The role of copper proteins in the ammonia oxidation pathway of Nitrosopumilus maritimus

Master of Science (by Dissertation) in Biochemistry

Megan Faye Downes

Department of Life Sciences

The University of Essex

October 2020

Abstract

The contribution of ammonia oxidising archaea (AOA) to the global nitrogen cycle has become apparent in recent years. Where previously thought to be driven by bacteria, archaea may outcompete bacteria, particularly in low nutrient environments like unfertilised soils. Ammonia oxidising bacteria (AOB) are well characterised, with ammonia oxidising pathways relying predominantly on proteins utilising haem c active sites. However, homologues for these proteins (or indeed genes for any proteins containing the characteristic haem c CXXCH motif) are absent in AOA, suggesting a completely different pathway in AOA. Novel Cu-containing multi copper oxidase homologue proteins in *Nitrosopumilus maritimus*, have been suggested as candidates for ammonia oxidation in AOA. Bioinformatic analysis of three of these proteins lead to the prediction that; Nmar1131 is a homotrimeric laccase-like 2dMCO, Nmar1354 a homotrimeric 3dMCO with an additional T1Cu site in the C-terminus cupredoxin domain, and Nmar1667 an NO-forming nitrite reductase. Purification of Nmar1131 was performed. Originally over-expression was attempted in a cytoplasmic vector, which proved unsuccessful. To echo the predicted native periplasmic nature of this protein, the *nmar1131* gene was cloned into pET26b, a vector containing a periplasmic signal peptide which yielded purified protein. Data from this study, could provide important insights into how to mitigate anthropogenic nitrogen inputs into the environment such as from modern agricultural practices.

Acknowledgements

I would like to thank everyone in Lab 5.08 for their friendship and guidance throughout this year. With special thanks to Marina for teaching me everything I know.

Thank you to Jonathan, Mike, Corinne, and Richard for massively increasing my confidence over this year and providing me with such a great opportunity. Your (mostly) positive feedback meant a lot.

Contents

Abstract 2

Acknowledgements 3

Abbreviations 8

Chapter 1 Literature Review 10

1.1. Nitrogen and the Environment 11

1.1.1. The Nitrogen Cycle 12

1.1.2. Research Context 13

1.2. Introduction to Archaea 13

1.2.1. Species Diversity 14

1.3. Ammonia Oxidising Bacteria 15

1.3.1. Ammonia Oxidation Bacteria Introduction 15

..... 16

1.3.2. Nitrification 16

1.3.3. Nitrifier Denitrification 17

1.3.4. AOB vs AOA Pathways 18

1.4. Key Proteins in Nitrogen Cycling 19

1.4.1. Ammonia Monooxygenase 19

1.4.2. Nitrite Reductases 20

1.4.2.1. Cupredoxin Domains 23

1.4.2.2. *N. viennensis* NiR 24

1.4.2.3. Nitrite Reductase Mechanism and Key Residues 25

1.4.3. NirK Cluster 26

1.4.4. Multicopper Oxidases 29

1.5. Environmental Variation in Function 31

1.5.1. *Nitrosopumilus maritimus* 31

1.6. Bioremediation 33

1.7. Techniques 33

1.7.1. Homology Searches and Sequence Alignments 34

1.7.2. Experimental Protein Structure Determination 34

1.7.3. Protein Purification Introduction 36

1.8. Project Aims 38

Chapter 2 Bioinformatics 39

2.1. Introduction 39

2.1.1. Homology Modelling 40

2.2. Methods 44
2.2.1. Sequence Retrieval and Alignment 44
2.2.2. Homology Modelling 44
2.2.3. Structural Homology Search 45
2.2.4. Protein Characterisation Programs 45
2.3. Results 46
2.3.1. Nmar1131 Homology Modelling 49
2.3.2. Nmar1131 Model Vetting 52
2.3.3. Nmar1131 Structural Homology Search 54
2.3.4. Nmar1354 57
2.3.5. Nmar1667 65
2.4. Discussion 72
2.4.1. Model Vetting 72
2.4.2. NiR Vs MCOs 73
2.4.3. Nmar1131 74
2.4.4. Laccases 76
2.4.5. Copper homeostasis 79
2.4.6. Nmar1354 80
2.4.7. Nmar1667 82
2.4.8. Conclusions 82
Chapter 3 Over-Expression of *N. maritimus* genes in *Escherichia coli* 85
3.1. Introduction to protein over-expression 86
3.2. Methods 86
3.2.1. Competent Cells Preparation: DH5α 86
3.2.2. pET28a Plasmids 86
3.2.3. pET26b Plasmids 87
3.2.4. pET28a/pET26b Transformation of XL1 Blue/BL21 DE3/DH5a/C43 Cells 87
3.2.5. Restriction Digest of Purchased Plasmids 87
3.2.6. Preparation of 1% Agarose Gel and DNA Samples 88
3.2.7. Large Scale Over-Expression of Nmar Proteins 88
3.2.8. Test Expression 89
3.2.9. Preparation of 15% SDS-PAGE Electrophoresis Gels 89
3.2.10. Preparation of SDS-PAGE Samples 89
3.2.11. Immobilised Metal Affinity Chromatography 90
3.2.12. Gel Filtration Chromatography 90
3.2.13. QuikChange Mutagenesis 91

3.2.14. Cloning into pET26b 91
3.2.15. Plasmid Sequencing 92
3.2.16. Sucrose Shock 93
3.2.17. Ion Exchange Chromatography 93
3.2.18. UV Vis Spectra 93
3.3. Results 94
3.3.1. Sequencing Purchased Plasmids 94
3.3.2. Test Expressions of Nmar genes 95
3.3.3. Test Expressions with 1mM $CuSO_4$ 96
3.3.4. Large Scale Expression of Nmar1131 96
3.3.5. Expressions with 0.1mM $CuSO_4$ 98
3.3.6. Test Expression in C43 (DE3) 99
3.3.7. Cloning into pET26b Vector 99
3.3.8. Overexpression of Nmar1131 in pET26b 103
3.3.9. Second Over-Expression of Nmar1131 in pET26b 105
3.3.10. UV Vis Spectroscopy 108
3.4. Discussion 109
3.4.1. Nmar1131_pET26b Over-Expression 110
3.4.2. Nmar1131 pET26b Characterisation 111
3.4.3. Environmental Distribution of AOA and AOB 112
3.4.4. Conclusions 112
3.5. Supplementary 114
3.5.1. Environmental Samples 114
3.5.2. *AmoA* Amplification in Environmental Samples 116
References 118

Abbreviations

AcNiR- *Achromobacter cycloclastes* Nitrite Reductase

AOA- Ammonia Oxidising Archaea

AOB- Ammonia Oxidising Bacteria

AMO- Ammonia Monooxygenase

APS- Ammonium Persulphate

AxNIR- *Alcaligenes xylosoxidans* Nitrite reductase

bp- Base Pairs

Cu - Copper

CV- Column Volumes

DMSO- Dimethyl Sulfoxide

E.coli - *Escherichia coli*

GF- Gel Filtration

IE- Ion Exchange

IMAC- Immobilised Metal Affinity Chromatography

IPTG- Isopropyl β-d-1-thiogalactopyranoside

kDa- Kilodaltons

LB- Lysogeny broth

MCO- Multicopper Oxdisase

MW- Molecular Weight

MWCO- Molecular Weight Cut Off

NiR- Nitrite Reductase

Nmar- Nitrosopumilus maritimus

NvNiR- *Nitrososphaera viennensis* Nitrite Reductase

O/N- Overnight

PDB - Protein Data Bank

SDS-PAGE- Sodium Dodecyl Sulphate–Polyacrylamide Gel Electrophoresis

T1Cu- Type I Copper Site

T2Cu- Type II Copper Site

TEMED- Tetramethy-ethylenediamine

TNC- Tri-Nuclear Cluster

Chapter 1 Literature Review

1.1. Nitrogen and the Environment

The nitrogen cycle plays a key role in sustaining life on earth. Nitrogen is a key building block to proteins and nucleic acids, thus the cycling of nitrogen as different inorganic species is vital to sustain life. Atmospheric nitrogen (N_2) is not an accessible form for most organisms due to its stability and innate inertness. Thus, nitrogen is naturally fixed into ammonia by nitrogen fixing bacteria and to a smaller degree by lightning. This more accessible form of nitrogen can then be utilised by a range of plants and microorganisms, either as an energy source, or assimilated into organic nitrogen compounds. (Arp, Sayavedra-Soto and Hommes, 2002)

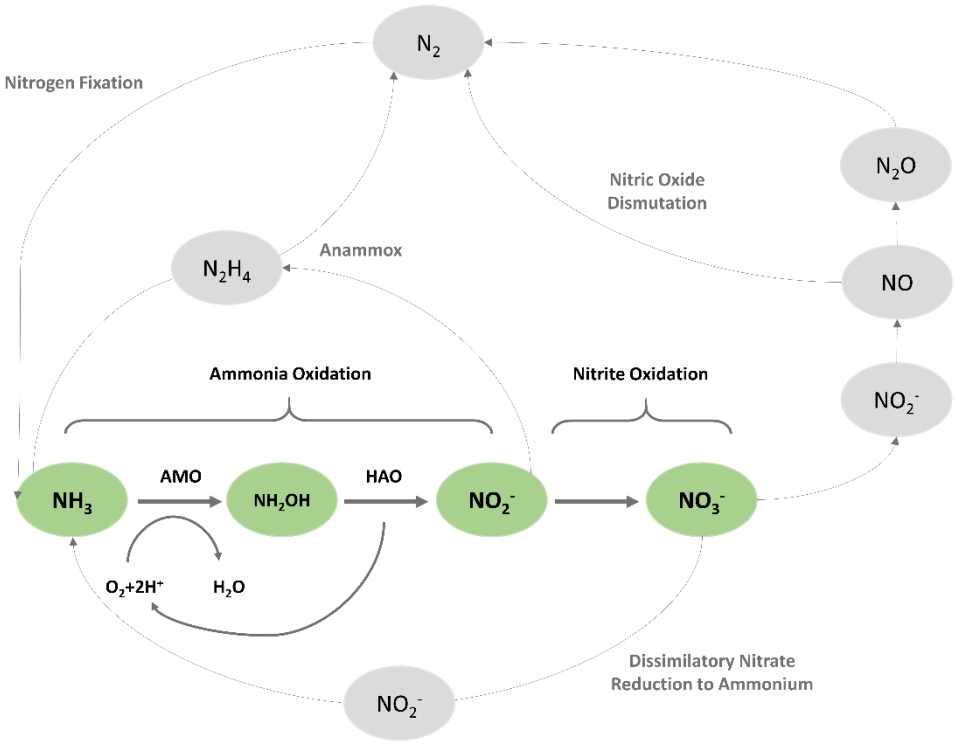


Figure 1 Simplified view of the nitrogen cycle showing oxic (green) and anoxic (grey) reactions

1.1.1. The Nitrogen Cycle

The nitrogen cycle is a complicated chain of pathways that can be carried out by a range of different organisms. A simplified cycle can be said to start with the N_2 fixation into ammonia (NH_3), which is then oxidised to nitrite (NO_2^-) and nitrate (NO_3^-) by ammonia and nitrite oxidising microorganisms. This is known as nitrification. These species are then reduced back to N_2 via nitric oxide (NO) and nitrous oxide (N_2O) by denitrifying organisms. There are many different pathways interlinking this cycle, including anaerobic ammonia oxidation (annamox), dissimilatory nitrate reduction to ammonium (DNRA), nitric oxide dismutation (where NO is converted directly back into atmospheric nitrogen) and nitrifier denitrification, ammonification (where organic nitrogen is converted to NH_4^+ by the breakdown of nitrogen containing compounds, such as amino acids) and comammox (the complete ammonia oxidation, where ammonia is converted straight to nitrate by one organism). A simplified N cycle concentrating on the processes relevant to this thesis is shown (Figure 1). The reality of nitrogen cycling depends greatly on environmental conditions and microorganisms present. (Rütting et al., 2011) (Mulder et al., 1995).

Typically, access to bioavailable nitrogen is tightly controlled, with most nitrogen in the unreactive N_2 state. However, anthropogenic contribution to reactive nitrogen species (RNS) (highly reactive nitrogen species such as NO, N_2O, NO_2^-) accumulation has increased drastically in recent years. From an estimated 3Tg $year^{-1}$ 2,000 years ago to 210Tg $year^{-1}$ in 2013 (Galloway et al., 2013). This increase is largely due to modern agricultural practices. With the advent of the Haber-Bosch process, artificial nitrogen fixation was available on a large scale, with 85% of produced ammonia is used for fertilisers. This also saw an increased cultivation of leguminous crops, increasing the abundance of symbiotic nitrogen fixing bacteria. Furthermore, the burning of fossil fuels in atmospheric N_2 and the conversion of organic nitrogen stored in fossils all contributes to this large increase in RNS. (Galloway et al., 2003)

Life evolved under nitrogen limited conditions and this increase in bioavailable nitrogen can negatively impact the environment (Lead et al., 2001). NO and N_2O are greenhouse gases, with N_2O being 300 times more potent than CO_2, 4 times stronger than methane, and with an expected lifetime of over 100 years in the stratosphere (IPCC et al., 2007), and thus contributing to global warming as well as acidification of soils and surface water (Galloway and Cowling, 2002). Furthermore, the increased concentrations of ammonium in cultivated soils gives rise to a greater concentration of nitrates through the action of nitrifying microorganisms. NO_3^- is more mobile than positively charged NH_4^+, meaning NO_3^- can leach from fertilised soils into lakes, rivers and ground water leading to complications such as eutrophication, such as algal blooms and contamination of drinking water (Arp, Sayavedra-Soto and Hommes, 2002).

1.1.2. Research Context

As RNS play a large role in global warming, it is important that the intricacies of these pathways and organism are understood if anthropogenic RNS contribution is to be mitigated. The oxidation of ammonia is one of the first steps available after N_2 fixation. For many years, this was thought to be carried out only by ammonia oxidising bacteria. However, the first ammonia oxidising archaea (AOA) was isolated in 2005 (Könneke et al., 2005) revealing a brand new aspect and further complicating this cycle.

1.2. Introduction to Archaea

Still a relatively newly discovered domain of life, archaea were once thought to only reside in extremophilic conditions, such as hydrothermal vents and salt plains, however the discovery of archaea in mesophilic, oxygenated waters and agricultural soils changed this perception.

Nitrification, as mentioned above, involves the oxidation of reduced nitrogen species by ammonia oxidising microorganisms (AOM). Where this was previously thought to be carried

out predominantly by ammonia oxidising bacteria (AOB), with well-established pathways (Denk et al., 2017; Soliman and Eldyasti, 2018)' , more recent studies have begun to show the contribution of ammonia oxidising archaea (AOA) in nitrogen cycling, making up for the higher nitrification rates in nutrient poor marine environments than expected from AOB alone (Martens-Habbena et al., 2009; Rees, Woodward and Joint, 2006).

1.2.1. Species Diversity

Archaea make up a considerable proportion of microorganisms in the mesopelagic (100-1,000m) and bathypelagic zones (1,000-4,000m). With almost 40% of microbial species belonging to archaea below 1,000m in North Pacific subtropical waters, competing with bacterial species in abundance, where bacteria were shown to be more prevalent at shallower depths (~80% below 100m) (Karner, Delong and Karl, 2001).

All known AOA belong to Thaumarchaeota and are clustered into groups, loosely dependant on ecological niche. Predominantly marine species, such as *Nitrosopumilus maritimus* are clustered into Marine Group 1.1a, where soil dwelling AOA are typically assigned to Marine Group 1.1a Associated, or Soil Group 1.1b, such as *Nitrososphaera viennensis*. Identified thermophilic AOA, have been assigned to their own group ThAOA (Hatzenpichler, 2012; Plasencia *et al.*, 2013). However, *Nitrosopumilus* species have also been found in a range of environments, suggesting that the typically marine cluster can be found in not only seawater, but a range of different environments, with varying nutrient availabilities (Bollmann, Bullerjahn and McKay, 2014).

Studies carried out in the Great Lake System in North America showed varying abundance of AOA and AOB in different freshwater lakes. Lake Superior; an oligotrophic environment, was dominated by AOA (Bollmann, Bullerjahn and McKay, 2014). AOA prevalent in Lake Erie consisted of those from the *Nitrososphaera* cluster (Soil Group 1.1b), whereas those found in Lake Superior were predominantly from the *Nitrosopumilus* cluster (Marine Group 1.1a)

and the *Nitrosotalea* cluster, (Bollmann, Bullerjahn and McKay, 2014). Whereas Lake Erie; a meso/eutrophic environment, likely caused by pollution from surrounding urban and agricultural environment (Baker and Richards, 2002), was dominated by AOB, with species showing much more diversity than the AOA in Lake Superior. This study showed a distinct difference in the environments of AOB and AOA, however adaptations of different AOA to their environmental niches have not thoroughly been examined, however, this subject is touched upon later.

1.3. Ammonia Oxidising Bacteria

1.3.1. Ammonia Oxidation Bacteria Introduction

Chemolithoautotrophs are microorganisms that use the oxidisation of inorganic compounds as their sole energy source. In the case of AOB, this is ammonia. All known AOB fix carbon from CO_2 through via the Calvin Cycle. Energy production i.e. ATP synthesis is typically carried out through an electron transport chain (ETC). There are many different techniques used by different bacteria, but a general pathway in chemolithoautotrophs is as such. Electrons are passed from an inorganic electron donor through a series of redox reactions at electron carriers called cytochromes. In aerobic respiration, O_2 is used as a terminal electron acceptor and is reduced to water by an oxidase and in anaerobic respiration, alternative terminal electron acceptors, such as nitrite are used. The reduction of nitrite to N_2O or N_2 is called denitrification. This is coupled to the synthesis of ATP through establishing a proton motive force (PMF) across the cytoplasmic membrane. The flow of hydrogen from the outside to the inside of the cytoplasmic membrane drives the synthesis of ATP from ADP.

There are two key pathways carried out by AOB, the first is nitrification, the initial oxidation of NH_3 to NO_2^-, and nitrifier denitrification, the reduction of NO_2^- to N_2O. In denitrification, nitrite can act as the final electron acceptor, and be reduced to N_2.

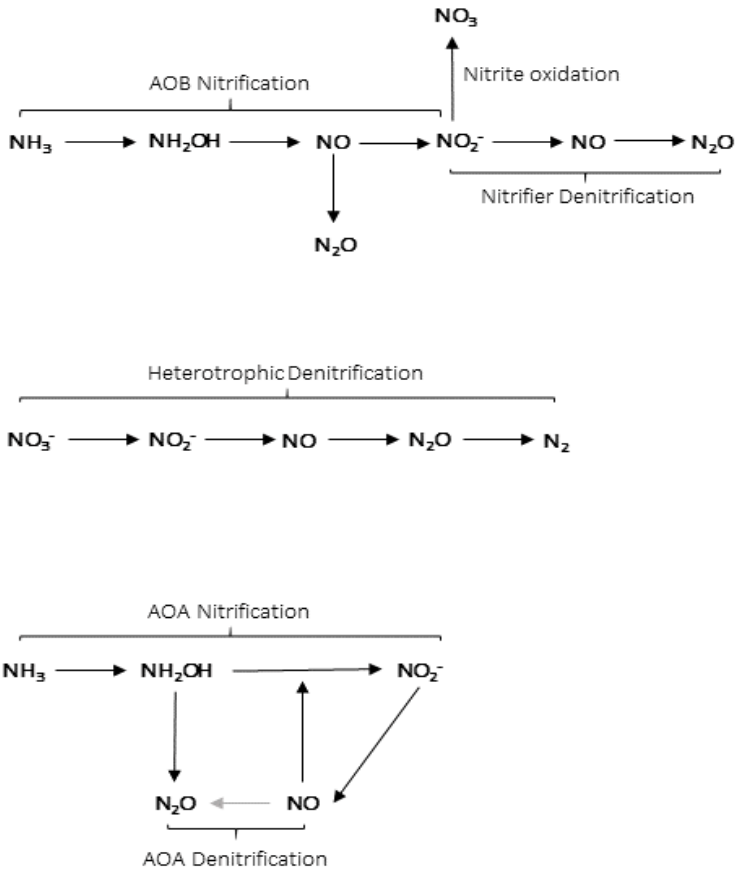


Figure 2 Key pathways of interest involved in nitrogen cycling. Grey arrows show non-enzymatic reactions.

1.3.2. Nitrification

In AOB, nitrification takes place in two steps. The first is ammonia oxidation to nitrite (NO_2^-). NH_3 is catalysed by ammonia monooxygenase (AMO), to hydroxylamine (NH_2OH). It was previously accepted that that NH_2OH was oxidised to NO_2^- by HAO, with two of four electrons released during this oxidation of NH_2OH are passed back for use in the initial ammonia oxidation reaction (Hollocher, Tate and Nicholas, 1981). However, more recent studies have shown that NO is the enzymatic product of HAO ($2NH_2OH + 1.5\ O_2 \rightarrow 2NO + 3H_2O$) (Caranto and Lancaster, 2017), with another enzyme (suggested to be NirK) further oxidising the NO to NO_2^-. Previous studies observed the release of NO_2^- from this reaction, however, this is likely due to a non-enzymatic oxidation of NO with O_2. This study shows that

NO is an obligate intermediate of AOB ammonia oxidisation (Caranto and Lancaster, 2017), This reaction releases three electrons in total, despite this reaction yielding slightly lower free energy change than the oxidisation to NO_2^-, the third step carried out by an unknown oxidoreductase, likely makes up for this loss in energy, with also releasing one electron. Showing a plausible alternate pathway, as this third step reaction makes up the four-electron complement shown to be passed to the ETC. From this two electrons are passed back for the initial AMO reaction, leaving two electrons to be passed to the ETC to generate a PMF, allowing the use of ammonia as the sole energy source for chemolithotrophic AOB (Kowalchuk and Stephen, 2001) (Caranto and Lancaster, 2017).

The production of NO from NH_2OH oxidation also helps to explain the N_2O emission from AOB under aerobic conditions. Under anaerobic conditions, this is mainly though to be due to nitrifier denitrification, or anaerobic oxidation of NH_2OH to N_2O via Cytochrome P460 (Caranto, Vilbert and Lancaster, 2016) (Figure 2). Under aerobic conditions, it has been suggested that NO is reduced to N_2O when the rate of NH_2OH oxidation outpaces the reduction of NO to NO_2^-, and thus excess NO or the reduced N_2O is emitted. Two mechanisms for the reduction of NO have been suggested. The first presumes that the reduction is carried out by NorBC, or that NO is scavenged by Cytochrome P460 or Cytochrome c554, two haem c type cytochromes involved in the ETC, (see below) which form N_2O. This mechanism also explains the increased N_2O output by AOB under low O_2 environments, the oxidation of NO to NO_2^- would presumably be seen. (Caranto and Lancaster, 2017)

1.3.3. Nitrifier Denitrification

In AOB a copper containing nitrite reductase (NirK) is used to reduce NO2- to NO ($NO_2^- + e^- + 2H^+ \rightarrow NO + H_2O$) in nitrifier denitrification. Nitrifier denitrification differs from typical denitrification in that it is not inhibited in aerobic environments, and that N2O (via NO) is

the final product instead of N2 (Beaumont et al., 2002; Dundee and Hopkins, 2001). (Figure 2). Although this pathway's function is unknown, it has been suggested that it could act as a way to conserve O2 in limited oxygen environments, a way to reduce nitrite toxicity, or potentially increase hydroxylamine kinetics by providing additional terminal electron acceptors (Beaumont et al., 2002; Lawton et al., 2013),. It has also been suggested that NirK oxidises NO produced from NH2OH to NO2-, as mentioned above (Caranto and Lancaster, 2017).

