_r. The inaccuracy in determining M_r using SEC is that the technique does not provide a discrete and accurate measurement of the protein particles in solution, especially those with slight differences in M_r. In $CynD_{pum}$, this potentially arises due to the presence of particles with potential single monomer or subunit differences within a single peak, leading to a range of masses rather than one discrete measurement. Therefore, we found that measuring the average fibre length on TEM micrographs was a more accurate method of determining the size of the protein particles and a better representation of the true quaternary structure of each sample.

2.5.3 Structural insights

Valuable insights into the structure of $CynD_{pum}$ and its thermostability were obtained from *in silico* SSM and biophysical analysis. From these experiments, we identified unexplored sites on the enzyme which are rich targets for engineering thermostability in $CynD_{pum}$ and gained further insights into the structure and thermostability of $CynD_{pum}$.

Regions of $CynD_{pum}$ that have not yet been investigated for their role in thermostability include the grooves and the surfaces. Most studies have focused on the interfacial regions through which the homo-oligomer is formed and the C-terminal tail (Sewell, Thuku et al. 2005, Crum, Park et al. 2015, Crum, Park et al. 2015, Park, Mulelu et al. 2016, Xu, Cai et al. 2018). Variants, E155R and N119R, located along the grooves, displayed enhanced thermostability and formed longer helical fibres than the wildtype enzyme. The structural basis of the increased thermostability resulting from mutations at the grooves was not immediately clear. Similarly, variants S29A, T260I, and T217I located at the exterior surface of the enzyme, exhibited a higher ΔΔG and formed fibres with an average length exceeding that of the wildtype $CynD_{pum}$.

2.5.3.1 The role of the N- and C-termini to protein folding and thermostability

In this study, we attribute the enhanced thermostability of the Q86R/H305K/H308K/H323K variant of $CynD_{pum}$ to interactions occurring between the C-terminal tail and the α4 helix at

the D-interface, rather than involving two R86 amino acids interacting across the groove through an arginine-arginine stacking interaction or salt bridge interactions. In the refined model (PDB ID: 8c5i), R86 appears to be involved in ionic and hydrophobic interactions with amino acids of the C-terminal tail of distant monomers (Figure 2.4.5 (c)). This new network of short-range interactions between distant monomers reinforces the structural integrity of $CynD_{pum}$, thereby enhancing its thermostability. We postulate that the thermodynamic favourability brought about by these new interactions, significantly lowers the energy required for the addition of subunits to a growing helical fibre during oligomerisation, thus resulting in the formation of dramatically elongated helical fibres.

We confirmed our interpretation by introducing the Q86R mutation to the $CynD_{pum}$-$CynD_{stu}$ chimera which possesses the $CynD_{stu}$ C-terminal tail (Chapter IV, Figure 4.4.5). This alteration potentially disrupted or weakened the interactions between the α4 helix and the C-terminal tail, which is divergent in $CynD_{stu}$, causing the oligomeric structure to transition from long fibres to short spirals resembling those of $CynD_{stu}$. Moreover, we found that in addition to the C-terminal tail, the nature and strength of the contacts at the D-interface are crucial for stabilising the enzyme, demonstrated by the inactivity of the Q86K variant and the Q86M which was less thermostable than the Q86R variant and formed much shorter helical fibres. Therefore, understanding the role of the D-interface in forming and maintaining the homo-

oligomeric assembly is essential to identifying synergistic mutations with the C-terminal tail that confer thermostability.

The importance of the C-terminal tail to the cyanide dihydratases was previously highlighted in the $CynD_{pum}$- $CynD_{stu}$ chimera (Sewell, Thuku et al. 2005, Crum, Park et al. 2015) where it was shown to be larger in size than the wildtype, active and thermostable. Confirming this observation in our study, we demonstrate that the $CynD_{pum}$-$CynD_{stu}$ chimera forms elongated helical fibres (Figure 2.4.9 (3b)), emphasising the contribution of the C-terminus to enhanced thermostability and oligomerisation.

The I4M mutation, located at the N-terminus was predicted to be thermostable. However, the variant was inactive. Since the N-terminus is distantly located to the active site and the A-interface, we postulate that the observed loss of activity is due to improper protein folding and, therefore, that the N-terminus is critical to the proper formation of the homo-oligomeric assembly.

2.5.3.2 Exterior surface amino acid contributions to thermostability and oligomerisation

The surface of the enzyme is another previously unexplored region which was found to be important for increased thermostability in this study. Three mutations, S29A, T217I, and T260I, resulted in helical fibres that were longer than those of the wildtype enzyme. These

variants had a favourable ΔΔG and potentially contribute to thermostability through increased hydrophobicity brought about by the reduced number of solvent-accessible amino acid sidechains which promotes protein folding (Vetriani, Maeder et al. 1998, Tang, Li et al. 2021).

All single- and multiple substitution thermostable variants formed long helical fibres at pH 7.4, displaying a positive linear correlation between fibre length and ΔΔG. We propose that the correlation between helical fibre formation and thermostability is brought about by the increased amount of energy required to break the new bonds and interactions occurring in the homo-oligomers existing as longer helical fibres and having higher M_r than the wildtype enzyme. In previous studies, a change in the oligomeric structure of $CynD_{pum}$ from short helices to long helical fibres, resulting from a pH change between 8 and 5.4 has been observed (Jandhyala, Berman et al. 2003, Mulelu 2017). The longer helices occurring at pH 5.4 were thermostable, emphasising the correlation between reinforced and elongated oligomeric structure and thermostability.

The chemical and structural basis for oligomerisation has previously been attributed to the C-interface, particularly the insertion regions found only in helical- and spiral-forming nitrilases (Thuku, Brady et al. 2009, Park, Mulelu et al. 2016). These insertion sites were previously

thought to occur at or near the C-interface based on homology modelling (Park, Mulelu et al. 2016). On the current atomic resolution model of $CynD_{pum}$, amino acids H62-E72 are located at the C-interface while the remaining amino acids, P55-I60 and E222- E235 are located at the core and surface, respectively. Mutation of amino acids in these insertion sites resulted in inactivity and formation of homo-oligomers of equal size to the $CynD_{pum}$ wildtype 18mer and smaller homo-oligomers when analysed using SEC (Park, Mulelu et al. 2016). Therefore, if the insertion regions unique to helical and spiral forming nitrilase, are potentially responsible for oligomerisation and are not solely located at the C-interface, we can conclude that the basis of oligomerisation is not strictly determined by interactions at the C-interface. Rather, amino acids at the surface and grooves also contribute to the formation of these large assemblies. The long helical fibres of the variants obtained in this study; S29A, T217, T260, N119R, E155R and the multi-site-substituted variant (S29A/Q86R/N119R) support this hypothesis. The S29A/Q86R variant was partially inactive (Figure 2.4.15). The basis of this loss of activity resulting from lack of synergy between the two mutations is unclear since both mutations are distantly located from each other and the substrate binding site. However, this observation is an indication of the importance of surface amino acids to the structure and function of $CynD_{pum}$.

Mutations occurring at the core of the enzyme (S102C and Q115C) did not result in a favourable change in ΔΔG experimentally. The length of the fibres of both variants was less than that of the wild-type enzyme, suggesting that these substitutions do not contribute to oligomerisation and thermostability.

2.5.3.3 Interfacial regions

The helices (α1 and α4) of the D-interface are reported to associate when the helix completes a turn, and together with the C-interface, dictate the helical rise and twist of the helical assembly (Thuku, Brady et al. 2009).

The R86 amino acid may adopt more than one possible conformation within the charge density. One conformation forms a transient electrostatic interaction with the sidechain of amino acid E90 on the α4-helix of the monomer across the groove and with E90 on the same chain (not shown). Another conformation enables R86 to interact with the amino acids of the C-terminal tail. Based on previous studies demonstrating a significant improvement in activity, thermostability and increased helical fibre length conferred by the Q86R mutation (Wang, Watermeyer et al. 2012) and the increased network of interactions R86 forms with the C-terminal tail, we postulate that the predominant conformation leading to the enhanced thermostability of the Q86R variant is due to interactions with the C-terminal tail rather than

across the groove. Additionally, our analysis of protein-protein binding at this interface showed that the formation of this interface is thermodynamically unfavourable (Table 2.4.1); therefore, interactions across the interface are unlikely to increase the thermostability of the enzyme.

The Q86M variant was thermostable and existed as helical fibres longer than the wild type, whereas the Q86K variant was inactive. We hypothesise that the increased thermostability and fibre length of the Q86M variant are due to hydrophobic interactions with amino acids of the C-terminal tail, while loss of enzyme activity in the Q86K variant is potentially due to interactions occurring across the groove, resulting from the elongated side chain of the lysine amino acid. Thus, we postulate that stabilising mutations in this region are those that reinforce interactions between the α4 helix and the C-terminal tail. Furthermore, based on the loss of activity resulting from the Q86K mutation, potentially due to misfolding, we hypothesise that this groove is crucial for enzyme conformational flexibility. This flexibility might have implications for the formation and stability of the C-interface, which influences the conformation of the substrate-binding pocket and, therefore, enzyme activity.

Mutations at the C-interface were either inactive (H71R, E79P) or unstable (H62Y, K75L) *in vitro*. Based on the H71R output model from FoldX, the arginine creates an additional

interaction with E275 across the interface (not shown). While this interaction may theoretically appear to increase contacts across the interface and therefore, enhance the thermostability of the enzyme by promoting oligomerisation, we speculate that it results in an altered helical twist, therefore, changing the size of the substrate-binding pocket, as shown in plant nitrilases (Mulelu, Kirykowicz et al. 2019). The D172N mutation was found to be inactive from activity analysis conducted on the crude lysate of the expressed protein. This mutation has previously been shown to contribute to increase the substrate affinity and thermostability of $CynD_{pum}$ (Crum, Sewell et al. 2016). The inactivity observed in this study may have been a false positive or the protein may have been insolubly expressed under the conditions used.

The thermostability of mutations H62Y and K75L predicted to have a ΔΔG higher than the reference model could not be validated. Both variants did not follow the expected two-state model of protein denaturation and therefore, the ΔΔG and C50 for both variants could not be determined. Based on negative stain TEM micrographs, both variants are smaller than the wildtype protein with significant heterogeneity indicative of structural instability. The observed instability is potentially due to dissociation of subunits from the assembly suggesting the importance of H62, which is located on region 2 (Figure 2.4.4) of the C-interface in stabilising the C-interface. Mutations created at this interface with the aim of

creating homo-oligomers that are tolerant to alkaline pH showed the importance of a positive charge at position 62, with substitutions H62A and H62D leading to insoluble and therefore improperly folded protein (Mulelu 2017). The aromatic side chain of tyrosine in H62Y potentially compensates for interactions contributed by the imidazole side chain because enzyme activity was retained eventhough the enzyme is clearly unstable and highly heterogeneous (Figure 2.4.9), confirming the requirement for a positive charge at this position as suggested by Mulelu (2017). This highlights that engineering thermostable variants with mutations at the C-interface is challenging due to its influence on both enzyme activity and oligomerisation.

Mutations at the A/dimerisation interface leading to inactivity were consistent with previous observations (Sewell, Thuku et al. 2005), and mutations (N177L) at this interface led to loss of activity. This was the result for mutations located more than 10 Å from the active site, indicating the structural and functional importance of the interface and its sensitivity to amino acid changes. Although mutations at the A-interface have previously been shown to result in enzyme inactivation (Sewell, Thuku et al. 2005), mutations located at this interface but distant from the active site were not expected to contribute to loss of activity. We postulate that this observation indicates that even slight changes to the chemistry of the A-interface affects the formation of the dimer subunit.

Chapter III

Structural analysis of the thermostable E155R variant of Cyanide Dihydratase from *Bacillus pumilus* C1

3.1 Introduction

In chapter II, we used structure guided *in silico* site saturation mutagenesis to identify $CynD_{pum}$ variants with enhanced thermostability. Of the six variants determined to be thermostable based on favourable experimental Gibb's free energy measurements, two variants (N119R and E155R) existed as fibres with relatively uniform length. The basis of their enhanced thermostability and increased fibre length was not immediately clear based on the current models of $CynD_{pum}$ and the Q86R/H305K/H308K/H323K variant. Therefore, to understand the structural and chemical basis of the observed thermostability and oligomerisation, we solved the structure of the E155R variant of $CynD_{pum}$ at atomic resolution using CryoEM and single particle analysis.

In this chapter we describe the structure of the E155R variant of $CynD_{pum}$ and use it to gain further insight into the structure and structural determinants of thermostability in $CynD_{pum}$.

3.2 Materials and methods

3.2.1 Protein expression and purification

The mutation was introduced as described in Chapter II section 2.3.4. The protein was recombinantly expressed, purified, and visualised by negative stain TEM as described in sections 2.3.5 and 2.3.8., respectively.

3.2.2 Cryogenic-electron transmission electron microscopy (CryoTEM)

Cryogenic electron microscopy and single particle analysis (SPA) were used to visualise protein samples at atomic resolution. The high molecular weight of the protein sample necessitated its structural determination using this technique. A schematic workflow of the steps followed is shown in Fig 3.2.1.

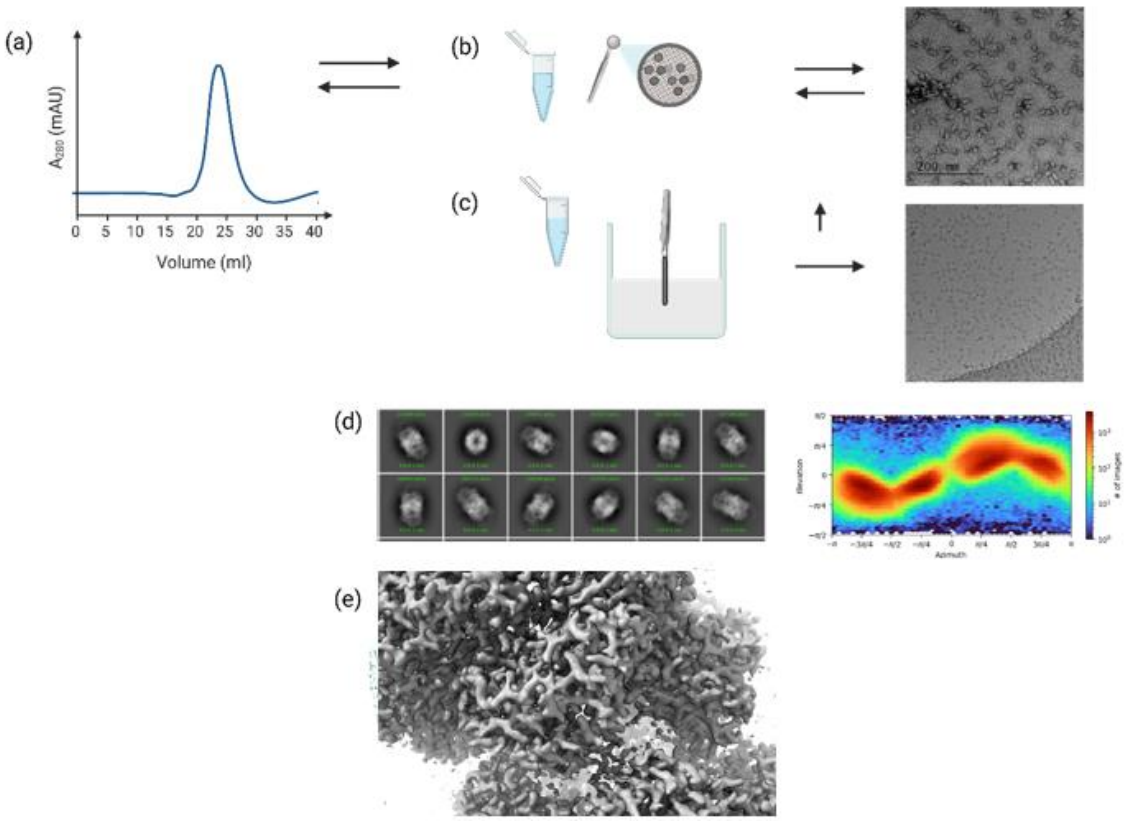


Figure 3.2.1: Schematic overview of the Cryo-EM sample preparation process used in this study. (a) Protein samples were first assessed for homogeneity and stability in the sample buffer via size exclusion chromatography and low-resolution negative-stain electron microscopy. (b) After evaluation, the sample was applied to a grid, plunge-frozen in liquid ethane, and vitrified. (c) The vitrified sample was then screened under cryogenic conditions to evaluate protein stability in vitrified ice, ensure optimal ice thickness, check for even particle distribution, and confirm appropriate protein concentration. (d) A high-resolution cryo-EM dataset was subsequently collected and processed. Particle 2D classes were analysed, and an initial map was generated to assess particle orientation and distribution. (e) The initial map was then refined to produce a high-resolution 3D map, which was evaluated for overall and local resolution, as well as potential alternative conformational states. Finally, a model was built using the charge-density map, refined, validated and interpreted.

3.2.2.1 Sample vitrification

Vitrification of the protein sample traps particles in various 3D conformations within a layer of vitreous ice by rapidly plunge freezing the grid loaded with sample into liquid ethane (Fig 3.1 (c)), preserving the protein in a near-native state. This technique enables accurate data

collection and visualisation of the protein's native structure without distortions or artefacts emanating from stains or crystalline ice.

The protein concentration was optimised for optimal particle distribution by vitrifying different concentrations of protein samples (0.5, 1, 1.5, 1.8, and 2 mg/mL). A 3.0 µL (1.8 mg/mL) protein sample was applied to a glow-discharged Holey carbon grid (R1.2/1.3) (Quantifoil, Germany), vitrified on a Vitrobot Mark II cryo-plunger (FEI, USA), and immediately blotted for 4 s with a blot force of 0 MPa at 100% humidity and 4 °C. The grid was then plunge frozen in liquid ethane and transferred to a cryo-grid box and stored in liquid nitrogen.

3.2.2.2 Image data collection

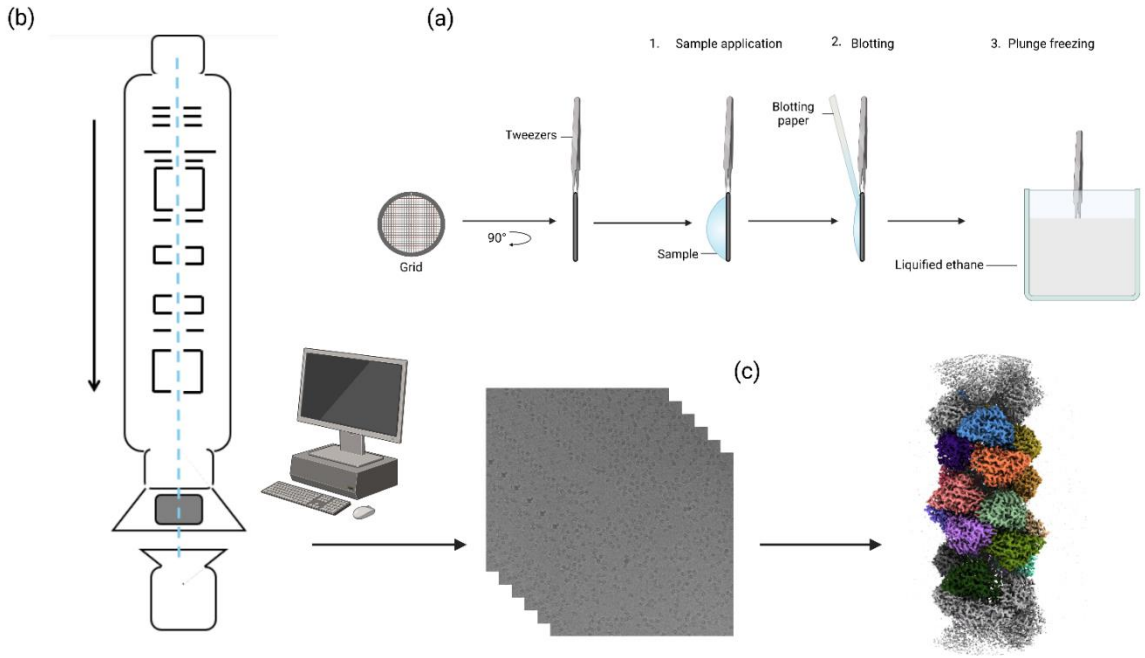


Figure 3.2.2: Summarised workflow for structural determination of protein structures by cryogenic electron microscopy followed in this study. (a) and (b)The sample is vitrified onto a grid support and inserted into the vacuum column of a cryo-TEM. (c) High resolution data is collected and processed to obtain a high-resolution map of the vitrified protein particles using the workflow show in Fig 3.3.

Image formation

In cryo-TEM, an image represents the projected electron scattering density of a biological molecule, modified by the contrast transfer function (CTF) of the microscope's objective lens.

Electron scattering

The vitrified grid, containing the protein sample embedded in vitreous ice, is inserted into the high-vacuum column of the cryo-TEM. An electron beam is accelerated down the column toward the sample (Figure 3.2 a and b). The wavelength of this beam, determined by the accelerating voltage, influences the achievable resolution of the final charge density map. As the electrons travel through the column and interact with the sample, they scatter, resulting in unscattered, elastically scattered, and inelastically scattered electrons. Elastically scattered electrons interact with the atoms in the sample and are deflected without losing energy, contributing to the formation of 2D images of the protein sample. In contrast, inelastically scattered electrons lose energy upon interacting with the sample, contributing to background noise, reducing the signal-to-noise ratio, and ultimately decreasing the achievable resolution by transferring energy as heat and X-rays which cause radiation damage (Figure 3.3) (Baker

and Rubinstein 2010). Unscattered electrons pass through the sample without interacting with it and pass through the objective lens of the microscope directly to the direct electron detector without being diverted. These electrons do not directly contribute to the contrast of the image but form the background intensity on which elastically and inelastically scattered waves create contrast for visualisation of the sample image (Orlova and Saibil 2011).

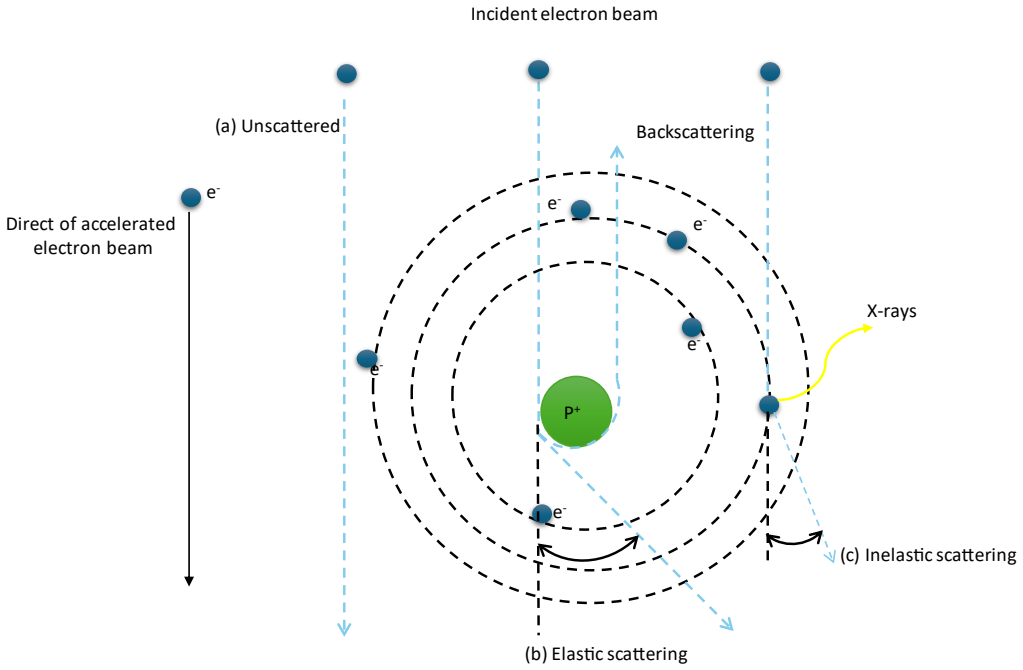


Figure 3.2.3: Interactions of the electron beam with the protein sample. (a) Some electrons pass through the sample specimen without scattering (b) while others interact with the sample resulting in elastic scattering which is characterised by a large angle of scattering. (c) Lastly some electrons interact with the sample and scatter inelastically resulting in loss of energy and a small angle of scattering the generation of heat and X-rays which damage the sample.

Phase contrast

In EM contrast refers to the different intensities of specific features on a sample to be visualised. The contrast in cryo-EM is described by depicting the electrons as waves with an

amplitude and phase thereby generating amplitude and phase contrast. Phase contrast is the main source of contrast in cryo-TEM images, enabling the visualisation of biological macromolecules, such as proteins, which are characterised by weak electron scattering due to being composed by light atoms to be visualised at atomic detail by converting phase shifts into observable contrast on the final images, therefore, revealing structural details. To enhance the contrast of the thin and weakly scattering atoms of the protein sample, images are collected slightly under focus and with an addition 90˚phase shift to increase contrast (Orlova and Saibil 2011). When the waves of elastically scattered electrons interact with the sample, they are deflected where the angle of deflection is proportional to the electron potential of the atom, and experience phase shifts that are converted into variations of intensity relative to the background of unscattered electrons. Electron waves in phase result in constructive interference and an intense signal which form the dark regions on the image. Out of phase electrons result in destructive interference which form the light regions on the image. The wave of elastically scattered electrons is dispersed in multiple angles which are focused on magnetic lenses, primarily the objective lens to form an image which is magnified by intermediate and projector lenses and then projected onto a detector (direct electron detector or charge-coupled device (CCD) camera) (Fig 3.2 (b)) The projection image which is a capture of the electron intensities across the protein sample is converted into a multi-frame

movie that is then processed as broadly illustrated in Fig 3.4 to obtain the final charge density map.

Phase contrast transfer function (CTF)

The CTF is a sine function in reciprocal space that describes how details of the object are transferred to the image during the imaging process (Cong and Ludtke 2010, Cheng, Grigorieff et al. 2015). It quantifies the extent to which lens aberrations, such as coma and astigmatism, and the phase shift distort the image. The CTF is described by Equation 3.1. Estimating and correcting for the CTF allows for the compensation of these optical defects on the final image, thereby improving image quality and accurate resolution estimation.

Equation 3.1

$$CTF(f) = \sin[-\pi(\Delta z \lambda q^2 + C_s \lambda^3 q^4/2)]$$

Where C_s = spherical aberration coefficient

ΔZ = defocus

q = spatial frequency

λ = electron wavelength

To identify the optimal protein concentration and select the best grid for high-resolution data collection, all grids were imaged at low magnification under cryogenic conditions.

The vitrified grids with different concentrations of the protein were screened on a Tecnai F20 (FEI, USA) with a DE16 direct electron detection camera using a single-tilt cryo-holder operated at 200 kV. Data were collected using SerialEM software (Mastronarde 2003) with the low dose functionality enabled. The ice thickness, presence of contaminants, and even distribution of particles and their orientations were assessed. Forty-five frame movies were taken at a dose of 45 e/Å^{2}s with an exposure time of 0.91 s. A magnification of 29 000x, pixel size of 1.53 Å, and defocus range of 2–4 µm were used.

High-resolution images were collected remotely at the electron Bio-Imaging Centre (eBIC) Diamond Light Source, UK, using EPU software (Thermo Fisher Scientific, USA); the parameters and equipment used are outlined in Table 3.1.1.