1.3.4. AOB vs AOA Pathways

Most of the key proteins in these AOB pathways use the haem c cofactor in their active site (Hooper *et al.*, 2004) However, the lack of HAO homologue, Nor, and c-type cytochromes, or indeed any genes containing the characteristic haem c CXXCH motif (Zajicek et al., 2009) (Hubertus J.E. Beaumont et al., 2004) (Hooper, 1968) in AOA suggests that they use an alternate pathways for nitrification. The only known link between the AOA and AOB ammonia oxidising pathways lies in AMO. Even at this, the AOA AMO homologue is highly divergent from the AOB AMO complex. (Walker et al., 2010; Vajrala et al., 2013)

Despite the low level of similarity between the archaeal and bacterial AMO. Archaeal AMO has been shown to produce hydroxylamine, as in the AOB pathway. It was also shown in *N. maritimus* that hydroxylamine is then oxidised to nitrite (Walker et al., 2010). Although more studies may have to be performed to determine whether NO is the actual product like in AOB.

Although the identity of many key proteins in the AOA ammonia oxidising pathway are unknown, studies have focused on a range of novel copper-containing proteins that have been proposed to replace the HAO, cytochrome c (haem c) and other key proteins not present in AOA.(Kozlowski et al., 2016; Walker et al., 2010).

1.4. Key Proteins in Nitrogen Cycling

1.4.1. Ammonia Monooxygenase

In AOB, AMO is a three-subunit membrane protein, belonging to a larger superfamily of methane, alkane, and ammonia monooxygenases. AMO catalyses the first step in nitrification by inserting an oxygen into ammonia ($NH_3 + O_2 + 2H^+ + 2e^- \rightarrow NH_2OH + H_2O$) (Ward, 2011). The two electrons for this reaction are thought to be supplied by the ubiquinol pool, which is later replenished by the oxidation of NH_2OH (see above). In AOB AMO is encoded by three genes: *amoA*, *amoB* and *amoC*. *amoA*, is commonly used as a functional marker gene for ammonia oxidisation, and has helped establish the wide ecological range of AOA and AOB in the environment. (Rotthauwe, Witzel and Liesack, 1997)

The closest relative to AMO a with a known crystal structure is a methane monooxygenase (pMMO) an enzyme catalysing the initial step in methane to methanol oxidation in methanotrophs; organisms that use methane or methanol as their sole energy and carbon source (Anthony, 1982; Bédard and Knowles, 1989). The active site of pMMO lies on the interface between the two domains of its β-subunit, with three histidines, which form the dinuclear copper co-ordination site.(Lawton et al., 2014) When this is aligned with AmoB subunit from AOB and AOA despite good alignment between AOB and the pMMO β subunit (PmoB), the AOA sequences were consistently seen to lack ~20kDa, attributed to the second domain of PmoB. Searches for the missing section of PmoB in AOA genomes showed no matches. (Tolar et al., 2017).

AmoB crystal structure from *Nitrosocaldus yellowstonii* AOA shows one of the copper co-ordinating histidines is disordered, meaning that the active site for AmoB would be disrupted if this were truly a homologous model. As the AmoB subunit has only been crystallised alone, without the other subunits. It is possible that stabilising inter subunit interactions are lost, disrupting the co-ordination site geometry.

Furthermore, a conserved helix of unknown function is found in many pMMO structures which contacts the γ subunit (PmoC). A homologous contact loop is not present in AOA AmoB. However, in a study of *N. viennensis*, an additional AMO gene of unknown function (*amoX*) was discovered. AmoX in AOA AMO may play a role in this mechanism, potentially completing the active site, suggesting a novel mechanism for AMO catalysis in AOA. (Kerou et al., 2016) (Tolar et al., 2017). (Sirajuddin and Rosenzweig, 2015).

Furthermore, in *N. viennensis*, six copies of *amoC* are present. The additional copies are thought to be part of a stress response, ensuring the stability of the AMO protein complex, aiding the recovery from ammonia starvation (Kerou et al., 2016).

AOB AMO has a broad specificity also allows the oxidation of other substrates, including aromatic and halogenated species, which may provide a mechanism for the removal of environmental contaminants other than excess inorganic nitrogen species. This wide range of substrates suggest that activation of O_2 rather than the substrate is responsible for the enzymatic reaction (Hooper et al., 1997). It has been shown that due to this promiscuity, pMMO is also capable of oxidising NH_3 to NH_2OH, strengthening the confidence in this being a suitable model for AMO (Bédard and Knowles, 1989).

There is high diversity among the AMO sequences, with those from soil and marine environments varying considerably. Likewise, marine *amoA* samples from shallow and deep waters also show defined characteristics, with potentially varying activities (Beman *et al.*, 2007).

1.4.2. Nitrite Reductases

NirK is a nitrite reductase that catalyses the reduction of NO_2^- to NO, found in both AOA and AOB. NirK is a copper containing protein, with two types of copper sites (T1Cu and T2Cu). Thought to act in the nitrifier denitrification pathway in AOB ($NO_2^- + e^- + 2H^+ \rightarrow NO + H_2O$) and an alternative NO production pathway in AOA. Recent studies showing NO production

by HAO have eluded to another function of NirK in AOB, to catalyse the final step in the oxidation of NH_3 to NO_2^- (Caranto and Lancaster, 2017).

Studies performed by Kozlowski et al. show that AOA are unlikely to perform the nitrifier denitrification pathway, with an alternate function for NirK proposed. After the initial oxidation of NH_3 to NH_2OH by AMO. It is proposed that NH_2OH is oxidised in a reaction with NO and H_2O, where NO acts as the electron acceptor ($NH_2OH + NO + 2H_2O \rightarrow 2NO_2^- + 7H^+ + 5e^-$). This NO is produced by the reduction of the subsequent NO_2^- likely by NirK, produced from NH_2OH oxidation. (Figure 2) This proposed mechanism is also seen in AOB where Cytochrome P460 carries out the same reaction. Here this is suggested to take place via a Cu containing protein. (Kozlowski et al., 2016) These results were also supported by a study on recombinantly expressed NirK from the AOA *N. viennensis* (Kobayashi et al., 2018). However, the production of N_2 was also seen, which is unexplained by these studies.

It was suggested that the ability to produce NO, a key intermediate in the AOA oxidation of ammonia(Kozlowski et al., 2016) through two different pathways gives *N. viennensis* an advantage for efficient ammonia oxidation under oxygen limited conditions, preserving the limited oxygen for use by AMO.

CuNiRs, encoded by the NirK gene are trimeric, with each monomer containing two cupredoxin (small copper protein-like) domains, adopting a β sandwich fold. Each monomer typically has a molecular weight of ~36kDa. AOB CuNiR on average have 2 Cu atoms per subunit (Kondo, Yoshimatsu and Fujiwara, 2012), Stability of the homotrimer is thought to be due to a c-terminal extension (from residue 324 in AcNIR) protruding from the second domain of one monomer, making contacts with the first domain of another(Godden et al., 1991; Horrell et al., 2017). CuNiR can be classified to two groups depending on the spectroscopic features of the T1Cu site redox partner. Electrons are fed into this mechanism from an external physiological electron donor as part of a respiratory electron transport

chain (ETC). Class I proteins are those who are reduced by azurin or cytochrome c such as AxNiR, and psuedoazurin for Class II such as AfNiR and AcNiR (Dodd et al., 1998). Furthermore, a substantially less well characterised Class III NiR exists, defined by the presence of a third domain containing an additional T1Cu or haem c site, such as *Hyphomicrobium denitrificans* (HdNiR) a hexameric NiR (Ellis et al., 2007) (Yamaguchi et al., 2004).

Here AxNiR is used as an example of the typical features and key residues found in CuNiRs.

AxNiR has two domains (domain I, residues 11 to 160; domain II, residues 173 to 336). Each subunit consists of a pair of eight-stand Greek key β-barrel (typical of cupredoxins) connected by a linker region (residues 161 to 172) (Figure 3A). Where Greek key β-barrels are typically made up of purely antiparallel strands, cupredoxin domains contain a single pair of parallel strands, an identifiable feature of this domain type (Figure 3B).

The AxNIR T1Cu site is found in Domain I. The co-ordinating ligands for the T1Cu site; His89, His139 and Cys130 form a trigonal planar geometry, distanced at 2.07 Å, 1.94 Å and 2.21Å from the Type I Cu atom respectively (Figure 3A). Typically, T1Cu also have an axial Methionine ligand (Met144) which has weaker co-ordination, due to its more distal position at 2.62Å.

The AxNiR T2Cu Site lies at the interface of Domain I and II, co-ordinated by His94, His129 and His300. His94 and His129 from Domain I, and His-300 from Domain II of the neighbouring chain (Figure 3A), residing 2.02 Å, 2.13 Å and 2.25Å away from the Type II Cu respectively (Dodd et al., 1998).

Residues Cys129 and His130 form the Cys-His bridge between the two copper sites and is thought to be the internal electron transfer route in this mechanism. With a peptide stretch made up of His89, Asp92, and His94, was deemed the sensing loop, which determines the presence of nitrite in the binding site. This aspartate is conserved in most known CuNiRs and

is thought to be essential for catalytic activity. However, *Thurmus scotoductus* (Ts)NiR a Class III NiR was recently discovered to have an sensing loop without this catalytic residue (Opperman et al., 2019).

The tower loop of AxNiR (Pro-180 - Thr206) forms a random coil which leads into an α helix consisting of four turns, towards the T1Cu site (Boulanger and Murphy, 2002; Dodd et al., 1998; Kataoka et al., 2000). The tower loop and defined hydrophobic patch on CuNiRs are thought to be important for protein-protein interactions with potential electron donors. Typical tower loops extend from second domain in one monomer, towards the T1Cu site in the N-terminus to help position electron donors near the T1Cu site hydrophobic patch (Boulanger and Murphy, 2002), where electrons are initially donated to the CuNiR (Strange et al., 1999).

1.4.2.1. Cupredoxin Domains

Cupredoxin domains differ from other Greek Key β-Barrel motifs , as they contain a mixture of anti-parallel and parallel β-strands, as opposed to the typical solely anti-parallel β-strands (Figure 3) seen in superoxide dismutases, for example. (Murphy, Lindley and Adman, 1997). These β-strands are connected by loops that aren't well conserved between different Greek Key β-barrels, in NiR the first of these loops forms an α-helix, with negatively charged residues, which contribute to the interaction surface with its electron donor (Kukimoto et al., 1996).

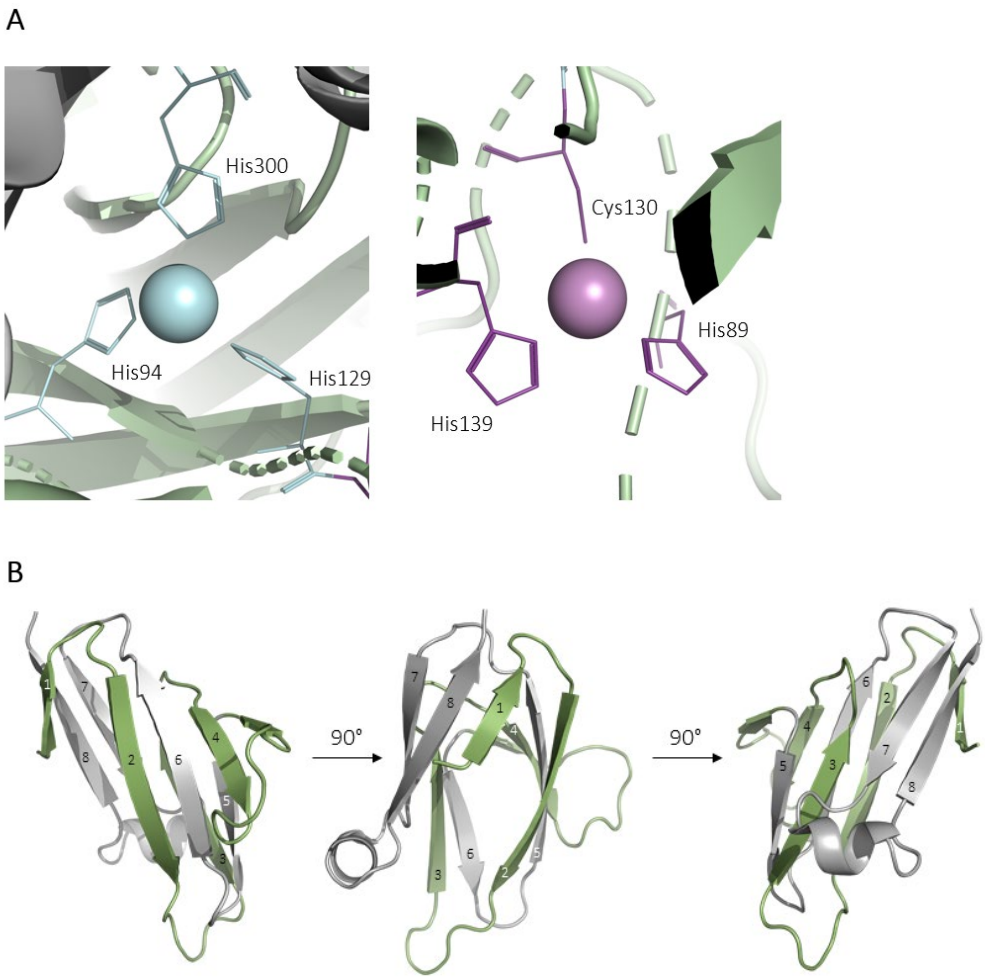


Figure 3 A) T1Cu (cyan) and T2Cu (magenta) sites in AxNiR. B) A typical cupredoxin Greek Key β- barrel domain from 3GDC, an MCO protein from N. europaea. The eight β strands are labelled showing the fold of the barrel. A mixture of parallel and anti-parallel strands can be seen (7 and 8).

1.4.3. *N. viennensis* NiR

Much like in the AOB NirK, equivalent residues co-ordinate the Cu ions at T1 (His101, Cys141, His152, and Met157) and T2 Cu sites (His106, His140, and His316) in the *N. viennensis* NirK (NvNiR) along with typical tower loop and linker loops. However, the NvNiR also has a ~26 residue c-terminus extension not seen in AOB structures, which may impact function, or the contacts AOA NirK can make. A signal peptide (Met1 to Ala24) was also present, pointing to the periplasmic nature of this protein, as seen with other NirK (Kobayashi et al., 2018).

One study recombinantly expression NirK from *N. viennensis* found the K_m value for NO_2^- reduction 287µM, and activity turnover of 3.1 s^{-1}. When this is compared to values for AOB, the K_m is significantly higher, and the activity much slower(Kobayashi et al., 2018).NirK from *Nitrosococcus oceani* for example has a K_m 53µM and a turnover rate of 1,600 s^{-1}. Likewise, for A. xylosoxidans with a K_m of 35µM and turnover rate of 445 s^{-1} (Abraham, Lowe and Smith, 1993; Kondo, Yoshimatsu and Fujiwara, 2012). This large increase in K_m seen in *N. viennensis* is likely typical for AOA in general, as an adaptation which allows them to reside in niche, low nutrient environments.

1.4.4. Nitrite Reductase Mechanism and Key Residues

A key part of the CuNiR mechanism involves an electron being passed from the T1Cu site received from an electron donor, to the T2Cu catalytic centre. However, it is disputed as to whether this occurs before or after the binding of the nitrite substrate, leading to the proposal of two different mechanisms. The Random Sequential Mechanism says that this order is variable depending on pH and substrate concentration. Where the Bind First Mechanism says the nitrite displaces an ordered water molecule in the active site, then T2Cu is reduced(Strange et al., 1999). However, the true nature of this mechanism is highly contested and will not be looked at in depth here, other to highlight key residues at these centres.

The involvement of a proximal Aspartate residue (Asp98 in AcNiR) has been suggested as fluctuations in its conformation from serial crystallography (a technique used to look at time-resolved structures) were seen throughout the catalytical cycle. Suggesting a possible role in T2Cu mechanism perhaps triggering the electron transfer from the reduced Cu_2^+ centre to reduce NO_2^- to NO (Ghosh et al., 2009) Changes in Met141 and Glu139 conformation were also seen, but no function was suggested (Horrell et al., 2016).

1.4.4.1. Other Nitrite Reductases

A second class of Nitrite reductase consists of NirS. A cytochrome cd_1-type nitrite reductase thought to carry out the same reaction as NirK in denitrifying bacteria. These two NiRs and typically do not co-exist (Coyne et al., 1989). Although primarily beyond the scope of this research, it is introduced here as a means of comparison later.

1.4.3. NirK Cluster

The *nirK* gene can be found in different types of gene clusters, depending on the phylogeny and characteristics of the species.(Cantera and Stein, 2007). In AOB, *nirK* is found in an operon with three n*irK* cluster genes (*ncg*) *ncgA*, *ncgB* and *ncgC*. *ncgBC* are two haem c type cytochromes, with di and mono haem centres respectively. *ncgA* is a predicted multicopper oxidase protein (also known as Blue Copper Oxidase (BCO)), with mild nitrite reductase activity. MCOs are discussed in greater detail below (Beaumont et al., 2005).

Mutant studies in *N.europea* showed that *ncg* are not required for functional Nirk expression, but to mitigate the toxic NO produced by NirK during NO_2^- reduction. Ncg mutants have higher NO_2^- susceptibility, but when combined with the induced expression on NirK, NO_2^- sensitivity and thus growth were more greatly impacted than these Ncg mutants alone (Beaumont et al., 2005).

In AOB, the expression of this cluster is controlled by NsrR, a transcriptional repressor, which regulates *nirK* expression via the upstream P_{nir} promoter in direct response to NO_2^- and NO (Figure 4) (Cantera and Stein, 2007). The increase in NO_2^- is thought to directly inhibit this repression, to allow the transcription of NirK in the presence of NO_2^-, it is also shown that a drop in pH also increases the transcription, however is thought to be via a more indirect pathway (Hubertus J. E. Beaumont et al., 2004).

NsrR is only one transcriptional repressor of the *nirK* gene clusters, NsrR are found in β-proteobacteria and γ-proteobacteria and Gram-positive bacteria. Others include DNR and

NnrR transcriptional activators, sensitive to NO. Binding sites for NnrR and DNR regulators are often found in denitrifying bacteria (Rodionov et al., 2005).

In *P. stutzeri*, when *nirS* is deactivated through a mutation, denitrification pathway is restored by the addition of nirK. As this enzyme is active *in vivo*, *P.stutzeri* must possess an appropriate electron donor for the CuNiR, however this was not the case for the in vivo expression in E.coli, suggesting no appropriate electron donor in this system (Glockner, Jüngst and Zumft, 1993).

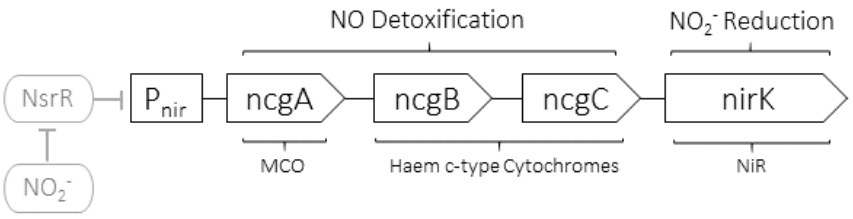


Figure 4 NirK Gene Cluster in N. europaea. Transcription is controlled from P_{nir} promoter, and inhibited by NsrR, this is reversed by the presence of NO_2^-.

Azurin functions as an electron donor for CuNiR in P.aureofaciens in vitro, however, P.stutzeri lacks azurin, suggesting that the nirK protein accepts electrons from another endogenous electron donor, potentially one of the multiple soluble Cytochrome c-type proteins found in P. stutzeri genome (Zumft et al., 1988).

The MCO encoded by the NcgA gene is also the most well characterised MCO (2dMCO) found in nitrifiers, so has also been used as a model(Lawton and Rosenzweig, 2011). Close relationship between nitrite reduction and ammonia oxidation. Function can be suggested by the surrounding genes. The presence of an MCO in the NirK cluster strongly suggests a role of MCOs in ammonia oxidation.

Copper nitrite reductases (CuNiR) are typically inactivated by the presence of oxygen. This is likely due to the role of NiR in anaerobic respiration, which is less efficient than aerobic respiration, thus in the presence of O_2, the higher energy yielding process should be

favoured. The NiR from *Nitrosomonas europaea* (NeNiR), a model nitrifying chemolithoautotrophic bacterium has been characterised (Lawton et al., 2013).

Sequence alignments of CuNiR from denitrifiers and NeNIR show Type I Copper sites (T1Cu) and Type II Cu sites (T2Cu) are conserved, as well as Histidine and Aspartic residues adjacent to the T2Cu centres. However, where typical T2Cu sites are accessed by one wide funnel-shaped channel, the NeNIR T2Cu channel has additional residues found in a backbone extending into the channel and a shorter loop in the channel opening, along with an additional Met128 residue, in place of a Valine or Isoleucine found in most other CuNIRs. This limits access to only one well-ordered water molecule in the channel. The sulphur of this Met128, as well as an Asp-86 O atom, lie within hydrogen bonding distance of the water molecule. Met128 lies 5.8Å of the T2Cu site, indicating could possible interaction with the bound substrate (Lawton et al., 2013).

Susceptibility to oxygen in denitrifier CuNiR is thought to be due to the binding of O_2 to T2Cu sites and its subsequent reduction to peroxide (Kakutani et al., 1981), however, it has been suggested that the presence of Met128 in NeNiR inhibits by blocking the formation of key intermediates in this reaction. NeNIR also has a secondary channel leading to the T2Cu site, which may act to account for solvent diversion resulting from the restricted primary channel. It is though that these features may attribute to the oxygen tolerant nature of NeNIR (Lawton et al., 2013). This may provide some insight to the structure of AOA CuNiR, as ammonia oxidation in these organisms occurs aerobically, and the production of NO by CuNiR is an essential intermediate in ammonia oxidation, as mentioned previously.

A very similar tower loop in the NeNiR and NeMCO suggests a shared electron donor. However, differences in charge of the hydrophobic patches in these two proteins suggest a different mode of interaction, or a different electron donor entirely (Lawton et al., 2013).

1.4.4. Multicopper Oxidases

Multicopper oxidases (MCOs) are a group of proteins that use Cu ions as cofactors to couple the oxidation of a broad range of substrates to the reduction of O_2. MCOs are descendants from small blue proteins (cupredoxins) and can be made up of two, three, or six cupredoxin domains described above.

MCOs can be divided into two groups determined by functionality. Despite all MCOs' ability to oxidise aromatic compounds, some have a propensity to also oxidise metal ions with high efficiency, such as CueO. These are termed metallo-oxidases (EC 1.16.3). Those which have a higher efficiency oxidising organic compounds are called laccases (EC 1.10.3.2) . (Stoj and Kosman, 2006)

A typical MCO ligates four Cu ions into two types of Cu centres. The first being a mononuclear T1Cu centre, and the second, a trinuclear Copper centre (TNC) consisting of a lone T2Cu centre and a dinuclear Type III Copper (T3Cu) Centre (Figure 5). The four Cu ions allow the transfer of one electron each, allowing the reduction of O_2 to two water molecules. Although NiR are not technically MCOs due to the lack of TNC, many similarities are seen between the two types of proteins. Much like in NiRs, substrate oxidation occurs at the T1Cu site, and electrons are passed down to the TNC for a O_2 reduction (Cys-His residues) (Kosman, 2010).

A range of MCOs are prevalent in AOA genomes with *N.viennensis* encoding four periplasmic MCOs and (Kerou et al., 2016). *N. maritimus* has three 2dMCOs and one 3dMCO. Although the function of MCOs in AOA ammonia oxidation is still not understood. There have been suggestions of their function based on phylogenetic analysis of AOA genomes.

One study sought to elucidate the ammonia oxidation pathway of AOA by establishing a core set of genes present in AOA. From this, one 2dMCO was found to co-occur in almost all AOA with a Zinc-Iron Permease (ZIP) encoding gene. ZIP proteins are responsible for the

transport of divalent metal ions, such as Cu^{2+}. Both genes were seen to have a shared evolutionary history, along with the fact that co-occurrence often suggests a link in function and regulation (Walker *et al.*, 2010). It is a fair assumption that the proteins they encode have a linked function, suggested to be copper sequestration in the periplasm. Alternatively, this MCO could still be a functional part of the ammonia oxidation pathway, as a module of archaeal hydroxylamine: ubiquinone redox module (HURM) (Kerou et al., 2016). The former was also seen in a study by Walker et.al where Nmar1131 and 1663; soluble periplasmic MCOs, were shown to be colocalised with members of DtxR metal regulator family of genes and ZIP genes respectively, which was suggested to show a role in metal homeostasis for these MCOs(Walker et al., 2010). NirK was also found to be well conserved in AOA genomes, as mentioned above.

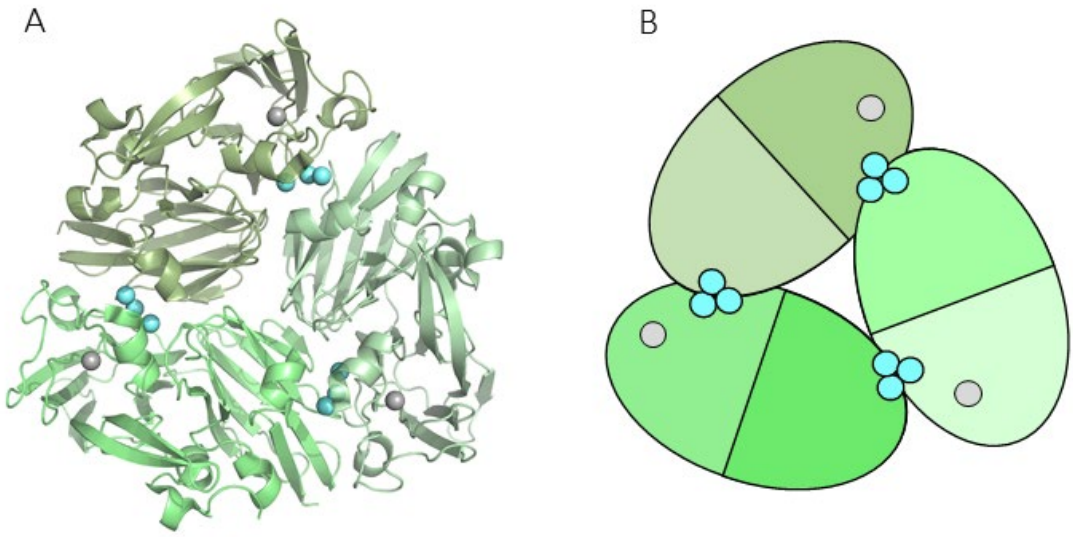


Figure 5 MCO structure. A) 3GDC an MCO from N. europaea B) Cartoon representation of a typical 2dMCO. The T1Cu site is found in domain I of each subunit. The TNC is found between two subunits. The residues co-ordinating the TNC coppers arise from domain I and trans acting residues in domain II of the adjacent subunit.

1.5. Environmental Variation in Function

1.5.1. *Nitrosopumilus maritimus*

Nitrosopumilus maritimus is a slow growing, chemolithoautotrophy which solely uses the oxidation of ammonia for growth. Originally isolated from a tropical fish tank in Seattle, grown in synthetic crenarchaeota medium (SCM) supplemented with 500µM ammonium, with a measured minimum generation of time of 21h at 28°C without shaking, as this was shown to impede growth. It is a marine group 1.1a thaumarchaea adapted for growth in oligotrophic open ocean environments where the average range for ammonium concentrations of <0.03–1 µM (and <0.03–100 µM in coastal waters) (Könneke et al., 2005). *N.maritimus* fixes carbon through the 3-hydroxypropionate/4-hydroxybutryrate pathway, typical of marine crenarchaeota (Walker et al., 2010).

N. viennensis, a terrestrial I.1b group of thaumarcheota isolated from soil shares many properties with the earlier isolated N. maritimus with a few key differences. *N. viennensis* (like *N. piranensis*) is able to grow on higher ammonia concentrations than *N. maritimus* and is also able to use urea as an energy source, and although capable of chemolithoautotrophic growth, *N. viennensis* requires the addition of pyruvate (as a carbon source) to achieve comparable growth rates to *N. maritimus* (Tourna et al., 2011).

Unlike *N. maritimus*, the growth of *N. viennensis* was not inhibited by shaking (Tourna et al., 2011), potentially showing the adaptations to growth in a terrestrial environment, as opposed to the pelagic open ocean of marine AOA. *N. viennensis* was the first AOA to be isolated from soil, and demonstrated that these AOA are also capable of aerobic ammonia oxidation to nitrite (Tourna et al., 2011).

Looking at archaeal NirK genes, there are distinct variations in NirK sequences, although these are less discrete differences than between different ecotypes in AmoA. However,

Phyre2 structure prediction of different archaeal NirK assigned the soil samples a laccase template and the marine sample a nitrite reductase (Tolar et al., 2017).

AOA *Amo* gene abundance is significantly more prevalent than AOB in marine and unfertilised soil environments. It has been suggested that substrate availability is responsible for this difference in abundance, with kinetic studies supporting this theory, showing N.maritimus (SCM1 strain) to have much lower K_m value than other known ammonia oxidising microorganism (~0.15μM) (Martens-Habbena et al., 2009). However, archaea are difficult to study due to their prevalence in low nutrient environments and thus difficulty in producing large scale pure cultures to study in a laboratory setting.

Where catalase genes (catalyse the decomposition of peroxide to water) are found in most AOB, this is not the case for AOA. Recent studies have shown that α-keto acids are used as a non-enzymatic peroxide detoxification mechanism, where these were previously thought to be used for carbon assimilation suggesting mixotrophic growth. The discarding of these genes in AOA genomes may be due to specialisation to their environment. The oligotrophic environments most known marine AOAs have been isolated from also have low concentrations of H_2O_2, which may have made these genes obsolete, especially in the presence of peroxide scavenging bacteria, which were shown to promote AOA growth in co-cultures lacking α-keto acids. However, the variation seen between AOA species, makes sweeping observations difficult to assume for all AOA species. As features such as this are likely due to adaptations to environmental niches. However, this study confirmed that AOA are most likely not capable of mixotrophic growth and grow chemolithoautotrophically (Kim et al., 2016).

Contrastingly to other AOA, both *N. viennensis* and *Nitrososphaera evergladensis* lack a homologue of the H^+, Na^+-translocating pyrophosphatase, suggesting another element to

the energy metabolism of these, or AOA in general, as this pump is an important part of the electron transport chain's generation of proton motive force (Kerou et al., 2016).

1.6. Bioremediation

Despite this, under controlled environments, there are possible beneficial applications of nitrification for use in bioremediation. Over fertilised soils and raw sewage, a NH_4^+ rich environment can be reclaimed through treatment with nitrifying and denitrifying organisms, allowing conversion back into non-bioavailable atmospheric N_2 (Arp, Sayavedra-Soto and Hommes, 2002) (Kowalchuk and Stephen, 2001).

N_2O production in wastewater is related to three processes. Nitrifier denitrification (the reduction of NO_2^- by pyruvate, ammonia or hydrogen) by AOB heterotrophic denitrification (the reduction of NO_3^- to N_2 via NO_2^-, N_2O, and NO), and hydroxylamine oxidation in AOB (potentially due to an imbalance in metabolic activity or hydroxylamine oxidation by nitrite). These are all though to take place because of an imbalance in metabolic processes, or limited availability of oxygen or nitrite (toxic) accumulation. As in these processes, NO_2^- is used as an alternative final electron acceptor to O_2 in the electron transport chain (Wunderlin et al., 2012)(Kampschreur et al., 2009) (Bock et al., 1995).

As oxygen typically inhibits denitrification activity, aerobic denitrifiers, are useful in bioremediations in biowaste and ground water applications, as conventional activated sludge plant tanks, and agricultural areas are typically aerobic (Otani, Hasegawa and Hanaki, 2004). This could suggest future use for AOA for industrial applications.

1.7. Techniques

Uniprot (Bateman *et al.*, 2021)un is a secondary database which pools a wealth of structural and functional information from other databases, such as GenBank (Clark *et al.*, 2016),

STRING (Search Tool for the Retrieval of Interacting Genes/Proteins) (Szklarczyk *et al.*, 2019), eggNOG (evolutionary genealogy of genes: Non-supervised Orthologous Groups) (Huerta-Cepas *et al.*, 2016) etc. to display a comprehensive collection of information on the chosen protein/gene.

1.7.1. Homology Searches and Sequence Alignments

One way to compare the possible functions of an unknown gene is to conduct a search for similar genes in different organisms. This was done by BLAST search by nucleotide and amino acid sequence. It is possible to filter this search for from all submitted sequences to the various reference databases. Or the search can be made for specific by for example filtering by evolutionary domains, or species, or for known protein structures from the Protein Databank (PDB). Sequences found from this search can then be analysed through alignment software such as MUSCLE to align equivalent residues/amino acids from different sequences. Alignment outputs can be further formatted using programmes such as Weblogo (Crooks *et al.*, 2004) and BoxShade. This allows conserved residues or nucleotides to be visualised, eluding to key residues or regions linked to protein function and phylogeny.

Further bioinformatic programs such as Phyre2 (Kelley et al., 2015), I-TASSER (Zhang, 2008)(Roy, Kucukural and Zhang, 2010;Yang et al., 2015), SWISS-MODEL (Waterhouse et al., 2018) can also be used to predict protein structure from amino acid sequence alone. These use the principle of homology modelling, which uses the amino acid sequence to search for template structures to model the amino acid sequence against. These three programs are discussed in greater detail below.

1.7.2. Experimental Protein Structure Determination

As useful as bioinformatics techniques can be at predicting the structure and function of unknown proteins. There is still a lot of room for error, and there are many factors that, at current, cannot be accounted for by these modelling programmes. There are a range of

different structural techniques such as Nuclear Magnetic Resonance Imaging (NMR), X-ray crystallography and cryo-electron microscopy (Cryo-EM). All these techniques have their own advantages and disadvantages, and have varying suitability to different of proteins, depending on size, and nature of the protein.

X-Ray Crystallography

For X-ray crystallography, the purified protein in crystallised, and X-ray beams are used to create diffraction within the crystal structure which produces a distinct diffraction pattern which is analysed to create a map of electron density. This, along with the known protein sequence can be modelled into an overall structure. Crystal quality is important for the reliability of the spectra created. Some proteins, typically those with little flexibility in their structure, crystallised more easily than others and are more suited to X-ray crystallography (Smyth and Martin, 2000).

Nuclear Magnetic Resonance Imaging

In NMR, the purified protein is placed in a magnetic field and subjected to radio waves, which produce specific spectra which can be interpreted to determine the relative positions of residues and produce an overall map of the protein structure. NMR is more suitable for smaller proteins (<35kDa), and is carried out in solution, meaning it can be used for proteins that are hard to crystallise, or those with flexible regions, although these regions will give a weaker signal than the less more static regions (Cavalli et al., 2007).

Cryo-Electron Microscopy

In Cryo-EM, a sample of pure protein is vitrified and imaged in 2D by a transmission electron microscope. The sample will contain many molecules of the protein at different angles, which can be put together to form a 3D image of the protein. As Cryo-EM does not require proteins to be crystallised, it can be used to observe proteins in their native environment

and in different conformations. Cryo-EM is also particularly useful for looking at large proteins and complexes and those with flexible regions. One of the main obstacles with Cryo-EM currently is the resolution, however this is being improved rapidly (Murata and Wolf, 2018).

Often, a combination of different methods can be used to produce an overall structure of a protein. Particularly with large assemblies, where smaller parts of the structure can be identified by X-ray crystallography or NMR and fit into a lower resolution model produced by Cryo-EM.

1.7.3. Protein Purification Introduction

A common theme in all these techniques is the need for purified protein. There are also a variety of protein purification techniques which take advantage of the biochemical properties of the target protein, again, making some techniques more suitable than others.

Protein purification typically begins with the overexpression and extraction of protein from these cells, typically expression cell lines of E.coli, such as BL21 (DE3). The cell lysate is then applied to a column, where the target protein will bind with high affinity, allowing none-specific protein present in the lysate to be removed. The way in which the target protein binds is what differs throughout the different techniques.

Immobilised Metal Affinity Column

One of the simplest means of purification is through an Immobilised Metal Affinity Column (IMAC). The addition of a tags that bind to immobilised metals in the column. Such as a poly Histidine tag with a Nickle column. The Histidine tag, typically around 6 residues long at the N or C terminus of the target protein allow it to bind strongly to Ni^{2+} ions in a column through the imidazole groups on Histidine. This interaction is later disrupted by the addition of free imidazole, which outcompetes and displaces the bound protein. This technique is capable of up to 95% purification in one step, but is also susceptible non-specific binding,

especially with lower protein concentrations, and to issues caused by protein folding, where the tag may become inaccessible, or inversely the tag may cause the protein to aggregate. This can be overcome by purifying the protein under denaturing conditions; however, this may cause problems for future experiments. Efficiency of this process can be improved by using a multi-affinity fusion system, where two different types of tags on one protein, for example a Histidine and a biotin tag (Bornhorst and Falke, 2000).

This concept is used in other types of purification, including Ion Exchange Chromatography. The protein is loaded to the column in a buffer with a pH above or below its isoelectric point, so the protein and the column are oppositely charged. With the addition of a second buffer which changes the charge of the protein, or outcompetes it with high salt concentration, the no longer binds and is eluted. For example, an anionic column will have a positively charged surface, thus overall negatively charged proteins will bind. If a protein has an isoelectric point of 5.0, in a buffer of pH >5.0 the protein will be negatively charged and will bind to the column. Lower the pH of the buffer to 4, and the protein will be positively charged and will elute from the column (Selkirk, 2004).

The robust properties of these purification methods mean they are suitable as the first step in purification of a crude cell lysate and can be followed up by further purification techniques.

Size Exclusion Chromatography

Following these methods, further chromatographic methods can be used to increase purity, such as Size Exclusion Chromatography (SEC). This method separates molecules by size. An aqueous solution containing the protein of interest is passed through the column packed with beads of different sizes depending on the column type used. The contents move in and out of the pores formed the beads in the column. Small molecules have access to pores

formed by these beads unavailable to larger molecules. This leads to the separation of molecules by size, and the removal of contaminants (Hagel, 2011).

1.8. Project Aims

This project aims to express and purify three archaeal proteins; Nmar1131, Nmar1354 and Nmar1667. As these are predicted copper-containing proteins, homology modelling was used to probe the nature of the copper co-ordinating sites and overall structure. This data allows predicted classification of these proteins, as a step towards elucidating their predicted role in archaeal ammonia oxidisation.

Chapter 2 Bioinformatics

2.1. Introduction

Bioinformatics plays a key role in identifying the function of unknown proteins. In the absence of structural data, homology modelling tools can be used to predict the structure of proteins as well as identify key features eluding to the protein's function. Furthermore, alignment software such as BLAST can be used to search for conserved DNA or amino acid sequences, which can be used to identify possible functional and phylogenetic relationships between genes or proteins from different organisms. Conservation of specific residues across different organisms often points to the key functional regions of the protein, whether this be the active site, a site of interaction, or structural motif.

Homology modelling can predict the structure of a protein from its amino acid sequence by comparison to a homologous protein with a known structure. This technique typically follows five steps. First, an appropriate template is selected from a database of experimentally produced protein structures, it is typically considered that proteins with a 35% minimum sequence identity are reliable templates. Secondly, the target sequence is aligned with the template and a model of the core target protein is produced, as this is the region where most confidence can be placed in the model. In two proteins where there is 50% sequence identity, the α-carbons of representative residues can typically be placed within 1Å for the core protein. Step three models the less predictable surface loops, where there is typically much more structural deviation than in the core of the protein. Again, this is driven by α-carbon placement. Modellers often have a selection of known loop α-carbon conformations, along with the neighbouring "stem" residues that connect the loop to the core structure. This database is extracted from the PDB and is used to model more generic structures onto the target with conformational agreement with the stem residues to the

target protein. Step four is where the side chains of the residues are added in sterically compatible positions relative to the side chains of the neighbouring residues, determined by the previous backbone modelling. Finally, in step five, the model is refined in terms of atom placement (Rhodes, 2006).

Even highly homologous proteins often adopt different folds, and the same protein folds can often be seen in proteins with low sequence homology(Zhang and Skolnick, 2005). This is important to consider when homology modelling and highlights the importance of model vetting as part of this process. There are many ways to test the quality of a model, not all of which are created equally. A range of different modelling software and quality parameters were used and are discussed in detail below.

The three proteins of interest; Nmar1131, Nmar1354, Nmar1667, predicted copper-containing proteins from *Nitrosopumilus maritimus*, are all of unknown structure and function. Here, a variety of bioinformatics techniques are applied to elucidate the roles of these proteins, and their importance in the nitrogen cycle of AOA.

2.1.1. Homology Modelling

Three different homology modelling programs were used to create comparative structures of each target protein. As these programs utilise different algorithms, and have their own individual strengths and weaknesses, multiple programs were tested to find the most reliable model possible, and to find structural consistencies in search for a consensus structure. These were; Protein Homology/Analogy Recognition Engine V 2.0 (Phyre2)(Kelley et al., 2015), Iterative Threading Assembly Refinement (I-TASSER) (Zhang, 2008)(Roy, Kucukural and Zhang, 2010)(Yang et al., 2015)and SWISS-MODEL(Waterhouse et al., 2018).

I-TASSER, is a homology modelling method which relies on the threading technique, also known as fold recognition for template selection. Phyre2 is the second iteration of the Phyre modelling software and specialises in using uses remote homology detection methods.

SWISS-MODEL carries out comparative modelling from an evolutionarily related homologue template, and is the only program out of those used that can predict multimeric assemblies (Waterhouse et al., 2018).

TM-Align is a program used to compare the structure of two proteins of different sequence (unlike TM-Score which is designed for those of the same sequence). TM-Align produces a structural similarity score between the two proteins, based on a pairwise alignment by structural similarity. This score is scaled by protein length and residue distances. A score of 1 suggests two proteins of identical structures. Higher than 0.5 suggests that the two proteins adopt the same fold, and a score of lower than 0.2 suggests there is no statistically relevant homology between the two proteins (Zhang and Skolnick, 2005).

Qualitative Model Energy Analysis (QMEAN) was another program used to test the quality of predicted models. The QMEAN Z-score is calculated in a way to normalise by protein length, by comparing the base QMEAN score with other scores from high resolution experimentally derived structures. This is referred to as the measure of “nativeness” and bases the score on the likelihood of these structures being found in nature, as a measure of how reliable the model is. QMEAN considers backbone geometry, solvent accessibility and residue contacts, factors which drive the native folding of proteins. The higher the QMEAN Z-score, the better the model quality. A score of 0 shows that the model is of similar quality or “nativeness” as an experimentally derived structure. Below -0.4 is considered a poor-quality model. Higher than 0 shows a very good model or in some cases a very stable protein, these are often from in thermophilic microorganisms (Benkert, Biasini and Schwede, 2011).

Verify3D is a program that compares the predicted 3D structure of a model (or experimentally derived structure) against its amino acid sequence. Verify3D uses a 3D profile of a protein to calculate 3D-1D scores for each residue. This score is dependent on the area of residues buried, fraction of side chain covered by polar atoms, and the local 2D

structure. This is used to assign a predicted fold that residue (1D) is likely involved in. This is then compared to other known proteins to establish if this residue is likely to be found in the predicted structural feature, and thus determine if the predicted model is likely to be accurate. This can be displayed as an overall S score which is the sum of all these 3D-1D scores, or as a graph of individual scores of all the residues in the protein. Looking at each residue individually makes the output less susceptible to a poor global score due to short protein length or areas of poor folding (Lüthy, Bowie and Eisenberg, 1992).

Molprobity was also used to calculate a range of values. The first of which is a clash score, which shows the number of steric clashes per 1,000 atoms, defined by a >0.4 Å overlap in each atom's Van der Waal's sphere. In a good quality structure, there should be no clashes as these are thermodynamically unfavourable and do not occur in nature.

Molprobity also calculated the percentage of Ramachandran favoured and unfavoured outlier angles. This parameter shows whether the predicted angles in the model fall in line with the allowed angles in native structures. This parameter has the target of Ramachandran favoured and unfavoured as 98% and 0.05% of residues, respectively. The ϕ, ψ angle pairs from the Ramachandran analysis are also compared to those in known structures according to secondary structure and residue identity. The Ramachandran Z-score (RamaZ) shows the degree of "nativeness" in the predicted angle pairs, and the likelihood of these being a reliable prediction of native protein fold in the target protein. Like with the QMEAN Z-score, 0 shows a good agreement between the predicted angle pairs and those seen in known structures, where lower than -4 shows poor agreement and an unreliable model (Hooft, Sander and Vriend, 1997).

Cβ deviations were also calculated by MolProbity. This measure considers the chirality of the Cα-Cβ bond in each residue. If these deviate from the accepted angles, it can suggest a misplaced side chain or poorly modelled backbone in this region. MolProbity calculates the

percentage of Cβ deviations >0.25Å and recommends 0% Cβ deviation for a reliable model (Williams et al., 2018).

Molprobity also measures all the angles in the model. Bad angles are defined as > 4σ deviations from the expected value as defined by the Phenix geometry libraries. It is suggested that 0.1% of angles would be considered as outliers for a good model (Williams et al., 2018).

Using the above these parameters, the "best" model for each protein was selected to be carried forward for use in structural homology searches. PDBeFold allows the predicted protein structure to be compared to the PDB to find proteins with homologous structures. This is a key step in function determination, as protein structure is closely linked to function.