Table 3.3.1: High resolution data collection parameters for $CynD_{pum}$ E155R variant

Data collection parameters	
Microscope	Titan Krios II
Direct electron detector	Ametek-Gatan Bio-quantum-K3
Data collection mode	Counted super resolution
Magnification	105 000 x
Accelerated voltage	300 keV
Pixel (px) size	0.825 Å/px
Exposure time	2.04 s
Dose per frame	0.999 e^- /fraction
Total dose	45 e^- /Å^2/s
Dose rate over vacuum	15 e^- /px/s
Data collection mode	Superres
Defocus range	-3, -2.5, -2.0, -1.5, -1
Number of frames	44.96

Number of squares	22
Number of grids	1
Total number of movies collected	31418

3.2.2.3 Image pre-processing

Image processing was performed using Relion 4.0.1 (Scheres 2012). Multi-frame dose-fractionated movies were corrected for beam-induced motion using Relion's implementation of the MotionCor2 algorithm with dose weighting. The contrast-transfer function (CTF) was estimated using CTFFIND 4.1 (Rohou and Grigorieff 2015). Micrographs were manually curated and selected based on the quality of CTF estimation for each motion-corrected micrograph. Initial particle picking was performed using template-based picking with an initial model obtained from processing a small subset of the data "on-the-fly" during high resolution data collection. Particles were extracted from the micrographs with 6x rescaling resulting in particles with a pixel size of 4.95 Å/ px to reduce computing time. The extracted particles were subjected to multiple rounds of 2D classification using the VDAM algorithm (Kimanius, Dong et al. 2021) with 200 mini-batches and a regularisation parameter (T) of 2. Images were aligned with an in-plane sampling of 6, offset search range of 5 px, and offset search step of 1 px.

3.2.2.4 Single particle analysis (SPA)

The particles were split iteratively into 50 and 100 classes to exclude junk particles, assess particle quality, and identify potential sample heterogeneity. An initial model was generated *ab initio* by back projection without imposing symmetry using the variable-metric gradient descent (VDAM) algorithm with 200 minibatches 4. Two 3D classes were generated using particles from the best 2D classes and the initial model used for template-based picking as a reference (T=4, 25 iterations). The low-resolution initial model was refined using 3D auto-refinement with C2 symmetry imposed, an initial low-pass filter of 60 Å was applied, and a mask diameter of 300 Å used. 3D reconstruction and resolution estimation were performed using Gold-standard Fourier Shell Correlation (GS-FSC) (Scheres 2012) with an FSC cutoff of 0.143, where particle data composing the low-resolution initial model are split in half with each half refined independently and cross-correlated iteratively until the resolution where both halves converge into a single solution.

Post-processing involved the creation of a solvent mask using Chimera (Pettersen, Goddard et al. 2004). This mask was applied in a subsequent iteration of 3D-autorefinement to eliminate unwanted signal and enhance the resolution of the map. Subsequently, the map underwent sharpening using automatic B-factor estimation in Relion. The CTF-refinement feature of Relion was used to estimate the anisotropic magnification, as well as the per-micrograph and

per-particle defocus. Additionally, estimation of beam tilt and trefoil was conducted to a minimum resolution of 30 Å.

Bayesian particle polishing (Zivanov, Nakane et al. 2019) was used to further estimate and correct the beam-induced motion correction for the individual particles composing the refined map using a subset of 10 000 particles. A helical symmetry search was performed in CryoSPARC 3.4.0 (Punjani, Rubinstein et al. 2017) by searching over a symmetry grid defined by helical rise with minimum and maximum helical rises of 14 Å and 16 Å, respectively. The minimum and maximum symmetry grid search values over the helical twist used were 76˚and 78˚, respectively.

3.2.2.5 Atomic model building, refinement, and validation

Model building was performed using the PredictAndBuild program in the PHENIX suite version 1.20.1-4487-000 (Afonine, Poon et al. 2018) using the refined and B-factor sharpened map and the amino acid sequence of the E155R variant of $CynD_{pum}$. The resulting model was analysed and manually refined using the real-space refinement function in Coot (Emsley and Cowtan 2004). Automatic refinement was performed using the real-space refinement program for CryoEM maps in PHENIX (Afonine, Poon et al. 2018) with global minimisation and flexible fitting using molecular dynamics in the ISOLDE plugin version 1.5 (Croll 2018) in UCSF Chimerax version 1.6 (Pettersen, Goddard et al. 2021) to correctly assign amino acid

geometries. Model fitting validation was performed using the real-space refinement program for CryoEM maps in PHENIX version 1.20.1-4487-000 (Afonine, Poon et al. 2018) and Molprobity structure validation software within the PHENIX suite (Chen, Arendall et al. 2010) by cross-correlating the model to the map. To monitor overfitting, the FSC of the model to map was assessed as the model was refined at three cut-off values: 0, 0.143, and 0.5.

3.3 Results

3.3.1 Image pre-processing and particle 3D reconstruction

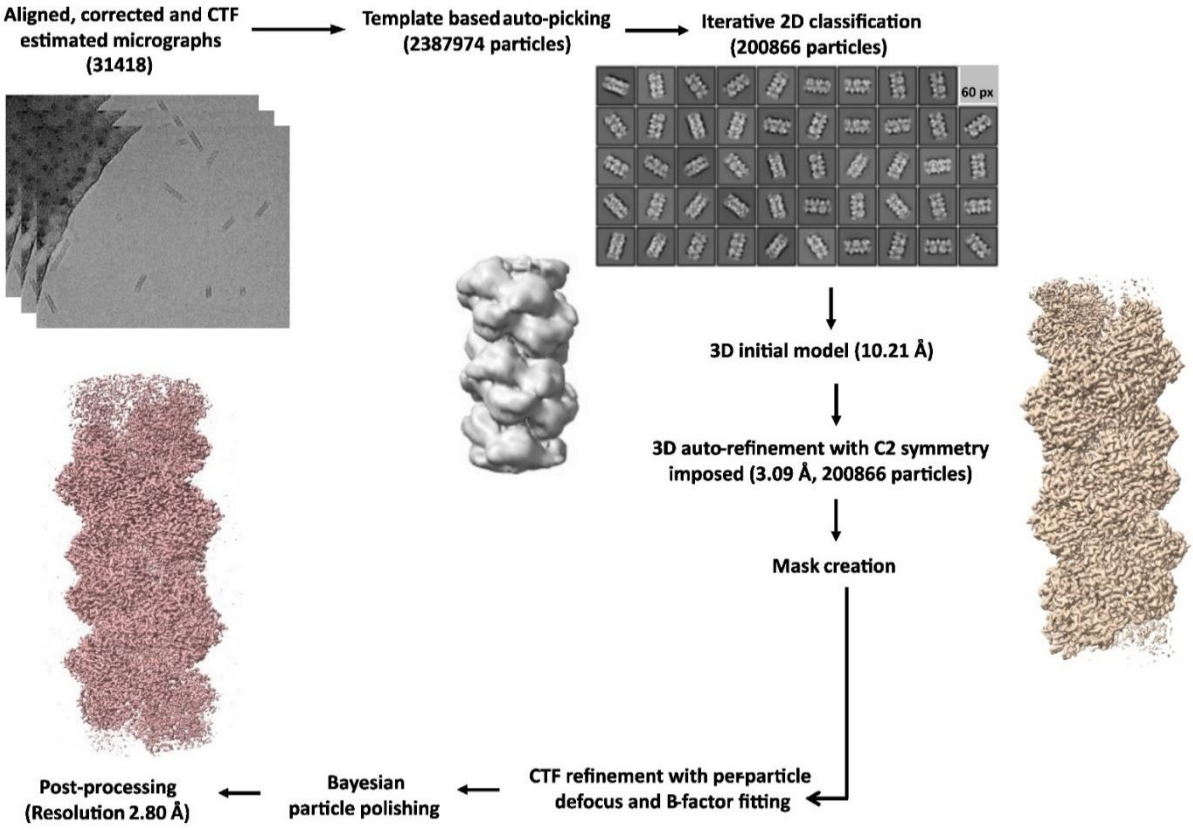


Figure 3.3.1: Schematic representation of image processing and 3D reconstruction used for the E155R variant of $CynD_{pum}$.

Table 3.3.2: Map reconstruction, model fitting, and refinement statistics.

3D Map reconstruction (EMD-19910)	
Initial number of particles	2387974
Final number of particles	200866
Box size (px)	360
Helical twist (Δϕ) (°)	-76.71
Helical rise (Δz) (Å)	15.62
Pitch (Å)	7.33
Number of subunits per turn	4.69
Resolution (Å)	2.8
Map sharpening B-factor (Å^2)	-101
Model composition (PDB ID 9ER3)	
Number of chains	19
Non-hydrogen atoms	94635
Amino acids (all/built)	340/318 (2-320)
Ligands	0
Fit-to-map	
Correlation coefficient (CC) (mask)	0.88
Correlation coefficient (CC) (box)	0.61
Correlation coefficient (CC) (peaks)	0.54
Correlation coefficient (CC) (volume)	0.83
FSC (model) 0, 0.143, 0.5 (masked)	2.27/ 2.45/ 2.8
Root mean square (RMS) deviations	
Bond lengths (Å)	0.003
Angles (°)	0.493
Model geometry	
Ramachandran Favoured (%)	97.94
Ramachandran Allowed (%)	2.06
Ramachandran Outliers (%)	0
Cβ outliers	0
Rotamer outliers	0.8
Cis proline/ general	11.8/0
Twisted proline/ general	0/ 0
Molprobity score	1.21
Clash score	6

3.3.2 Model fitting analysis and resolution validation

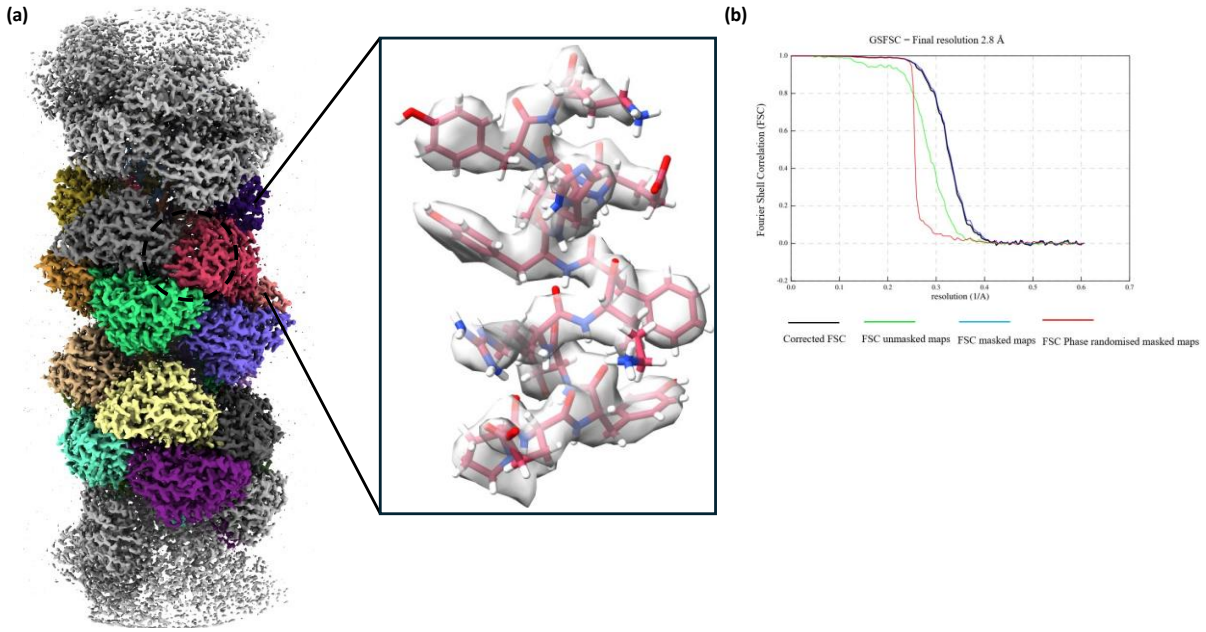


Figure 3.3.2: Assessment of model fitting and resolution validation (PDB ID 9ER3). (a) The quality of the model's fit into the charge density map is depicted in the zoomed-in version of the map. (b) The Fourier Shell Correlation (FSC) between the two half maps is illustrated, with the estimated resolution based on the phase-randomised masked map found to be approximately 2.8 Å.

3.3.3: Structural analysis

Overall structure

The E155R variant of CynDpum forms a left-handed helix with D1 symmetry, comprising 19 monomers resolved within the available density. The helix spans approximately 290 Å in length and has a diameter of about 125 Å. The subunits align along the central axis, resulting in a hollow central core measuring ~25 Å in diameter. Amino acids 2-320 are clearly resolved in each monomer, these correspond to 318 of the 340 amino acids comprising $CynD_{pum}$.

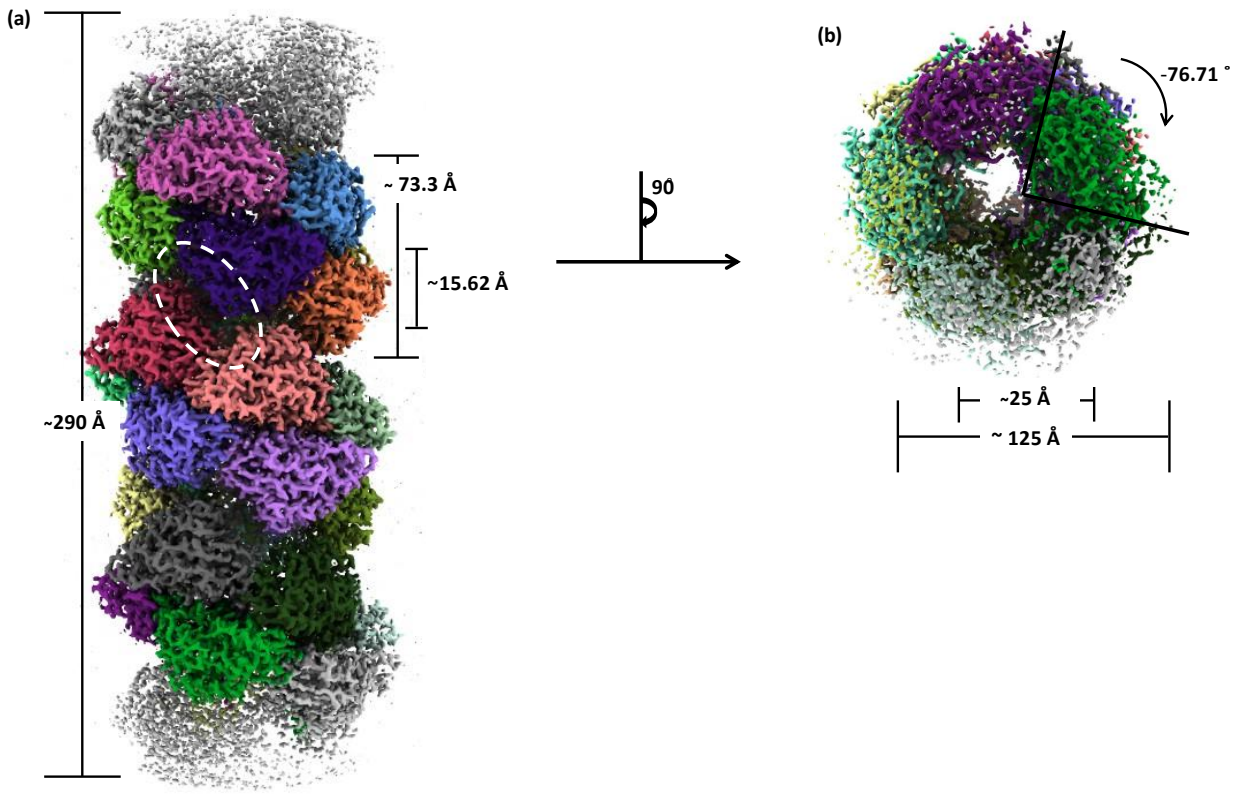


Figure 3.3.3: Overall atomic resolution map of the E115R variant of $CynD_{pum}$ at pH 7.4 solved at 2.8 Å resolution (EMDB ID EMD 19910). The different colours represent the different monomers resolved in the atomic resolution model. (a) Side-view. The variant forms a left-handed helix with D1 symmetry composed of 19 individual monomers. The length of the helix is ~290 Å, with a helical rise (Δz) of ~15.62 Å, a pitch of 7.33 Å and 4.69 subunits per turn and a helical twist (Δϕ) of -76.71˚. The dashed outline highlights the groove at which the E155R mutation occurs. (b) Top/end-on view. The helical assembly has a diameter of ~125 Å, the subunits are arranged along a central axis forming a central pore of ~25 Å.

Interfacial regions and the E155R mutation

Individual subunits are related to one another across interfaces A and C. These interfaces are stabilised by interactions with the C-terminal tail along the central core of the helix. Specifically, the C-terminal tail forms a four-strand anti-parallel β-sheet at the core of the assembly. The two middle strands interact with two monomers related to one another across

the A-interface thus stabilising the subunit at the A-interface. Two loops of the C-terminal tail occur perpendicular to the C-interface, stabilising the interface from within the interior of the helical assembly (Figure 3.3.4). The C-terminal tail is also characterised by a short α-helix (unresolved in chain M) and a loop that traverses most of the length of the D-interface from the interior of the helical assembly (Figure 3.3.6).

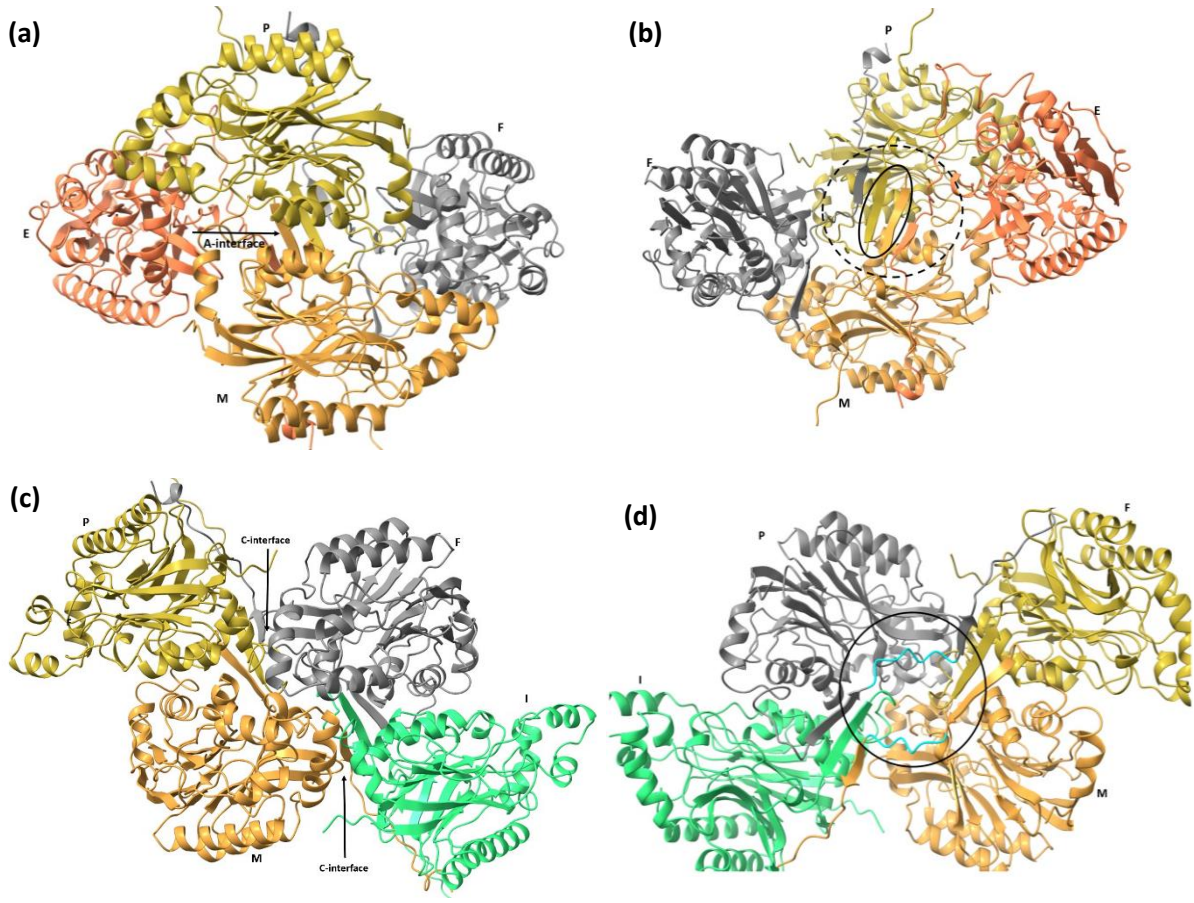
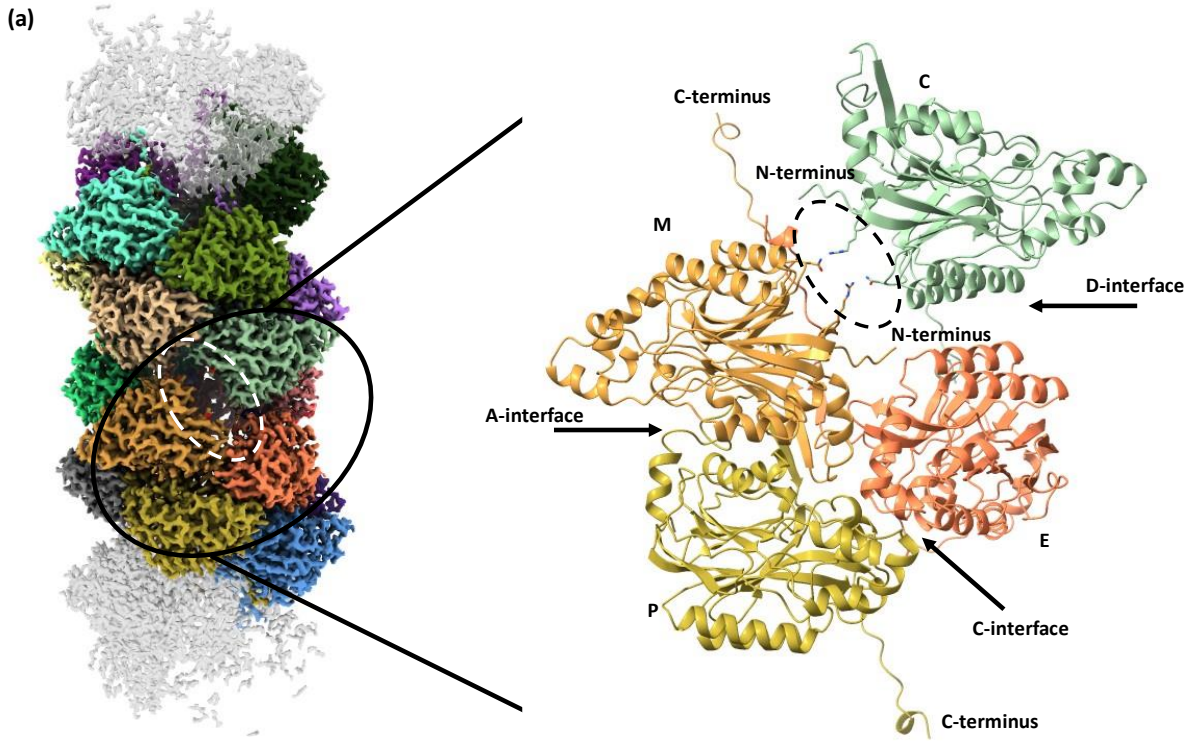


Figure 3.3.4: The structure of the E155R variant is formed and stabilised by interactions occurring between interfaces A and C (PDB 9E3R). (a) Exterior view of monomers M and P which associate across the A-interface. (b) Interior view of the A-interface. The four-strand β-sheet (dashed outline) stabilises the A-interface predominantly through interactions with two β strands (solid outline) belonging to the

two monomers associated across the interface. (c) Exterior view of subunits (MP and IF) associated across the C-interface. (d) Interior view of the C-interface. Loops from monomers, P and M (cyan) stabilise the interface from within the core of the helical assembly.

The E155R mutation is located along the groove between two subunits that are related to one another by a rotation of approximately 306.84 ° along the central axis near the C interface. Hydrogen bonding and ionic interactions occur between the guanidium side chain of R155 and the carboxyl and amino groups of N121 (Figure 3.3.5(b)).



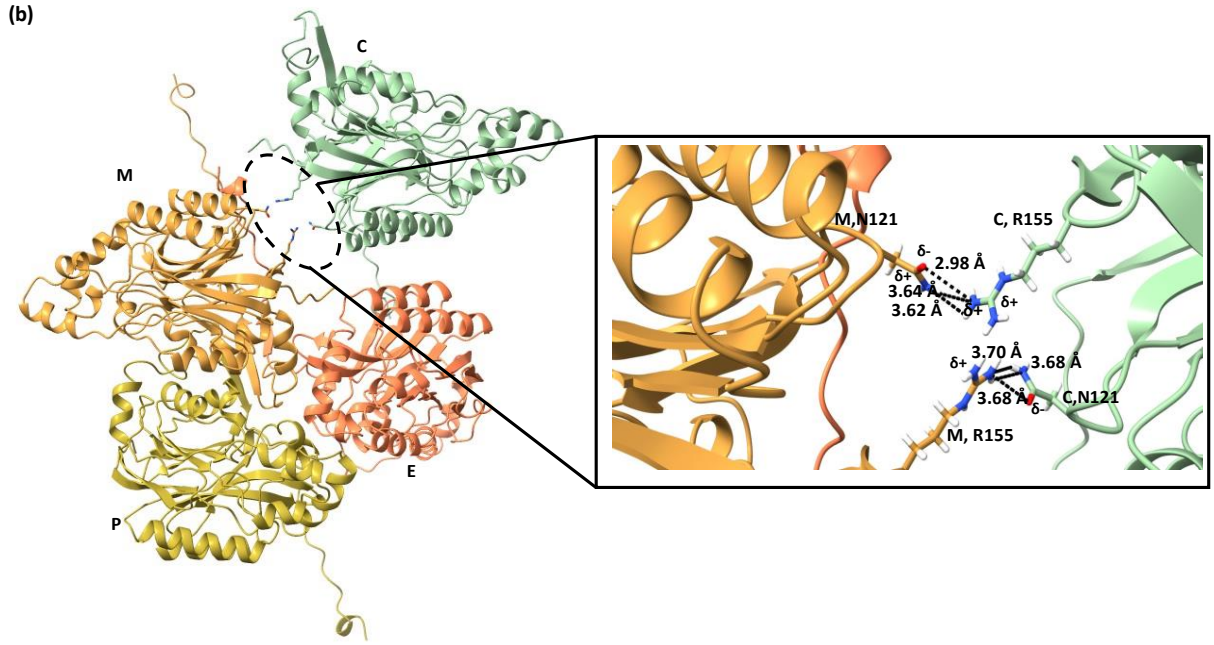


Figure 3.3.5: Interactions brought about by the E155R mutation across the groove. (a) The E155R mutation occurs across the groove between two subunits. (b) Ionic and hydrogen bonding interactions occur between the sidechains of R155 and N121 located on the monomer across the groove.

The N and C-termini

The C-terminal tail of the E155R variant of $CynD_{pum}$ has an additional α-helix and a short loop. The α-helix consists of amino acids Y316, E317 and D318. The remaining two amino acids resolved in the current model I319 and Q320 form part of a short loop. Both regions traverse part of the length of the D-interface (Figure 3.3.7).

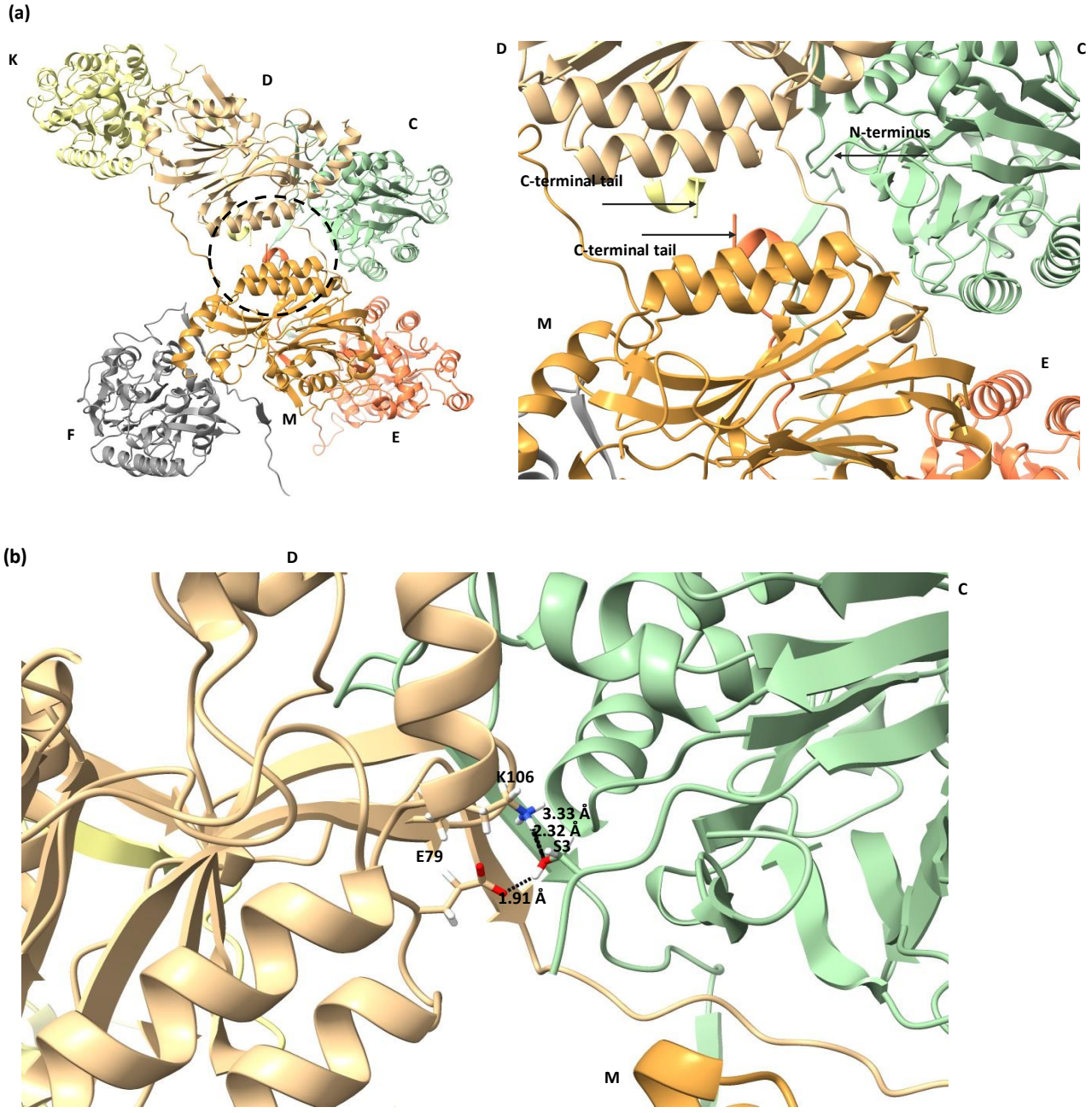


Figure 3.3.6: The N and C-termini of the CynD$_{pum}$ E155R variant. (a) and (b) The termini of the enzyme are closely related across a groove. (c) Potential interactions occur across the groove with amino acids of the N-terminus. Amino acids E79 and K106 form short range electrostatic and hydrogen bonding interactions with the lone pairs of electrons on the mainchain carboxyl group of the N-terminal S3.

The C-terminal tail of the E155R variant of $CynD_{pum}$ has an additional α-helix and a short loop. The α-helix consists of amino acids Y316, E317 and D318. The remaining two amino acids resolved in the current model I319 and Q320 form part of a short loop. Both regions traverse part of the length of the D-interface (Figure 3.3.7).

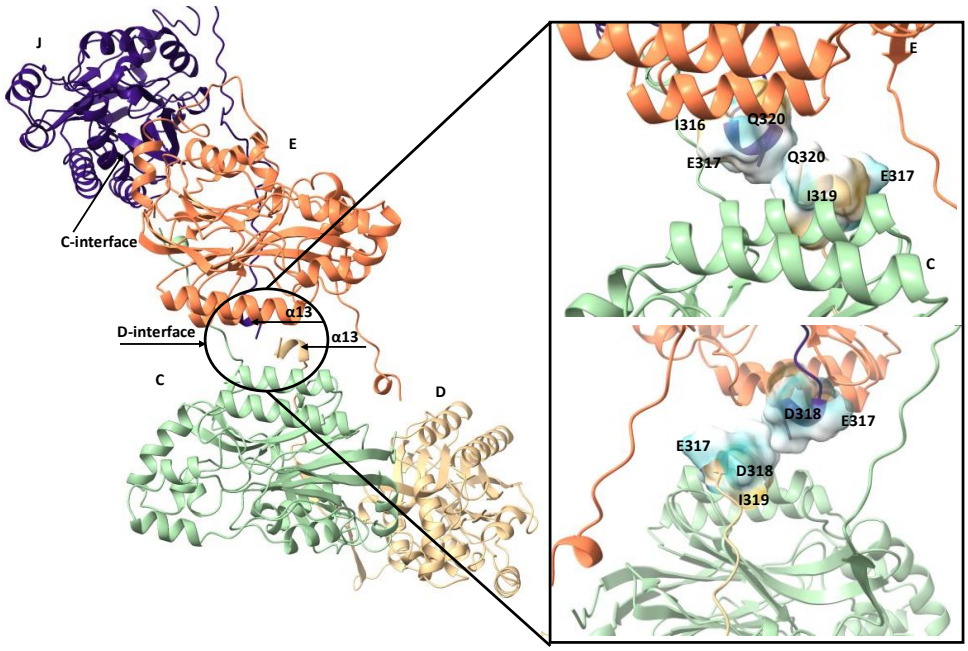


Figure 3.3.7: The C-terminus of the $CynD_{pum}$ E155R variant. The C-terminus in the current model has an additional α-helix (α13) comprising amino acids Y316, E317 and D318 at the C-terminus. An addition short loop that traverses part of the length of the D-interface is comprised of amino acids I319 and Q320.

Comparison of the atomic resolution structures of $CynD_{pum}$ and thermostable $CynD_{pum}$ variants Q86R/H305K/H308K/H323K and E155R

The $CynD_{pum}$ variants exist as larger helical assemblies than the wild-type. However, wild-type $CynD_{pum}$ and the Q86R/H305K/H308K/H323K variant have the same helical twist (-77°) which

is slightly smaller in the E155R variant (-76.71°). Moreover, the diameter and helical rise of the E155R variant are notably smaller than those of the wild type and Q86R/H305K/H308K/H323K variant.

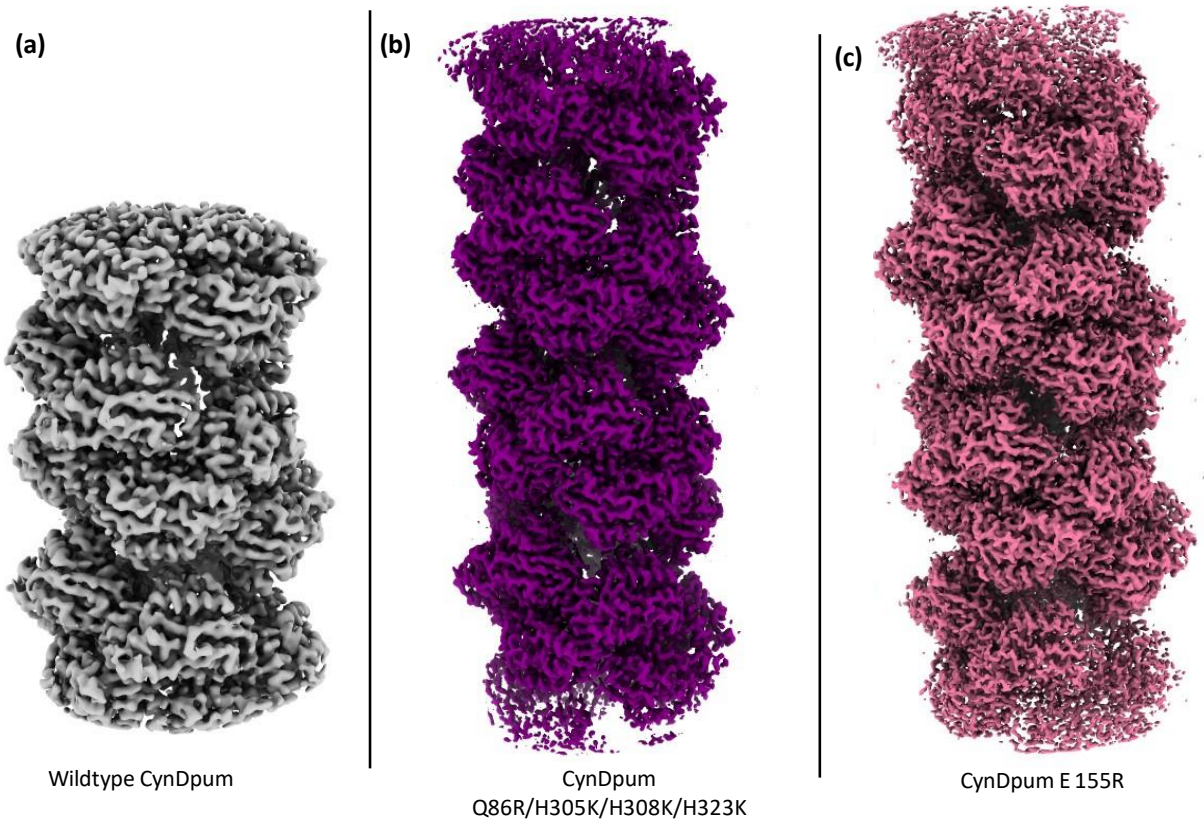


Figure 3.3.8: Comparison of the overall atomic resolution structures of $CynD_{pum}$ (EMD-17409) and $CynD_{pum}$ variants E155R (EMD-19910) and Q86R/H305K/H308K/H323K (EMD-16437). The overall charge density maps of the three proteins are depicted. Wild-type $CynD_{pum}$ is illustrated in grey, the Q86R/H305K/H308K/H323K variant in purple, and the E155R variant in pink.

Table 3.3.3: Comparison of the structural and helical parameters and the solvent conditions of $CynD_{pum}$ and $CynD_{pum}$ variants Q86R/H305K/H308K/H323K and E155R

	Wildtype CynDpum	Q86R/H305K/H308K/ H323K	E155R
EMDB ID	EMD-17409	EMD-16437	EMD-19910
PDB ID	8P4I	8C5I	9ER3
Resolution (Å)	3.83	3.15	2.78
Diameter (nm)	13.0	12.2	12.0
Helical rise Δz (Å)	16.7	16.9	15.62
Helical twist ϕ (˚)	-77	-77	-76.71
Pitch (Å)	75	75	73.3
Number of subunits per turn	4.6	4.6	4.69
Length of helix (nm)	18.9	30.0	29.0
pH	5.4	5.4	7.4

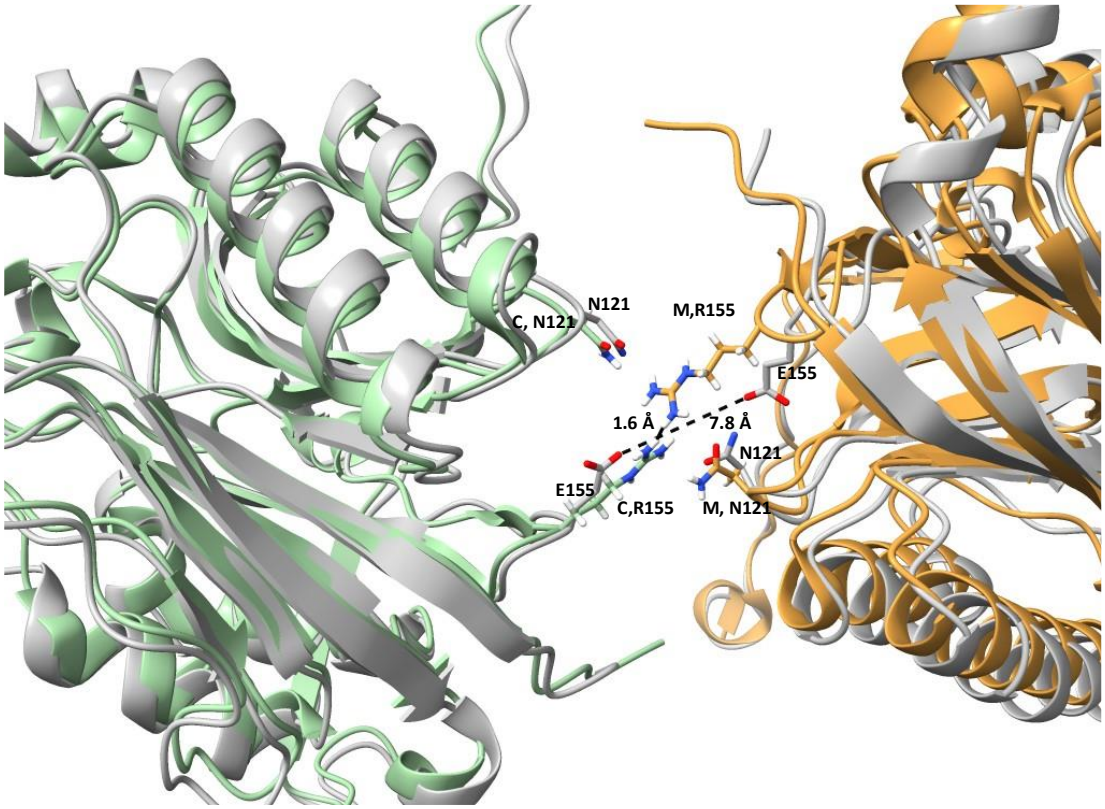


Figure 3.3.9: Comparison of position 155 between wildtype CynD$_{pum}$ and the E155R variant. Wildtype CynD$_{pum}$ (grey) and CynD$_{pum}$ E155R (green and brown) are shown superimposed, where the E155R mutation occurs. E155 on wild-type CynDpum is shorter, hindering any interactions across the groove from occurring. For clarity only the chain IDs belonging to the variant are labelled. Due to the altered helical twist of the E155R variant the two models do not exactly align.

Attempts to create a variant with mutations E155R and Q86R with potentially increased thermostability resulted in insoluble protein (Chapter II, section 2.4.3.3). The E155R variant leads to a reduced. The atomic resolution structure of the E155R variant of $CynD_{pum}$, has a smaller helical twist compared to the wildtype. One of the structural consequences of the reduced helical twist is reduction of the distance between the helices of the D-interface and the C-terminal tail. The reduced space at this region means that the space available for the introduction of an arginine is reduced and that addition of an amino acid with a bulky sidechain such as arginine would result in steric clashes at this region, potentially leading to inactivity.

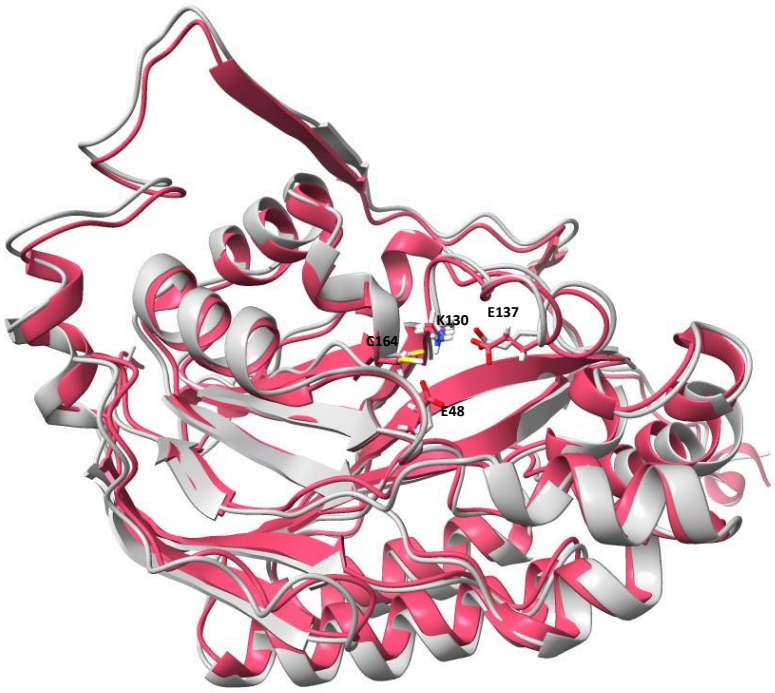


Figure 3.3.10: Comparison of the substrate binding site and catalytic amino acids of wildtype $CynD_{pum}$ and its E155R variant. Wildtype CynDpum (grey) and CynDpum E155R (purple and magenta) are shown superimposed. The conformation of the substrate binding site and the conformations of the catalytic amino acids remain unchanged.

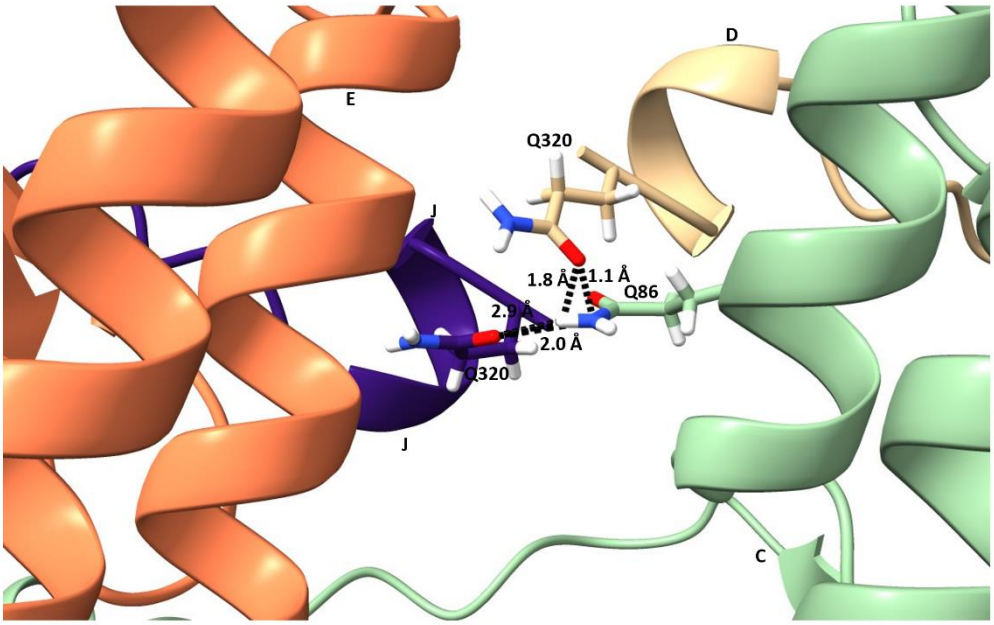


Figure 3.3.11: Potential basis for the observed insolubility of the Q86R/E155R variant of $CynD_{pum}$. Electrostatic and hydrogen bonding interactions occur between Q86 of the α4 helix on the D-interface and Q320 at the C-terminal tail.

3.4 Discussion

The thermal stability of the E155R variant is attributable to the introduction of a network of electrostatic and hydrogen bonding interactions between the two R155 amino acids and N121 on of the two each monomer interacting across the groove (Figure 3.3.5(b)). These interactions result in the formation of a new network of hydrogen bonds and ionic interactions across the groove, which in turn reduces the helical twist of the enzyme by 0.29 degrees. As a result, two distant monomers are brought closer together, leading to a narrower distance

across the groove and a more compact assembly with a smaller diameter of approximately 125 Å compared to the 130 Å diameter of the wild-type.

The increased number of interactions stabilising the assembly relative to the wildtype, along with the added structural rigidity of the assembly account for the observed thermostability. Moreover, the mutation leads to the formation of a larger homo-oligomer, characterised by a longer average fibre length and thus, molecular mass, similar to what has been observed for the Q86R variant (Wang, Watermeyer et al. 2012, Mulelu 2017). The increased size of the molecule may contribute to its increased resistance to thermal stress, since there is now a higher energy barrier to overcome due to the additional interactions stabilising the larger homo-oligomer. In contrast, on the wild-type, the groove is wider, and the two E155 amino acids are located too far apart for any interaction with the N121 on the monomer across the groove to occur (Figure 3.3.9). The Q86R/H305K/H308K/H323K variant has the same helical twist (-77 °) as wild-type $CynD_{pum}$, positioning E155 at the same location as in the wild-type.

The reduced helical twist (-0.29 °) does not affect the substrate-binding pocket or the arrangement of the catalytic tetrad amino acids in space (Figure 3.3.11). However, this finding suggests the potential for further exploration of helical twist alterations in CynDpum, which may lead to the enzyme catalysing different substrates, as has been demonstrated in certain plant nitrilases (Woodward, Trompetter et al. 2018, Mulelu, Kirykowicz et al. 2019). This

presents a promising avenue for engineering microbial nitrilases with altered substrate specificities. Several $CynD_{pum}$ mutations occurring at the C-interface, have been shown to result in long fibres with altered helical twist, and were inactive towards cyanide (Mulelu 2017).

The $CynD_{pum}$ E155R reconstruction provides more insight into the C-terminal tail region of the enzyme, revealing a short α-helix (Y316-I319) in addition to the β-sheet, that has been observed in the only other atomic resolution structure of a nitrilase that has been solved, Nit4 from *Arabidopsis thaliana* (PDB-ID 6I5T) (Mulelu, Kirykowicz et al. 2019). The aliphatic portion of amino acids Y316 and I319 project into the groove, forming a hydrophobic environment. The sidechains of E317 and D318 project towards the hollow core of the assembly interacting with the amino acids on α4 helix of the D-interface, the N-terminus, and the C-terminal tail (Figure 3.3.7). The N-terminus is related to the C-interface and C-terminal tail (Figure 3.3.6 (b)). In the current structure, a hydrogen bonding and ionic interaction exist across the groove between S3 and E79 at the C interface, joining the groove near the C-interface with the N-terminus.

The introduction of E155R to the same expression construct as the Q86R mutation resulted in an insoluble protein, suggesting improper folding (Chapter II section 2.4.3.3). Although the S29A/Q86R variant was partially inactive, the introduction of N119R restored its activity.

Therefore, we hypothesised that mutations Q86R and E155R are non-synergistic and disrupt proper formation of the quaternary structure. According to the atomic resolution structure of both variants, this may be due to two plausible reasons. First, potential steric clashes occurring between the α4 helix, where the Q86R mutation is located, and the C-terminal tail due to the introduction of a bulkier sidechain (R86) to this region, which is narrower in the E155R variant due to the reduced helix twist (3.3.12). Secondly, Q86R introduces rigidity at the conventional D-interface, which can hinder the flexibility required for the formation of an assembly with an altered helical twist brought about by the E155R mutation. This rigidity may impede proper folding once a sufficient number of subunits have been added to create the interactions between the C-terminal tail and R86.

Chapter IV

Structural characterisation of cyanide dihydratase from *Pseudomonas stutzeri* AK61 (*Stutzerimonas stutzeri*)

4.1 Introduction

The bacterial strain *Pseudomonas stutzeri* AK61, (recently reclassified to *Stutzerimonas stutzeri*) capable of converting cyanide to ammonia and formic acid was initially isolated from the wastewater of a metal-plating plant in Japan. Cyanide dihydratase (CynD) or cyanidase was shown to be the enzyme responsible for this conversion. The cyanide dihydratase from *Pseudomonas stutzeri* AK61 ($CynD_{stu}$) has an approximate monomeric size of 38 kDa and optimal pH and temperature at 7.5 and 37 ˚C, respectively (Watanabe, Yano et al. 1998). Another study reported the same optimal temperature and thermal stability of 23 ˚C, with a gradual loss of activity between 42 °C and 55 °C (Jandhyala, Willson et al. 2005).

$CynD_{stu}$ shares functional, sequence, and structural similarity with other nitrilases (Watanabe, Yano et al. 1998, Watanabe, Yano et al. 1998). Cyanide dihydratase from *Bacillus pumilus* C1 ($CynD_{pum}$) and $CynD_{stu}$ are homologous, sharing 79.8% sequence identity with sequence variability at the C-terminal tail. The conserved catalytic cysteine found in nitrilases is also present in $CynD_{stu}$ (C163), and its substitution has been shown to inactivate the enzyme, emphasising its catalytic significance (Watanabe, Yano et al. 1998). The catalytic amino acids, in addition to C163, were identified as K129 and E47 (Sewell, Berman et al. 2003).

Subsequent research using negative-stain transmission electron microscopy (TEM) and 3-dimensional (3D) reconstruction by single-particle analysis (SPA) revealed a left-handed spiral

homo-oligomer with two-fold symmetry, composed of 14 distinct monomers (Sewell, Berman et al. 2003). The monomers interact across interfaces A, C, D, and E (shown in Chapter I (Figure 1.2 (c)). The E-interface, unique to this nitrilase, is formed by specific electrostatic interactions between amino acids 266EID268 and 92RKNK95. These interactions cause a reduction in the spiral's outer radius from 6.2 nm to 5.1 nm at its termini, impeding the addition of further subunits due to steric hindrance, thus resulting in the self-termination of the spiral. Additionally, two insertion regions corresponding to amino acids 56-68 and 219-232 were identified as crucial to the quaternary structure of the enzyme and contribute to the left-handedness of the spiral, these insertion sites were postulated to be part of the C-interface (Sewell, Berman et al. 2003, Thuku, Brady et al. 2009).

To better understand the interfacial regions through which the quaternary structures of CynD are formed and stabilised, several variants of $CynD_{pum}$ and $CynD_{stu}$ have been generated (Sewell, Thuku et al. 2005, Crum, Park et al. 2015, Park, Mulelu et al. 2016). A notable discovery by Sewell, Thuku et al. (2005), was the retained activity of a $CynD_{pum}$ - $CynD_{stu}$ chimera, where the C-terminal tail of $CynD_{pum}$ was replaced with that of $CynD_{stu}$ (amino acids 287-340 of $CynD_{stu}$) while the inverse chimera $CynD_{stu}$ – $CynD_{pum}$ (amino acids 287- 330 of $CynD_{pum}$) was inactive.

The $CynD_{pum}$-$CynD_{stu}$ chimera was subsequently shown to be more thermostable than either $CynD_{stu}$ or $CynD_{pum}$ wildtype enzymes. The $CynD_{pum}$-$CynD_{stu}$ chimera retained more than 10% of its activity after incubation at 42°C for 40 hours and was active at pH 9.0, where both wild-type enzymes were inactive (Crum, Park et al. 2015). The improved thermostability was attributed to an amino acid motif, 312GERDST317 found in $CynD_{stu}$ but not on $CynD_{pum}$. Replacing corresponding amino acids in $CynD_{pum}$ with the 312GERDST317 motif resulted in 50% improvement in the thermostability of $CynD_{pum}$ (Abou Nader 2012).

4.2 Rationale

Although the structure of $CynD_{stu}$ was solved at a low resolution (20 Å), it provided important information about the enzyme's structure. However, the absence of an atomic-resolution structure prevents the visualisation of individual amino acids and their chemical interactions, particularly at the C-terminal tail. Additionally, the interest in $CynD_{stu}$ in this study was due to its reported self-termination, existing as determinate 14mer spirals induced by electrostatic interactions at the E-interface. Previous studies have hypothesised that the quaternary structure of microbial nitrilases consisting of long helical assemblies (or fibres) may be an indicator of thermostability and activity at alkaline pH (Wang, Watermeyer et al. 2012, Crum, Park et al. 2015, Mulelu 2017). Consequently, the objective of this chapter is to gain further insight into the structure of $CynD_{stu}$ and the specific interactions stabilising the interfacial regions through which the spiral assembly is formed specifically the C-terminal tail region, with the aim of understanding the basis of thermostability in spiral and helical nitrilases. This chapter addresses objectives 4 and 5 of this study.