PDBeFold produces a range of parameters to determine the quality of each result, such as RMSD, percentage sequence identity and Q-Score ranked on a scale of 0-1, shows the quality of Cα alignment in reference to the RMSD and aligned length. A score of 1 shows a perfect match between structures, and 0 no similarity. Likewise, two statistical outputs P-score and Z-score are provided, showing the statistical significance of these matches. The higher the P score, the more statistically significant the match in relation to the RMSD, number of gaps and number of residues aligned, and a high Z-score shows that the matched protein structure is of statistical significance as determined by Gaussian distribution

2.2. Methods

2.2.1. Sequence Retrieval and Alignment

Nucleic acid and amino acid sequences for Nmar1131 (ABX13027.1), Nmar1354 (ABX13250.1) and Nmar1667 (ABX13563.1) were gathered from the Uniprot database. These were used to create multiple sequence alignments by Multiple Sequence Comparison by Log- Expectation (MUSCLE) and formatted by BoxShade (ExPASy) using the RTF_New output.

2.2.2. Homology Modelling

The amino acid FASTA sequence of each target protein was uploaded to each of these three modelling programs. In I-TASSER and SWISS-MODEL, no changes were made to the default options. In Phyre2, the intensive modelling program was selected. This produced PDB files for each model, which were visualised and manipulated in PyMOL. RMSD calculations were performed in PyMOL to the respective template using the “align, cycles = 0” command. Typically, this command rejects outliers in the alignment, so setting “cycles” to 0 here removes this outlier rejection to produce a more useful measure of RMSD.

Where multimeric assemblies were predicted, monomers produced by the modelling software were aligned with the template sequence or another multimeric protein (if the template did not match the predicted assembly type anticipated from other analyses) in PyMOL using the align command. Where no appropriate template was available based on the modelled residue positions, minor manipulations of Cu placement were performed using the drag command.

Model quality was assessed by various programs. To calculate QMEAN, 3D-1D, and all MolProbity-calculated scores, the model PDB file was uploaded to each respective website. The Structure Assessment tool from SWISS-MODEL was used to calculate QMEAN score and produce Ramachandran plots. Verify3D was used to calculate the 3D-1D % and produce the

corresponding plot of residue scores. MolProbity was used to calculate clash score, Hydrogen Flipped clash score, Ramachandran favoured and outlier residues, Cβ-deviations, and bad angles. All of these were calculated by uploading the model PDB files and running the "analyse all-atom contacts geometry" program. Where no hydrogens were included in the uploaded structure, these were added using the "add hydrogens" option, where no flips were performed, using electron cloud bond lengths. To calculate the hydrogen flipped clash score, the hydrogens were removed, and Asn/Gln/His flips option chosen when re-adding the hydrogens and the clash score recalculated as previously.

All TM-Align scores were calculated using the respective template protein for each model. Both PBD files were copied into the program input, and the score aligned by the template length was chosen.

2.2.3. Structural Homology Search

The chosen model for each target protein was uploaded and searched against the whole PDB archive by 3D pairwise alignment. This was carried out with normal precision setting and the suggested 70% lowest acceptable match for both model and template. However, no structural homology matches were found for the Nmar1354 SWISS 6HBE model using these parameters. So, the lowest acceptable match percentage was lowered from 70%:70% to 50%:50%.

2.2.4. Protein Characterisation Programs

The amino acid sequences for each target protein were uploaded to Protparam and SignalP 5.0 server, to calculate the molecular weight and signal peptide cleavage site respectively.

For SignalP (Almagro Armenteros et al., 2019), the archaea organism group and long output options were selected.

2.3. Results

Proteins of known structure from the PDB with similar sequences to Nmar1131 were found by performing a BLAST search with the Nmar1131 amino acid sequence against the PDB database. These were aligned to see the conserved residues in Nmar1131. Most of these PDB top hits were NiR and MCO proteins and were arranged as such in the alignment outputs. Among the residues conserved were those for a T1Cu site and a TNC, which are highlighted in Figure 6.

Further BLAST searches were also carried out which were not limited to the PDB database, but for all sequence data available. These searches showed that extremely similar proteins are present in a range of known AOA, such as *Nitrososphaera viennensis* and *Nitrosopumilus adriaticus* (Figure 7) with 90.31% and 88.95% sequence identity for 100% and 98% sequence coverage, respectively.

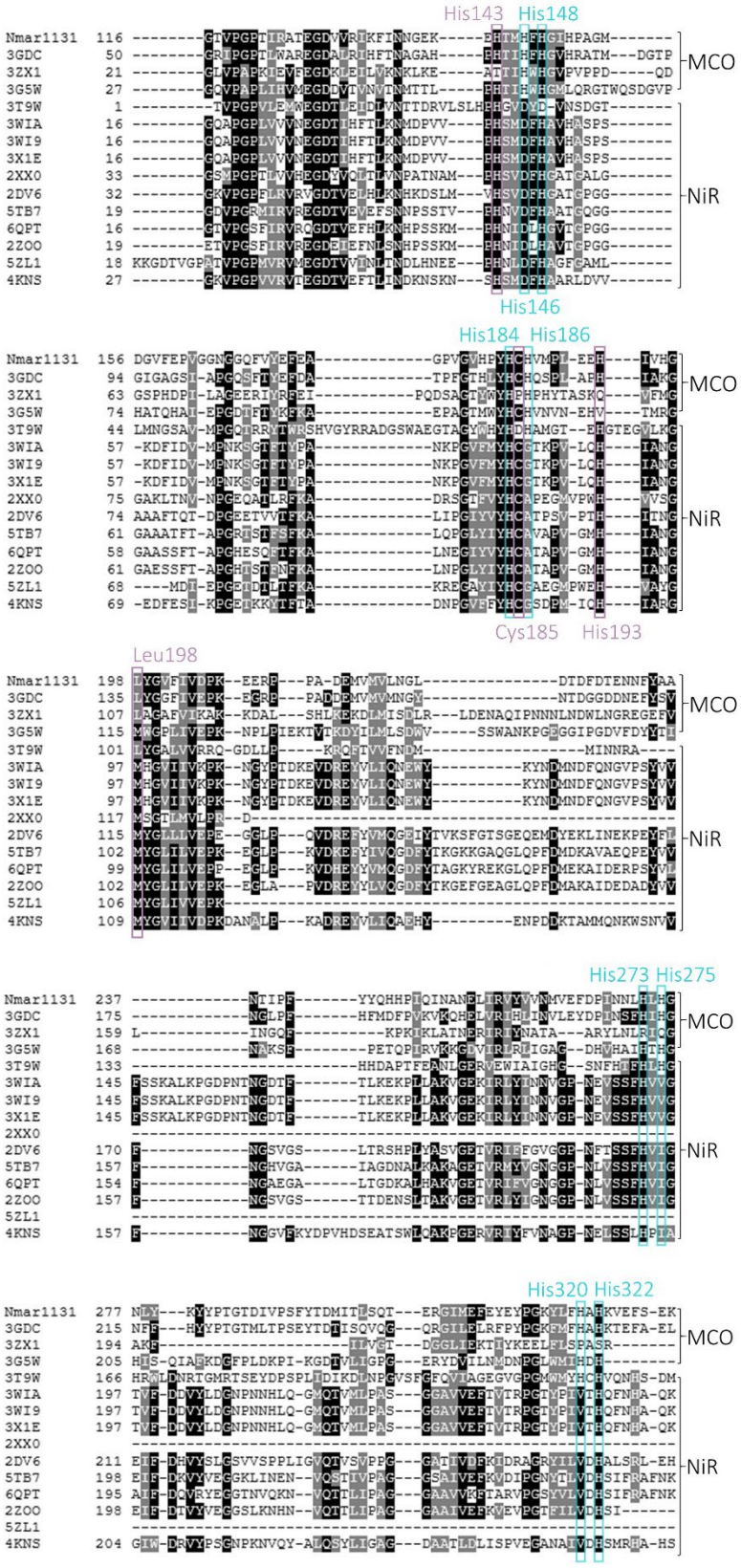


Figure 6 Sequence Alignment of Nmar1131 and Top PDB Hits. Blue boxes show the conserved residues in the TNC (His146, His148, His184, His186, His 273, His275, His320, His322), and violet boxes show T1Cu site (Cys185, Leu198, His193, His143).

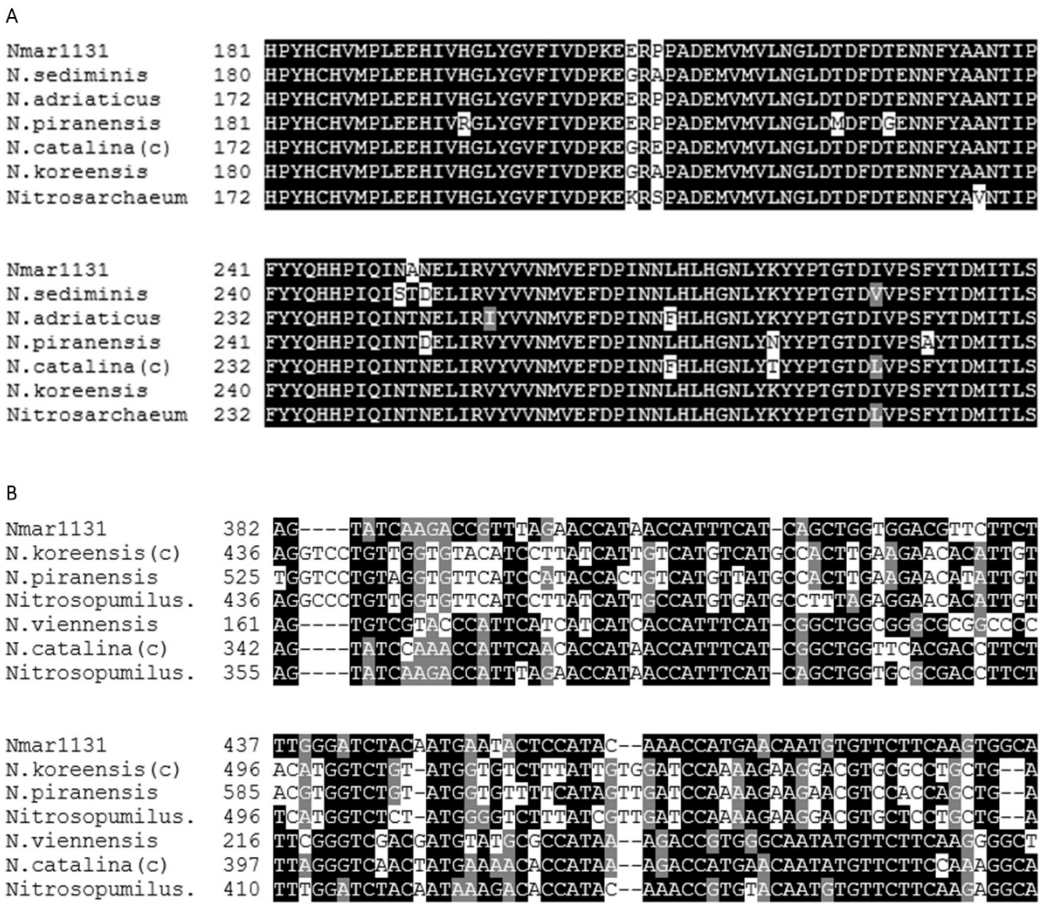


Figure 7 Alignment of top BLAST hits for Nmar1131. A) Amino acid sequence, B) nucleic acid sequences.

2.3.1. Nmar1131 Homology Modelling

I-TASSER and Phyre2 only produce monomers and do not predict the multimeric assembly of the protein, where SWISS-MODEL can produce multimers dependant on the multimeric assembly of the template protein. From the features of these models and the template proteins, along with other top PDB hits from the BLAST search, it was predicted that Nmar1131 is also homotrimeric, and was modelled and displayed as such using PyMOL. SWISS-MODEL is also capable of predicting metal binding sites and co-ordinating residues. As these predictions did not align with those made though other analysis, models were manipulated as above, and the original SWISS-MODEL outputs were shown separately (Figure 8D1&D2).

For each of the second Nmar1131 models, the templates suggested by the Phyre and SWISS-MODEL programs were monomeric three domain laccases (3SQR and 5LWW). However, the TNC residues in the resulting models still resembled those from the trimers made from the 3GDC template, thus the images were produced as above, with adjustments made to monomer and Cu placement based around completing the TNC sites.

All models produced predicted a structure reminiscent of 2dMCOs. Where the TNC is placed between adjacent subunits, in close proximity to the T1Cu site. Phyre2 (3GDC) model predicted a long N-terminal α-helical extension, which extends down from the structure. However, this region in cleaved in the other models.

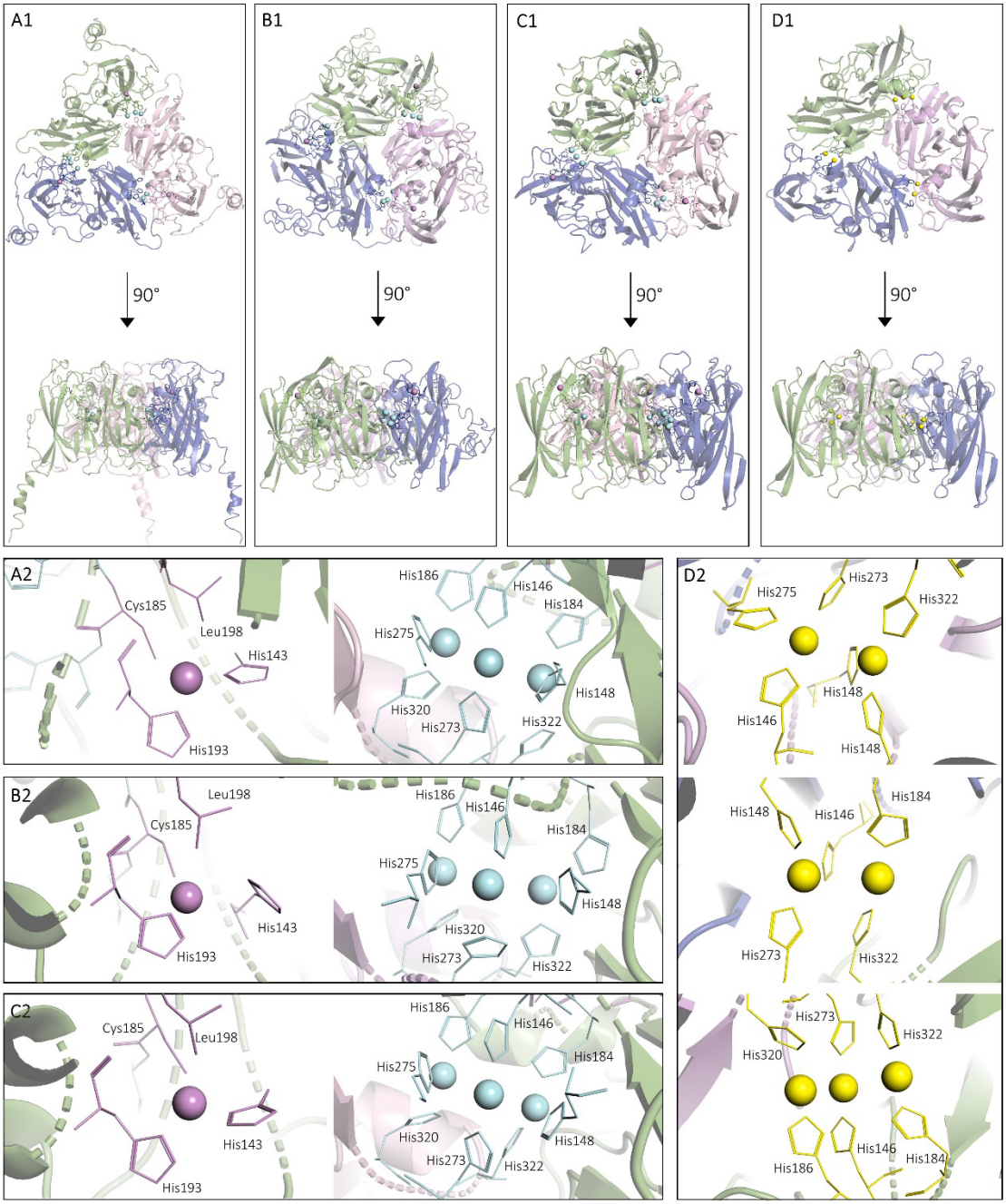


Figure 8 A comparison of the overall structure and the T1Cu (violet) and TNC (blue) copper sites in Nmar1131 depicted by different modelling software using 3GDC as a template. All predicted copper ions (except those in shown in D) were positioned by aligning the predicted and 3GDC structure, showing the predicted analogous position. A) Phyre, B) I-TASSER, C) Swiss Model D) Swiss Model, showing the predicted copper centres and residues by this program without remodelling around 3GDC copper ligands.

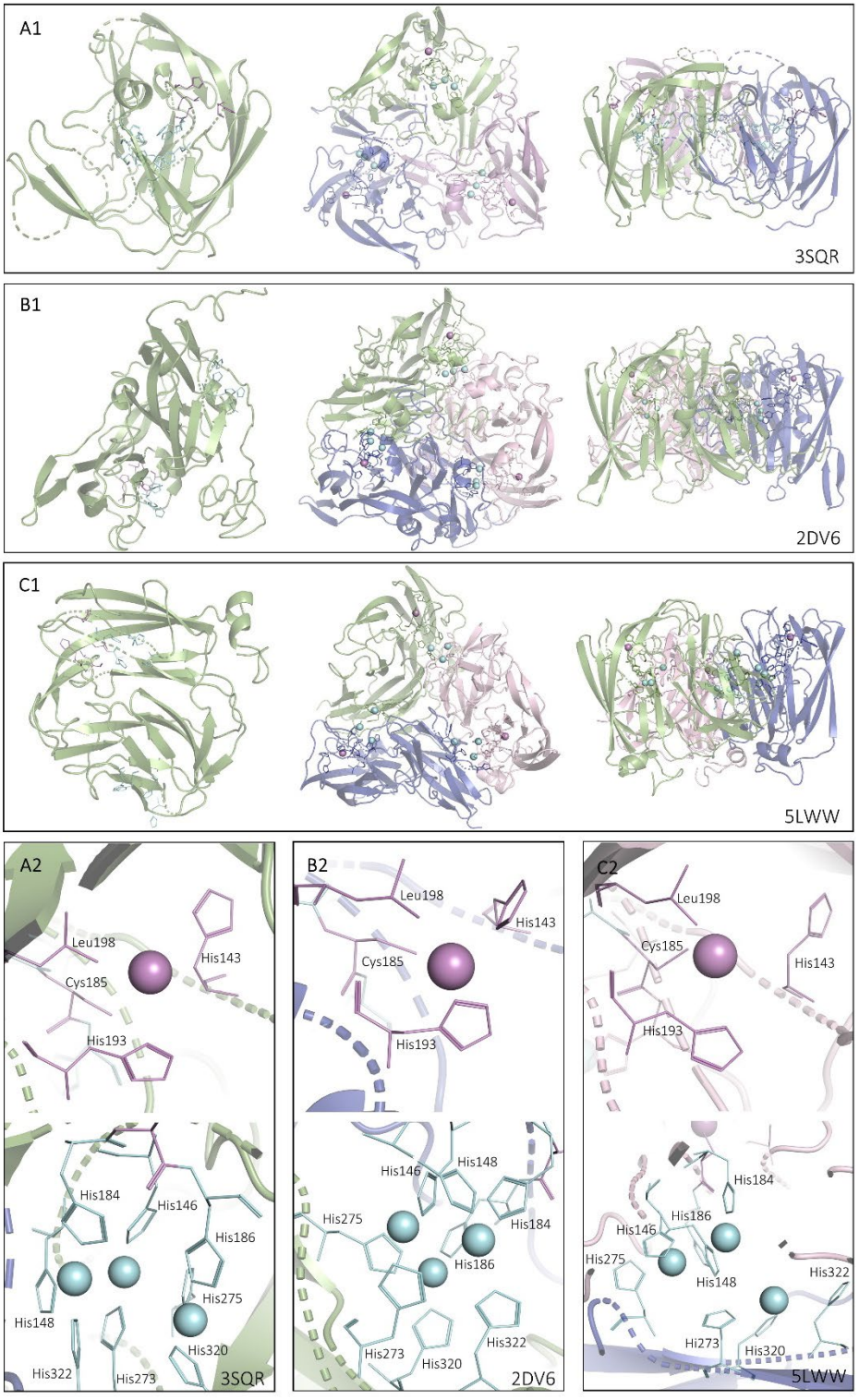


Figure 9 Second Models Nmar1131 comparison of the overall structure and the T1Cu (violet) and TNC (blue) copper sites in Nmar1131 depicted by different modelling software using alternative templates. Copper ions were positioned from the predicted placement of the co-ordinating ligands and in reference to another 2dMCO (3GDC), showing the predicted analogous position. A) Phyre, B) I-TASSER, C) Swiss Model.

2.3.2. Nmar1131 Model Vetting

Model quality was assessed using a range of parameters (Table 1). I-TASSER uses multiple templates to create the models, for ease of comparison with the other methods it has been referred to by the top model/s identified by this program, as clear architecture similarities can be seen between the single templates and the model. For this reason, RMSD and TM-Align score are likely to be poorer than for the other methods, however for consistency these are still included.

Verify3D determines that a model have at least 80% of residues that score higher than 0.2 in 3D-1D score to pass as a reliable model. The Nmar1131 3GDC models produced by I-TASSER and SWISS- MODEL (3GDC) passed this test with 84.98% and 90.48% respectively. Where Phyre2 (3GDC) scored 73.50%, failing this test. Likewise, the second models for Phyre2 (3SQR) and SWISS-MODEL (5LWW) also failed according to the 3D-1D score, with 33.33% and 68.73% respectively. The second I-TASSER model marginally passed this test with 80.34% (Table 1).

QMEAN was also used to assess the reliability of these models, with a minimum target score of -4. All models for Nmar1131 failed according to this parameter, baring the SWISS-MODEL 3GDC model with a score of -1.69. The Phyre2 3GDC model was close to the suggested cut off scoring -4.2 (Table 10).All models except the I-TASSER (2DV6) model performed well according to this parameter, all with scores above 0.94. The 2DV6 model scored 0.75 which is still indicative of a good quality model. As this is above 0.5, which shows that the model and the template share the same fold.

In regards to RMSD, SWISS-MODEL performed the worst out of the three 3GDC models with a RMSD of 1.9Å, and I-TASSER with the lowest RMSD of 0.64Å. The second set of models for each program also performed poorly with RMSD in the range of 14-20 Å.

Table 1 Quality parameters for Nmar1131 Models. Minimum target values for accepted models: 3D-1D(80%), Qmean (-4). 3D-1D (Verify3D), QMEAN (SWISS-MODEL), TM-Align, (Zhang Lab), RMSD (PyMOL), Clash Score-Bad Angles (MolProbity). Hydrogens (H) flipped by MolProbity. A Cβ deviation is defined as a difference of more than 0.25Å from the allowed conformation. Bad angles defined as more than 4σ deviations from the expected value. As I-TASSER uses multiple templates per model, the top template hit (shown in brackets) was used to calculate TM-Align score and RMSD for consistency.