4.3 Materials and methods

4.3.1 Cloning

The $CynD_{stu}$ gene insert (GenBank ID D82961.1) was synthesised and cloned into the *pet26b* plasmid with an N-terminal His_6 tag. Synthesis of the expression construct was performed by Genscript (www.genscript.com).

4.3.2 Recombinant protein expression and purification

Recombinant protein expression

The *pET26b-cyndstu* recombinant plasmid was transformed into BL21 (DE3) cells as outlined in Section 2.3.4. A single colony was used to inoculate 5mL of LB medium supplemented with 50 µg/mL kanamycin, and grown overnight at 37°C. The initial culture was used to inoculate 1 L of LB medium supplemented with 50 µg/mL kanamycin, and the cell culture was incubated in an Innova 44 incubator (New Brunswick, USA) at 37°C with shaking at 200 rpm until an optical density (OD) at 600 nm of 0.4-0.6 was reached. Target protein expression was induced by the addition of 100 mM isopropyl β-d-thiogalactopyranoside (IPTG), followed by further incubation for 18 h at 20°C with shaking at 200 rpm. Following recombinant protein expression, cells were harvested by centrifugation using a Beckman Coulter Avanti J26 XP centrifuge fitted with a JLA 8.1 rotor, at 4°C for 15 min at 4500 × g. The cell pellet was

resuspended in binding buffer (20 mM imidazole, 50 mM Tris-HCl, and 350 mM NaCl, pH 7.4) and stored at -20°C.

Cell lysis

The cells were lysed by sonication using a Vibracell ultrasonic processor (Sonics, USA) with a blunt-tip horn. Cell disruption was conducted for 4 min and interspersed with 15 s of cooling and 10 s of lysis. The cells were kept on an ice slurry during sonication.

Protein purification

Protein purification was performed using a two-step protocol. The target protein was isolated using immobilised metal affinity chromatography (IMAC), followed by size exclusion chromatography (SEC) as the final polishing step. All purification procedures were performed at 4°C to maintain protein stability.

Immobilised metal-ion affinity chromatography (IMAC)

IMAC was used to isolate the His_6-tagged target protein from the cell lysate using a HisTrap HP nickel affinity chromatography column (Cytiva, USA). The column was initially equilibrated with 25 mL (5 CV) binding buffer (25 mM Imidazole, 50 mM Tris-HCl, 300 mM NaCl, pH 7.4) and the sample was loaded at a flow rate of 1 mL/min. Subsequently, the column was washed with 200 mL of wash buffer (30 mM imidazole, 50 mM Tris-HCl, and 350 mM NaCl, pH 7.4) at

a flowrate of 5 mL/min to remove non-specifically bound proteins. Elution of the target protein from the resin was achieved by adding elution buffer (50 mM Tris-HCl, 300 mM NaCl, and 500 mM imidazole, pH 7.4) using a step gradient and flowrate of 3 mL/min. The eluted fractions were identified by monitoring the absorbance at a wavelength of 280 nm. Each fraction was collected, and a 25 µL aliquot of each fraction was analysed for protein purity and expected monomeric size using sodium dodecyl sulphate-polyacrylamide gel electrophoresis (SDS-PAGE). Subsequently, fractions containing the target protein were pooled and concentrated to 5mL using an Amicon centrifugal filter (Merck, USA) with a 100 kDa membrane size limit.

Size exclusion chromatography (SEC)

SEC was performed by loading the 5 mL sample purified using IMAC onto a HiPrep™ 16/10 Sephacryl™ S300 HR gel filtration column (Cytiva, USA) with SEC buffer (360 mM NaCl, 10 mM Tris-HCl, pH 8.0) as the mobile phase. Following SEC, fractions containing the target protein were pooled and concentrated using an Amicon centrifugal filter (Merck, USA) with a 100 kDa membrane filter cut-off. The same centrifugal filter was used for buffer exchange into citrate buffer (10 mM citrate, 50 mM NaCl, pH 5.4) by washing the concentrated sample three times with citrate buffer. The sample was then diluted 1:40 into the pH 8.0 SEC buffer following a previous protocol (Sewell, Berman et al. 2003).

The purification of $CynD_{stu}$ was subsequently optimised by changing the pH of all buffers from 8.0 to 7.4. Following IMAC, the sample was purified using SEC with SEC buffer (50 mM Tris-HCl, 50 mM NaCl, pH 7.4) as the mobile phase. Fractions corresponding to an increase in absorbance at 280 nm were analysed for purity using SDS-PAGE as detailed in Chapter II section 2.3.6.1, using the SeeBlue Plus2 protein molecular weight standard (Thermo Fisher Scientific, USA) as a reference. The protein concentration was determined according to the procedure outlined in Chapter II section 2.3.6.2.

4.3.3 Activity and biophysical analysis

Enzyme activity and biophysical analyses were conducted as described for $CynD_{pum}$ and $CynD_{pum}$ variants in sections, 2.3.7.1 and 2.3.7.2, respectively.

In addition to computing the thermodynamic parameters for $CynD_{stu}$, thermal denaturation by nanoscale differential scanning calorimetry (nanoDSF) was used to assess the quality of $CynD_{stu}$ protein preparations.

4.3.4 Negative-stain transmission electron microscopy (TEM)

Protein samples (0.5 mg/mL) were loaded onto glow-discharged continuous carbon-coated grids (Agar Scientific, UK), incubated for 30s and the excess sample was blotted onto the edge of a wedge of filter paper. The sample was stained by adding 2% (v/v) uranyl acetate and

excess stain was removed from the grid. Images were captured using a Jeol JEM-2100 TEM (Jeol, Japan) at a magnification of 30 000x.

4.3.5 Determining the molecular weight (M_r) of $CynD_{stu}$ at pH 7.4

Size exclusion chromatography (SEC)

SEC was used as an initial crude estimate of the M_r of $CynD_{stu}$ by monitoring its retention time and comparing it to that of gel filtration standards (Figure A1). A SEC calibration curve was not constructed. M_r standards (50 µL) (Biorad, USA) were loaded onto a HiPrep™ 16/10 Sephacryl™ S300 HR (Cytiva, USA), pre-equilibrated with 240 mL (2CV) SEC buffer (50 mM Tris-HCl, 50 mM NaCl, pH 7.4).

Size exclusion chromatography multi-angle light scattering (SEC-MALS)

SEC-MALS was used to further characterise and determine the M_r of $CynD_{stu}$.

The run was calibrated using BSA (2 mg/mL) diluted in the SEC buffer. The purified sample was diluted to 2.5, 3.0 and 4.0 mg/mL in SEC buffer (50 mM Tris-HCl, 50 mM NaCl, pH 7.4) to a volume of 160 µL and loaded simultaneously onto a Superose ™ Increase 10/300 gel filtration column (Cytiva, USA) pre-equilibrated with 48 mL of SEC buffer at pH 7.4. The experiment was performed using an OMNISEC multi-detector SEC system equipped with a multi-angle light

scattering (MALS) detector (Malvern Panalytical, USA). SEC buffer (50 mM Tris-HCl and 50 mM NaCl, pH 7.4) was used as the mobile phase.

4.3.6 Cryogenic electron microscopy (CryoEM)

Sample preparation

Purified protein (100 µL) was prepared for vitrification. Potential aggregates were removed by centrifugation at 41 000 xg for 30 min at 4°C on a Benchtop Ultracentrifuge with a TLA-55 rotor (Beckman Coulter, USA). A 75 µL sample was taken from the centrifuged sample and vitrified.

Vitrification

Vitrification was carried out using a Vitrobot Mark I cryo-plunger (FEI, USA). A 4 µL protein sample was applied to a glow-discharged Holey carbon grid (R1.2/1.3) (Quantifoil, Germany), and immediately blotted for 2.5 s with a blot force of 2 MPa at 100% humidity and 4°C. The grid was then plunge frozen in liquefied ethane, clipped, and stored in liquid nitrogen. Particle distribution was optimised by vitrifying the protein sample at various concentrations (2.5, 3 and 4.0 mg/mL) and selecting the conditions with the best distribution, the 2.5 mg/mL protein concentration had the best distribution of protein particles and was used for high resolution data collection

Grid screening

Images of the vitrified grids were captured using a Glacios 2 cryo-TEM (Thermo Fisher Scientific, USA) equipped with a Falcon 4 direct electron detector operated at 200 kV and at 105 000x magnification.

High resolution data collection

Images of the vitrified sample were captured using a Titan Krios II cryo-TEM (Thermo Fisher Scientific, USA) equipped with a Bio-quantum-K3 direct electron detector and operated at 300kV using the parameters listed below.

Table 4.3.2: High resolution data collection parameters for $CynD_{stu}$

Data collection parameters	
Microscope	Titan Krios II
Direct electron detector	Bio-quantum-K3
Detector mode	Counting super resolution
Accelerated voltage	300 keV
Magnification	105 000 x
Pixel size	0.831 Å
Exposure time	2.27 s
Dose per frame	1 e^-
Total dose	50 $e^-/Å^2/s$
Dose rate	15.2 e^-/s
Defocus range	-3, -2.7, -2.4, -2.1, -1.8, -1.5
Number of frames	50
Number of squares	21
Number of holes	33741
Number of grids	1
Number of movies collected	25 100

Grid were screened and high-resolution data were collected at the electron Bio-Imaging Centre (eBIC) Diamond Light Source, UK, using EPU software (Thermo Fisher Scientific, USA).

4.3.7 3D reconstruction and single particle analysis (SPA)

Single particle analysis and 3D reconstruction were performed using CryoSPARC (Punjani, Rubinstein et al. 2017). A general outline of the procedure followed is shown in Figure 4.2.1.

Image pre-processing

Beam-induced motion correction, image alignment, and estimation of the contrast transfer function (CTF) were performed using Patch Motion Correction and Patch CTF correction, respectively. Manual exposure/ micrograph curation was based on the CTF statistics, relative ice thickness, and suitable defocus.

Single particle analysis, 3D reconstruction and map post-processing

A small subset of particles was picked manually and used to generate a 2D class template that was used to pick particles from the remaining micrographs using the template-based particle picking function in CryoSPARC. The blob picker function with a specified minimum and maximum particle diameter of 100 and 200 Å, respectively was also used for automated indiscriminate particle picking.

Picked particles were subsequently extracted using a box size of 256 Å, which was sufficient to encompass the longest reported axis of the $CynD_{stu}$ particle (20 Å). To generate homogeneous particle data and exclude low quality particles, we conducted multiple iterations of two-dimensional (2D) classification with varying class sizes. Following satisfactory 2D classification, three initial models were built *ab initio* using a class similarity value of 0.1. *Ab inito* reconstruction was also used to further eliminate junk particles. The *ab initio* models were refined using heterogeneous refinement. To identify conformational heterogeneity, 3D classification was performed for five classes. The number of particles at each stage of SPA are shown in Figure 4.4.6.

3D refinement was initially performed using homogeneous refinement and the Gold Standard Fourier Shell Correlation (GS-FSC) used to determine the alignment resolution of the two half-maps generated during the refinement.

Subsequently, refinement was performed using non-uniform refinement with the maximum alignment resolution for refinement manually set to 6 Å, without imposing symmetry (C1 symmetry) and without defocus or global CTF refinement. The resulting map was sharpened using the Autosharpen function in PHENIX version 1.20.1-4487-000 (Afonine, Poon et al. 2018).

Model building, refinement and validation

The initial model (single monomer) was built using AlphaFold (Jumper, Evans et al. 2021). The AlphaFold model was then fitted into the $CynD_{stu}$ cryo-EM density map using rigid-body fitting on UCSF Chimera (Pettersen, Goddard et al. 2004), followed by automated refinement using the real-space refinement function on PHENIX version 1.20.1-4487-000 (Afonine, Poon et al. 2018), and manual refinement on Coot (Emsley and Cowtan 2004). Additional refinement, including correction of rotamer assignments, clashes, and Ramachandran outliers, was performed using the ISOLDE plugin in USCF ChimeraX (Croll 2018, Pettersen, Goddard et al. 2021). The validity and quality of the model's fit to the map were evaluated using map-model cross-correlation at three different FSC cutoff values; 0,0.143 and 0.5 using PHENIX version 1.20.1-4487-000

4.4 Results

4.4.1 Protein purification and characterisation of $CynD_{stu}$

$CynD_{stu}$ was initially purified according to previously reported protocols (Sewell, Berman et al. 2003). In this study, purification of $CynD_{stu}$ in a pH 8.0 buffer yielded two distinct peaks of approximately 158 kDa and 670 kDa, as indicated in the elution profile of protein standards (Figure A1). Notably, neither of these peaks corresponded to the reported 14mer (532 kDa). The ~670 kDa fraction was unstable, rapidly dissociating into the smaller and predominant form (~158 kDa). Subsequent buffer exchange into citrate buffer (pH 5.4) after SEC at pH 8.0, resulted in the formation of a larger homo-oligomer than the expected 14mer.

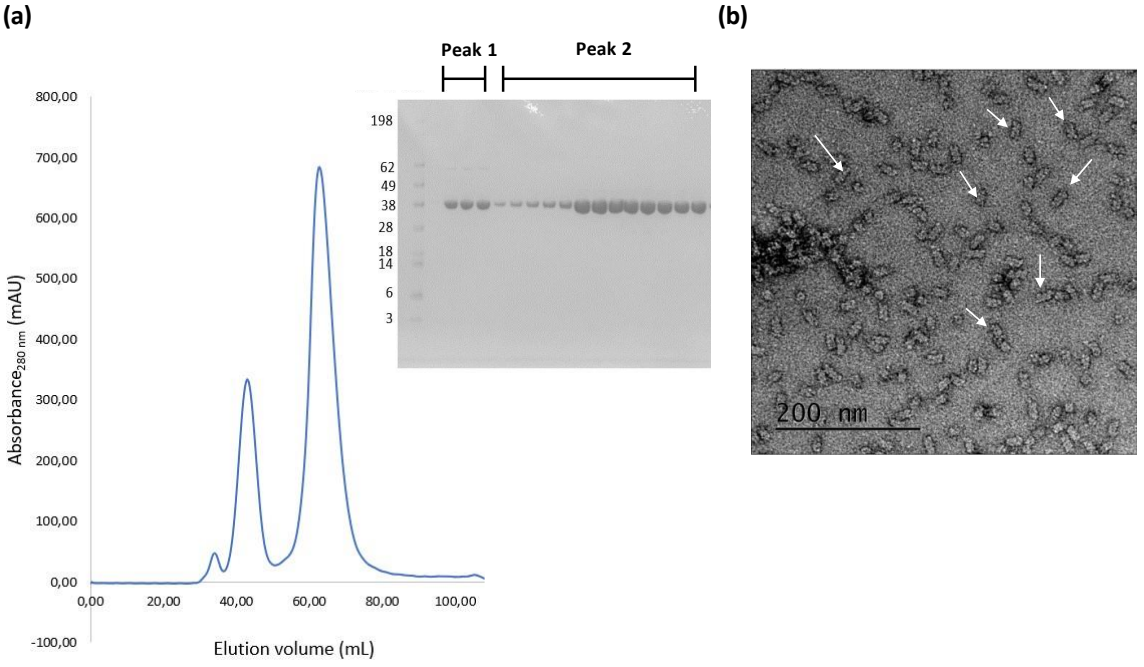


Figure 4.4.1: Purification of $CynD_{stu}$ at pH 8.0, followed by buffer exchange (pH 5.4 buffer) and dilution in pH 8.0 buffer. Initial protein purification was conducted at pH 8.0, leading to the formation of two distinct peaks in the SEC elution profile (a). SDS-PAGE analysis revealed a single monomeric species corresponding to ~38 kDa. (b) Negative stain TEM micrograph of the sample after incubation at pH 5.4 revealed larger multimer than the expected 14mer.

All purification steps were subsequently conducted at pH 7.4 which was reported by Jandhyala, Willson et al. (2005) as the optimal pH for $CynD_{stu}$ activity. Purification at pH 7.4 resulted in the elution of protein particles in a single peak corresponding to ~670 kDa. The enzyme was active at this pH (Chapter II, Figure 2.4.8).

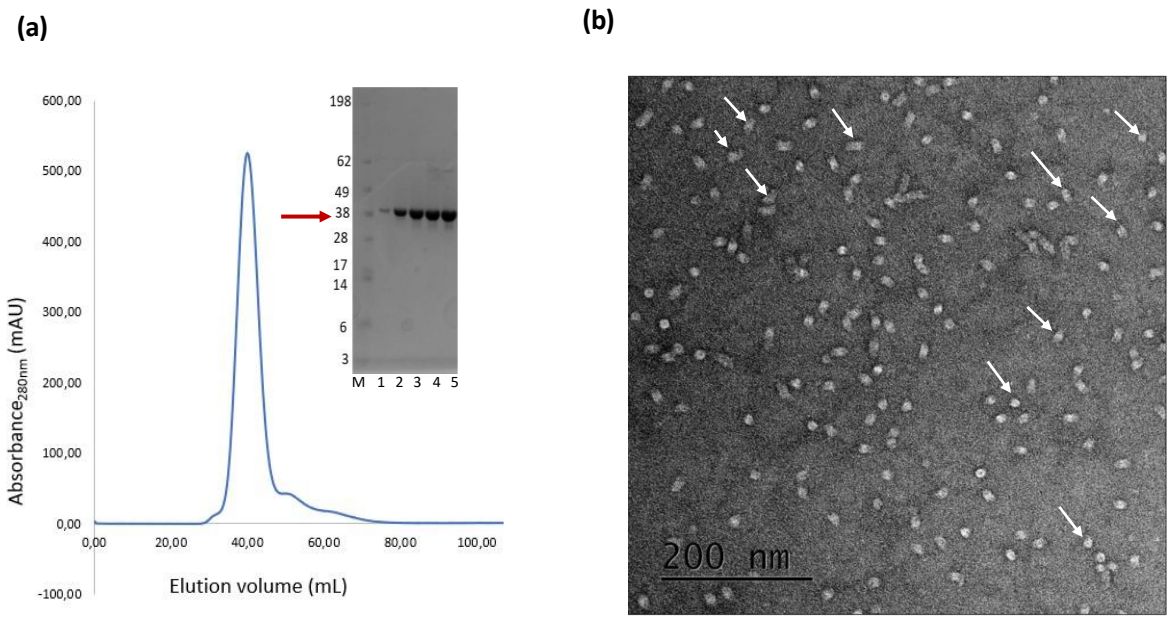


Figure 4.4.2: Purification of CynDstu at pH 7.4.(a) SEC elution profile with corresponding SDS-PAGE analysis and SEC fractions (inset). (b) Negative stain micrograph of a sample from the middle of the peak representing the most dominant oligomeric form in the sample.

$CynD_{stu}$ was found to adopt multiple oligomeric forms that are brought about by changes to solvent pH even under the conditions wherein it is reported to exist as 14mers. Therefore, to obtain a more precise estimate of its M_r at pH 7.4, size-exclusion chromatography with multi-angle light scattering (SEC-MALS) was employed. The SEC-MALS elution profile indicates that $CynD_{stu}$ primarily exists as a 22-mer at pH 7.4, with a potential smaller oligomer present indicative of inherent heterogeneity even at optimal pH.

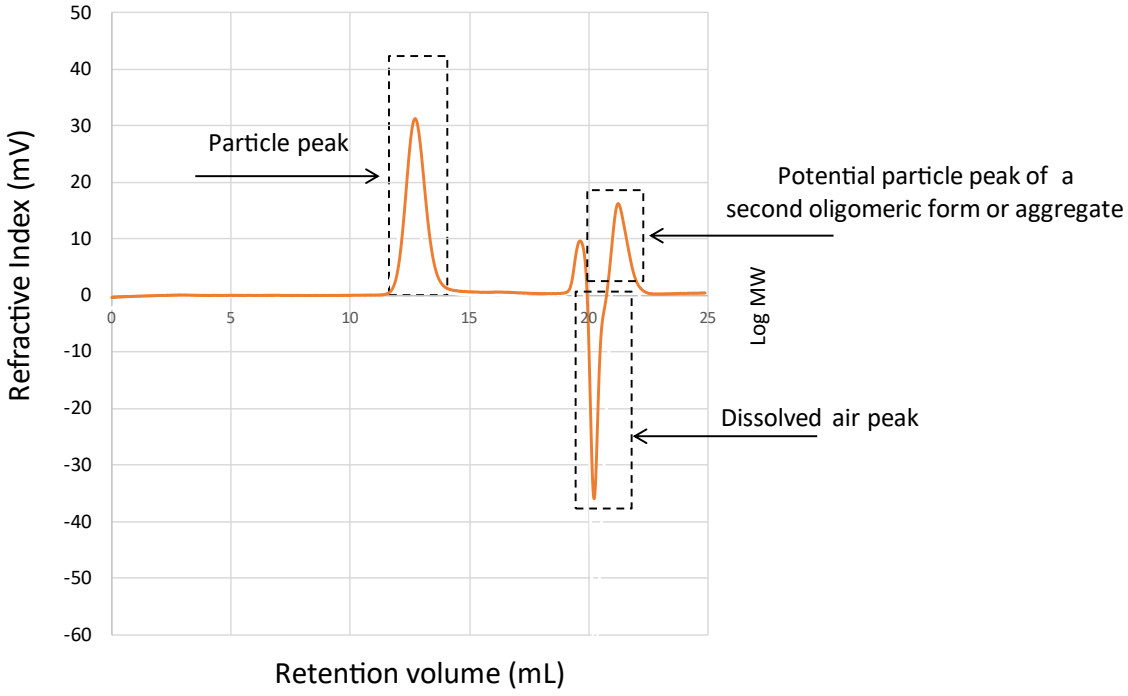


Figure 4.4.3: SEC-MALS refractive index trace and elution profile for $CynD_{stu}$ (3 mg/mL). SEC-MALS data revealed a dominant peak corresponding to a M_r of ~832 kDa (22mer) with a smaller peak of a potentially smaller oligomeric form.

To investigate whether $CynD_{stu}$ demonstrates the pH-dependent oligomerisation observed in its homologue $CynD_{pum}$ which exists as short helices at pH 8.0 and thermostable, inactive long

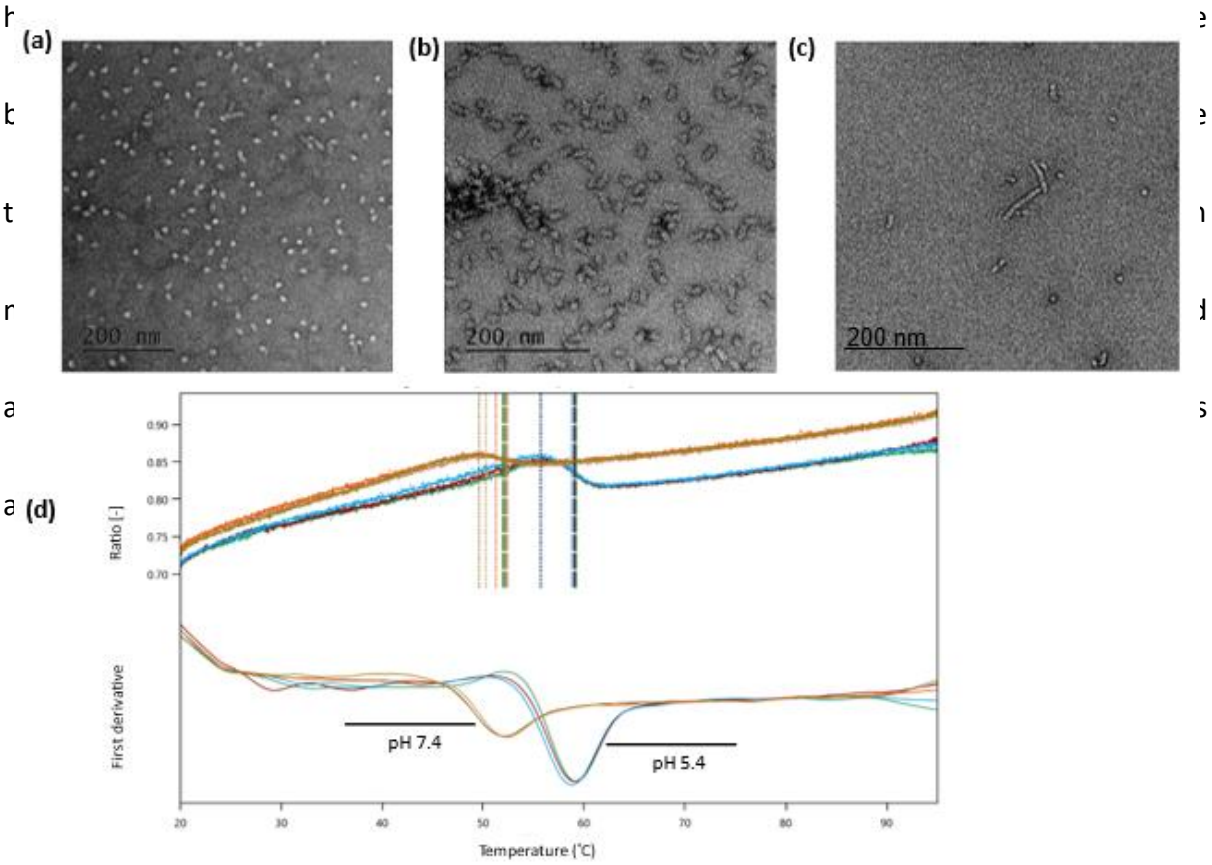


Figure 4.4.4: Negative stain TEM images of $CynD_{stu}$ at pH 7.4, 5.4 and diluted in pH 8.0 buffer and thermal denaturation of CynDpum at different pH. Negative stain TEM images of $CynD_{stu}$ at pH 7.4

(a), pH 5.4 (b), pH 5.4 and as described in previous protocols (c). (d) Thermal denaturation melting curve of $CynD_{stu}$ samples. The top panel shows a scatter plot of the fluorescence ratio FI350nm/FI330 nm vs. temperature. The bottom panel is a plot of the first derivative vs. temperature, indicating inflection points where changes in protein conformation occur, thus defining T_m. The denaturation experiment for each sample was performed in triplicate.

$CynD_{stu}$ undergoes pH-dependent oligomerisation, resulting in the formation of larger homo-oligomers at pH 5.4 (Figure 4.4.4 (c)) than at optimal pH. At pH 5.4, $CynD_{stu}$ exhibits a higher emitted fluorescence signal than at pH 7.4, which suggests the presence of additional tryptophan amino acids and therefore, additional monomers. Additionally, $CynD_{stu}$ has a melting temperature ($T_{m)}$ of 52.3 and 59.9 °C at pH 7.4 and 5.4, respectively. Both thermal melting curves exhibited single inflection points and smooth scattering, which is indicative of conformational stability in the solvent buffer.

The quaternary structure of the thermostable $CynD_{pum}$-$CynD_{stu}$ chimera was visualised using negative-stain TEM to determine whether it exists as short or long helical fibres. The Q86R mutation was also introduced to the $CynD_{pum}$-$CynD_{stu}$ chimera to investigate potential changes to its oligomeric structure.