Modelling		**Quality Parameter**										
Tool	**Template**	**3D-1D >0.2 %**	**QMEAN**	**TM-Align**	**RMSD**	**Clash Score**	**Clash Score (H Flipped)**	**Ramachandran Outliers /%**	**Ramachandran Favoured /%**	**RamaZ-score**	**Cβ deviations/ %**	**Bad Angles /%**
Phyre2	3GDC	73.50	-4.2	0.95530	1.789	66.18	65.43	2.87	91.98	-2.34±0.39	0.63	1.86
I-TASSER	(3GDC)	84.98	-9.91	0.98174	0.644	131.02	130.19	7.09	83.15	-3.57±0.30	9.06	2.48
SWISS	3GDC	90.48	-1.69	0.96297	1.933	0.68	0.68	3.24	92.81	-2.08±0.47	2.35	1.00
Phyre2	3SQR	33.33	-6.57	0.97000	14.686	125.87	124.84	7.94	85.71	-3.65±0.43	9.83	5.98
I-TASSER	(2DV6)	80.34	-11.09	0.75184	19.244	39.31	325.58	5.33	87.78	-3.34±0.27	8.97	1.99
SWISS	5LWW	68.73	-5.44	0.94550	15.723	21.51	15.49	5.49	86.45	-2.25 ± 0.49	2.80	1.75
Target		>80	>-4	1	0	0	0	<0.05	>98	>-4	0	<0.1

2.3.3. Nmar1131 Structural Homology Search

Nmar1131 SWISS-MODEL (3GDC) model was uploaded to PDBeFOLD. The top hit was 3GDC with a Q score of 0.92 and 100% secondary structure match, much like the results from the sequence homology search. This is to be expected as this was the structural template used for this model. The other top results albeit similarly dominated by MCO and NiR, introduce some new homologous proteins not seen in the BLAST search.

One protein of interest repeatedly appearing in homology searches is CueO, a copper efflux protein from *E.coli.* CueO has an unusual fifth copper site where Cu(I) is oxidised to Cu(II) (Figure 11D). To investigate whether Nmar1131 possessed a homologous site an amino acid sequence alignment was performed (Figure 10). From this alignment no conserved residues were seen baring leucine (198) residue. From this, CueO and Nmar1131 (SWISS 3GDC) model were aligned to see the predicted arrangement of these residues. However, from the alignment, it was clear that the CueO fifth copper site residues did not form a plausible copper binding site in Nmar1131 (Figure11 A, B &C).

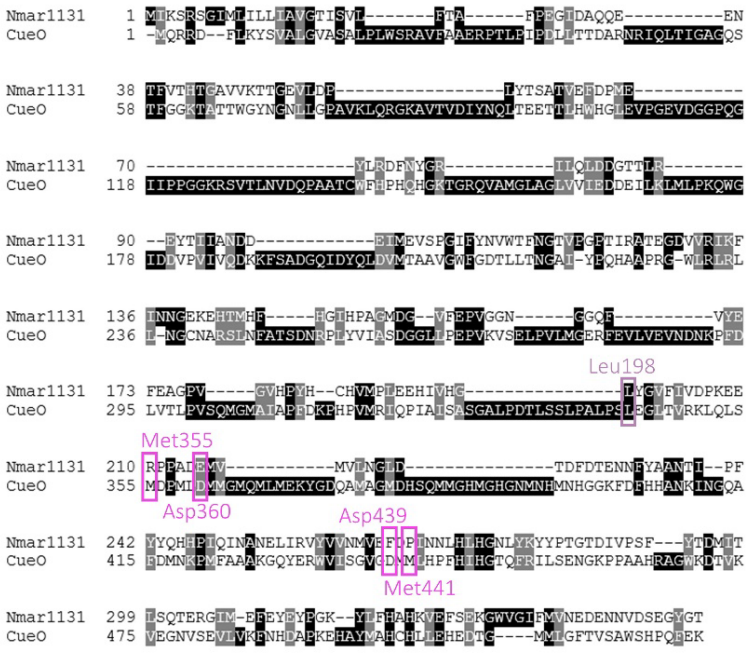


Figure 10 Nmar1131 and CueO alignment. Fifth copper site (magenta). Axial T1Cu leucine (violet).

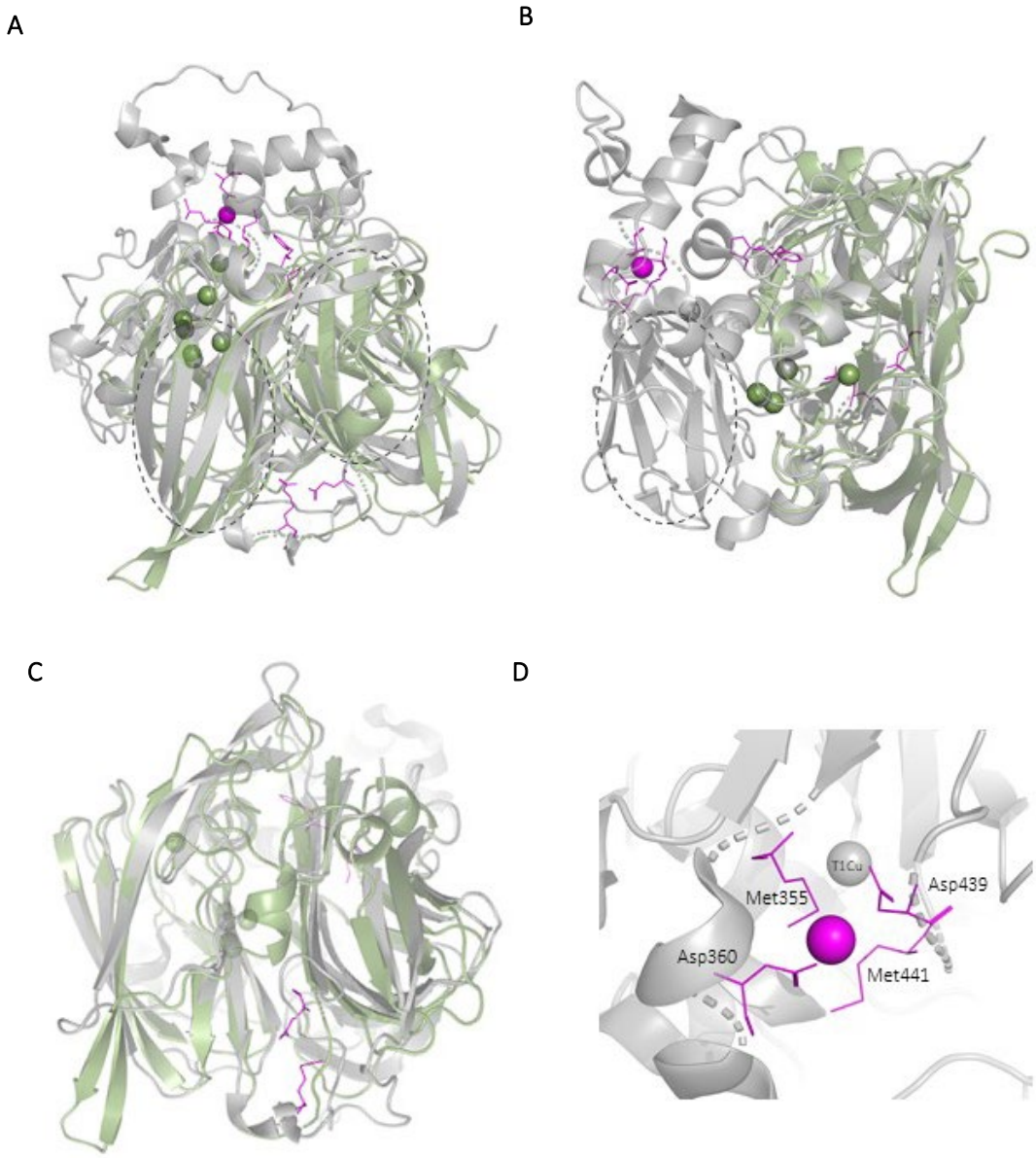


Figure 11 Swiss Nmar1131 (Swiss 3GDC) (green) aligned with 3OD3 (grey) metal regulation laccase CueO from E. coli. Cupredoxin domains circled (A&B). Fifth copper binding site residues (magenta) shown in CueO (3OD3) and the corresponding residues in Nmar1131 (Arg210, Glu215, Phe266, Pro268). Homologous fifth copper side residues in Nmar1131 (C) are compared to the arrangement of the fifth copper site residues in CueO (D).

An interesting result from the PDBefold is 3TA4 , a small (two domain) laccase from the *Amycolatopsis* genus which has a structural similarity of 94% (3TA4 target in Nmar1131 query)/71% Nmar1131 query in 3TA4 target), with 20% sequence similarity. Likewise, the amino acid sequences for 3TA4 and Nmar1131 were aligned. Where the TNC residues were well aligned, the T1Cu sites were in varying positions in Nmar1131 and 3TA4 (Figure 12). The structure of 3TA4 was aligned with the chosen Nmar1131 (SWISS 3GDC) (Figure 13). This visualised the high level of structural similarity but showed that the relative positions of the Cu sites are inverted relative to the rest of the structure. In the Nmar1131 model, the T1Cu site is placed towards the outside of the structure, but in 3TA4 it is towards the core of the trimer.

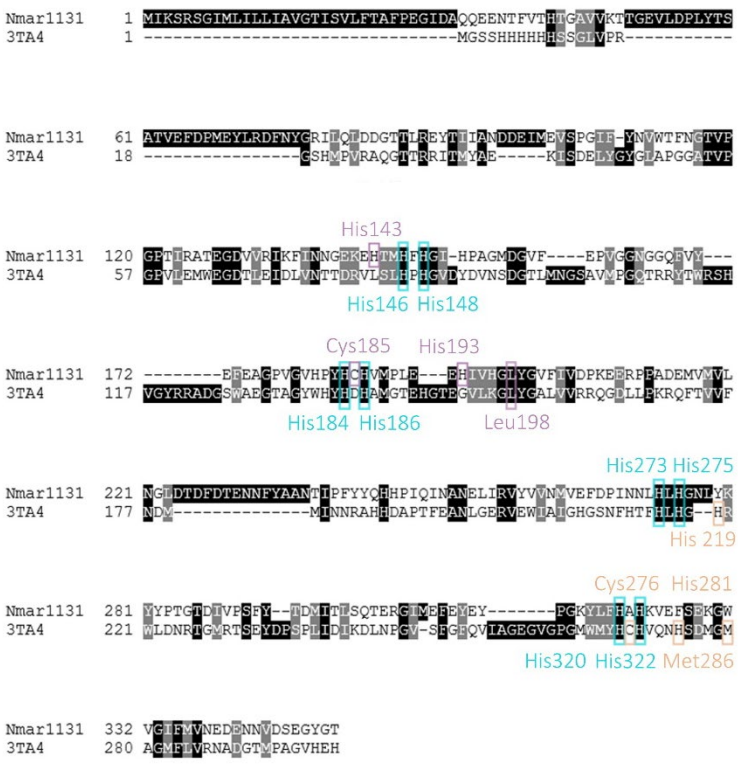


Figure 12 Nmar1131 and Laccase 3TA4 alignment showing TNC (blue), Nmar1131 T1Cu site (violet), 3TA4 T1Cu site (orange)

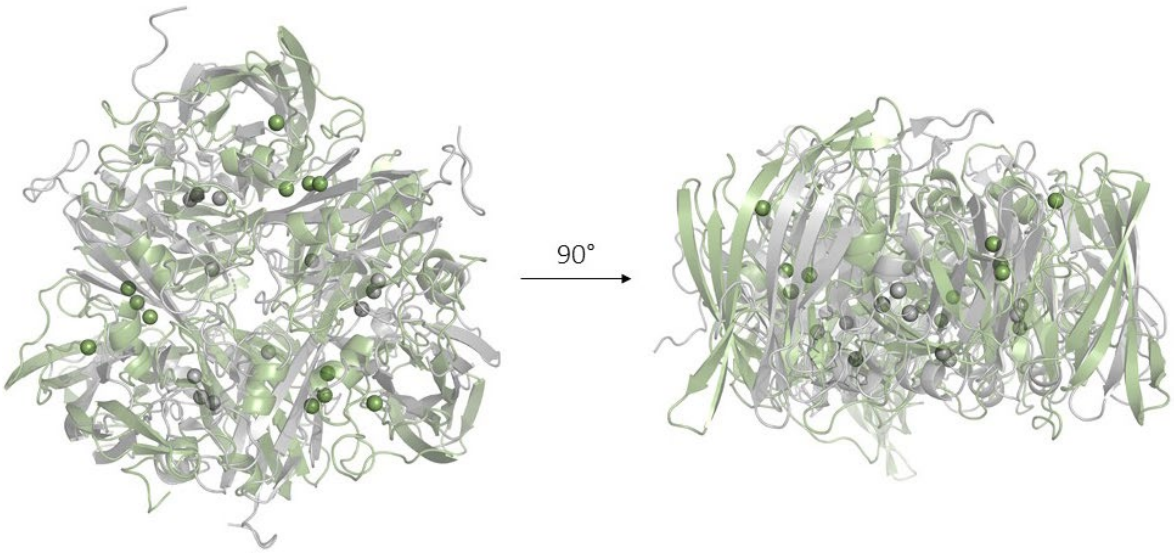


Figure 13 Swiss Model of Nmar1131 (3GDC)(green) as a template aligned with 3TA4 (grey) a laccase from Amycolatopsis sp.

2.3.4. Nmar1354

Nmar1354 is a 444 amino acids long, Protparam was also used to calculate the molecular weight of Nmar1354 at 49.64kDa, and a theoretical pI of 4.55 (Gasteiger et al., 2005). Furthermore, SignalP also predicted a Sec signal peptide from residue 1-35 (possibility 0.3381).

Again, a BLAST search for Nmar1354 was performed within the PDB to find homologues of known structure. This search presented a range of MCO and NiR much like with Nmar1131. These were aligned and the copper co-ordinating residues identified (Figure 14A) T1Cu site (His154, His105, Cys146, Met149) and TNC (His108, His110, His145, His147, His236, His238, His283, His285) were seen, suggesting that Nmar1354 is an MCO.

Interestingly, one of the top PDB hits was 2DV6, a NiR with a third cupredoxin domain containing an additional T1Cu site (Yamaguchi et al., 2004). From a further literature search, two representative three domain NiR were chosen. *Hyphomicrobium denitrificans (*HdNiR*,* 2DV6) and *Thermus scotoductus* NiR (TsNiR, 6HBE), with an additional N-terminal and a C-

terminal cupredoxin domain, respectively. These were aligned with Nmar1354, and a potential additional C-terminal T1Cu site was found (6HBE was not one of the top PDB targets but was added into the alignment in Figure 14) to show the placement of the T1 and T2Cu sites.) This aligned site in Nmar1354 was missing a histidine residue typically seen in T1Cu. This His390 in 6HBE is present in a surface loop which is not seen in Nmar1354, however a neighbouring His397 in Nmar1354 was predicted to fulfil this role (Figure 15), and was a key region of interest in the predicted models.

The traditional T1Cu site in 6HBE is found in Domain I, and the additional T1Cu site in Domain III. These make up the periphery of the trimer, where Domain II is found in the core. The additional T1Cu site is co-ordinated by His390, His431, Cys428 and Met434, these four ligands are typical of T1Cu sites. The traditional T1Cu site in contrast is co-ordinated by His75, His125, Cys115, and an unusual Gln130. The T2Cu site is co-ordinated by three histidine residues, His80, His114 (Domain I) and a trans-acting His267 (Domain II) (Opperman et al., 2019). All the T2Cu and classic T1Cu site residues are seen in the alignment (Figure 14) in homologous positions to Nmar1354.

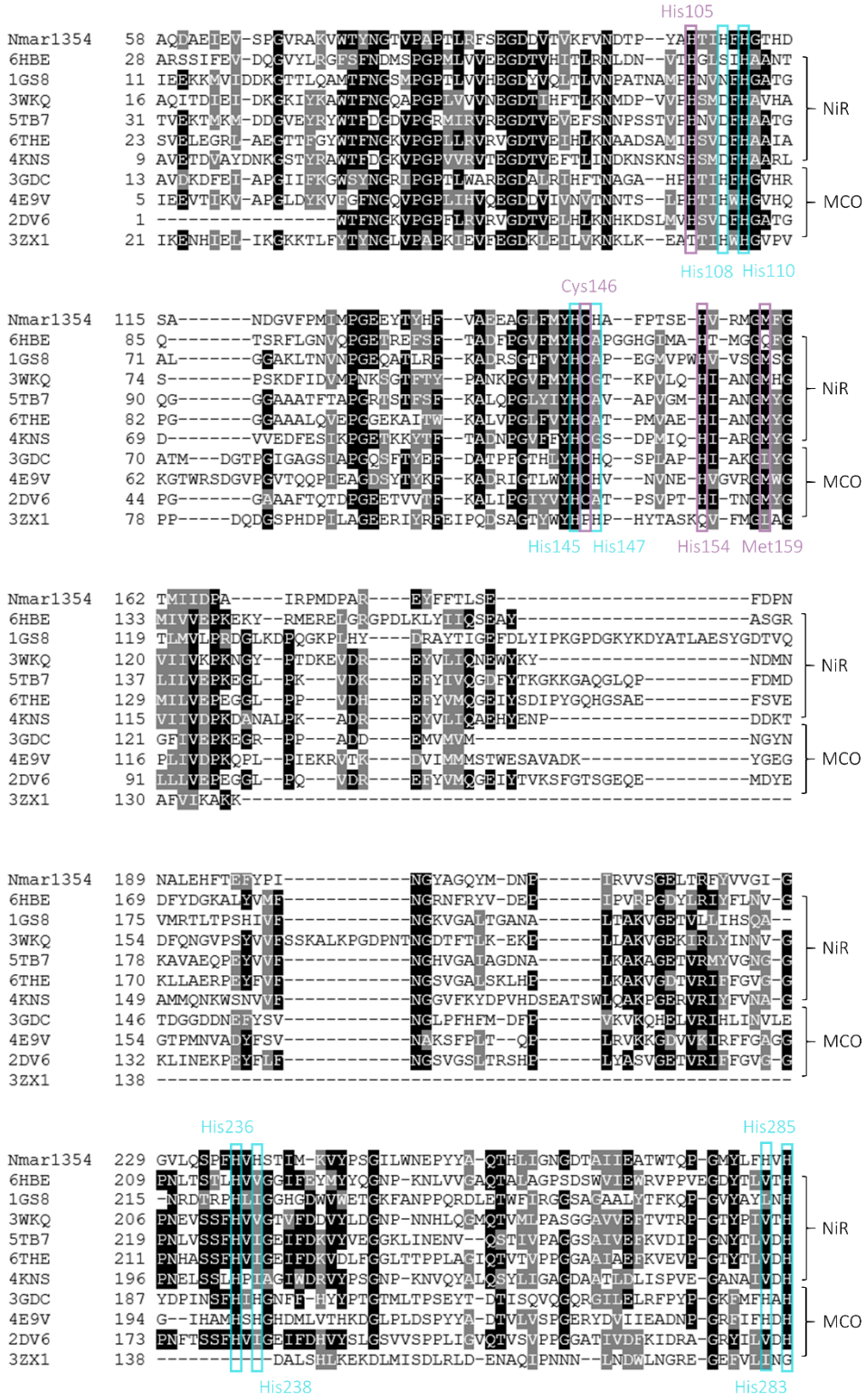


Figure 14 Nmar1354 alignment with top PDB hits. Blue boxes show the conserved residues in the TNC (His108, His110, His145, His147, His236, His238, His283, His285), and violet boxes show T1Cu site (His154, His105, Cys146, Met149).

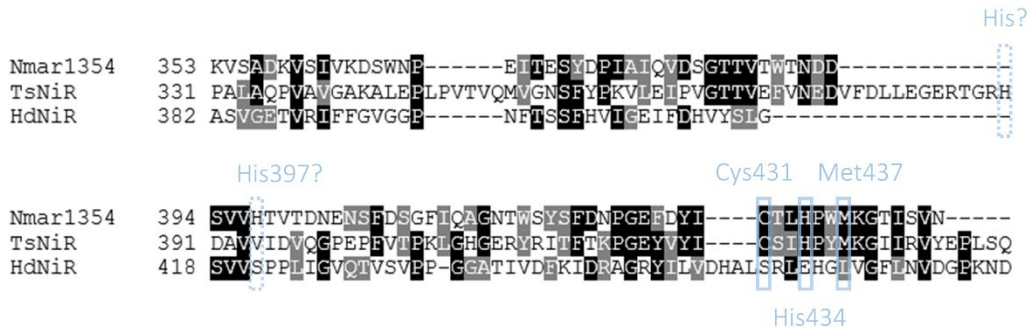


Figure 15 Nmar1354 alignment with TsNiR (6HBE) and HdNiR (2DV6). Light blue boxes show the additional T1Cu site in TsNiR and Nmar1354 (His234, Cys231, Met237). The fourth ligand (His390 in TsNiR) is not seen in this alignment, the TsNiR and possible analogous ligand in Nmar1354 are highlighted (dashed box).

Phyre2 used 6HBE as the template for the first model. As 6HBE is homotrimeric, and the model's TNC residue positioning suggested a multimeric assembly. The monomers were aligned to the 6HBE trimer, which produced plausible TNC sites. However, there is a large clash between His283 and His236, where these are seen to completely overlap (Figure 17A). As these are part of the same monomer this is likely due to poor modelling of the TNC from a NiR with only a T2Cu in the homologous position.

I-TASSER uses multiple templates to produce a model, both models shown here used 6HBE, 3DIV, 2XU9, 3SQR and 5JRR as templates.

Again, the TNC residues were split across two Domain I and II, at distal positions in the monomer, suggesting intersubunit co-ordination. However, the residues in Domain I were occluded by Domain III, which would impede the formation of a complete TNC. When aligning the first I-TASSER model to 6HBE, the TNC half sites were positioned within roughly 10 Å of each other (Figure 16B). Although this positioning was not convincing enough for a prediction of viable TNC cluster. As 6HBE is a NiR not an MCO, only a mononuclear T2Cu site is seen in the homologous position. Likewise, the other templates are monomeric laccases, so the poor modelling of the TNC is not to be taken as evidence against the presence of this site in Nmar1354.

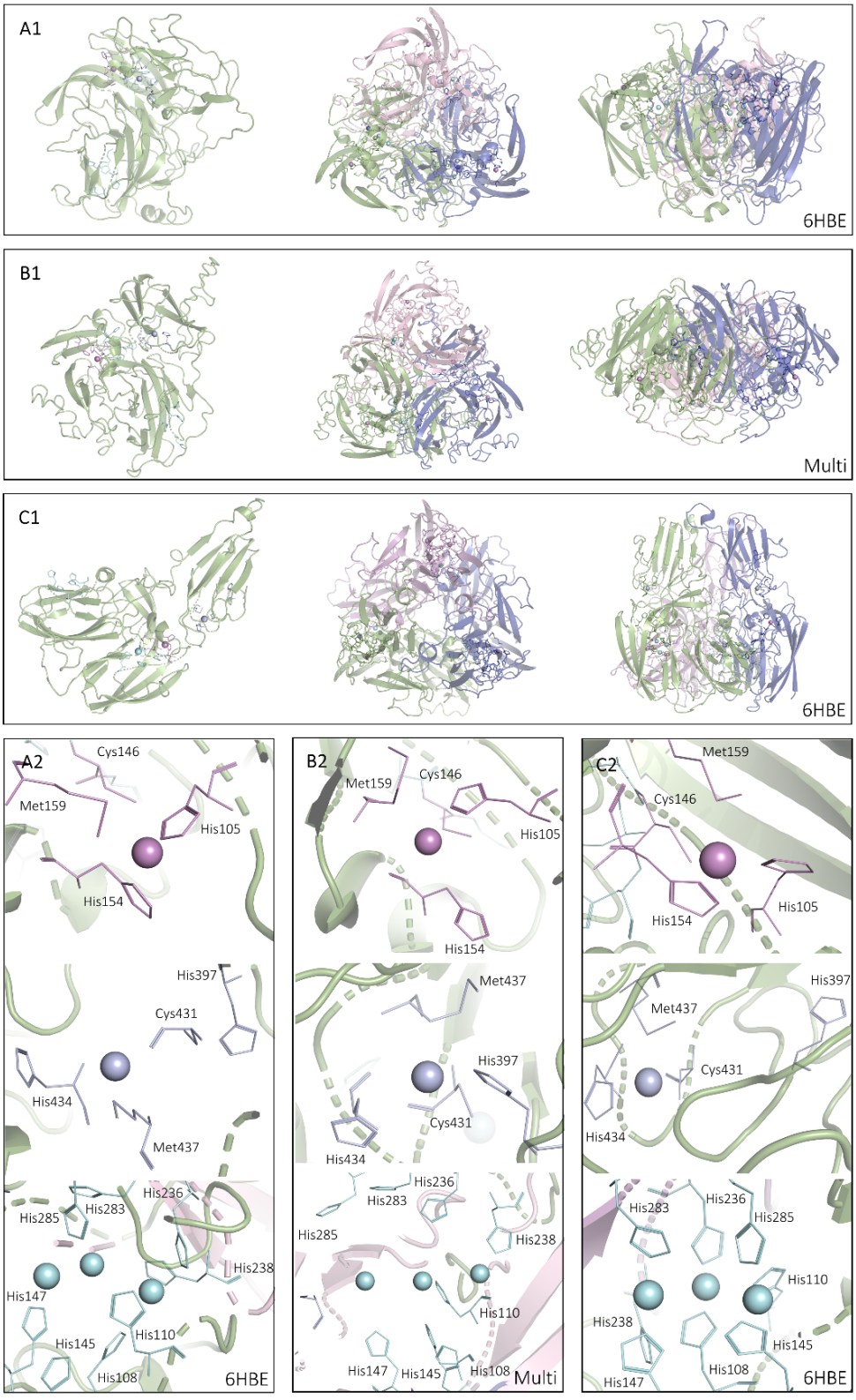


Figure 16 First models of Nmar1354. Comparison of the overall structure and the T1Cu (violet), TNC (blue) copper and additional T1Cu (lilac) sites in Nmar1354 depicted by different modelling software using alternative templates. Copper ions were positioned from the predicted placement of the co-ordinating ligands and in reference to each template protein, with manual adjustments where needed showing the predicted analogous position. A) Phyre, B) I-TASSER, C) Swiss Model. The template used in each model is displayed in the bottom right of each panel.

.

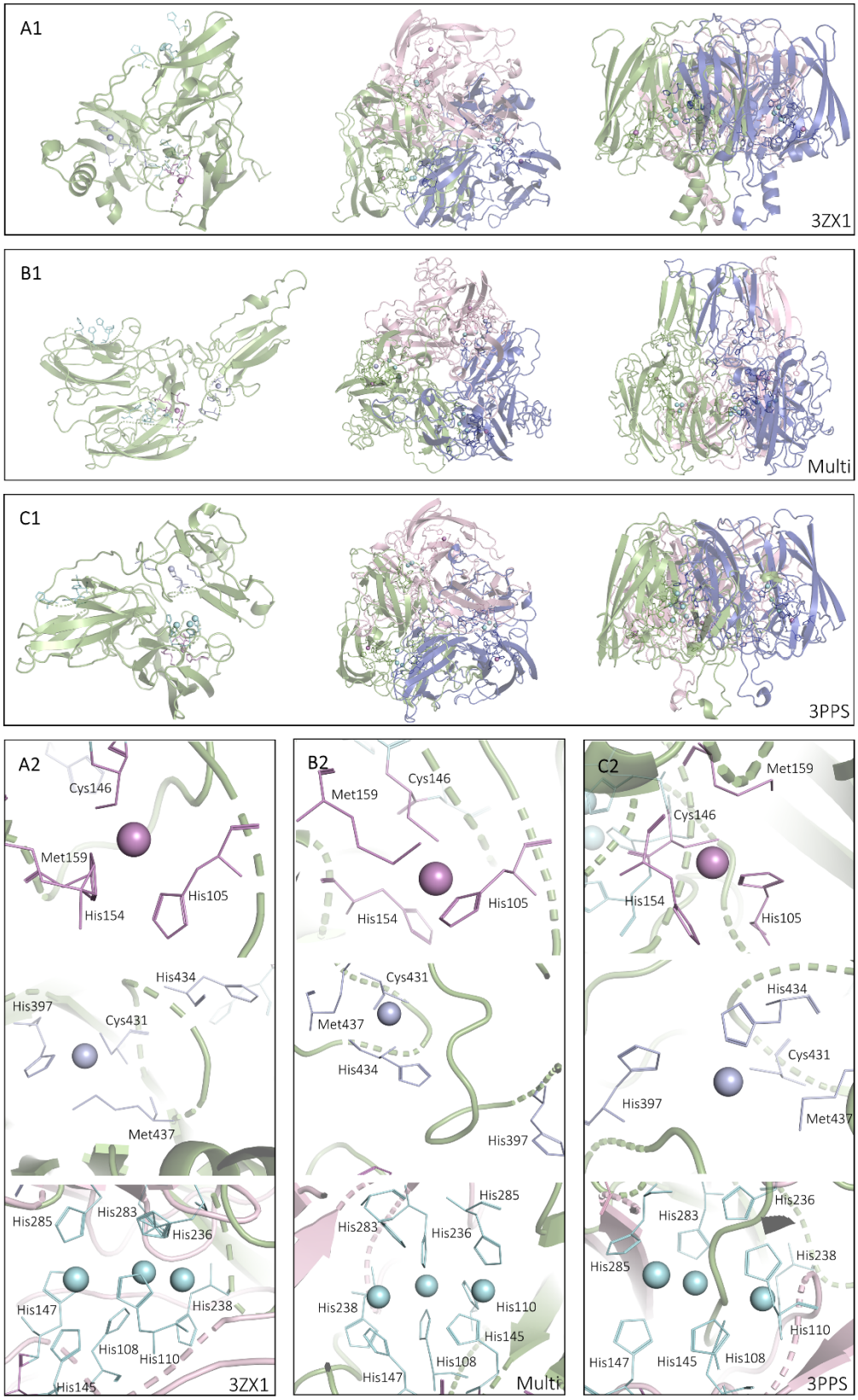


Figure 17 Second models of Nmar1354. Comparison of the overall structure and the T1Cu (violet), TNC (blue) copper and additional T1Cu (lilac) sites in Nmar1354 depicted by different modelling software using alternative templates. Copper ions were positioned from the predicted placement of the co-ordinating ligands and in reference to each template protein, with manual adjustments where needed showing the predicted analogous position. A) Phyre, B) I-TASSER, C) Swiss Model. The template used in each model is displayed in the bottom right of each panel.

Table 2 Quality parameters for Nmar1354 models. Minimum target values for accepted models: 3D-1D(80%), Qmean (-4). 3D-1D (Verify3D), QMEAN (SWISS-MODEL), TM-Align, (Zhang Lab), RMSD (PyMOL), Clash Score-Bad Angles (MolProbity). Hydrogens (H) flipped by MolProbity. A Cβ deviation is defined as a difference of more than 0.25Å from the allowed conformation. Bad angles defined as more than 4σ deviations from the expected value. As I-TASSER uses multiple templates per model, the top template hit (shown in brackets) was used to calculate TM-Align score and RMSD for consistency

Modelling		**Quality Parameter**										
Tool	**Template**	**3D-1D >0.2 %**	**QMEAN**	**TM-Align**	**RMSD**	**Clash Score**	**Clash Score (H Flipped)**	**Ramachandran Outliers /%**	**Ramachandran Favoured /%**	**RamaZ-score**	**Cβ deviations /%**	**Bad Angles /%**
Phyre2	6HBE	59.23	-7.42	0.56342	17.662	123.47	122	4.75	86.65	-2.51 ± 0.37	3.86	2.82
I-TASSER	(6HBE)	63.74	-11.23	0.55090	17.626	16.02	10.59	13.57	67.87	-6.57 ± 0.30	12.80	3.07
SWISS	6HBE	76.46	-2.53	0.89262	6.385	13.91	13.58	2.80	90.33	-1.61 ± 0.41	2.18	1.43
Phyre2	3ZX1	52.90	-8.06	0.85621	19.856	162.42	161.45	5.85	87.80	-2.84 ± 0.38	8.81	5.89
I-TASSER	(6HBE)	70.05	-13.43	0.88102	7.059	37.78	25.43	16.52	61.76	-6.68 ± 0.31	15.70	4.81
SWISS	3PPS	63.48	-5.37	0.68191	18.317	6.25	6.25	3.94	86.70	-3.11 ± 0.36	2.89	1.03
Target		>80%	>-4	1	0	0	0	<0.05	>98	>-4	0	<0.1

The additional T1Cu site is modelled fairly well in the first I-TASSER model (Figure 16)

B) and is much more convincing than the second I-TASSER model (Figure 17B). Likely because a homologous site is seen in 6HBE (minus His397).

SWISS-MODEL predicted a monomeric structure based on both the 6HBE and 3PPS templates. However, this leaves the TNC site incomplete, with the predicted Copper co-ordinating residues positioned in two different regions of the protein. However, these residues are not found on the periphery of this monomer, but rather are occluded by the third domain, much like in the first I-TASSER model. Decreasing the likelihood that this site could be completed by forming a multimer. However, dimeric, and trimeric arrangements still were attempted by aligning the model with 3PPS and 6HBE, respectively.

From the SWISS (6HBE) model in the trimeric assembly , the TNC residues are predicted to form a surprisingly convincing albeit sterically hindered TNC site (Figure 16C2), especially given that a T2Cu site, not a TNC is seen in the homologous 6HBE position. The modelled site alone would produce too many steric clashes to be viable in nature, however, visualises a potential site with more refined modelling. With the 3PPS (SWISS) model no clear completion of this site was seen when assembled into multimers, although the trimeric form was promising in terms of orienting the eight TNC residues. However, this site seems unlikely in this model due to the predicted fold of the surrounding residues. Again, 3PPS is a monomeric laccase, thus a TNC is unlikely to be modelled correctly using this as a template.

In the 6HBE SWISS model, a plausible additional T1Cu site can be seen, again with His397 laying slightly out of the expected orientation. However, the presence of this additional T1Cu site is supported by the SWISS 3PPS model. Although slightly further from the copper than is expected in a generic T1Cu site, all the residues, and most interestingly His397 are in an appropriate orientation to co-ordinate a copper at this site.

With well-formed copper binding sites, despite the unexpected His139 residue (Figure 16C) and with the lowest percentage of Cβ deviations (2.18%), Ramachandran outliers (2.80%), and the highest QMEAN score (-2.53), 3D-1D allowed residues (76.46%) and Ramachandran Z-score (-1.61) (Table 2) SWISS 6HBE model was chosen as the representative Nmar1354 model for further experiments.

2.3.5. Nmar1667

Again, a BLAST search was performed with Nmar1667 against the PDB. Unlike Nmar1131 and 1354 the top results were all NiR. These amino acid sequences were aligned and a clear T1Cu site (His134, Cys173, His184, Met189) and T2Cu site (His139, His172, His372) were seen (Figure 18) All predicted models were 2dNiR-like proteins, with exposed T2Cu site residues suggesting multimerisation in the native state. When aligned to their cognate templates, the T2Cu sites were completed well, in all but the Phyre models from 4AX3 (Figure19A) and a lesser extent from 2ZOO (Figure 20A), where His139 is further from the expected position of a typical co-ordinating residue, although both His139 residues are present in surface loops, where residue orientation is more difficult to predict. From these models it is plausible that the T2Cu site is completed with minimal adjustment of His139 positioning.

Furthermore, SWISS-MODEL predicted two homotrimeric NiR based on the 1NIA and 3XIE templates. Copper placement of both models was reasonable, given the predicted positions of the co-ordinating residues. Thus, no adjustments were made in for these models.

The T1Cu and T2Cu sites were consistently seen in all Nmar1667 models, increasing confidence of the presence of these sites in Nmar1667.

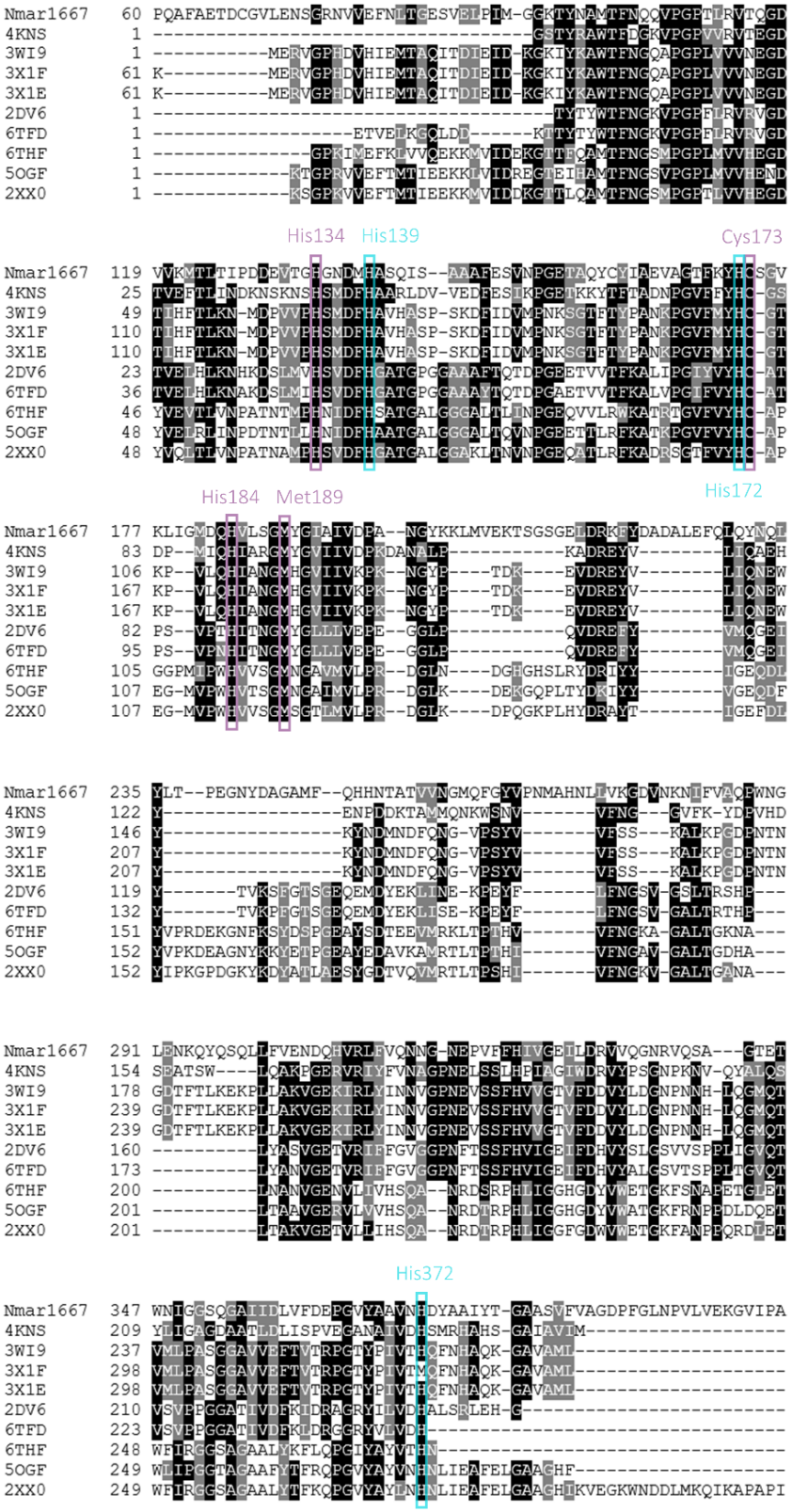


Figure 1819 Nmar1667 Top PDB BLAST hit alignment. T2Cu site residues are shown in blue boxes (His139, His172 and His372), and T1Cu sites in violet (His134, Cys173, His184 and Met189).

Both I-TASSER models (templates 3ZIY, 1MZZ, 2ZOO, 1KBV, 5AKR, 1J9Q) predicted long N-terminus extensions primarily forming α-helices that wrap around the other subunits, potentially forming stabilising interactions (Figure 19B & Figure20B). This is also seen to a lesser extent in Phyre models (Figure 19A & Figure 20A) and SWISS 1NIA model (Figure 19C). Alternatively, in SWISS 3X1E model, this extension is cleaved (Figure 20C). These models represent two possibilities for this region. This is further investigated using a signal peptide prediction program Signal P, which predicted a cleaved Sec signal peptide from residue 1-27 probability of 0.3644). This further supports the structure modelled in the SWISS 3X1E model.

The T1Cu site of the Nmar1667 models is consistently placed towards the periphery of the trimer, as seen in most NiR, allowing for electron transfer from large proteinaceous electron donors. Unlike those commonly seen in laccases.

SWISS 1NIA is the highest ranked model by QMEAN score, followed closely by SWISS 3X1E. Both also with the lowest MolProbity clash scores of 3.70 and 7.81 respectively. Likewise with the lowest percentage of residues with unacceptable Cβ deviations (1.83% and 2.73%) and bad angles (1.24 and 1.04%) (Table 3). From this point SWISS 1NIA was used as the representative model. Although Phyre 2ZOO had the best Ramachandran Z-score, the high clash score of 104.29 and Cβ deviations of 14.02% negates the small difference in this score compared to SWISS 1NIA model. However, albeit the model with the lowest score, these results are still far from ideal when compared to the target scores (shown as a minimum for an excepted model).

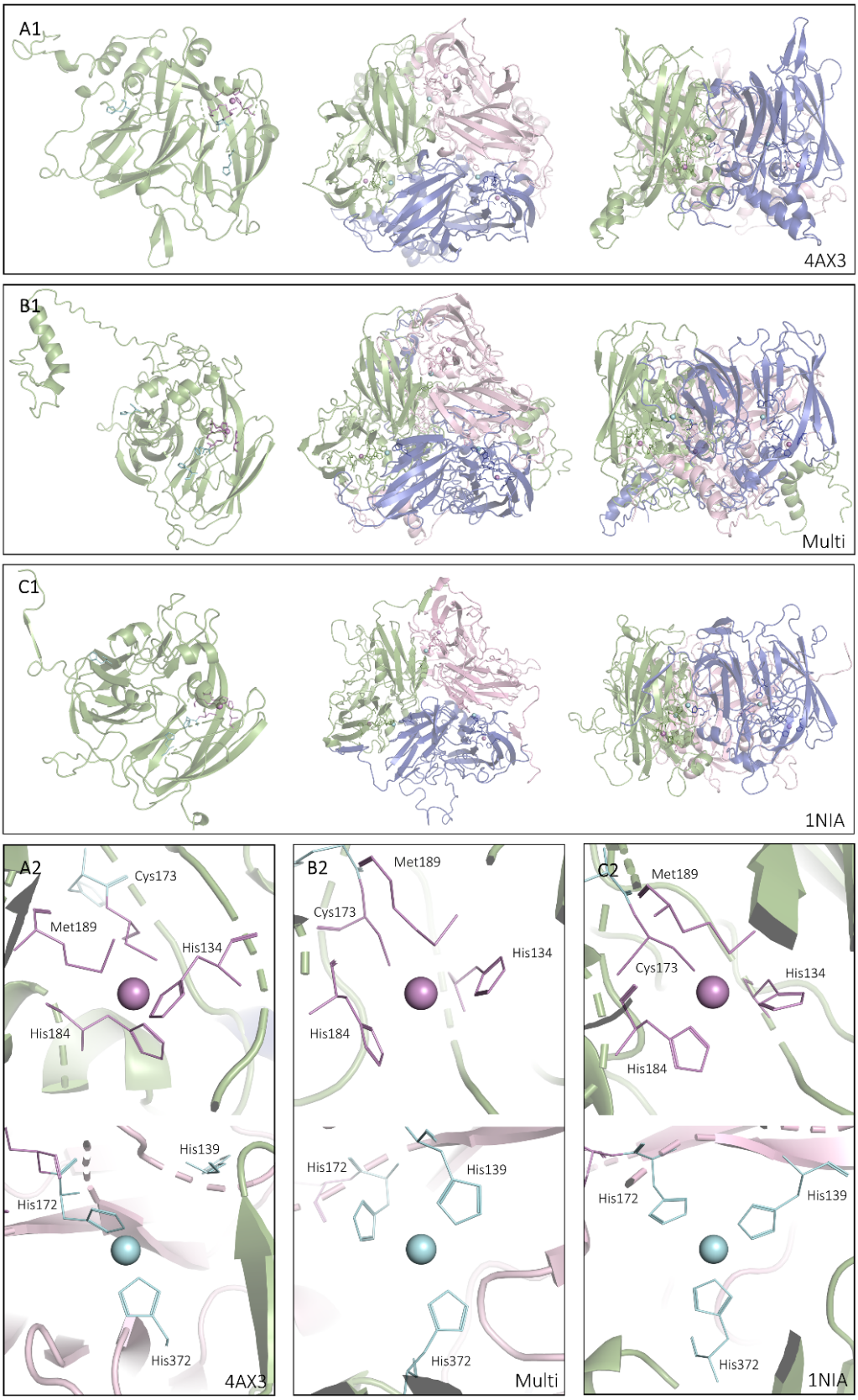


Figure 1920 Nmar1667 First models Comparison of the overall structure and the T1Cu (violet) and T2Cu (blue) copper sites in Nmar1354 depicted by different modelling software using alternative templates. Copper ions were positioned from the predicted placement of the co-ordinating ligands and in reference to each template protein, with manual adjustments where needed showing the predicted analogous position. A) Phyre, B) I-TASSER, C) Swiss Model. The template used in each model is displayed in the bottom right of each panel.