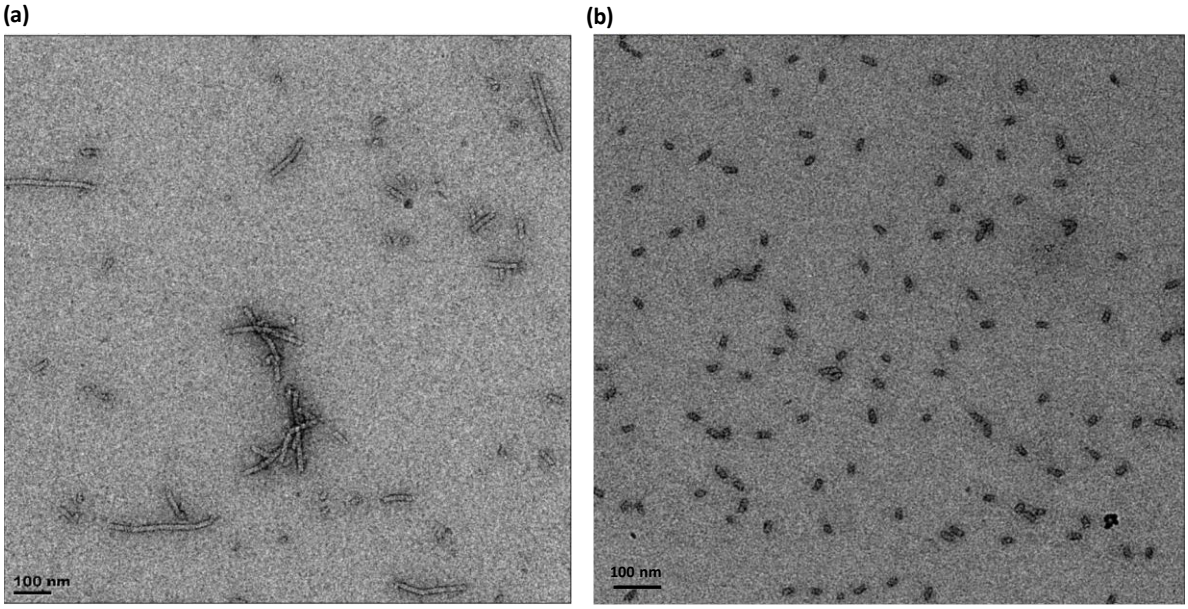


Figure 4.4.5: $CynD_{pum}$-$CynD_{stu}$ chimeric protein at pH 7.4. (a) The $CynD_{pum}$-$CynD_{stu}$ chimeric protein exists as long helical fibres at pH 7.4. (b) The introduction of the Q86R mutation leads to loss of the elongated helical quaternary structure, with the enzyme adopting an oligomeric form similar to that of wild-type $CynD_{stu}$. Images were captured at 19 000x magnification.

The chimeric protein exists as long helical fibres at pH 7.4, this quaternary structure is abolished by the introduction of the Q86R mutation which occurs on the $CynD_{pum}$ portion of the chimeric protein.

4.4.2 $CynD_{stu}$ 3D reconstruction

The atomic-resolution structure of active $CynD_{stu}$ at pH 7.4 was resolved using CryoEM, to investigate the chemical and structural factors contributed by the C-terminal tail of $CynD_{stu}$ to the thermostability of the $CynD_{pum}$ - $CynD_{stu}$ chimera, study its structure and the basis of its thermostability observed in Chapter II.

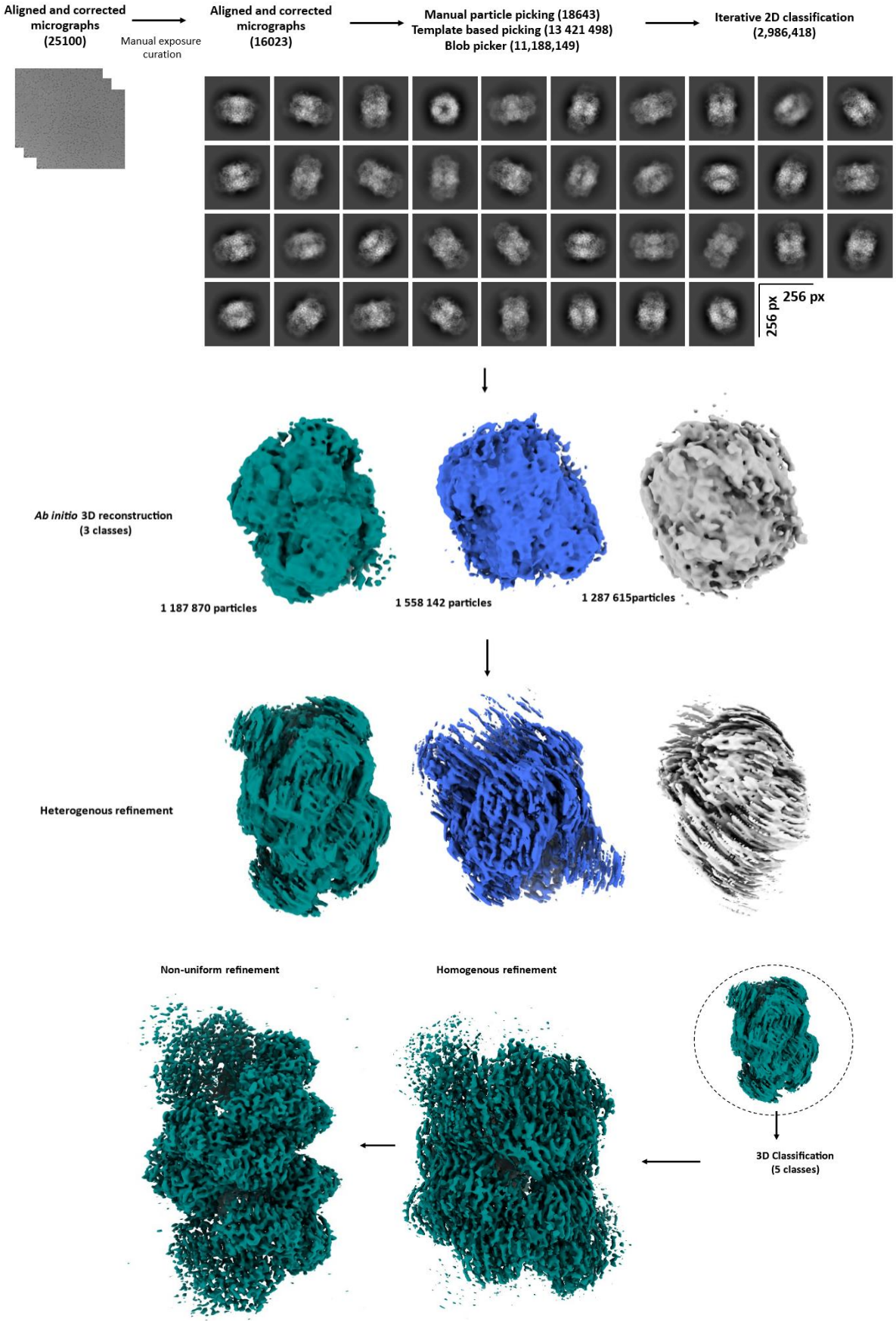

Aligned and corrected micrographs (25100)
Manual exposure curation
Aligned and corrected micrographs (16023)
Manual particle picking (18643)
Template based picking (13 421 498)
Blob picker (11,188,149)
Iterative 2D classification (2,986,418)
256 px
256 px
Ab initio 3D reconstruction (3 classes)
1 187 870 particles
1 558 142 particles
1 287 615particles
Heterogenous refinement
Non-uniform refinement
Homogenous refinement
3D Classification (5 classes)

Figure 4.4.6: Workflow of the SPA and 3D reconstruction approach used in this study. Maps shown in teal correspond to maps used in the processing and refinement of the final map used in this study.

Ab initio reconstruction of selected particles using CryoSPARC yielded three similar low-resolution maps (Figure 4.4.6). Notably, two of these maps had significant stretching along the central axis and flattening, indicative of potential anisotropy possibly attributed to dominant signal from end-on views or overfitting. Consequently, particles belonging to these classes were excluded from further analysis. The remaining class (depicted in teal on Figure 4.4.6) underwent additional refinement.

Heterogeneous and uniform 3D refinement, with and without restricting the alignment resolution of the best map, produced a stretched and flattened final map with an overestimated resolution. Despite attempts to address this issue by limiting the alignment resolution, overestimation persisted with both heterogeneous and homogeneous 3D refinement approaches. To ensure that the observed stretching of the refined map was not due to the inclusion of junk or broken protein particles introducing noise and reducing the amount of useful signal, multiple particle picking methods were explored. This included manual particle selection and 2D and 3D template-based picking (using the best *ab initio* model). Blob picking was also performed to minimise potential biases in particle selection.

Notably, the 2D classes generated from all particle picking approaches were identical, leading to the conclusion that biased particle picking were not the cause of the observed anisotropy. 3D classification was also conducted to assess whether sample heterogeneity contributed to the observed anisotropy. However, no significant differences were observed between the resulting classes.

Given the direction of stretching, we hypothesised that this effect was induced by the dominant signal from particles adopting an end-on view orientation. Subsequent reconstruction efforts aimed to mitigate the potential orientation bias associated with the end-on views by excluding them from further analysis. Since the spiral is formed by identical monomers arranged along the helical axis, the end-on orientation of the spiral can be obtained if all side-views are present, generating a useful final 3D reconstruction. However, this approach resulted in a substantially poor map with no clear secondary structure. Finally, non-uniform refinement was attempted, with maximum alignment resolution limits of 10 Å, 8 Å, and 6 Å. Reconstruction with a limited alignment resolution of 6 Å yielded a map with the secondary structure elements of 16 subunits that were clearly resolved. Although residual anisotropic effects persist in the current map, significant improvement in map quality resolution has been achieved through this refinement strategy.

Table 4.4.1: 3D reconstruction, model refinement and validation statistics for $CynD_{stu}$

Image processing	
Movies total/ curated	25100
Initial number picked particles	11,188,149
Final number of particles	1 187 870
Symmetry imposed	C1
Particle box size	256
Map resolution	3.26
Resolution determination method	FSC 0.143 cutoff
Map sharpening B-factor (Å^2)	- 117.56
Model composition	
Number of chains	16
Non-hydrogen atoms	94635
Amino acids (all/built)	315/340
Ligands	0
Water	0
Fit-to-map	
Correlation coefficient (CC) (mask)	0.72
Correlation coefficient (CC) (box)	0.52
Correlation coefficient (CC) (peaks)	0.44
Correlation coefficient (CC) (volume)	0.63
FSC (model) 0, 0.143, 0.5 (masked)	3.2/ 3.2/ 3.3
Root mean square (RMS) deviations	
Bond lengths (Å)	0.002
Angles ()	0.564
Model geometry	
Ramachandran Favoured (%)	97.94
Ramachandran Allowed (%)	2.06
Ramachandran Outliers (%)	0
Cβ outliers	0
Rotamer outliers	0.95
Cis proline/ general	11.8/0
Molprobity score	1.21
Clash score	3.76

4.4.3 The overall structure of $CynD_{stu}$

$CynD_{stu}$ exists as a left-handed spiral with C2 symmetry. The length of the spiral in the current model is ~ 19.2 nm (Figure 4.4.7(a)), the subunits associate along the central axis creating a

star-shaped central pore with a diameter of ~11.5 nm Figure 4.4.7(b)). The whole assembly has an outer diameter of ~ 12.9nm. Two monomers interact to form a dimer across the A-interface, this dimer (subunit) becomes the repeating unit along the length of the spiral with individual dimers interacting across the C-interface thus forming the homo-oligomer (Figure 4.4.7).

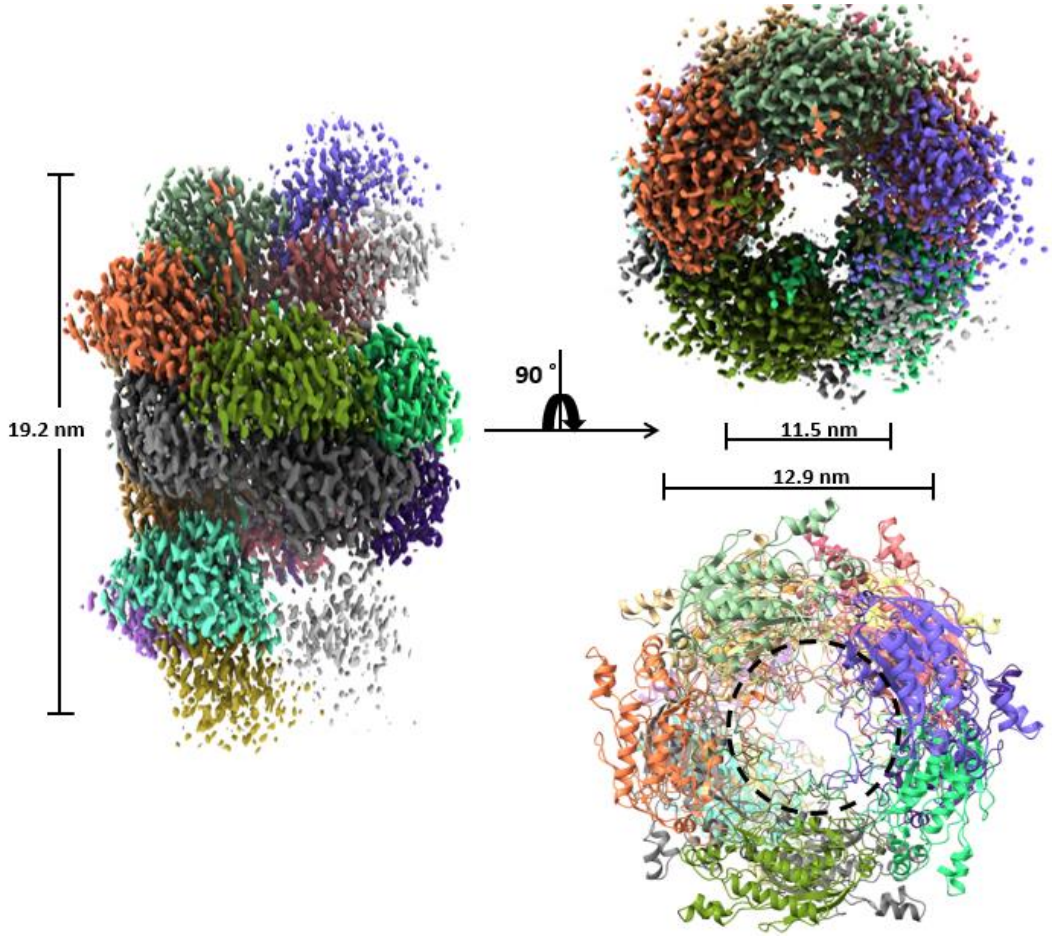


Figure 4.4.7: Charge density map of the overall structure of the $CynD_{stu}$ 16mer. $CynD_{stu}$ is a left-handed spiral with two-fold symmetry and is composed of 16 monomers (individually coloured) in the current reconstruction. (a) Side view of the spiral. Subunits are arranged along the central axis forming a central pore of ~ 17.8 Å (b) composed of the C-terminal tails (dashed outline).

The A-interface

The A-interface is comprised of helices α6, α7, and α8, and loops L11 and L19 (Figure.4.4.8 (b)). This interface is predominantly stabilised by hydrophobic and van der Waals interactions (Figure. 4.4.8(c)). Additionally, hydrophobic interactions from amino acids on the C-terminal tail contribute to the stability of the interface (Figure.4.4.8(d)).

The C-interface

The C-interface is predominantly stabilised by robust electrostatic interactions along its interface, with some hydrophobic interactions also contributing, particularly at positions close to the enzyme core. The specific regions forming this interface are shown in Figure 3.4.8 (b). Notably, the C-terminal tail primarily contributes electrostatic interactions that are crucial in establishing and maintaining this interface, as illustrated in Figure 4.4.9 (c).

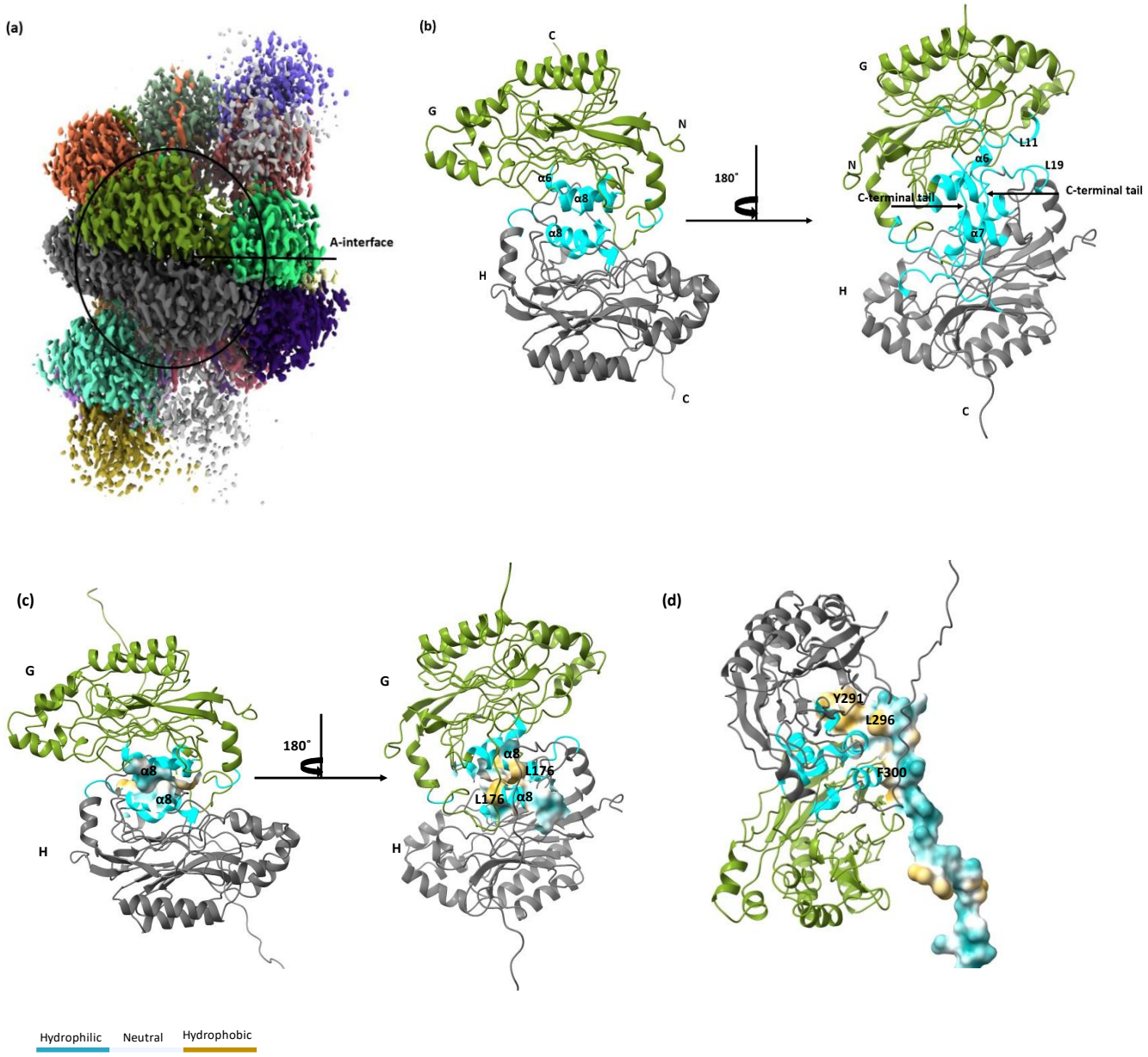


Figure 4.4.8 The A-interface. (a) The A-interface, across which two monomers interact to form a single dimer subunit shown here between monomers G and H. (b) The A-interface in CynDstu is formed by helices α6 and α8 on the exterior of the spiral, at the interior of the interface helix α6, and loops L11 and 19 comprise amino acids that interact with one another across the interface forming the dimer. (c) The C-terminal tails of each of the monomers interacting across the interface contribute to its formation and stabilisation. (d) The interface is further stabilised by hydrophobic interactions across with the C-terminal tail.

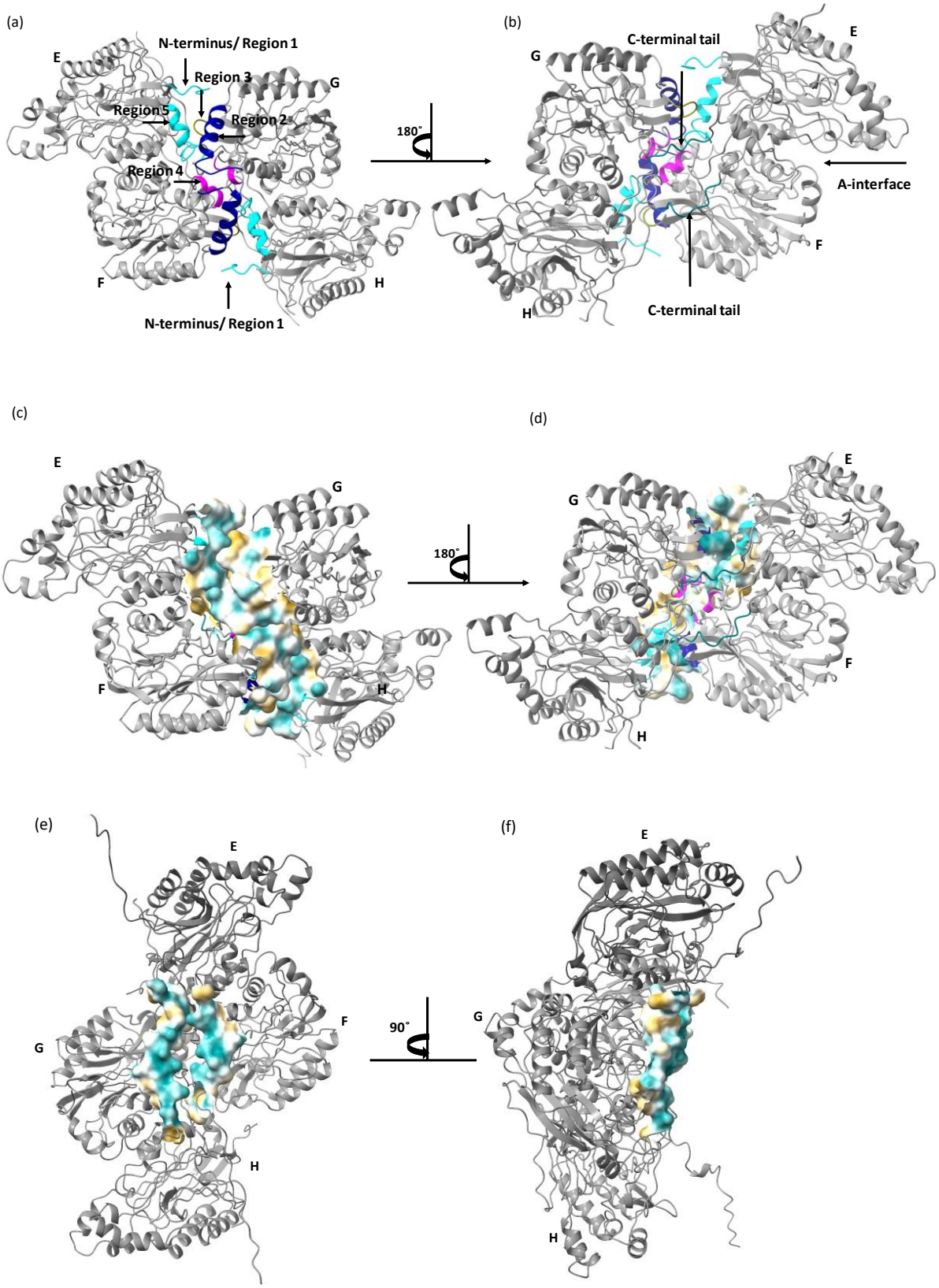

(a)
N-terminus/ Region 1
E
Region 3
G
Region 5
Region 2
Region 4
F
H
N-terminus/ Region 1
180°
(b)
C-terminal tail
E
G
A-interface
F
H
C-terminal tail
(c)
E
G
F
H
180°
(d)
E
G
F
H
(e)
E
F
G
H
90°
(f)
E
G
H

Figure 4.4.9: The C-interface. (a) The C-interface across which a pair of dimer subunits interact to form the $CynD_{stu}$ spiral is defined by five distinct regions. (b) These regions are mirrored across the 2-fold symmetry axis at the A interface. Region 1 comprises the N-terminus. Region 2 is composed of a long helix α5 (amino acids P68-N81). Region 3 is a short β-hairpin loop composed of amino acids D112-S115. Region 4 is formed by two short alpha helices (α2 and α5) corresponding to amino acids F62-F64, a short loop (amino acids L65-H67) while region 5 (helix α11), which is composed of amino acids I275-H290. (b, e and f) Two loops (S303-I311) (teal) occur perpendicular to the C-interface along the central helical axis, stabilising the interface through predominant electrostatic interactions. (c and d) The interface is stabilised by weak electrostatic and hydrophobic interactions, particularly at regions close to the core.

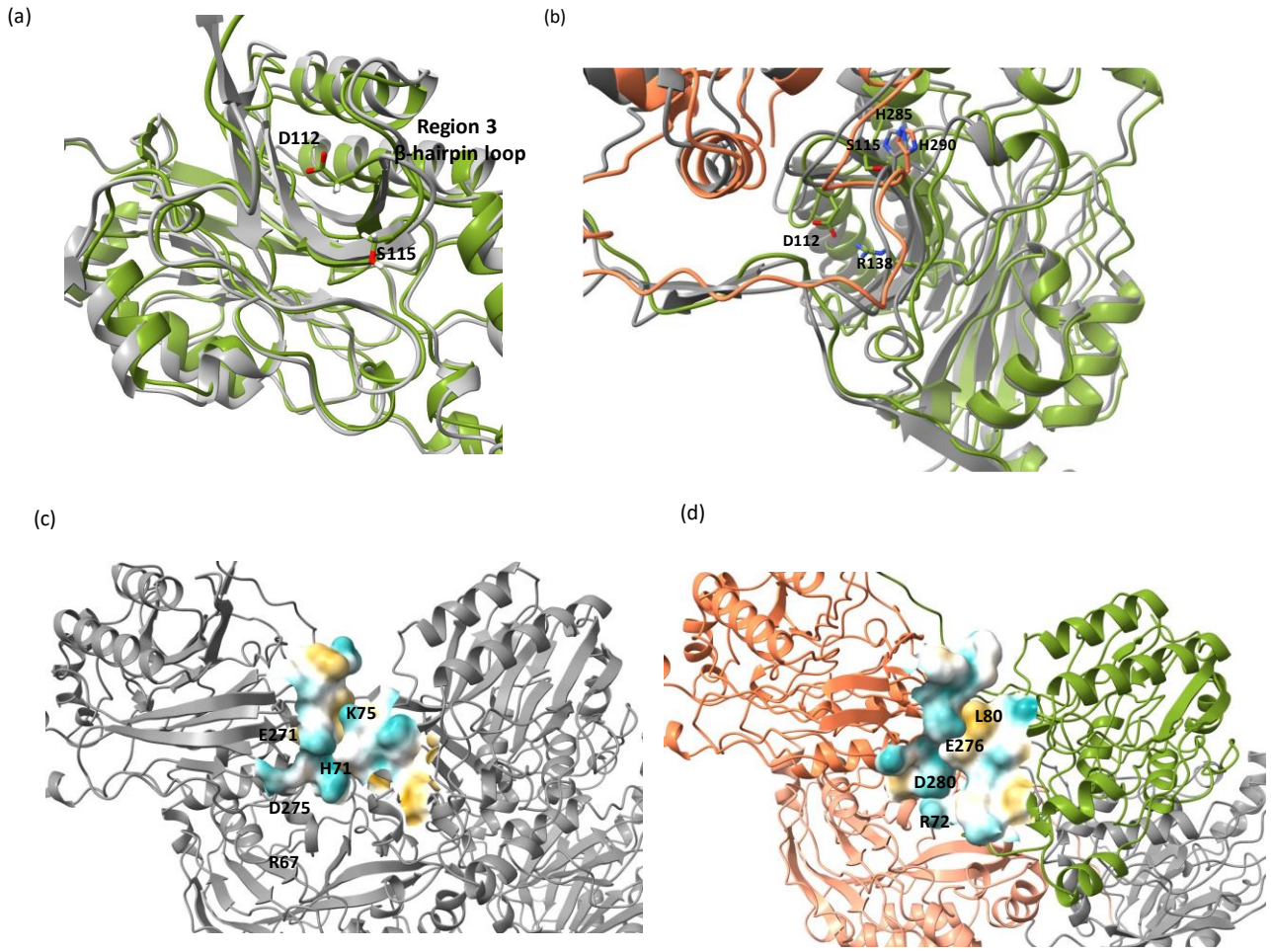
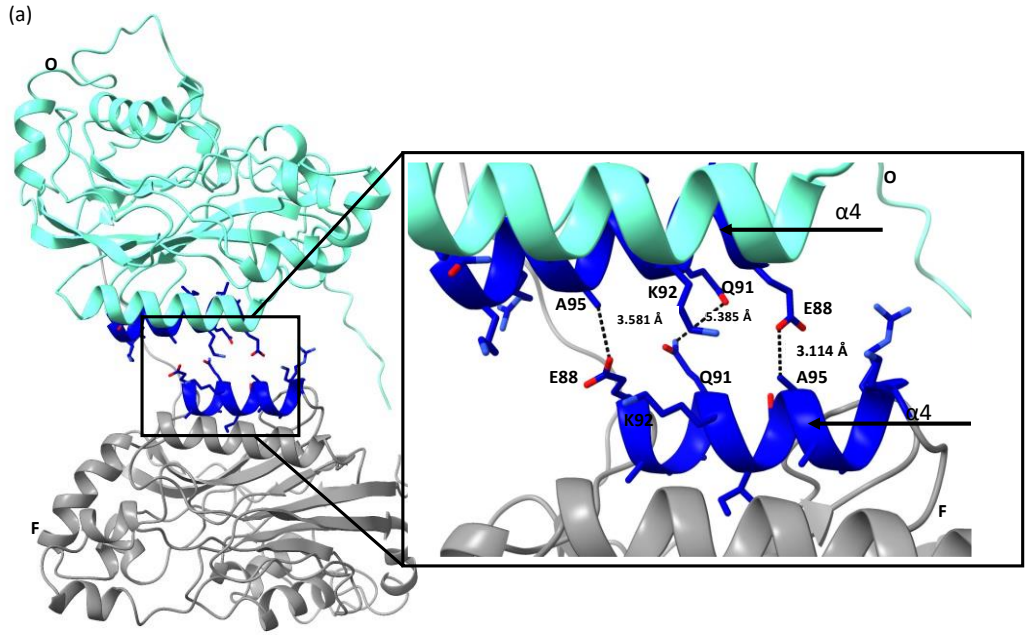


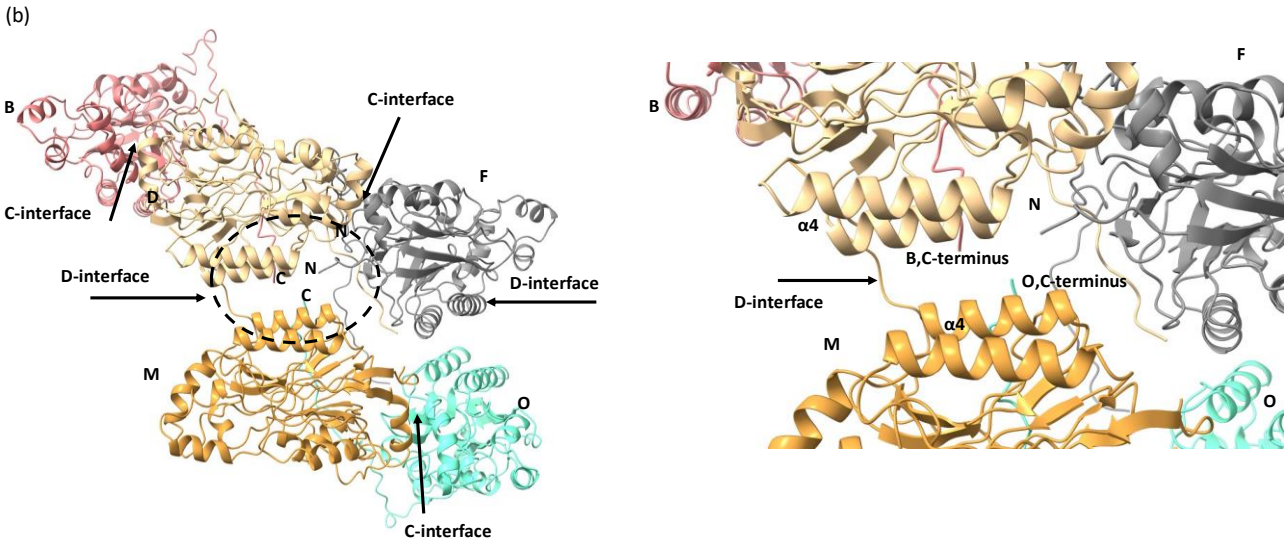
Figure 4.4.10: Comparison of the C-interface of $CynD_{stu}$ and $CynD_{pum}$. (a) Side by side comparison of the C-interface of $CynD_{pum}$ (left) and $CynD_{stu}$ (right). The C-interface comprises amino acids that are largely conserved between the two homologues, except for the aliphatic L80 and positively charged K75 at the top of the interface on $CynD_{pum}$. (b) Superimposed models of $CynD_{stu}$ (green) and $CynD_{pum}$ (grey). The β-hairpin loop is conserved in both homologous structures, but that of $CynD_{stu}$ (green) consists of polar and charged amino acids. S115 appears to interact with H290 via apparent van der Waals and pi-cation interactions across the interface. D115 forms intra-chain contacts, interacting primarily with R138. These interactions are not possible with the glycine amino acids in $CynD_{pum}$ (grey).

While amino acids at the C-interface are mostly conserved between $CynD_{pum}$ and $CynD_{stu}$, subtle differences are present. The β-hairpin loop that forms region 3 on the interface consists of S115 which interacts with H290 across the interface and D112 on this loop also forms intra-chain contacts that may contribute to stabilising the core of each monomer. These amino acids (D112 and S115) correspond to glycine amino acids G108 and G109 in $CynD_{pum}$. L80 creates a core packing effect, 'sealing' the interface through hydrophobic interactions across the chain with M7, A8 and Y10. The interface is more open in $CynD_{pum}$, where the corresponding amino acid is K75.

The D-interface

The D-interface is defined by helix α4 on $CynD_{stu}$. Interactions across this region are potentially unlikely due to the sidechains of amino acids being too far apart to facilitate inter-chain interactions where the functional groups of the interacting sidechains are capable of interacting. In instances where the sidechains are within interaction distance, the functional groups of adjacent sidechains are chemically incompatible (e.g. A95 and E88), thereby preventing the formation of interactions (Figure 4.4.11(a)). At this region, two C-terminal tails of monomers related to those associated across the D-interface by the C-interface meet near the α4 helices of the D-interface and the N-terminus (Figure 4.4.11(b)).


(a)
O
α4
A95
K92
Q91
E88
3.581 Å
5.385 Å
3.114 Å
E88
Q91
A95
K92
α4
F
F


(b)
B
C-interface
C-interface
D
F
N
D-interface
C
N
C
D-interface
M
O
C-interface
F
B
N
α4
B,C-terminus
O,C-terminus
D-interface
α4
M
O

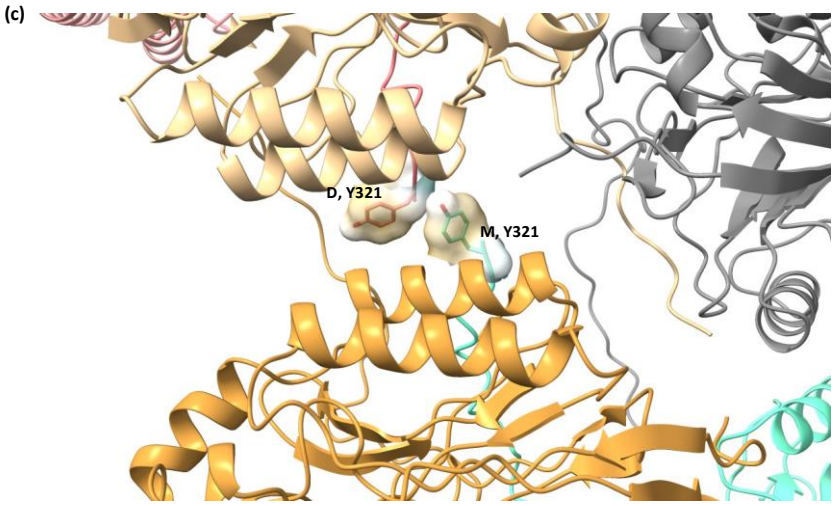


Figure 4.4.11: The D-interface. The exterior view of the D-interface is shown. (a) The D-interface is located between two α4 helices (shown in blue) in $CynD_{stu}$. The structural and chemical nature of the side chains of the amino acids, and their conformations on both helices is such that interactions across the groove are unlikely. Although the two Q91 amino acids across the interface can form an ionic interaction, the sidechains are too far apart for the interaction to occur (>5 Å). Those that occur close enough for chemical interaction to occur are chemically incapable of interacting (A95 and E88). (b) Two C-terminal tails of adjacent monomers and the N-terminus appear to meet at this region. The aliphatic N-terminal amino acids (MAHYPKF) potentially stabilise this region by increasing hydrophobic contacts with the aliphatic portion of Y321 across the groove. (c) The two C-terminal tails at this region clearly interact with one another, forming interactions between distantly located monomers. In the current model the two Y321 amino acids clearly interact through a hydrophobic interaction between the aliphatic portions of the sidechains.

The C-terminal tail

The C-terminal tail of $CynD_{stu}$ is composed of four loops that traverse the length of the A-interface along the central pore of the homo-oligomeric assembly (Figure 4.4.12). As it traverses the interior of the assembly, it interacts with interfaces A and C through hydrophobic and electrostatic interactions, respectively. Further hydrophobic interactions occur where the

C-terminal tail meets the α4 helix of the D-interface and the N-terminus, potentially stabilising the entire homo-oligomeric assembly (Figure 4.4.11 (b) and Figure 4.4.13).

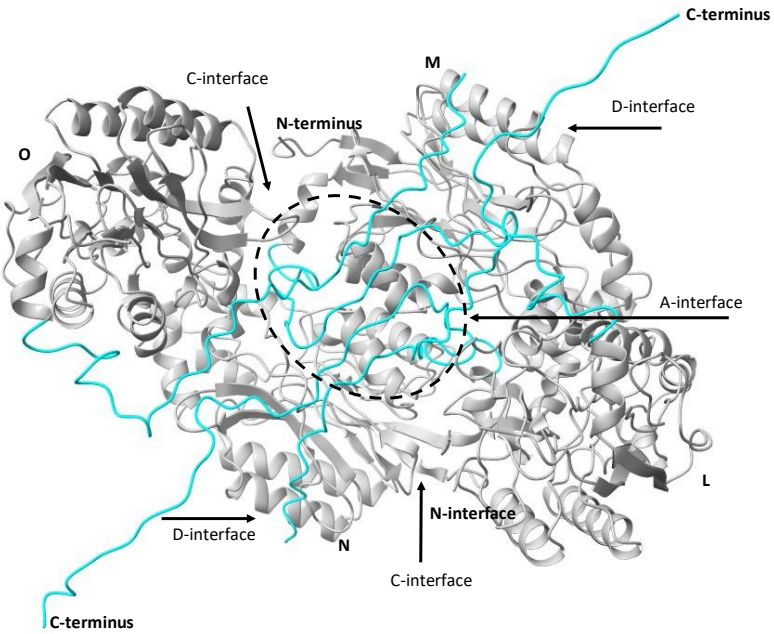


Figure 4.4.12: Interior view of the C-terminal tail. The C-terminal tails of monomers of $CynD_{stu}$ (cyan), traverse the length of the interior of the spiral. The four loops occur along the length of the A-interface (encircled).

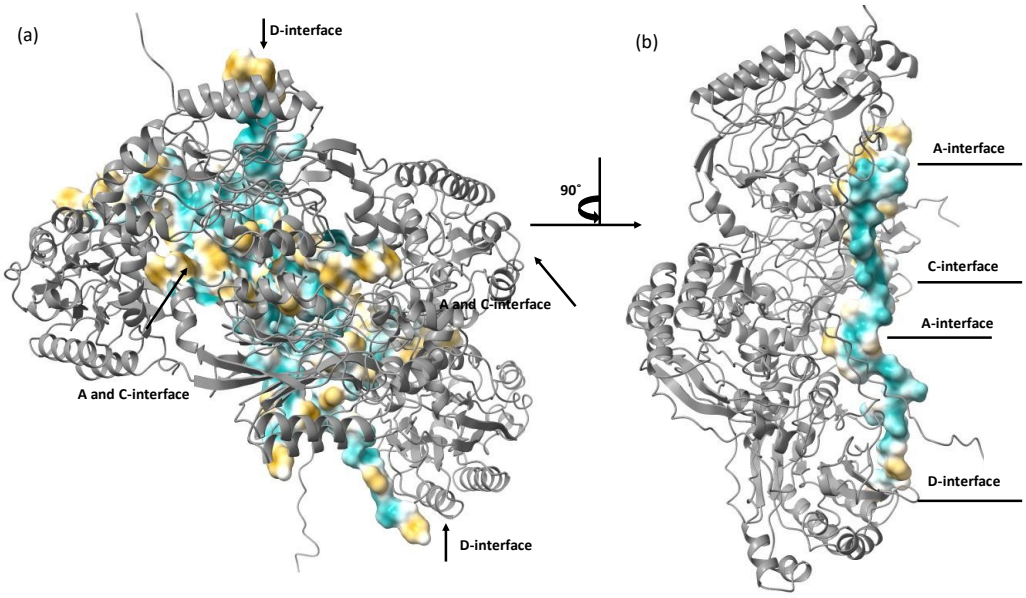
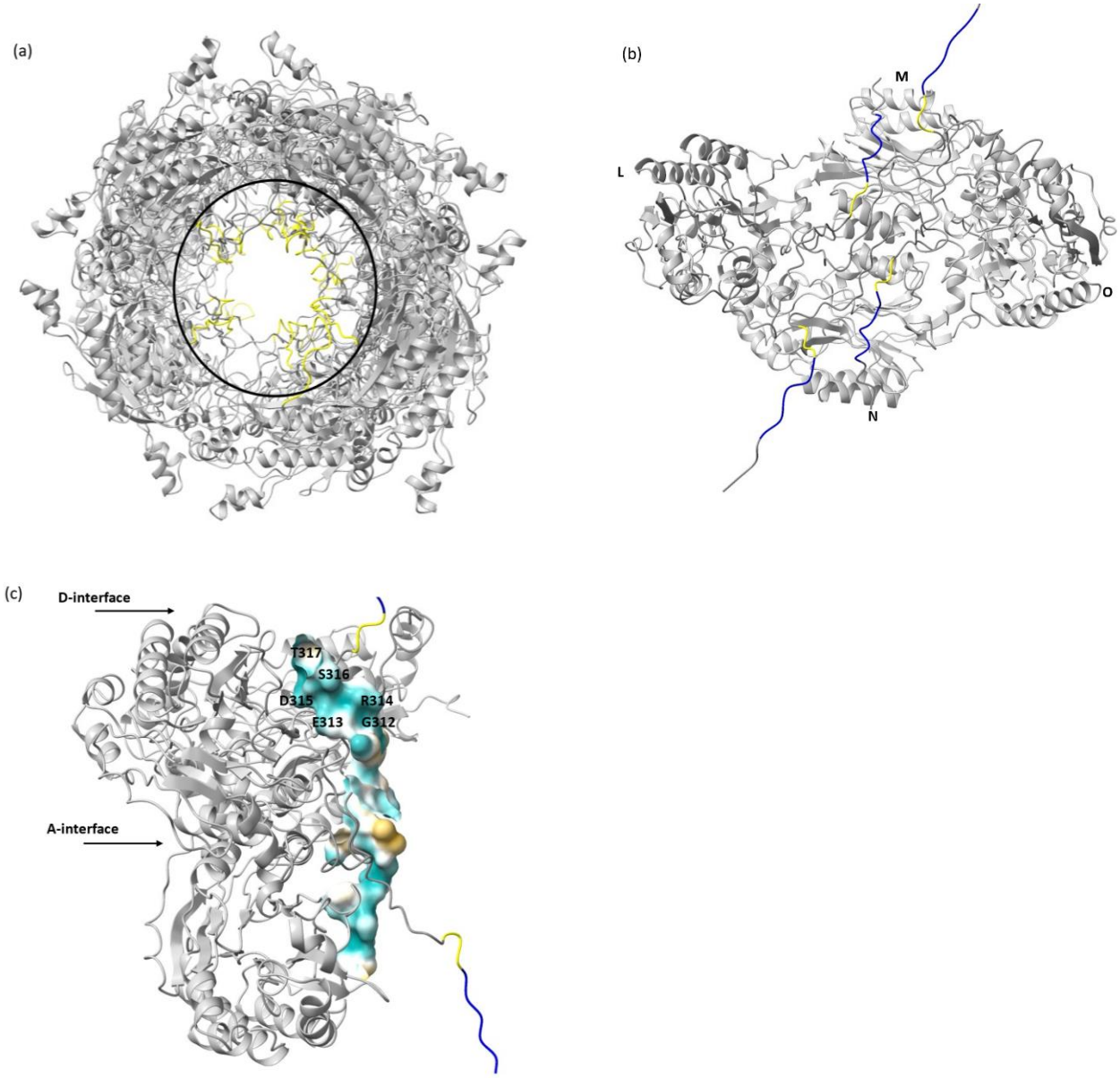


Figure 4.4.13: The C-terminal tail stabilises all other interfaces through specific interactions. Interfaces A and C-interface through electrostatic interactions. It interacts through hydrophobic interactions at the α4 helix of the D-interface.

The C-terminal tail amino acid sequence of $CynD_{stu}$ is dissimilar to that of $CynD_{pum}$ at two regions; 309RKIGERDST317 which corresponds to 304KHLNHQNE312 in $CynD_{pum}$ (outlined in green on Figure 4.4.16). The second region is 322DDLNSLVSDEEPVVRSLRK340, the last five amino acids of this region (335RSLRK340) make the C-terminal tail of $CynD_{stu}$ slightly longer than that of $CynD_{pum}$ these amino acids are unresolved in the current structure of $CynD_{stu}$.

In a previous study it was hypothesised that the enhanced thermostability observed in the $CynD_{pum}$-$CynD_{stu}$ chimeric protein stems from the amino acid sequence 312GERDST317 at the C-terminal tail of $CynD_{stu}$, which is distinct to this enzyme. This region was also identified as crucial for maintaining enzyme activity in $CynD_{stu}$, with its absence in the converse hybrid ($CynD_{pum}$-$CynD_{stu}$) postulated to be the cause of its inactivity (Abou Nader 2012, Crum, Park et al. 2015, Crum, Park et al. 2015).



4.4.14: The C-terminal tail region (309RKIGERDST317) is divergent between homologs $CynD_{stu}$ and $CynD_{pum}$. (a) The divergent amino acid sequence that is different in $CynD_{stu}$ is highlighted in yellow and occurs along the central pore of the homo-oligomeric assembly. (b) The 312GERDST317 motif is highlighted in blue within the divergent region (yellow). It occurs along the interior or backbone of individual monomers related to one another across the A-interface. (c) The interactions contributed by this region are predominantly electrostatic.

4.4.4. The $CynD_{stu}$ active site

$CynD_{stu}$ shares the conserved tetrad of catalytic amino acids characteristic of members of the nitrilase superfamily. In the current model, including the N-terminal His_6 tag, these amino acids correspond to C163, K135, E53, and E142. We identify E142 as the fourth amino acid based on comparison with the active site of $CynD_{pum}$ and alignment of the amino acid sequences of the two homologues. The active site, in line with other nitrilases whose structures have been elucidated, is located close to the A-interface.

The substrate binding pocket is primarily formed by loops L11 and L19 and contribute to the formation of the A-interface. Additionally, amino acids from helix α2, designated as region 2 of the C-interface, may also contribute to the substrate binding pocket of $CynD_{stu}$.

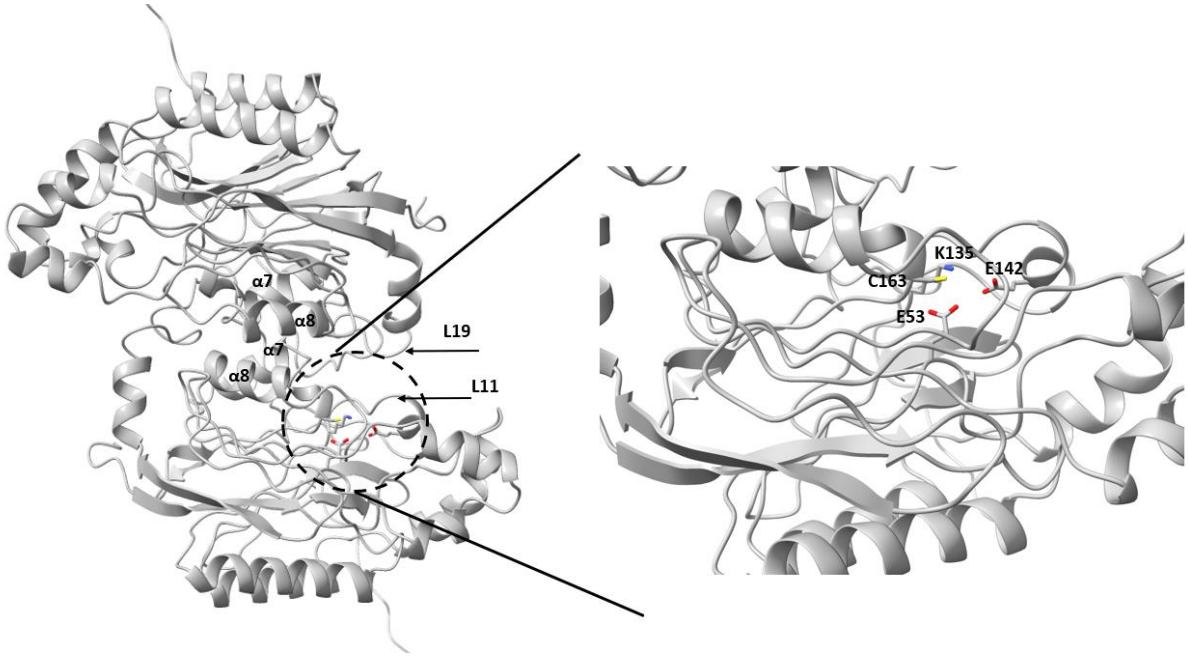


Figure 4.4.15: The $CynD_{stu}$ active site. The catalytic tetrad of amino acids is located near the A- and C-interfaces and is structurally homologous to other solved nitrilase structures.

4.4.5 pH-dependent oligomerisation and self-termination of the $CynD_{stu}$ spiral

The observed pH-dependent oligomerisation in $CynD_{pum}$ is hypothesised to be attributed to alterations in the ionisation state of histidine amino acids across the pH range of 8.0 to 5.4 (Mulelu 2013, Mulelu 2017). Mutating these histidine amino acids in $CynD_{pum}$ resulted in the formation of different oligomeric forms, indicating their importance to oligomerisation. These histidines correspond to H128, H184 and H241 located at the core and H285 at the C-interface. These histidines are conserved in $CynD_{stu}$ (Figure 4.4.16) corresponding to H133, H189, H246 and H290. The conserved nature of these amino acids suggests their structural importance. Based on the previously proposed hypothesis and current model, the histidines responsible for the observed changes in oligomeric state are likely to be located at the core and surface of the $CynD_{stu}$ homo-oligomer.

The $CynD_{stu}$ homo-oligomeric assembly was reported to be self-terminating with interactions occurring at the E-interface contributing to the termination of the assembly. The E-interface was found to occur between amino acids 266EID268 and 92RKNK95 (Sewell, Berman et al. 2003, Thuku, Brady et al. 2009). In the current model amino acids forming the E-interface 92RKNK95 and 266EID268 are located on helix α4 and groove at the D-interface and the

exterior surface, respectively. The two regions are distantly located, preventing any interactions from occurring.

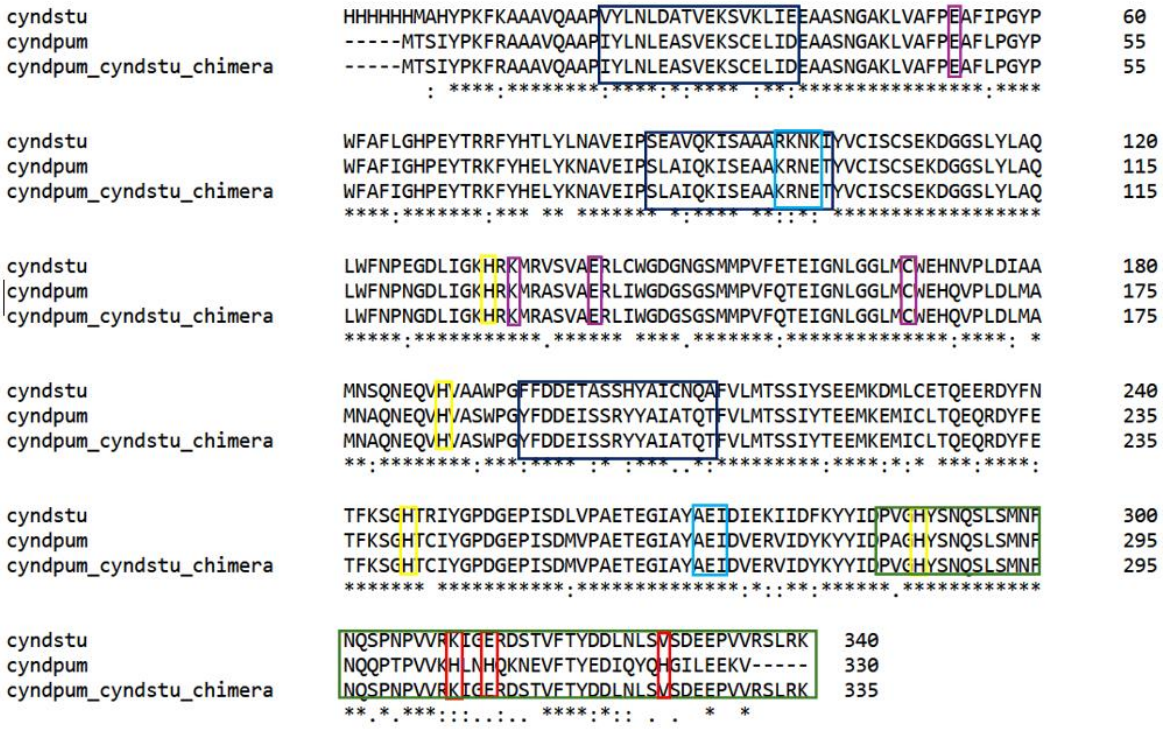


Figure 4.4.16: Multiple amino acid sequence alignment of $CynD_{stu}$, $CynD_{pum}$ and the $CynD_{pum}$-$CynD_{stu}$ chimera. Histidine amino acids postulated to play a role in pH-dependent oligomerisation are highlighted in yellow. Amino acids highlighted in red are histidines at the C-terminal tail of $CynD_{pum}$, which are postulated to play a role in pH-induced oligomerisation of $CynD_{pum}$. They are not conserved in $CynD_{stu}$ and therefore do not occur in the chimera. The amino acids highlighted in blue are regions with some variability between the two enzymes, some of which are located at the C and D interfaces. Highlighted in green are the C-terminal tail amino acids that were swapped to form the chimeric protein. The amino acids whose interaction is postulated to result in the self-termination of the $CynD_{stu}$ spiral are highlighted in a lighter blue shade. Amino acids in the catalytic tetrad (CEEK) are highlighted in purple. Catalytic amino acids are highlighted in purple.