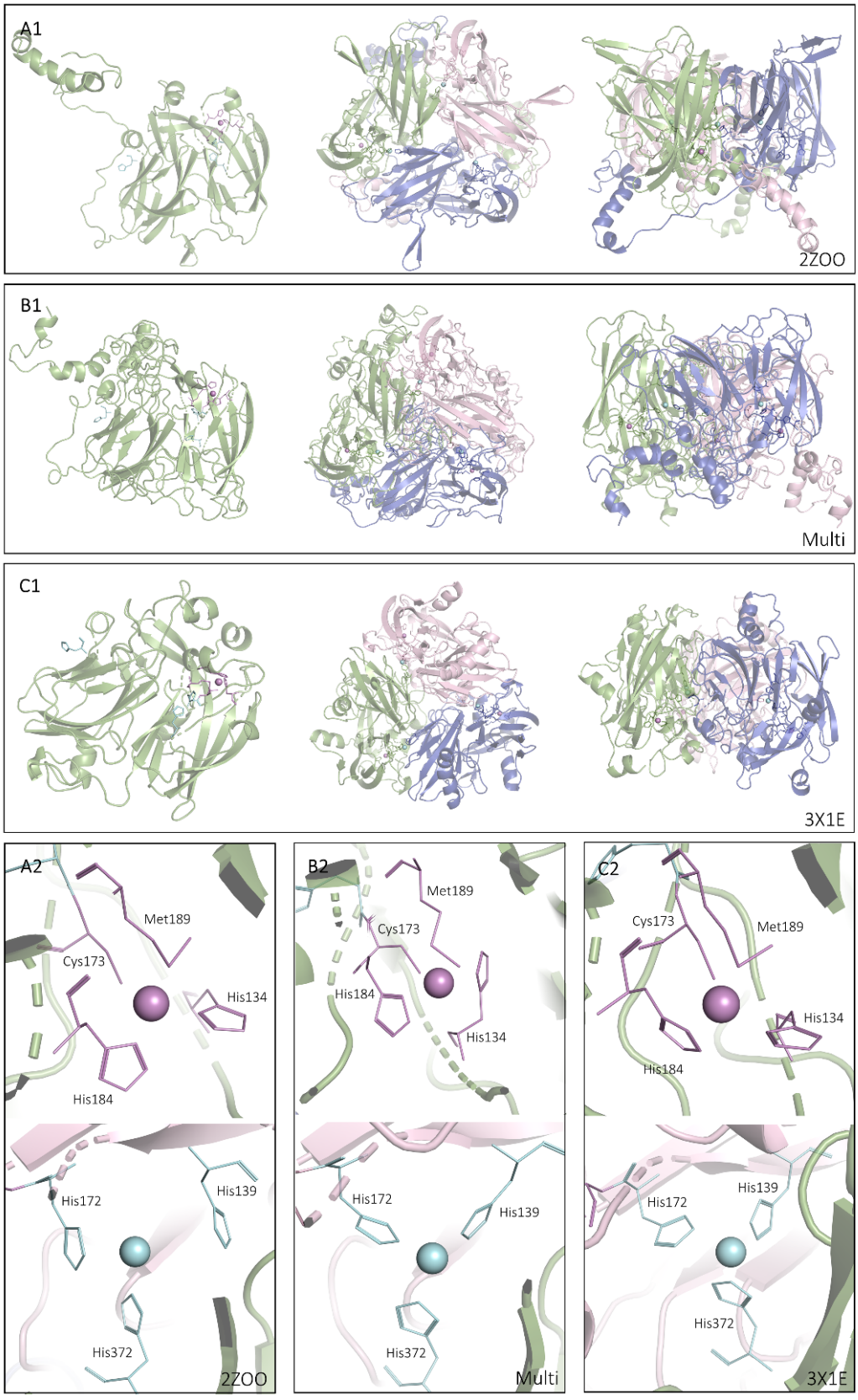


Figure 20 Nmar1667 second models. First models Comparison of the overall structure and the T1Cu (violet) and T2Cu (blue) copper sites in Nmar1354 depicted by different modelling software using alternative templates. Copper ions were positioned from the predicted placement of the co-ordinating ligands and in reference to each template protein, with manual adjustments where needed showing the predicted analogous position. A) Phyre, B) I-TASSER, C) Swiss Model. The template used in each model is displayed in the bottom right of each panel.

Table 3 Quality parameters for Nmar1667 models. Minimum target values for accepted models: 3D-1D(80%), Qmean (-4). 3D-1D (Verify3D), QMEAN (SWISS-MODEL), TM-Align, (Zhang Lab), RMSD (PyMOL), Clash Score-Bad Angles (MolProbity). Hydrogens (H) flipped by MolProbity. A Cβ deviation is defined as a difference of more than 0.25Å from the allowed conformation. Bad angles defined as more than 4σ deviations from the expected value. As I-TASSER uses multiple templates per model, the top template hit (shown in brackets) was used to calculate TM-Align score and RMSD for consistency.

Modelling		**Quality Parameter**										
Tool	**Template**	**3D-1D >0.2 %**	**QMEAN**	**TM-Align**	**RMSD**	**Clash Score**	**Clash Score (H Flipped)**	**Ramachandran Outliers /%**	**Ramachandran Favoured /%**	**RamaZ-score**	**Cβ deviations /%**	**Bad Angles /%**
Phyre2	4AX3	42.22	-6.15	0.78557	9.994	124.35	123.29	3.18	89.39	-2.43 ± 0.39	10.62	5.38
I-TASSER	(3ZIY)	49.57	-15.33	0.74637	17.082	59.80	45.99	16.24	63.46	-6.54 ± 0.29	18.82	5.36
SWISS	1NIA	53.73	-4.51	0.92217	11.230	3.70	3.70	2.84	87.45	-2.76 ± 0.22	1.83	1.24
Phyre2	2ZOO	48.77	-6.30	0.77989	2.858	104.84	104.29	3.58	89.81	-1.43 ± 0.43	14.02	5.40
I-TASSER	(3ZIY)	52.55	-18.30	0.82058	13.315	58.95	45.14	16.88	61.32	-6.70 ± 0.28	17.41	5.20
SWISS	3X1E	71.27	-4.53	0.95068	5.717	7.81	7.81	3.25	87.77	-2.71 ± 0.23	2.73	1.04
Target		>80	>-4	1	0	0	0	<0.05	>98	>-4	0	<0.1

Table 4 PDBeFold results showing the representative model for each protein of interest. Percentage of secondary sequence match (%SSE) shown for the model in the structural search match protein and visa versa.

Protein (Model)	Match	Q-score	P-score	Z-Score	RMSD/ Å	Number of gaps	% Sequence identity	%SSE (model in match)	%SSE (match in model)
Nmar1131 (SWISS 3GDC)	4KNU	0.59	24.2	14.9	1.45	12	28	83	91
	3WIA	0.58	25.5	15.2	1.58	9	26	88	88
	3WI9	0.55	23.6	15.1	1.65	10	25	92	85
Nmar1354 (SWISS 6HBE)	4KNS	0.49	15.0	15.0	1.39	9	29	57	77
	3WIA	0.49	19.6	15.8	1.26	7	29	57	67
	3WI9	0.44	19.3	15.8	1.30	7	29	57	65
Nmar1667 (SWISS 1NIA)	2BWI	0.68	32.2	17.5	0.89	9	32	78	72
	1ZV2	0.66	27.2	15.8	0.94	10	27	74	77
	1KBW	0.62	28.8	16.3	1.25	11	23	83	79

2.4. Discussion

2.4.1. Model Vetting

It is important to consider the usefulness of each quality check method to properly determine which model could be considered the most reliable. All quality parameters have advantages and disadvantages. For example, a major problem with RMSD is that all the differences in distance between atoms in the model and the template are weighted equally in the overall score. As mentioned previously, where the core of two homologous proteins often have little structural deviation, this is not the case for surface loops. Likewise, the comparison of two different length proteins can often skew the results to give a higher RMSD. Overall, RMSD does not account for the length of the homologous sequence covered by the alignment. Which could be interpreted as the proteins being less similar in structure than they are. This problem is overcome with the use of Template Modelling (TM)-Score as a comparative measure, as this attempts to normalise for the different lengths of the model and template proteins, and weighs close residue pairs much more heavily than those more distant pairs, which leaves this measure as more useful for looking at the global similarity of two proteins (Zhang and Skolnick, 2004; Zhang and Skolnick, 2005).

Furthermore, RMSD is based on how well the model aligns to the template. Although proteins of a similar function often have similar structures, this does not mean that a lower RMSD score is indicative of a higher quality model, but can still be used in consideration how well a program has modelled an unknown protein to a template. Likewise, the TM-Align score is a measure of how the model deviates from the template and is not necessarily the most useful in determining the true structure of the target protein.

Using other, more template-independent measures of quality, such as Verify3D, QMEAN z-score and Ramachandran analysis can be considered as a more accurate or at least useful method, as

these scores do not wholly depend on the templates. These figures are a measure of the likelihood of each residue being present in the predicted fold. As such, these programs are driven more by the properties of the residues rather than attempting to fit an unknown protein into a single template, making this a more reliable measure of model quality. This is shown in Nmar1131, how all models scored similarly well (except for 2DV6) on the TM-Align score but had very different QMEAN z-scores.

Overall, this shows the importance of using multiple quality parameters to test the reliability of a model. By using a range of these tests, and importantly parameters that evaluate both template dependant and independent features, more robust models can be produced.

Furthermore, although the primary sequence of a protein can be used to predict the 3D structure of a protein, two proteins of a similar 3D structure can have very different primary sequences. A further way to test the reliability of this model, and search for possible function is by performing a structural homology search using PDBeFold. This superimposes the predicted 3D structure onto those in the PDB. More potential homologues can be found than by searching in the sequence alone. As the function of a protein is often dependent on its 3D structure, this is a very useful technique to begin to understand the function of an unknown protein, as proteins of unelated sequence can be structurally homologous.

2.4.2. NiR Vs MCOs

MCOs and NiRs are both classes of blue copper proteins, with distant homology to cupredoxins; a single domain protein identified by the Greek Key β Barrel fold and T1Cu site, which produces the eponymous blue colouring(Majumdar et al., 2014). NiR are two domain proteins, in that each monomer is made up of two cupredoxin domains, which assemble into a homotrimer. MCOs can be made of two, three or six domains, and are typically organised into monomers (three and six domain) or homotrimers (two domain) Approximately 2dMCOs are 350 amino

acids long, 3dMCOs are 550 amino acids long and 6dMCOs more than 1,000 amino acids long. However, the presence of extensions beyond the cupredoxin domains can skew these numbers. (Lawton and Rosenzweig, 2011). Both MCOs and NiR have two distinct copper sites, a T1Cu site and a TNC or T2Cu site.

A key feature that defines an MCO is the presence of a TNC. At this site eight histidines co-ordinate three copper atoms at the interface between two subunits (four from domain I, four from domain II). This is absent in NiR and a mononuclear T2Cu centre is seen in the homologous position, co-ordinated by three histidines (two from domain I, and one from domain II).

MCOs are also defined by the presence of four key motifs; two HXH motifs, one HXXHXH and one H<u>C</u>HX_{3-6}<u>H</u> (Lawton and Rosenzweig, 2011). These residues make up the TNC, baring those underlined, which co-ordinate the T1Cu site.

The T1Cu site in blue copper proteins, accepts a single electron from an external donor, which is passed to the TNC/T2Cu site through a Cys-His/Cys-His-Cys bridge where the substrate is reduced. NiRs accept an electron from an external proteinaceous electron donor, such as azurin or pseudoazurin, and reduce NO_2^- at the T2Cu site. Where MCOs oxidise a broad range of substrates such as aromatic compounds and phenols, which is coupled to the four-electron reduction of molecular oxygen to H_2O at the TNC.

2.4.3. Nmar1131

Sequence alignment (Figure 6) shows the conservation of TNC residues in Nmar1131. As labelled, the first four hits are from known MCOs in the PDB, whereas the others are from NiRs. One key difference between MCOs and NiRs is the presence of the TNC, containing the eight histidine residues (His146, His148, His184, His186, His 273, His275, His320, His322) that co-ordinate the three Cu atoms at this site. The equivalent site in NiRs is a mononuclear T2Cu

centre, with three histidines (His148, His 184 and His322). This is the first piece of evidence to suggest that Nmar1131 is an MCO.

In the TNC the "T2Cu" is co-ordinated His146 and His273 as well as an external $^{-}$OH from water acting as the third co-ordinating ligand. In the TNC this copper ion is only co-ordinated by two histidines, rather than three Histidines as seen in NiRs. The other two coppers are co-ordinated by three histidines each, thus totalling the eight histidines seen in the TNC.

The key MCO motifs are also present in Nmar1131 as follows; ^{320}**H**A**H**^{322}, ^{193}**H**IV**H**G^{197}, ^{273}**H**L**H**^{275}, ^{184}**H**C**H**VMPLEE**H**^{193}. The key residues in the motif are shown in bold, those involved in the T1Cu site are also underlined, the others are part of the TNC.

Furthermore, His273 is a fourth well conserved histidine in both NiRs and MCOs. Acting as one of the two T2Cu co-ordinating Histidines in MCO and is consistently found close to the T2Cu in NiRs, albeit more distal (3.8 Å) than the co-ordinating ligands (2.0-2.2 Å) in 4KNS. This histidine is not involved in copper co-ordination in NiRs but rather acts as the proton donor in this mechanism.

The SWISS-MODEL (3GDC) model also predicted the presence of seven TNC-like Cus in the trimer (Figure 8D1&D2). Where the predicted co-ordinating residues and Cu ions lay within the expected distances typically seen in TNC sites. The unusual co-ordination spheres where re-modelled using PyMol to display more typical TNC conformation, as the entire compliment of classic TNC residues were present in a close proximity to the site and provided a more convincing copper site (Figure 8C2)

T1Cu sites are typically co-ordinated by two histidines, one cysteine and an axial methionine. However, in Nmar1131, and homologous proteins in archaea, an axial leucine (198) is typically seen in place of this methionine. This can be seen in other classes of proteins also, where a methionine, leucine, or phenylalanine acts as the axial residue. This is discussed in more detail below.

From the combined results from all of these analyses (Table 1) the Nmar1131 structure produced by SWISS-MODEL (3GDC) was deemed to be the most reliable, with the best 3D-1D and QMEAN z-scores, lowest Clash Score, highest percentage of Ramachandran favoured conformation and lowest percentage of outliers. Although this model has a higher RMSD and lower TM-score value, these measurements can be considered as less useful in determining the true structure of the protein as mentioned above. Furthermore, the values for the Ramachandran outliers and a unfavoured Cβ deviations, although are not the best out of the three models tested, are not a great deal worse, and are made up for in the superior scores from the other tests.

2.4.4. Laccases

Searching for structural homologues of the Nmar1131 model (3GDC, SWISS) using PDBeFold revealed a structural similarity of 94% with 3TA4, a small laccase from the Amycolatopsis genus. Laccases are a widespread group of MCOs that catalyse the one electron transfer from a wide range of aromatic substrates to the 4 electron reduction of O_2 to H_2O (Shraddha et al., 2011)(Martins et al., 2015). 3TA4 is the substrate-bound structure of *Amycolatopsis* laccase (AMLAC) (PDB: 3T9W).

Laccases are typically monomeric and consist of three cupredoxin domains. Domain I and III house the T1Cu and TNC sites, and Domain II typically forms a substrate binding cleft. However, there is a group of small laccases which have two domains (2dMCO) and form homotrimers, such as 3TA4.

3TA4 and *Streptomyces coelicolor* small laccase (SLAC, PDB 3CG8) are homotrimeric laccases, unusually each monomer is made up of two domains (2dMCO) and contain four copper atoms. Having a structure which more closely resembles NiRs than other large laccases (Skálová et al., 2009), makes these an interesting target for further understanding the role of Nmar1131.

When 3TA4 sequence was aligned with Nmar1131 (Figure 11), the TNC residue positions were conserved but the T1Cu site was seen elsewhere in the alignment. With the key motifs present as such ^{276}H<u>C</u>HVQN<u>H</u>^{281},^{90}**H**P**H**^{92}, ^{144}**H**D**H**^{146}, ^{219}**<u>H</u>**TF**H**L**H**^{224}.

The structural alignment of Nmar1131 and 3TA4 also showed the positions of the Cu sites were inverted in Nmar1131 as the sequence would suggest. This could be an indication that 3TA4 shows the true orientation of these sites, as due to the nature of these modelling programs, orientation of folds can often be misinterpreted. However, this is unlikely to be the case. As in 3TA4 the T1Cu site is in the C-terminus (Figure 10) which is common for laccases but this does not align with the T1Cu site in Nmar1131 where the T1Cu site is in the N-terminus (Domain I).

The position of the T1Cu site is important in considering the function of these proteins. In 3TA4 and (SLAC) both small laccases, the T1Cu site is towards the centre of the trimer, compared to NiRs where it is still buried, but much closer to the periphery of the protein. Having the T1Cu site close to surface in NiRs allows electron transfer from large electron donors, such as azurin and pseudoazurin. In SLAC the T1Cu site is close to core of the trimer where access is limited to smaller electron donors which can penetrate further the structure that large donors (Skálová et al., 2009).

It is likely that the T1Cu site in Nmar1131 is in the predicted position, homologous to NiRs and other MCOs. Which could potentially suggest that it interacts with a large electron donor, typical of electron transport chains, rather than small, molecular electron donors as laccases do. This further shows the importance of looking at structural, and not just sequence homologues when looking at the function of an unknown protein.

Furthermore, laccases can be classified due to the identity of the axial ligand in their T1Cu site. Typically, the axial residue is methionine in the T1Cu site, however, this can also be replaced with a leucine or phenylalanine, both hydrophobic residues. These residues do not co-ordinate

the T1Cu resulting in a trigonal planar structure, instead of the distorted tetrahedral geometry. The variation in the axial residues alters the redox potential of the T1Cu site(Jones and Solomon, 2015). This variation may be caused by the potential required for electron transport from the T1Cu site to the TNC or tuning to the external electron donor.

Three classes of laccases were proposed to reflect this variation in T1Cu redox potential. Class I, with a methionine in the axial position, is suggested to have the lowest redox potential, followed by Class II (Leu) and Class III (Phe) with the highest redox potential (Eggert et al., 1998). For example, an L499M mutation in 3SQR, a laccase from Botrytis aclada fungus gave a 140mV decrease in redox potential, thought to be solely due to this mutation (Osipov et al., 2014).

This means that the leucine seen in Nmar1131 gives the T1Cu site a higher redox potential than typical T1Cu sites. An axial leucine was also seen in a selection of homologous proteins from AOA (Figure 7) and CueO (Figure 10) and 3TA4 (Figure 11). Likely due to the different protein-protein interactions made in AOA compared to AOB. It is also typical for laccases to also have an additional axial Ile in the T1Cu site, this does not seem to be present in Nmar1131.

Laccases are also of interest in the biotechnology world, with the potential to be used as electron transport systems, without the need of an external co-factor. They have potential uses for bioremediation, possibly being able to mitigate the effect of using fertilisers and contamination from industry etc. (Martins *et al.*, 2015).

NO has been suggested to play a role in laccase regulation. It has been shown that the T2Cu of tree laccase can be reduced by NO. Furthermore, this reduced T1Cu site can be further re-oxidised by T2Cu-bound NO. This redox reaction is also seen in cytochrome c oxidase, although not with fungal laccases, likely due to their higher redox potentials. Again, possibly linking the function of laccases to the nitrogen metabolism in AOA, here cytochrome c-containing proteins are absent. (Torres et al., 2002) (Martin et al., 1981) However, the importance of this is still to

be determined, especially given the presence of laccase-like T1Cu sites (with an axial Leucine) in Nmar1131.

NO can reversibly inhibit cytochrome c oxidase by reducing the copper and haem sites, interrupting the electron transport chain. A similar reaction can be seen in laccases, where NO binds and reduced the T2Cu in the TNC. However, this does not inhibit the action of the laccase, as the partial reduction allows O_2 to bind and be reduced to water as the terminal electron acceptor and oxidising NO to NO_2^-. In tree laccases this reaction is also reversible, where NO can re-oxidise the reduced sites (Torres et al., 2002) (Martin et al., 1981). Perhaps playing a role in the wider regulation of TNC-containing proteins, and nitric oxide metabolism (forming HNO_2) (Torres et al., 2002).

2.4.5. Copper homeostasis

Copper homeostasis is important in many organisms, as even a low amount of free copper, especially Cu(I) is toxic. Copper efflux oxidase (CueO), is a regulatory laccase that plays a key part in copper homeostasis in E.coli. Seemingly the key function of CueO is not to oxidise aromatic compounds but to oxidise periplasmic Cu^+ to Cu^{2+} at a fifth binding site (Met355, Asp360, Asp439, Met441)(Figure 12) (Singh et al., 2004; Zhang et al., 2020).

Although this fifth copper site is not seen in Nmar1131, in CueO, an axial leucine is seen in the T1Cu site, much like in Nmar1131. Also, the methionine rich loop, which controls the electronic environment of the T1Cu site and is suggested to help co-ordinate 5th Cu (Reiss et al., 2013) is not seen. Even though the Nmar1131 gene is also colocalised with metal regulatory (DtxR) and, metal transporter (ZIP family) genes (Walker et al., 2010), the lack of CueO-like fifth binding site suggests that Nmar1131 is not directly involved in copper detoxification in N. maritimus. At least, not via the oxidation on Cu(I) such as CueO.

2.4.6. Nmar1354

A BLAST search was performed as previously and sequence alignments of Nmar1354 and the top PDB hits were aligned. This alignment interestingly showed the conservation of two T1Cu sites and a TNC.

Homology modelling of Nmar1354 interestingly produced a Class III NiR-like MCO protein. Every model contained three distinct cupredoxin domains, with the classic T1Cu and TNC sites, and an additional T1Cu site in the third domain, as seen with a small number of Class III NiR. 3dMCOs do not typically have a third copper site and those that do, such as CueO are typically monomeric, unlike their 2dMCO counterparts.

As discussed above, two representative Class III NiRs were selected. HdNiR (2DV6) has a N-terminal cupredoxin extension and TsNiR (6HBE) has a C-terminal cupredoxin extension (Figure 15) HdNiR's third domain (N-term) T1Cu does not interact with the typical T1 and T2Cu sites seen in NiRs, likewise in other in other Class III NiRs such as RpNiR and PhNiR. However, in 6HBE the additional T1Cu site in 6HBE is ∼14 Å away from the core T1Cu site, almost twice as close as the additional T1Cu site in HdNiR. the additional T1Cu site is close enough to interact with the T1Cu and T2Cu sites. A similar distance of around 14 Å was seen in the Nmar1354 models. It is predicted that electron shuttling from the additional T1Cu site, to T1Cu and T2Cu site is possible in TsNiR, unlike in HdNIR (Opperman et al., 2019). This may also be the case in Nmar1354, potentially suggesting a novel electron transfer mechanism in this protein.

Alternatively, the additional T1Cu site could have metallo-oxidase activity as seen in CueO. However, the identity of this additional site in Nmar1354 is more similar to that of a typical T1Cu site, unlike those seen in CueO.

HdNiR and TsNiR are roughly 50kDa(Yamaguchi et al., 2004)(Opperman et al., 2019), which is mirrored in the Protparam predicted weight of Nmar1354 of 49.64kDa (Gasteiger et al., 2005).

This is in contrast to the typical NiR molecular weight of 36kDa (Horrell et al., 2017). This further supports the prediction that Nmar1354 is a 3dNiR-like MCO.

3ZX1 is a 3dMCO suggested to act primarily as a cuprous oxidase, much like CueO. However, unlike CueO, 3ZX1 does not contain a known third copper site, where Cu(I) oxidation takes place in CueO, suggesting this takes place at the main T1Cu site or a currently undefined copper binding site. (Silva et al., 2012)

In the alignment, 3ZX1 TNC residues aligned with those from the other MCOs and NiRs, however, the T1Cu residues did not (Figure 14). In 3ZX1, the T1Cu site (His407, His468, Cys463, Met473) is found in Domain III, as it typical of 3dMCO, and the TNC residues at the interface of Domain I and III. Where the T1Cu site is in Domain I of Nmar1354, in a homologous position to this site in NiRs and some 2dMCOs (4E9Vand 3GDC in this alignment), where the T1Cu site is variable.