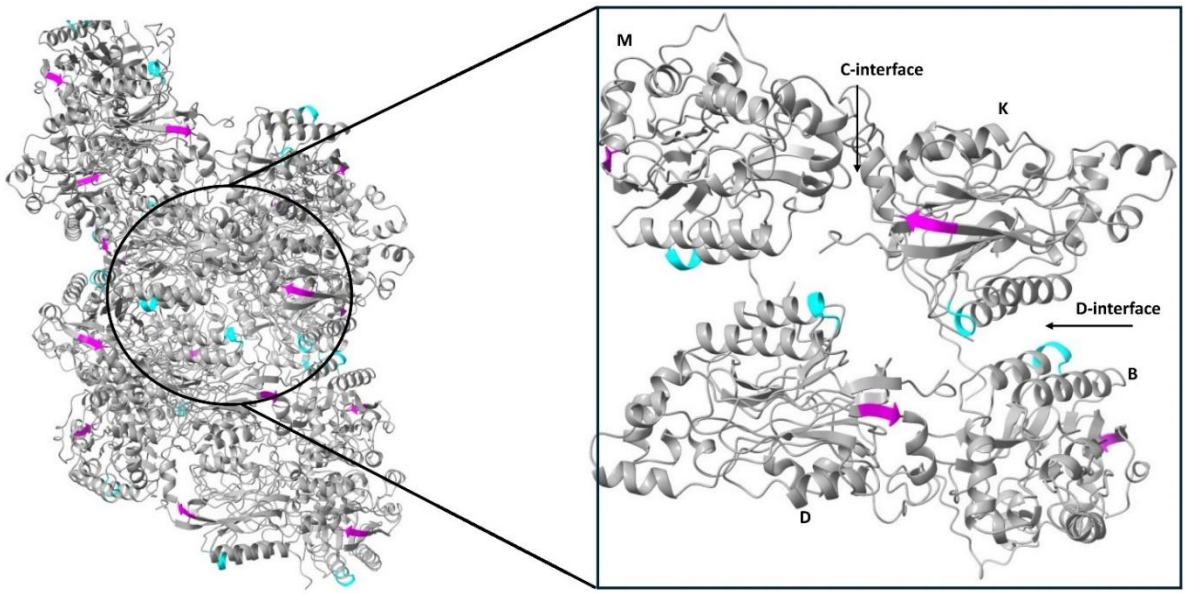


Figure 4.4.17: The E-interface of $CynD_{stu}$. The E-interface was reportedly formed through electrostatic occurring between 266EID268 (shown in magenta) and 92RKNK95 (shown in cyan).

4.5 Discussion

4.5.1 Overall structure of $CynD_{stu}$

The overall structure of $CynD_{stu}$ is structurally homologous to that of $CynD_{pum}$. The homo-oligomeric assembly is stabilised by interactions across interfaces A and C. The C-terminal tail stabilises the A and C interfaces through hydrophobic and electrostatic interactions, respectively.

Based on the findings of this study, we do not observe evidence supporting that $CynD_{stu}$ undergoes self-termination, nor are we able to discern the presence of the E-interface. In the current model, the proposed interaction sites reported to be responsible for forming the E-interface are situated too distantly from each other to facilitate any interaction. Furthermore, they are located at the surface of individual monomers instead of a region where two monomers or subunits meet. It is possible that artefacts brought about by the low-resolution negative stain TEM map, of the previous $CynD_{stu}$ map (Sewell, Berman et al. 2003) may have contributed to the appearance of the spiral closing and terminating at the terminal monomers on the assembly, but this does not occur *in vitro.*

We observed that the quaternary structure of $CynD_{stu}$ is notably sensitive to subtle changes in pH, degradation or dissociation at pH 8.0, as previously reported (Jandhyala, Willson et al. 2005), where a marked drop in enzyme activity was observed at this pH. This sensitivity to pH

disadvantages the use of wildtype $CynD_{stu}$ at alkaline pH. However, it remains fully active at its optimal pH, forming approximately 22 subunits in solution, with the current CryoEM structure revealing 16 resolved subunits at a resolution of 3.26 Å, with poor density at the terminal ends indicating the presence of additional but unresolved monomers. Moreover, we note that a decrease in pH correlates with the formation of larger homo-oligomers with partial activity as reported for $CynD_{pum}$ (Jandhyala, Berman et al. 2003, Mulelu 2017).

4.5.2 Thermostability

$CynD_{stu}$ demonstrates a T_m of approximately 53 °C at pH 7.4, indicative of a significant capacity to withstand thermal stress. Moreover, our investigation reveals a more thermodynamically favourable change in Gibb's free energy (ΔΔG) for $CynD_{stu}$ (-97 kJ/mol) compared to $CynD_{pum}$, further underscoring its thermostability. The findings of this study are consistent with those Jandhyala, Willson et al. (2005). Furthermore, our analysis aligns with the observed thermostability of the chimeric protein (Crum, Park et al. 2015), which forms fibres at optimum pH, has a T_m of 58.6 °C and an experimental ΔΔG of -16.8 kJ/mol (Chapter II, Figure 2.4.8).

While the sequence and structures of $CynD_{pum}$ and $CynD_{stu}$ share significant sequence and structural homology, the C-terminal tail is variable. The C-terminal tail of $CynD_{stu}$ is longer

than that of $CynD_{pum}$ by five amino acids. Our analysis suggests that the elongated C-terminal tail contributes to the stability of both the chimeric protein and wildtype $CynD_{stu}$. We hypothesise that the C-terminal tail spans across helices α4, traversing the D-interface and extending into the C-interface from within the interior of the enzyme, thereby stabilising the oligomer at the D-interface, reinforcing the stability of the C-interface and potentially the N-terminus, through strong electrostatic interactions in addition to those contributed by the GERDST region which traverses the backbone of each monomer. Therefore, we postulate that the increased thermostability conferred by the C-terminal tail of $CynD_{stu}$ onto the $CynD_{pum}$-$CynD_{stu}$ chimera is primarily due to increased contacts along the central pore of the enzyme. These are weaker in the C-terminal tail of $CynD_{pum}$, potentially leading to improper folding of $CynD_{stu}$ and inactivity in the $CynD_{stu}$- $CynD_{pum}$ chimera. We further postulate that the observed thermostability of $CynD_{stu}$ is due to this reinforcement of interactions at the backbone of monomers across the central pore of the enzyme.

Due to the absence of potential interactions across the D-interface, we propose that this region is a groove in $CynD_{stu}$. This hypothesis is supported by the introduction of the Q86R mutation to the chimera leading to changes in the oligomeric form of the enzyme from a fibre to a slightly longer spiral similar to the structure of $CynD_{stu}$. This suggests that the C-terminal

tail interacts with the helices defining the D-interface, along the interior of the enzyme, altering interactions at the C-interface, thus changing the oligomeric form of the enzyme.

The pH 5.4 homo-oligomer of $CynD_{stu}$ is considerably more thermostable than the 22mer at pH 7.4, indicating a degree of correlation between fibre formation and thermostability.

Chapter V

Discussion, conclusion, and future work

5.1 Discussion

5.1.1 Overview

Cyanide dihydratases are promising biocatalysts for the bioremediation of cyanide waste and have potential applications in industrial processes. A thorough understanding of their structure and function, and the relationship between the two, is crucial for optimising their use. Both $CynD_{pum}$ and $CynD_{stu}$ experience a significant loss of activity beyond pH 7.8 and 42°C (Jandhyala, Willson et al. 2005) making the use of their recombinantly expressed forms unsuitable for use in industry. Furthermore, improved thermostability has been linked to greater tolerance to alkaline pH, a crucial characteristic for the biotechnological application of this recombinant enzyme in industrial cyanide wastewater remediation.

We investigated the chemical and structural factors enhancing the thermostability of $CynD_{pum}$ for the engineering of industrially applicable thermostable and potentially operationally stable variants of the enzyme. This was achieved by analysing the structure of the quadruple $CynD_{pum}$ variant (Q86R/H305K/H308K/H323K) (Mulelu 2017, Sewell, Frangakis et al. 2017) *in silico* site saturation mutagenesis to identify stabilising mutations, validation of these mutations *in vitro* using site directed mutagenesis, nanoscale differential fluorimetry and

elucidation of the chemical and structural basis of the enhanced thermostability using negative-stain and cryogenic electron microscopy.

5.1.2 Structural determinants of enhanced thermostability

5.1.2.1 Interfacial regions

Our investigation revealed that both homologs are formed and stabilised by interactions at interfaces A and C. The C-terminal tail further stabilises each interface from along the interior pore of the enzyme. Previous reports suggested that the D-interface occurs when the $CynD_{pum}$ helix or $CynD_{stu}$ spiral completes a turn, with electrostatic interactions occurring across the groove between helices α1 and α3. However, recent findings by Mulelu (2017) indicated that these interactions primarily occur between helices α3. In the atomic resolution models of $CynD_{pum}$ and its variants analysed in this study (PDB IDs 8p4i, 8c5i, 9er3), the helix at this region is identified as helix α4.

Interactions between amino acids on helix α4 at the D-interface and the C-terminal tail are crucial for enzyme thermostability and proper protein folding. This is evidenced by the Q86R mutation which leads to increased thermostability and tolerance to alkaline pH (Wang, Watermeyer et al. 2012). This arises from a network of electrostatic, hydrogen bonding, and hydrophobic interactions facilitated by the aliphatic methylene group (C_3H_6) and basic

guanidino group of the arginine residue. In this study we found that the mutation Q86M also results in increased thermostability by creating hydrophobic interactions with amino acids of the C-terminal tail. Additionally, we suggest that this region serves as a groove crucial for enzyme conformational flexibility, with mutations creating contacts across the groove (Q86K) potentially disrupting proper protein folding and leading to enzyme inactivity.

The current model of $CynD_{stu}$ is a spiral consisting of 16 monomers, and *in vitro* experiments suggest that it can form assemblies with up to ~22 monomers. This finding indicates that $CynD_{stu}$ can form larger assemblies than the previously reported 14mer. Furthermore, the amino acid sequences that form the E-interface, which was postulated to be responsible for the self-termination of the $CynD_{stu}$ spiral, are distantly located, with one of them (266EID268) located at the surface of the enzyme. Based on these observations, it is concluded that $CynD_{stu}$ does not form self-terminating spirals and that the E-interface through which this self-termination is thought to occur is not possible.

5.1.2.2 The N- and C-termini

The N-terminus of CynD homologs appears to play a crucial structural role. We propose that the amino acids at the N-terminus interact with amino acids on grooves near the D-interface and with the C-terminal tail. These interactions have the effect of "closing-off" the homo-oligomeric assembly and maintaining proper protein folding. The E79P mutation, located on

a groove, resulted in inactivity. Visualisation of this region in the atomic resolution structure of the E155R variant of $CynD_{pum}$ demonstrates an ionic interaction between the main chain amino group of S3 and the side chain carboxyl of E79, indicating that the N-terminus forms stabilising interactions and is potentially critical to oligomerisation of both $CynD_{pum}$ and $CynD_{stu}$ CynD. Comparison of this region in the E155R variant of $CynD_{pum}$ and wildtype $CynD_{pum}$ reveals that this interaction is not brought about by the reduced helical twist of the E155R variant. The I4M mutation, predicted to be thermostable by the FoldX algorithm, was also inactive, and we postulate that the inactivity was caused by improper protein folding since I4 is distantly located from the active site and substrate binding site.

As previously mentioned, the C-terminal tail stabilises all other interfaces through which the homo-oligomeric assembly is formed and creates additional interactions with the α4 helix, thereby creating contacts between distantly located monomers.

5.1.2.3 Groove and surface regions

The grooves are previously unexplored regions for the enhancement of the thermostability of $CynD_{pum}$. Our findings demonstrate that mutations at the grooves affect protein structure with mutations E79P and E96G, leading to inactive proteins potentially due to improper folding, and others (N119R and E155R) lead to the formation of longer and more thermostable helical fibres than the wild-type enzyme. This also suggests the possibility that mutations removing

the electrostatic properties of the grooves lead to improper protein folding, whereas those that favourably increase them lead to the formation of thermostable oligomeric assemblies. For example, the atomic resolution structure of the E155R variant of $CynD_{pum}$ revealed a network of ionic and hydrogen bonding interactions at the groove, resulting in a helical assembly with reduced helical twist and increased structural rigidity, which contributed to the observed enhanced thermostability, whereas mutations E79P and E96G lead to inactivity.

Furthermore, we identified that the substitution of outer-surface amino acids with hydrophobic amino acids also leads to increased fibre length and enhanced thermostability by increasing core packing. Mutations, S29A, T217I and T260I, which also exist as fibres, are located at the outer surface with hydrophobic substituted mutations projecting towards the core of the enzyme. Interestingly, the S29A/Q86R variant was partially inactive, indicating a lack of synergy between the two mutations although they are distantly located on the enzyme. The underlying chemical basis for this observation are unclear.

5.1.3 Oligomerisation and thermostability

We discovered that $CynD_{stu}$ undergoes pH-dependent oligomerisation, similar to $CynD_{pum}$, and that several histidines located at the core of both $CynD_{pum}$ and $CynD_{stu}$ that are reported to contribute to the structural transition occurring between pH 8.0 and 5.4, are conserved between the two homologs, indicating their critical involvement in the observed fibre

formation. These correspond to H128, H184, H241 and H285 on $CynD_{pum}$ which correspond to H133, H189, H246 and H299 in $CynD_{stu}$.

Furthermore, we noted a correlation between fibre length and increased thermostability. This association is theoretically feasible, as larger assemblies necessitate more energy to disrupt their numerous bonds and interactions compared to smaller counterparts with a similar overall fold. However, the precise mechanisms governing fibre formation remains unclear and subject to ongoing investigation. Previous studies have indicated that mutations at the C-interface can induce alterations in oligomeric states and helical structures with modified helical twist (Mulelu 2017). Our findings, demonstrating fibre formation brought about by mutations at the grooves and surface regions, imply that structural changes leading to fibre formation are not solely dictated by interactions at the C-interface. Instead, they appear to be influenced by various regions across the enzyme.

5.1.4 Structural implications for thermostability

Our analysis suggests that the C-terminal tail region is critical for engineering the thermostability of $CynD_{pum}$. Rather than this being due to specific differences in the amino acid sequences of the tails, we postulate that it potentially occurs due to the elongation of the tail, which creates more contacts along the interior of the monomers, as observed in $CynD_{stu}$ and the $CynD_{pum}$-$CynD_{stu}$ chimera. The importance of the C-terminal tail to thermostability in

$CynD_{pum}$ has been previously established (Crum, Park et al. 2015, Crum, Park et al. 2015). Moreover, the interactions between the C-terminus, helix α4 of the D-interface groove, and N-terminus are important for increasing the thermostability of $CynD_{pum}$.

Additionally, we propose that increased hydrophobicity at the surface enhances the thermostability of $CynD_{pum}$ by promoting protein folding and oligomerisation, thereby improving its tolerance to thermal stress. The increased thermostability may also be a result of increased core packing since the amino acid side chains project into the interior core of the enzyme. Increased core packing has been shown to enhance the thermostability of other mesophilic proteins and to occur more frequently in thermophilic proteins (Baldwin and Matthews 1994, Abraham, Abraham et al. 2005, Glyakina, Garbuzynskiy et al. 2007, Van Dijk, Hoogeveen et al. 2015).

The introduction of interactions between the grooves reinforces the rigidity of the structure; therefore, increased interactions across the grooves are a major determinant of thermostability in $CynD_{pum}$. Although mutations at the C-interface have also been shown to result in the formation of larger helices with potentially increased thermostability, the risk of introducing mutations in this region is the potentially altered helical twist, which changes the size of the substrate-binding pocket, rendering the enzyme incapable of degrading cyanide. From the atomic resolution structure of the E155R variant of $CynD_{pum}$, we found that the

altered helical twist did not change the substrate-binding pocket or orientation of the catalytic amino acids. This indicates that the grooves are potentially more suitable regions for engineering thermostable $CynD_{pum}$ variants that are capable of degrading cyanide.

5.2 Conclusion and future work

In conclusion, our study has elucidated key structural and chemical determinants of thermostability in CynD, highlighting that mutations at interfaces and on the surface significantly enhance thermostability. These mutations contribute to improved stability primarily by promoting oligomerization, resulting in a more robust and heat-resistant protein structure. These findings offer valuable insights for designing thermostable CynD variants and advancing our understanding of enzyme optimization for environmentally friendly industrial applications.

Additionally, our research revealed that introducing contacts across the grooves of CynD alters its helical twist, suggesting potential for engineering cyanide-degrading enzymes with modified substrate specificity. This opens new opportunities for developing cyanide dihydratases capable of catalysing the conversion of non-cyanide substrates, thereby broadening their biotechnological applications. The interfacial analysis performed in this study will guide future directed evolution efforts by identifying regions most likely to produce variants with desirable properties, including enhanced thermostability.

Furthermore, our study has provided new insights into the role of the N- and C-termini, advancing our fundamental understanding of nitrilase structure. Future research should focus on elucidating the detailed mechanisms by which these mutations affect the thermodynamics of oligomerization and stability. Atomic resolution structures of the CynDpum-CynDstu chimera could offer further insights into improving thermostability and understanding enzyme structure. In this study, multi-site substituted variants, including those with multiple mutations across the groove, were found to be insolubly expressed. Optimizing their expression conditions and assessing their thermostability could be the next step in developing a thermostable variant of CynDpum. Additionally, exploring strategies to protect the functional groups of the catalytic tetrad from chemical inactivation is essential. This complements existing research on maintaining activity through structural integrity under non-optimal conditions, particularly at alkaline pH, where functional groups of the catalytic tetrad are inactivated.

Reference list

Abou Nader, M. (2012). Directed evolution of cyanide degrading enzymes, Texas A&M University.

Abraham, T., T. Abraham, S. Pil Pack and Y. Je Yoo (2005). "Stabilization of Bacillus subtilis Lipase A by increasing the residual packing." Biocatalysis and Biotransformation **23**(3-4): 217-224.

Afonine, P. V., B. K. Poon, R. J. Read, O. V. Sobolev, T. C. Terwilliger, A. Urzhumtsev and P. D. Adams (2018). "Real-space refinement in PHENIX for cryo-EM and crystallography." Acta Crystallographica Section D: Structural Biology **74**(6): 531-544.

Ahmad, F. (2022). "Protein stability [determination] problems." Frontiers in Molecular Biosciences **9**: 880358.

Almatawah, Q. A. and D. A. Cowan (1999). "Thermostable nitrilase catalysed production of nicotinic acid from 3-cyanopyridine." Enzyme and microbial technology **25**(8-9): 718-724.

Almatawah, Q. A., R. Cramp and D. A. Cowan (1999). "Characterization of an inducible nitrilase from a thermophilic bacillus." Extremophiles **3**: 283-291.

Álvarez-Cao, M.-E., R. González, M. A. Pernas and M. L. Rúa (2018). "Contribution of the Oligomeric State to the Thermostability of Isoenzyme 3 from Candida rugosa." Microorganisms **6**(4): 108.

Anning, C., J. Wang, P. Chen, I. Batmunkh and X. Lyu (2019). "Determination and detoxification of cyanide in gold mine tailings: A review." Waste Management & Research **37**(11): 1117-1126.

Arfi, T. and V. K. Nigam (2020). "Studies on a thermostable nitrilase from Staphylococcus Sp and its In-silico characterisation." Biologia **75**(12): 2421-2432.

Azizi, A. and R. Ghaedrahmati (2015). "Optimizing and evaluating the operational factors affecting the cyanide leaching circuit of the Aghdareh gold processing plant using a CCD model." Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences **471**(2184): 20150681.

Baker, L. A. and J. L. Rubinstein (2010). Radiation damage in electron cryomicroscopy. Methods in enzymology, Elsevier. **481:** 371-388.

Baldwin, E. P. and B. W. Matthews (1994). "Core-packing constraints, hydrophobicity and protein design." Current Opinion in Biotechnology **5**(4): 396-402.

Bell, E. L., W. Finnigan, S. P. France, A. P. Green, M. A. Hayes, L. J. Hepworth, S. L. Lovelock, H. Niikura, S. Osuna and E. Romero (2021). "Biocatalysis." Nature Reviews Methods Primers **1**(1): 46.

Bessonnet, T., A. Mariage, J.-L. Petit, V. Pellouin, A. Debard, A. Zaparucha, C. Vergne-Vaxelaire and V. de Berardinis (2021). "Purification and Characterization of Nit phym, a Robust Thermostable Nitrilase From Paraburkholderia phymatum." Frontiers in Bioengineering and Biotechnology **9**: 686362.

Bork, P. and E. V. Koonin (1994). "A new family of carbon-nitrogen hydrolases." Protein Science **3**(8): 1344-1346.

Brenner, C. (2002). "Catalysis in the nitrilase superfamily." Current opinion in structural biology **12**(6): 775-782.

Broderius, S. J. (1981). "Determination of hydrocyanic acid and free cyanide in aqueous solution." Analytical Chemistry **53**(9): 1472-1477.
Buß, O., J. Rudat and K. Ochsenreither (2018). "FoldX as protein engineering tool: better than random based approaches?" Computational and structural biotechnology journal **16**: 25-33.
Cellmer, T., E. R. Henry, J. Kubelka, J. Hofrichter and W. A. Eaton (2007). "Relaxation rate for an ultrafast folding protein is independent of chemical denaturant concentration." Journal of the American Chemical Society **129**(47): 14564-14565.
Chaudhary, M. and R. Gupta (2012). "Cyanide detoxifying enzyme: rhodanese." Current Biotechnology **1**(4): 327-335.
Chauhan, S., S. Wu, S. Blumerman, R. Fallon, J. Gavagan, R. DiCosimo and M. Payne (2003). "Purification, cloning, sequencing and over-expression in Escherichia coli of a regioselective aliphatic nitrilase from Acidovorax facilis 72W." Applied microbiology and biotechnology **61**: 118-122.
Chen, V. B., W. B. Arendall, J. J. Headd, D. A. Keedy, R. M. Immormino, G. J. Kapral, L. W. Murray, J. S. Richardson and D. C. Richardson (2010). "MolProbity: all-atom structure validation for macromolecular crystallography." Acta Crystallographica Section D: Biological Crystallography **66**(1): 12-21.
Chen, Z., H. Chen, Z. Ni, R. Tian, T. Zhang, J. Jia and S. Yang (2015). "Expression and characterization of a novel nitrilase from hyperthermophilic bacterium Thermotoga maritima MSB8." Journal of Microbiology and Biotechnology **25**(10): 1660-1669.
Cheng, Y., N. Grigorieff, P. A. Penczek and T. Walz (2015). "A primer to single-particle cryo-electron microscopy." Cell **161**(3): 438-449.
Chi, E. Y., S. Krishnan, T. W. Randolph and J. F. Carpenter (2003). "Physical stability of proteins in aqueous solution: mechanism and driving forces in nonnative protein aggregation." Pharmaceutical research **20**: 1325-1336.
Cong, Y. and S. J. Ludtke (2010). "Single particle analysis at high resolution." Methods in enzymology **482**: 211-235.
Costanzo, J. A., C. J. O'Brien, K. Tiller, E. Tamargo, A. S. Robinson, C. J. Roberts and E. J. Fernandez (2014). "Conformational stability as a design target to control protein aggregation." Protein Engineering, Design & Selection **27**(5): 157-167.
Croll, T. I. (2018). "ISOLDE: a physically realistic environment for model building into low-resolution electron-density maps." Acta Crystallographica Section D: Structural Biology **74**(6): 519-530.
Crum, M., J. Park, B. Sewell and M. Benedik (2015). "C-terminal hybrid mutant of Bacillus pumilus cyanide dihydratase dramatically enhances thermal stability and pH tolerance by reinforcing oligomerization." Journal of applied microbiology **118**(4): 881-889.
Crum, M. A.-N., J. M. Park, A. E. Mulelu, B. T. Sewell and M. J. Benedik (2015). "Probing C-terminal interactions of the Pseudomonas stutzeri cyanide-degrading CynD protein." Applied microbiology and biotechnology **99**: 3093-3102.
Crum, M. A., B. T. Sewell and M. J. Benedik (2016). "Bacillus pumilus cyanide dihydratase mutants with higher catalytic activity." Frontiers in Microbiology **7**: 1264.
Cutulle, M. A., G. R. Armel, J. T. Brosnan, M. D. Best, D. A. Kopsell, B. D. Bruce, H. E. Bostic and D. S. Layton (2014). "Synthesis and evaluation of heterocyclic analogues of bromoxynil." Journal of agricultural and food chemistry **62**(2): 329-336.

Deller, M. C., L. Kong and B. Rupp (2016). "Protein stability: a crystallographer's perspective." Acta Crystallographica Section F: Structural Biology Communications **72**(2): 72-95.
Dent, K. C., B. W. Weber, M. J. Benedik and B. T. Sewell (2009). "The cyanide hydratase from Neurospora crassa forms a helix which has a dimeric repeat." Applied microbiology and biotechnology **82**: 271-278.
Eicher, J. J. (2007). pH-dependence of the Quaternary structure of the Cyanide Dihydratase from Bacillus pumilus. Master of Science, University of Cape Town.
Emsley, P. and K. Cowtan (2004). "Coot: model-building tools for molecular graphics." Acta crystallographica section D: biological crystallography **60**(12): 2126-2132.
Engler, H. S., K. C. Spencer and L. E. Gilbert (2000). "Preventing cyanide release from leaves." Nature **406**(6792): 144-145.
Fernandes, B. C., C. Mateo, C. Kiziak, A. Chmura, J. Wacker, F. van Rantwijk, A. Stolz and R. A. Sheldon (2006). "Nitrile hydratase activity of a recombinant nitrilase." Advanced Synthesis & Catalysis **348**(18): 2597-2603.
Fisher, F. and J. Brown (1952). "Colorimetric determination of cyanide in stack gas and waste water." Analytical Chemistry **24**(9): 1440-1444.
Gail, E., S. Gos, R. Kulzer, J. Lorösch, A. Rubo, M. Sauer, R. Kellens, J. Reddy, N. Steier and W. Hasenpusch (2000). "Cyano compounds, inorganic." Ullmann's Encyclopedia of Industrial Chemistry.
Glyakina, A. V., S. O. Garbuzynskiy, M. Y. Lobanov and O. V. Galzitskaya (2007). "Different packing of external residues can explain differences in the thermostability of proteins from thermophilic and mesophilic organisms." Bioinformatics **23**(17): 2231-2238.
Goldenzweig, A. and S. J. Fleishman (2018). "Principles of protein stability and their application in computational design." Annual review of biochemistry **87**: 105-129.
Gong, J.-S., Z.-M. Lu, H. Li, J.-S. Shi, Z.-M. Zhou and Z.-H. Xu (2012). "Nitrilases in nitrile biocatalysis: recent progress and forthcoming research." Microbial cell factories **11**: 1-18.
Gong, J.-S., J.-S. Shi, Z.-M. Lu, H. Li, Z.-M. Zhou and Z.-H. Xu (2017). "Nitrile-converting enzymes as a tool to improve biocatalysis in organic synthesis: recent insights and promises." Critical reviews in biotechnology **37**(1): 69-81.
Grifantini, R., C. Pratesi, G. Galli and G. Grandi (1996). "Topological Mapping of the Cysteine Residues of N-Carbamyl-D-amino-acid Amidohydrolase and Their Role in Enzymatic Activity (∗)." Journal of Biological chemistry **271**(16): 9326-9331.
Gromiha, M. M. (2010). Protein bioinformatics: from sequence to function, academic press.
Henschler, D. (2003). "Hydrogen cyanide, potassium cyanide and sodium cyanide [MAK Value Documentation, 2003]." Occupational toxicants: Critical data evaluation for MAK values and classification of carcinogens **19**: 189-210.
Hewitt, D., P. Breuer and C. Jeffery (2012). "Cyanide detoxification of gold cyanidation tails and process streams." Mineral Processing and Extractive Metallurgy **121**(4): 228-236.
Hinzmann, M., H. Yavuzer, M. Bittmann and H. Gröger (2023). "Novel approach toward industrial aromatic nitriles via biocatalysis: From rational enzyme design to sustainable, energy-saving technology platform." Chem Catalysis **3**(4).
Jandhyala, D., M. Berman, P. R. Meyers, B. T. Sewell, R. C. Willson and M. J. Benedik (2003). "CynD, the cyanide dihydratase from Bacillus pumilus: gene cloning and structural studies." Applied and Environmental Microbiology **69**(8): 4794-4805.

Jandhyala, D. M., R. C. Willson, B. T. Sewell and M. J. Benedik (2005). "Comparison of cyanide-degrading nitrilases." Applied microbiology and biotechnology **68**: 327-335.
Jumper, J., R. Evans, A. Pritzel, T. Green, M. Figurnov, O. Ronneberger, K. Tunyasuvunakool, R. Bates, A. Žídek and A. Potapenko (2021). "Highly accurate protein structure prediction with AlphaFold." Nature **596**(7873): 583-589.
Kazlauskas, R. (2018). "Engineering more stable proteins." Chemical Society Reviews **47**(24): 9026-9045.
Kazlauskas, R. J. and U. T. Bornscheuer (2009). "Finding better protein engineering strategies." Nature chemical biology **5**(8): 526-529.
Khatik, A. G., A. K. Jain and A. B. Muley (2022). "Preparation, characterization and stability of cross linked nitrilase aggregates (nitrilase–CLEAs) for hydroxylation of 2-chloroisonicotinonitrile to 2-chloroisonicotinic acid." Bioprocess and Biosystems Engineering **45**(9): 1559-1579.
Kimani, S. W., V. B. Agarkar, D. A. Cowan, M.-R. Sayed and B. T. Sewell (2007). "Structure of an aliphatic amidase from Geobacillus pallidus RAPc8." Acta Crystallographica Section D: Biological Crystallography **63**(10): 1048-1058.
Kimanius, D., L. Dong, G. Sharov, T. Nakane and S. H. Scheres (2021). "New tools for automated cryo-EM single-particle analysis in RELION-4.0." Biochemical Journal **478**(24): 4169-4185.
Kiziak, C., D. Conradt, A. Stolz, R. Mattes and J. Klein (2005). "Nitrilase from Pseudomonas fluorescens EBC191: cloning and heterologous expression of the gene and biochemical characterization of the recombinant enzyme." Microbiology **151**(11): 3639-3648.
Kobayashi, M., H. Komeda, N. Yanaka, T. Nagasawa and H. Yamada (1992). "Nitrilase from Rhodococcus rhodochrous J1. Sequencing and overexpression of the gene and identification of an essential cysteine residue." Journal of Biological Chemistry **267**(29): 20746-20751.
Kumar, P. (2017). "Pharmacology of specific drug groups." Pharmacology and Therapeutics for Dentistry, 7th ed.; Elsevier: Amsterdam, The Netherlands: 457-487.
Kumwenda, B., D. Litthauer, Ö. T. Bishop and O. Reva (2013). "Analysis of protein thermostability enhancing factors in industrially important thermus bacteria species." Evolutionary Bioinformatics **9**: EBO. S12539.
Laemmli, U. K. (1970). "Cleavage of structural proteins during the assembly of the head of bacteriophage T4." nature **227**(5259): 680-685.
Lai, M. and E. Topp (1999). "Solid-state chemical stability of proteins and peptides." Journal of pharmaceutical sciences **88**(5): 489-500.
Layh, N., J. Parratt and A. Willetts (1998). "Characterization and partial purification of an enantioselective arylacetonitrilase from Pseudomonas fluorescens DSM 7155." Journal of Molecular Catalysis B: Enzymatic **5**(5-6): 467-474.
Legood, S., I. G. Boneca and N. Buddelmeijer (2021). "Mode of action of lipoprotein modification enzymes—Novel antibacterial targets." Molecular Microbiology **115**(3): 356-365.
Liu, Z.-Q., M.-M. Lu, X.-H. Zhang, F. Cheng, J.-M. Xu, Y.-P. Xue, L.-Q. Jin, Y.-S. Wang and Y.-G. Zheng (2018). "Significant improvement of the nitrilase activity by semi-rational protein engineering and its application in the production of iminodiacetic acid." International journal of biological macromolecules **116**: 563-571.
Lodish, H. F. (2008). Molecular cell biology, Macmillan.

Lu, X. F., X. L. Tang, C. H. Xu, H. J. Diao, Z. M. Wu, R. C. Zheng and Y. G. Zheng (2023). "Engineering residues on C interface to improve thermostability of nitrilase for biosynthesis of Pregabalin precursor." AIChE Journal **69**(11): e18211.
Lundgren, S., B. Lohkamp, B. Andersen, J. Piškur and D. Dobritzsch (2008). "The crystal structure of β-alanine synthase from Drosophila melanogaster reveals a homooctameric helical turn-like assembly." Journal of molecular biology **377**(5): 1544-1559.
Lutz, S. and U. T. Bornscheuer (2012). Protein engineering handbook, John Wiley & Sons.
Makhatadze, G. I. (1999). Thermodynamics of protein interactions with urea and guanidinium hydrochloride, ACS Publications. **103:** 4781-4785.
Manning, M. C., J. Liu, T. Li and R. E. Holcomb (2018). "Rational design of liquid formulations of proteins." Advances in protein chemistry and structural biology **112**: 1-59.
Martínková, L., N. Kulik, A. Sedova, B. Křístková and P. Bojarová (2023). "Recent Progress in the Production of Cyanide-Converting Nitrilases—Comparison with Nitrile-Hydrolyzing Enzymes." Catalysts **13**(3): 500.
Marziaz, M. L., K. Frazier, P. B. Guidry, R. A. Ruiz, I. Petrikovics and D. C. Haines (2013). "Comparison of brain mitochondrial cytochrome c oxidase activity with cyanide LD50 yields insight into the efficacy of prophylactics." Journal of Applied Toxicology **33**(1): 50-55.
Mastronarde, D. N. (2003). "SerialEM: a program for automated tilt series acquisition on Tecnai microscopes using prediction of specimen position." Microscopy and Microanalysis **9**(S02): 1182-1183.
Meyers, P. R., P. Gokool, D. E. Rawlings and D. R. Woods (1991). "An efficient cyanide-degrading Bacillus pumilus strain." Microbiology **137**(6): 1397-1400.
Meyers, P. R., D. E. Rawlings, D. R. Woods and G. Lindsey (1993). "Isolation and characterization of a cyanide dihydratase from Bacillus pumilus C1." Journal of Bacteriology **175**(19): 6105-6112.
Moore, V., A. Kanu, O. Byron, G. Campbell, M. J. Danson, D. W. Hough and S. J. Crennell (2011). "Contribution of inter-subunit interactions to the thermostability of Pyrococcus furiosus citrate synthase." Extremophiles **15**: 327-336.
Mulelu, A. E. (2013). Factors involved in the oligomerisation of cyanide dihydratase from Bacillus pumilus C1, University of Cape Town.
Mulelu, A. E. (2017). Factors involved in the oligomerisation of the cyanide dihydratase from Bacillus pumilus C1, University of Cape Town.
Mulelu, A. E., A. M. Kirykowicz and J. D. Woodward (2019). "Cryo-EM and directed evolution reveal how Arabidopsis nitrilase specificity is influenced by its quaternary structure." Communications Biology **2**(1): 260.
Nagasawa, T., J. Mauger and H. Yamada (1990). "A novel nitrilase, arylacetonitrilase, of Alcaligenes faecalis JM3: Purification and characterization." European Journal of Biochemistry **194**(3): 765-772.
Nagasawa, T., M. Wieser, T. Nakamura, H. Iwahara, T. Yoshida and K. Gekko (2000). "Nitrilase of Rhodococcus rhodochrous J1: conversion into the active form by subunit association." European Journal of Biochemistry **267**(1): 138-144.
Nakai, T., T. Hasegawa, E. Yamashita, M. Yamamoto, T. Kumasaka, T. Ueki, H. Nanba, Y. Ikenaka, S. Takahashi and M. Sato (2000). "Crystal structure of N-carbamyl-D-amino acid amidohydrolase with a novel catalytic framework common to amidohydrolases." Structure **8**(7): 729-738.

Nigam, V. K., T. Arfi, V. Kumar and P. Shukla (2017). "Bioengineering of nitrilases towards its use as green catalyst: applications and perspectives." Indian Journal of Microbiology **57**: 131-138.
Novo, C., S. Farnaud, R. Tata, A. Clemente and P. R. Brown (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." Biochemical Journal **365**(3): 731-738.
Orlova, E. and H. R. Saibil (2011). "Structural analysis of macromolecular assemblies by electron microscopy." Chemical reviews **111**(12): 7710-7748.
Pace, C. N. and K. L. Shaw (2000). "Linear extrapolation method of analyzing solvent denaturation curves." Proteins: Structure, Function, and Bioinformatics **41**(S4): 1-7.
Pace, H. C. and C. Brenner (2001). "The nitrilase superfamily: classification, structure and function." Genome biology **2**(1): 1-9.
Pace, H. C., S. C. Hodawadekar, A. Draganescu, J. Huang, P. Bieganowski, Y. Pekarsky, C. M. Croce and C. Brenner (2000). "Crystal structure of the worm NitFhit Rosetta Stone protein reveals a Nit tetramer binding two Fhit dimers." Current biology **10**(15): 907-917.
Park, J. M., A. Mulelu, B. T. Sewell and M. J. Benedik (2016). "Probing an interfacial surface in the cyanide dihydratase from Bacillus pumilus, a spiral forming nitrilase." Frontiers in Microbiology **6**: 1479.
Patel, J. M. (2009). "Biocatalytic synthesis of atorvastatin intermediates." Journal of Molecular Catalysis B: Enzymatic **61**(3-4): 123-128.
Peter, J. K., R. Singh, A. K. Yadav, R. Kothari and P. K. Mehta (2023). "Toxicity of nitriles/amides-based products in the environment and their enzymatic bioremediation." Journal of Hazardous Materials Advances: 100389.
Pettersen, E. F., T. D. Goddard, C. C. Huang, G. S. Couch, D. M. Greenblatt, E. C. Meng and T. E. Ferrin (2004). "UCSF Chimera—a visualization system for exploratory research and analysis." Journal of computational chemistry **25**(13): 1605-1612.
Pettersen, E. F., T. D. Goddard, C. C. Huang, E. C. Meng, G. S. Couch, T. I. Croll, J. H. Morris and T. E. Ferrin (2021). "UCSF ChimeraX: Structure visualization for researchers, educators, and developers." Protein Science **30**(1): 70-82.
Prasad, S., A. Misra, V. P. Jangir, A. Awasthi, J. Raj and T. C. Bhalla (2007). "A propionitrile-induced nitrilase of Rhodococcus sp. NDB 1165 and its application in nicotinic acid synthesis." World Journal of Microbiology and Biotechnology **23**: 345-353.
Punjani, A., J. L. Rubinstein, D. J. Fleet and M. A. Brubaker (2017). "cryoSPARC: algorithms for rapid unsupervised cryo-EM structure determination." Nature methods **14**(3): 290-296.
Raczynska, J. E., C. E. Vorgias, G. Antranikian and W. Rypniewski (2011). "Crystallographic analysis of a thermoactive nitrilase." Journal of structural biology **173**(2): 294-302.
Randolph, T. W., E. Schiltz, D. Sederstrom, D. Steinmann, O. Mozziconacci, C. Schöneich, E. Freund, M. S. Ricci, J. F. Carpenter and C. S. Lengsfeld (2015). "Do not drop: mechanical shock in vials causes cavitation, protein aggregation, and particle formation." Journal of pharmaceutical sciences **104**(2): 602-611.
Rees, D. C. and A. D. Robertson (2001). "Some thermodynamic implications for the thermostability of proteins." Protein Science **10**(6): 1187-1194.
Rohou, A. and N. Grigorieff (2015). "CTFFIND4: Fast and accurate defocus estimation from electron micrographs." Journal of structural biology **192**(2): 216-221.

Roldán, M. D., A. Olaya-Abril, L. P. Sáez, P. Cabello, V. M. Luque-Almagro and C. Moreno-Vivián (2021). "Bioremediation of cyanide-containing wastes: The potential of systems and synthetic biology for cleaning up the toxic leftovers from mining." EMBO reports **22**(11): e53720.
Sanavia, T., G. Birolo, L. Montanucci, P. Turina, E. Capriotti and P. Fariselli (2020). "Limitations and challenges in protein stability prediction upon genome variations: towards future applications in precision medicine." Computational and structural biotechnology journal **18**: 1968-1979.
Scheres, S. H. (2012). "RELION: implementation of a Bayesian approach to cryo-EM structure determination." Journal of structural biology **180**(3): 519-530.
Schymkowitz, J., J. Borg, F. Stricher, R. Nys, F. Rousseau and L. Serrano (2005). "The FoldX web server: an online force field." Nucleic acids research **33**(suppl_2): W382-W388.
Sedova, A., L. Rucká, P. Bojarová, M. Glozlová, P. Novotný, B. Křístková, M. Pátek and L. Martínková (2021). "Application potential of cyanide hydratase from Exidia glandulosa: Free cyanide removal from simulated industrial effluents." Catalysts **11**(11): 1410.
Sewell, B., M. Berman, P. Meyers, D. Jandhyala and M. Benedik (2003). "The cyanide degrading nitrilase from Pseudomonas stutzeri AK61 is a two-fold symmetric, 14-subunit spiral." Structure **11**(11): 1413-1422.
Sewell, B., R. Thuku, X. Zhang and M. Benedik (2005). "Oligomeric structure of nitrilases: effect of mutating interfacial residues on activity." Annals of the New York Academy of Sciences **1056**(1): 153-159.
Sewell, B. T., A. Frangakis, A. Mulelu and J. Reitz (2017). "The structure of the cyanide dihydratase (CynD) from Bacillus pumilus." Acta Crystallogr. Sect. A **73**: C1296.
Sharma, N. N., M. Sharma and T. C. Bhalla (2011). "An improved nitrilase-mediated bioprocess for synthesis of nicotinic acid from 3-cyanopyridine with hyperinduced Nocardia globerula NHB-2." Journal of Industrial Microbiology and Biotechnology **38**(9): 1235-1243.
Sheldon, R. A. and J. M. Woodley (2018). "Role of biocatalysis in sustainable chemistry." Chemical reviews **118**(2): 801-838.
Simeonova, F. P., L. Fishbein and W. H. Organization (2004). Hydrogen cyanide and cyanides: human health aspects, World Health Organization.
Stevenson, D., R. Feng, F. Dumas, D. Groleau, A. Mihoc and A. Storer (1992). "Mechanistic and structural studies on Rhodococcus ATCC 39484 nitrilase.“." Biotechnology and Applied Biochemistry **15**(3): 283-302.
Sun, D., Y. Yamada, S. Sato and W. Ueda (2017). "Glycerol as a potential renewable raw material for acrylic acid production." Green Chemistry **19**(14): 3186-3213.
Tang, K., J. Yi, K. Huang and G. Zhang (2009). "Biphasic recognition chiral extraction: a novel method for separation of mandelic acid enantiomers." Chirality **21**(3): 390-395.
Tang, S., J. Li, G. Huang and L. Yan (2021). "Predicting protein surface property with its surface hydrophobicity." Protein and Peptide Letters **28**(8): 938-944.
Thimann, K. V. and S. Mahadevan (1964). "Nitrilase: I. Occurrence, preparation, and general properties of the enzyme." Archives of Biochemistry and Biophysics **105**(1): 133-141.
Thuku, R., D. Brady, M. Benedik and B. Sewell (2009). "Microbial nitrilases: versatile, spiral forming, industrial enzymes." Journal of applied microbiology **106**(3): 703-727.

Thuku, R. N., B. W. Weber, A. Varsani and B. T. Sewell (2007). "Post-translational cleavage of recombinantly expressed nitrilase from Rhodococcus rhodochrous J1 yields a stable, active helical form." The FEBS journal **274**(8): 2099-2108.
Tiberti, M., T. Terkelsen, K. Degn, L. Beltrame, T. C. Cremers, I. da Piedade, M. Di Marco, E. Maiani and E. Papaleo (2022). "MutateX: an automated pipeline for in silico saturation mutagenesis of protein structures and structural ensembles." Briefings in bioinformatics **23**(3): bbac074.
Turner, P., G. Mamo and E. N. Karlsson (2007). "Potential and utilization of thermophiles and thermostable enzymes in biorefining." Microbial cell factories **6**: 1-23.
Van Dijk, E., A. Hoogeveen and S. Abeln (2015). "The hydrophobic temperature dependence of amino acids directly calculated from protein structures." PLoS computational biology **11**(5): e1004277.
van Putten, P. v. L. (1979). "Mandelic acid and urinary tract infections." Antonie van Leeuwenhoek **45**(4): 622-623.
Vargas-Serna, C. L., M. L. Carmona-Orozco and A. J. Panay (2020). "Biodegradation of cyanide via recombinant cyanide dehydratase from Bacillus pumilus expressed heterologously in Escherichia coli." Revista Colombiana de Biotecnología **22**(1): 27-35.
Vetriani, C., D. L. Maeder, N. J. Tolliday, H. H. Klump, K. S. Yip, D. W. Rice and F. T. Robb (1998). "Improving Enzyme Thermostability: The Thermococcus litoralis Glutamate Dehydrogenase Model." New Developments in Marine Biotechnology: 221-225.
Wang, L., J. M. Watermeyer, A. E. Mulelu, B. T. Sewell and M. J. Benedik (2012). "Engineering pH-tolerant mutants of a cyanide dihydratase." Applied microbiology and biotechnology **94**: 131-140.
Watanabe, A., K. Yano, K. Ikebukuro and I. Karube (1998). "Cloning and expression of a gene encoding cyanidase from Pseudomonas stutzeri AK61." Applied microbiology and biotechnology **50**: 93-97.
Watanabe, A., K. Yano, K. Ikebukuro and I. Karube (1998). "Cyanide hydrolysis in a cyanide-degrading bacterium, Pseudomonas stutzeri AK61, by cyanidase." Microbiology **144**(6): 1677-1682.
Watanabe, A., K. Yano, K. Ikebukuro and I. Karube (1998). "Investigation of the potential active site of a cyanide dihydratase using site-directed mutagenesis." Biochimica et Biophysica Acta (BBA)-Protein Structure and Molecular Enzymology **1382**(1): 1-4.
Watson, J. L., D. Juergens, N. R. Bennett, B. L. Trippe, J. Yim, H. E. Eisenach, W. Ahern, A. J. Borst, R. J. Ragotte and L. F. Milles (2023). "De novo design of protein structure and function with RFdiffusion." Nature **620**(7976): 1089-1100.
Way, J. L., P. Leung, E. Cannon, R. Morgan, C. Tamulinas, J. Leong-Way, L. Baxter, A. Nagi and C. Chui (2007). The mechanism of cyanide intoxication and its antagonism. Ciba Foundation Symposium 140-Cyanide Compounds in Biology: Cyanide Compounds in Biology: Ciba Foundation Symposium 140, Wiley Online Library.
Weber, B. W., S. W. Kimani, A. Varsani, D. A. Cowan, R. Hunter, G. A. Venter, J. C. Gumbart and B. T. Sewell (2013). "The mechanism of the amidases: mutating the glutamate adjacent to the catalytic triad inactivates the enzyme due to substrate mispositioning." Journal of Biological Chemistry **288**(40): 28514-28523.

Woodward, J., B. Weber, M. Scheffer, M. Benedik, A. Hoenger and B. Sewell (2008). "Helical structure of unidirectionally shadowed metal replicas of cyanide hydratase from Gloeocercospora sorghi." Journal of structural biology **161**(2): 111-119.
Woodward, J. D., I. Trompetter, B. T. Sewell and M. Piotrowski (2018). "Substrate specificity of plant nitrilase complexes is affected by their helical twist." Communications Biology **1**(1): 186.
Xiong, N., P.-J. Lv, J.-W. Song, Q. Shen, Y.-P. Xue and Y.-G. Zheng (2022). "Engineering of a nitrilase through consensus sequence analysis and conserved site substitution to improve its thermostability and activity." Biochemical Engineering Journal **184**: 108475.
Xu, Z., T. Cai, N. Xiong, S.-P. Zou, Y.-P. Xue and Y.-G. Zheng (2018). "Engineering the residues on "A" surface and C-terminal region to improve thermostability of nitrilase." Enzyme and microbial technology **113**: 52-58.
Xue, L. C., J. P. Rodrigues, P. L. Kastritis, A. M. Bonvin and A. Vangone (2016). "PRODIGY: a web server for predicting the binding affinity of protein–protein complexes." Bioinformatics **32**(23): 3676-3678.
Xue, Y.-P., Y.-P. Wang, Z. Xu, Z.-Q. Liu, X.-R. Shu, D.-X. Jia, Y.-G. Zheng and Y.-C. Shen (2015). "Chemoenzymatic synthesis of gabapentin by combining nitrilase-mediated hydrolysis with hydrogenation over Raney-nickel." Catalysis Communications **66**: 121-125.
Xue, Y.-P., S.-Z. Xu, Z.-Q. Liu, Y.-G. Zheng and Y.-C. Shen (2011). "Enantioselective biocatalytic hydrolysis of (R, S)-mandelonitrile for production of (R)-(−)-mandelic acid by a newly isolated mutant strain." Journal of Industrial Microbiology and Biotechnology **38**(2): 337-345.
Yamamoto, K. and K.-i. Komatsu (1991). "Purification and characterization of nitrilase responsible for the enantioselective hydrolysis from Acinetobacter sp. AK 226." Agricultural and biological chemistry **55**(6): 1459-1466.
Yanase, H., T. Sakai and K. Tonomura (1983). "Purification, crystallization and some properties of β-cyano-L-alanine-degrading enzyme in Pseudomonas sp. 13." Agricultural and Biological Chemistry **47**(3): 473-482.
Yu, S., P. Yao, J. Li, J. Feng, Q. Wu and D. Zhu (2019). "Improving the catalytic efficiency and stereoselectivity of a nitrilase from Synechocystis sp. PCC6803 by semi-rational engineering en route to chiral γ-amino acids." Catalysis Science & Technology **9**(6): 1504-1510.
Zhang, L., B. Yin, C. Wang, S. Jiang, H. Wang, Y. A. Yuan and D. Wei (2014). "Structural insights into enzymatic activity and substrate specificity determination by a single amino acid in nitrilase from Syechocystis sp. PCC6803." Journal of structural biology **188**(2): 93-101.
Zhang, X.-H., C.-Y. Wang, X. Cai, Y.-P. Xue, Z.-Q. Liu and Y.-G. Zheng (2020). "Upscale production of (R)-mandelic acid with a stereospecific nitrilase in an aqueous system." Bioprocess and biosystems engineering **43**: 1299-1307.
Zhao, B., L. Zhao, P. Gao, Y. Han, Y. Li, Z. Dong and Y. Song (2023). "Cyanide detoxification and iron mineral recycle of cyanide tailings by the sustainable utilization technology of oxidation-reduction roasting: Process optimization and mechanism study." Journal of Environmental Chemical Engineering **11**(3): 109861.
Zivanov, J., T. Nakane and S. H. Scheres (2019). "A Bayesian approach to beam-induced motion correction in cryo-EM single-particle analysis." IUCrJ **6**(1): 5-17.

Appendix

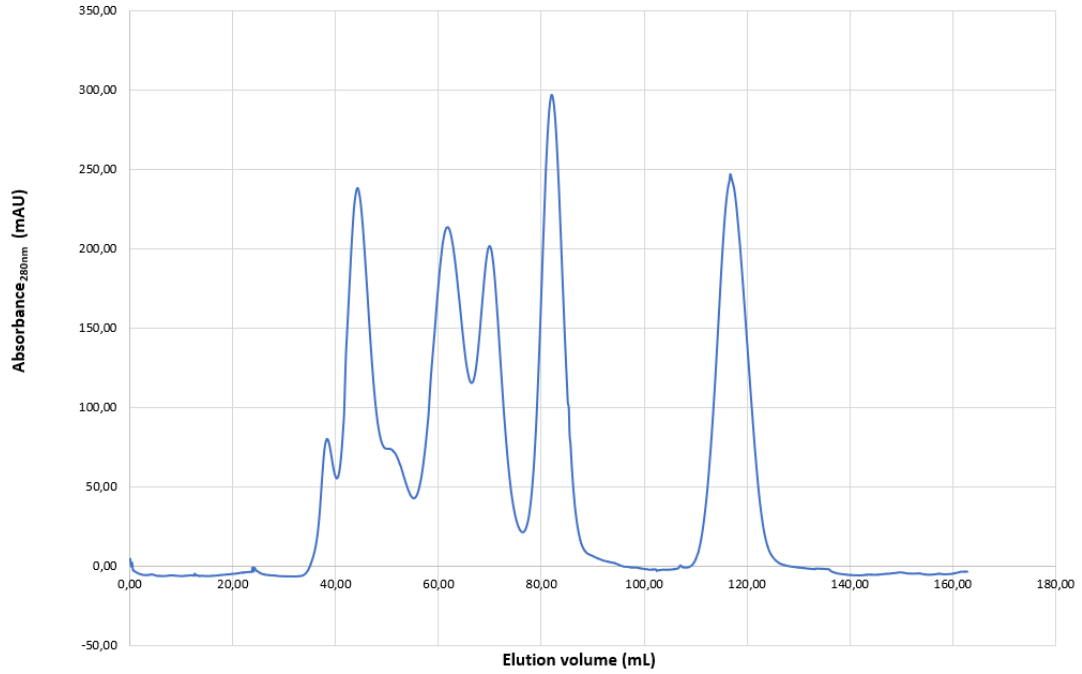


Figure A1: Elution profile of gel filtration standards used to calibrate the HiPrep™ 16/10 Sephacryl™ S300 HR column.

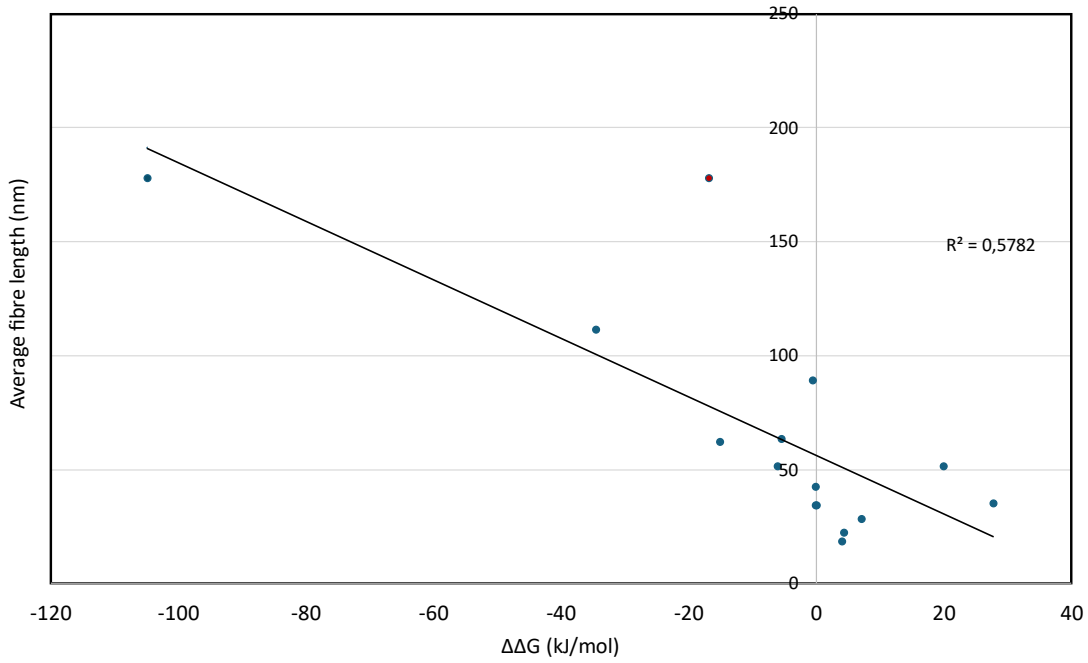


Figure A2: Correlation between the ΔΔG and average fibre length including the $CynD_{pum}$- $CynD_{stu}$ chimera.