Interestingly, when the 3ZX1 Nmar1354 model and 3ZX1 were aligned the additional T1Cu site in Nmar1354 was shown to be modelled in the site of the conserved T1Cu site in 3ZX1. This gave some interesting insight into the possible structure of the Nmar1354 additional T1Cu site. Where His397 was predicted to be the second histidine ligand at the site, other models did not model this site convincingly. This is to be somewhat expected as the templates did not have a T1Cu site here, and the location in the outer loops of the protein are less reliably modelled than those in the core. However, this model provides a well- formed additional T1Cu site (Figure 17A), at the expense of the main T1Cu site which is more poorly modelled, as a homologous site is not seen in the same position in 3ZX1.

The 3ZX1 Phyre Nmar1354 model shows a well-structured additional T1Cu site, with His397 in an appropriate position to co-ordinate the copper at this site. Increasing confidence in this prediction.

2.4.7. Nmar1667

All Nmar1667 models predicted a NiR-like homotrimer. The presence of the T1Cu site and T2Cu site are homologous with known NiRs (Godden et al., 1991; Horrell et al., 2017). Likewise, the lack of additional TNC residues confirms that Nmar1667 is not an MCO.

With all models, the T1Cu site is situated near the periphery of the protein, potentially allowing for interactions with large proteinaceous electron donors much like NiR in ETC.

The N-terminal extensions seen to various degrees in all the Nmar1667 models baring SWISS 3X1E where region is cleaved. This suggests two possibilities for this region. Often signal peptides are cleaved after exportation, or alternatively these regions can be inserted into the cytoplasmic membrane. Furthermore, SignalP also predicted a cleavage site at residue 27, much like what is seen in the SWISS 3X1E model. As this was the "best" model for Nmar1354 out of those produced here, the SWISS 3X1E model supports the presence of an N-terminal leader sequence that is post-translationally cleaved.

2.4.8. Conclusions

From BLAST searches and homology modelling, it is predicted that Nmar1131 is a homotrimeric laccase-like 2dMCO. Nmar1131 possesses the key MCO defining motifs and is predicted to contain two Greek β key fold. The T1Cu site is predicted to be positioned towards the exterior of the structure, as in CuNiRs, likely allowing for electron transfer from a large proteinaceous electron donor to the T1Cu site. The TNC lays between adjacent subunits (domain 1 and domain 2 of a neighbouring subunit).

Furthermore, Nmar1131 is not involved in Cu(I) detoxification, at least not through the same mechanism as CueO. From the results gathered here, it was predicted that Nmar1131 is a laccase with strong structural homology to CueO.

Alternatively, Nmar1354 is a predicted homotrimeric 3dMCO with an additional T1Cu site in the C-terminus cupredoxin domain. Similar C-terminal extensions in 6HBE, a NiR, are able to interact with the other T1Cu and TNC sites, potentially playing a role in Nmar1354 function. Typically 3dMCO are monomeric, however each model placed TNC residues in positions which would suggest multimerisation of the native protein for completion of these sites. Nmar1354 could form a novel class of trimeric 3dMCO, homologous in structure to 3dNiRs, with the additional TNC residues.

Furthermore, Nmar1667 is predicted to be an NO-forming nitrite reductase. Unlike Nmar1131 and Nmar1354, the additional TNC residues are not present, instead a mononuclear T2Cu site is seen at the interface of domain I and II. The T1Cu site is located on the periphery of the predicted homotrimer, as is seen in CuNiR, possibly allowing for interaction with a large proteinaceous electron donor. This peripheral T1Cu site could suggest a role of Nmar1667 in a metabolic electron transport chain, however, this is a highly speculative suggestion from these results.

The production of different models using the same input sequence and template protein shows the importance of predicting unknown protein structure with multiple homology modelling programs. Likewise, it is important to consider the importance of structural homology searches performed with these models. Typically, a protein with less than 35% sequence similarity is deemed inappropriate for homology modelling, however, as seen in the PDBeFold results, many proteins with predicted homologous structures, have less than 35% sequence identity with the target sequence. Showing the importance of these structural homology searches in tandem with sequence based techniques.

Finally, through amino acid BLAST searches, it was shown that extremely similar proteins to Nmar1131 are present in a range of known AOA, such as *Nitrososphaera viennensis* and

Nitrosopumilus adriaticus (Figure 7) with 90.31% and 88.95% sequence identity for 100% and 98% sequence coverage, respectively. These top results are a mixture of soil and marine AOA. Showing the conservation of a small laccase-like MCOs across a range of AOA.

Chapter 3 Over-Expression of *N. maritimus* genes in *Escherichia coli*

3.1. Introduction to protein over-expression

In order to study the target proteins, large amounts of pure protein are needed. As archaea are slow growing organisms, that are particularly hard to maintain in a pure culture, because of this recombinant protein expression in *Escherichia coli* was used. This required the optimisation of the genes for *E.coli* expression system, and insertion into an appropriate plasmid vector. In these experiments, pET28a and pET26b were used. These vectors use the T7 expression system where specific protein overexpression can be induced by the addition of IPTG. Following overexpression, the protein can be harvested and purified using techniques such as immobilised metal affinity chromatography, size exclusion chromatography, ion exchange chromatography. The purified protein can then be used for a variety of structural and functional experiments.

3.2. Methods

3.2.1. Competent Cells Preparation: DH5α

Overnight cultures of 10ml LB were prepared from an aliquot of DH5α competent cells, 2ml was used to inoculate 100ml flask of LB and grown at 37°C, 180rpm until OD_{600} 0.6-1.0. Cultures were split equally into two 50ml falcon tubes and incubated on ice for 20 minutes, then centrifuged at 2,500*g* 4°C for 10 minutes. Each pellet was resuspended in 10ml ice-cold 0.1M $CaCl_2$, combined, and incubated on ice for 30 minutes. The combined resuspension was centrifuged at 2,500*g* for 10 minutes at 4°. The pellet was again resuspended in 4ml ice-cold 0.1M $CaCl_2$ with 15% glycerol and transferred into 100µl aliquots and stored at -80°C.

3.2.2. pET28a Plasmids

Plasmids containing *Nmar1131*, *Nmar1354* and *Nmar1667* genes were purchased from Twist Bioscience. This plasmid utilises the IPTG inducible T7 lac promoter to specifically over-express the target genes. These plasmids were designed for recombinant protein expression and thus

were codon optimised for expression in *E.coli* in order to adjust to the specific tRNA makeup of *E.coli.*

3.2.3. pET26b Plasmids

After a series of unsuccessful over-expression experiments with the pET28a plasmid. *Nmar1131*, *Nmar1354*, and *Nmar1667* were cloned into a different vector pET26b. This series of experiments is discussed in greater detail below. pET26b also uses the IPTG inducible T7 lac promoter to specifically over-express target genes, however, this plasmid also adds a *pelB* signal peptide to the N-terminus of the expressed protein which targets the unfolded protein to the periplasm instead of the cytoplasm.

3.2.4. pET28a/pET26b Transformation of XL1 Blue/BL21 DE3/DH5a/C43 Cells

Transformation of the desired plasmids into XL1 Blue, BL21 DE3, DH5a and C43 E.coli cell lines was performed by the addition of 2µl of plasmid or 10µl of PCR product to each 1.5ml microcentrifuge tube of 50µl competent cells and incubated on ice for 30 minutes. The cultures were then heat shocked at 42°C for one minute then incubated on ice for 3 minutes. 500µl of sterile LB was added to each tube and incubated for 45 minutes at 37°C, 180rpm. 100µl of sample was plated on 1.5% agar Luria broth (LB) plates with 50µg/ml Kanamycin and incubated overnight at 37°C.

3.2.5. Restriction Digest of Purchased Plasmids

Restriction digests were carried out using 3µl of minipreped DNA, 2µl H_2O, 2µl 10x Buffer R and 2µl NdeI and 1µl HindIII (Thermo Scientific™) to give a total sample volume of 10µl. Samples were incubated at 37°C for 1.5 hours and then ran on a 1% (w/v) agarose together with a 1kb DNA ladder (Invitrogen) and stained using GelRed to determine the success of the digestion.

3.2.6. Preparation of 1% Agarose Gel and DNA Samples

Agarose gels for DNA resolution were prepared by heating 0.5g of agarose with 50ml of 1xTris-Borate-EDTA buffer (TBE). The boiled mixture was poured into a taped gel casket with an 8-well comb and left to set for 15 minutes. DNA samples were prepared by mixing 20µl of PCR product, 5µl of GelRed® (Biotium) and 5µl of 6x loading dye (Thermo Scientific™) . For the ladder preparation, 5µl of GelRed® was added to 10µl of 1kb ladder (Invitrogen) and the total sample loaded. Gels were ran for 45-60 minutes at 90 V.

3.2.7. Large Scale Over-Expression of Nmar Proteins

To produce the desired protein from the specific plasmid large Erlenmeyer flasks were used. 750ml of 2xYT was prepared in 2L flasks and autoclaved. Freshly transformed colonies (BL21 DE3) were picked and used to inoculate 10ml of LB with 50µg/ml Kanamycin and incubated overnight at 37 °C, 180 rpm. Flasks were warmed for 15 minutes and 50 µg/ml Kanamycin and 1mM/0.1mM $CuSO_4$ (when used) added. 3ml of overnight culture was added to each flask and incubated at 37 °C, 220 rpm, until OD_{600} 0.6-0.8 was reached as measured on a Cary60 (Agilent) UV-vis spectrophotometer. 500 µl of a Pre-Induction sample was taken prior to Isopropyl β- d-1-thiogalactopyranoside (IPTG) (to 250µM) addition. Flasks were grown at 37 °C, 220 rpm for 5 hours, then another sample taken prior to overnight incubation at 17 °C, 220prm. A final overnight sample was taken after ~16 hours. All samples were spun at 14,100*g*, and the supernatant discarded and pellets were frozen at -20° until use. Cells were harvested by centrifuging at 4,000*g* at 4 °C for 15 minutes. The supernatant was discarded, and the pellet resuspended in 20ml of IMAC Buffer A (50mM Tris/HCl, 500mM NaCl, 20mM imidazole, pH 7.5) and stored at -80°C.

3.2.8. Test Expression

To test protein expression in an alternative growth medium, 35ml of 2xYT broth was prepared in 100ml flasks and autoclaved. Previously transformed colonies (BL21 DE3) were picked and grown in 10ml flasks overnight at 37°C, 180rpm. The large-scale expression protocol was followed as above, until the final overnight sample was taken. As this experiment was only used to determine protein expression, cells were no harvested and the cultures discarded.

3.2.9. Preparation of 15% SDS-PAGE Electrophoresis Gels

A Biorad gel system was used to prepare and run gels. The resolving gel was prepared using 2.3 ml ddH_2O, 2.5 ml 1.5 M Tris/HCl pH 8.8, 5 ml acrylamide (30 % acrylamide/Bis-acrylamide), 100 µl 10% SDS, 100 µl 10% APS, 10 µl TEMED and then left to set for 15 minutes between casting plates. The stacking gel was prepared using 2.25 ml ddH_2O, 1 ml 0.5M Tris/HCl pH 6.8, Acrylamide 666 µl, 40 µl 10% SDS, 40 µl 10% APS, 5 µl TEMED and pipetted on to the top of the resolving gel. A 8/12-well comb was inserted and left to set for 15 minutes. For both components TEMED and APS were added immediately before pouring as these catalyse the polymerisation reaction.

3.2.10. Preparation of SDS-PAGE Samples

Protein expression samples were prepared by resuspending in 50 µl of cracking buffer (0.05 mM Tris pH6.8, SDS 1 % w/v, Glycerol 25 % w/v, DTT(few crystals), 0.05 % Bromophenol Blue, followed by further dilution with 4 µl of resuspended sample in 5 µl of cracking buffer and 11 µl of H_2O. IMAC/IE samples were prepared as follows: Cell lysate (5µl lysate, 10µl H_2O, 5µl cracking buffer), flow though and elution samples (10 µl flow through/sample, 5 µl H_2O, 5 µl cracking buffer). Gel Filtration samples (15 µl sample and 5 µl cracking buffer). After dilution all samples were boiled at 90°C for 10 minutes, spun at 14,100*g* for 15 seconds, and 20 µl of the samples and 10 µl of PAGERuler™ Unstained Protein Ladder (ThermoFisher) were loaded onto an SDS-

PAGE gel and ran at 200V for 45 minutes. Gels were then stained in Coomassie Blue for 1 hour and destained in 10% acetic acid, 50% methanol for 2 hours, then transferred to water for further destaining

3.2.11. Immobilised Metal Affinity Chromatography

Harvested cells were thawed then transferred to a large beaker and stirred for 10 minutes until smooth. An EmulsiFlex-C5 cell disrupter (Avestin) was washed 3 times with water and then IMAC Buffer A. The pump frequency was adjusted until pressure around 10,000 psi and the cell paste added through a sieve to remove any lumps. This was processed three times to rupture cell membranes until a homogenous paste was produced. The resulting lysed cells were spun at 40,000*g* at 4 °C for 45 minutes. The supernatant was collected and kept at 4 °C until use.

Two 1ml HiFliQ Ni-NTA FPLC (Generon) columns were joined in tandem and equilibrated with 5 column volumes (CV) of miliQ water and 5 CV of IMAC Buffer A. The supernatant was then loaded and the flow through collected. The column was then washed with 10 CV of Buffer A and loaded onto an ÄKTA Prime and eluted on a gradient of 100 % Buffer A to 100 % IMAC Buffer B (50 mM Tris/HCl, 500 mM NaCl, 500 mM imidazole, pH 7.5, filtered) and 2ml fractions were collected. 50µl aliquots were taken from fractions under the peaks on the A_{280} spectrum and ran on SDS-PAGE to detect if the target protein was present.

3.2.12. Gel Filtration Chromatography

A G75 column (GE-healthcare) was equilibrated with Gel Filtration (GF) Buffer (50 mM Tris/HCl, 100 mM NaCl, pH7.5, filtered). Selected IMAC fractions were pooled and concentrated to 2ml using a spin column (Sigma Vivaspin 10,000 Da MWCO) and loaded onto the G75 column. 100ml of GF Buffer was used to elute the sample, and 2 ml fractions collected, these were tested as previously for the target protein.

3.2.13. QuikChange Mutagenesis

To close the NcoI restriction site in *Nmar1354* and *Nmar1667*, QuikChange mutagenesis was used to mutate a cytosine to thymine as part of the Pro196 codon (CCC -> CCT) in Nmar1354 (5'GGCCATTCGCCCtATGGACCCGGCG), and a thymine to cytosine in His154 (CAT -> CAC) in Nmar1667 (5'CAGGTGACCGGCCAcGGAAATGATATG), these primers were designed to close the NcoI site without changing the identity of the subsequent amino acid. A 50 µl reaction mixture was used, consisting of; 1 µl of 10 µM forward and reverse primer, 1 µl plasmid DNA, 5 µl Dimethyl sulfoxide (DMSO), 5 µl 10x cloned Pfu Buffer (Aglient), 1 µl dNTPs, 1 µl PfuTurbo (Aglient) and 37 µl H_2O. The PCR reaction was carried out using the following cycle: 95°C 3min, (95°C 1min, 56-61°C (gradient) 1min, 72°C 8 min) x 16 cycles, 72°C 15minutes. An annealing temperature of 58 °C and 56 °C was successful for *Nmar1354* and *Nmar1667*, respectively. Remaining template DNA was digested with 1 µl of DnpI, and 3.5 µl of Buffer Tango (ThermoFisher) per 30µl of PCR product and incubated for 2 hours at 37°C. The PCR product was then transformed into DH5a cells and minipreped as above. Minipreped DNA was then retransformed into XL1 Blue cells to produce larger gene stocks.

3.2.14. Cloning into pET26b

Target genes were cloned into new vectors by adding the restriction sites NcoI (C'CATGG) and HindIII (A'AGCTT)) to the C-terminus and N-terminus of the of the target proteins, respectively. A reaction volume of 50µl was used, consisting of; 1µl plasmid DNA, 1 µl of 10 µM forward and reverse primer, 1µl dNTPs, 5µl 10x Pfu Buffer+ $MgSO_4$, 5µl DMSO, 0.5µl Pfu polymerase and 35.5µl ddH_2O. The amplification reaction was carried out using a thermocycler using the following program; 95 °C 5 min, (95 °C 1 min, 62 °C 1 min, 72 °C 1 min) x 35 cycles, 72 °C 8 min. An annealing temperature of 62°C was used for *Nmar1131*. 5 µl of each PCR product was ran on a 1% agarose gel to confirm PCR success. This PCR reaction was repeated to produce a larger volume of product.

The primers used for each target gene were as follows: *Nmar1131* (5' ATTA CCATGG GC TTG ATC GCT GTA GGC (Forward), 5' TAAT AAGCTT TCA CGT ACC GTA ACC TTC CGA (Reverse)), *Nmar1354* (5' ATTA CCATGG GC ATCGGTTTCTCTATCGCT (Forward), 5' TAAT AAGCTT TCATCCACGGGTGCAGGG (Reverse)), *Nmar1667* (5' ATTA CCATGG GC ACGCTGTTCGGGTC (Forward), 5' TAAT AAGCTT TTAACTGGTCAAAATTTCGGCG (Reverse)).

A restriction digest using HindIII and NcoI was performed on the PCR products and miniprepped pET26b plasmids. 40 µl PCR/miniprep product was mixed with 8µl Tango Buffer, 8µl HindIII, 4µl NcoI (Thermo Scientific™), 60 µl H_2O and incubated at 37 °C for 1 hour. The total product was run on a 1 % agarose gel and extracted by cutting the bands out under UV light and cleaned using a Nucleospin™ Gel and PCR Clean Up Kit (FisherScientific) to manufacturer's instructions.

With the target genes and the pET26b plasmids double digested, the sticky ends were ligated. Two ratios of reactants were used, 1:1 and 1:0.5 of plasmid to fragment carried out in parallel. 5 µl of plasmid, 5/2.5 µl fragment, 1.5 µl 10x T4 ligase buffer, 1 µl T4 ligase, 2.5/5 µl H_2O were mixed and incubated at room temperature for 1 hour. 7.5 µl of this product was transformed into DH5a cells, the remaining sample was left to incubate overnight to be transformed into DH5a the next day, if the initial transformation failed. The transformed colonies were miniprepped and ran on a 1 % indicative agarose gel (5 µl ligation miniprep, 10 µl GelRed 3 µl 6x loading buffer, 100V for 45 minutes) to check the success of the ligation before being sent for sequencing.

3.2.15. Plasmid Sequencing

All plasmid sequenced by Eurofins using the TubeSeq service from the T7 promoter and terminator. Sequencing results were visualised using Benchling, each sequence was aligned with a predicted *in-silico* clone to check the success of the ligation.

3.2.16. Sucrose Shock

Cell pellets from expression were weighed and resuspended in Sucrose Shock (SS) Buffer (50 mM Tris/HCl, 1mM EDTA, 20% sucrose, pH8, filtered) in place of IMAC Buffer A. This resuspension was stirred at 4°C for 1.5 hour (first 50 µl sample taken) with 60 µl 1M $MgSO_4$ per gram of pellet added after the first hour. The resuspension was centrifuged at 40,000*g* for 20 minutes at 4 °C and the supernatant collected (second sample taken) and stored at 4 °C. The pellet was resuspended in ice-cold water 1/30th volume of original expression culture (third sample). This was stirred for 1 hour at 4 °C then centrifuged at 40,000*g* for 20 minutes at 4 °C. The second supernatant was collected and stored at 4 °C (fourth sample) and the pellet discarded (fifth sample). The two supernatants were combined and dialysed overnight (8,000 kDa MWCO dialysis tubing (Thermo Fischer Scientific) in 2L Ion Exchange (IE) Buffer A (Tris 50 mM, NaCl 5mM, pH7.0). Dialysed sample was recovered and stored at 4 °C until needed.

3.2.17. Ion Exchange Chromatography

2x1ml HiTrap™Capto™Q columns (GE Healthcare) were combined in tandem and washed with 5CV miliQ water, then 5CV IE Buffer A. The dialysed supernatant was spun at 14,100*g* at 4 °C for 20 minutes to remove impurities then loaded onto the column using a peristaltic pump. The sample was washed with 2CV of IE Buffer A, the column was then attached to an ÄKTA Prime for elution using a salt gradient to 100% IE Buffer B (Tris 50 mM, NaCl 1 M, pH7.0) and collecting 2ml fractions. SDS-PAGE of these samples was run, and the promising fractions taken forward to GF, performed as above.

3.2.18. UV Vis Spectra

Selected GF fractions were combined and concentrated to 2ml using centricons (Sigma Vivaspin 10,000Da MWCO) at 4 °C. Protein concentration was calculated by measuring OD_{280} absorption of a 5x dilution and applying Beer Lamberts Law with the extinction coefficient of Nmar1131

(42,290 $M^{-1}cm^{-1}$) calculated from ProtParam (Gasteiger et al., 2005). Copper titrations were performed by scanning a x100 dilution of the concentrated protein sample from 200-800nm, with medium scan control and baseline correction. A stock $CuSO_4$ solution was then titrated in until 19 equivalents were added. Each addition was incubated for 2 minutes before the spectrum was rescanned. All measurements were made in a quartz cuvette (Hellma-Analytica), with a pathlength of 1cm.

3.3. Results

3.3.1. Sequencing Purchased Plasmids

The three target genes were codon optimised for expression in *E.coli* by Twist Bioscience, however, no sequencing information was provided. A restriction digest was first performed to confirm the identity of the restriction sites based on the *in-silico* data supplied to Twist Bioscience. The plasmids were first transformed in *E.coli* XL1 Blue cell line and miniprepped, before digestion with NdeI and HindIII restriction enzymes (Fermentas). Unexpectedly, only one band was seen on the agarose gel for all three plasmids suggesting an unsuccessful digest of either one or both sites. This was repeated and another negative result was obtained, so the samples were sent for sequencing using the T7 promoter and terminator sequence primers. Sequencing analysis using Benchling showed that a restriction site for NdeI was present, but the HindIII site had been replaced with XhoI, explaining the negative restriction digest result with NdeI and HindIII. Benchling was used to perform an in-silico digestion and ligation of the sequencing results into the pET28a plasmid, using the NdeI and HindIII restriction sites. This was used as the reference sequence for future experiments, as this showed the codon optimised sequence for expression in *E.coli* as produced by Twist.

3.3.2. Test Expressions of Nmar genes

Plasmids containing each *Nmar* gene were transformed into BL21 (DE3) *E. coli* cells, a cell line optimised for protein overexpression. To determine the optimum growth conditions, 100 ml flasks with 30 ml of LB and 2xYT medium were prepared and samples taken before and after induction with IPTG. The level of target protein expression was identified by SDS-PAGE.

Following induction with IPTG, Nmar11321 and Nmar1667 showed a proportionally larger expression band at the expected molecular weight (Nmar1131; 40.68kDa and Nmar1667; 51.17kDa) compared to the pre-induction samples. A clear expression band for Nmar1354 (49.64kDa) was difficult to distinguish due to gel overloading, however a potential band was seen around 50kDa, so this sample was carried over for further experiments. Slightly stronger expression bands for some of the 2xYT cultures were seen so this media was chosen for proceeding experiments (Figure 21).

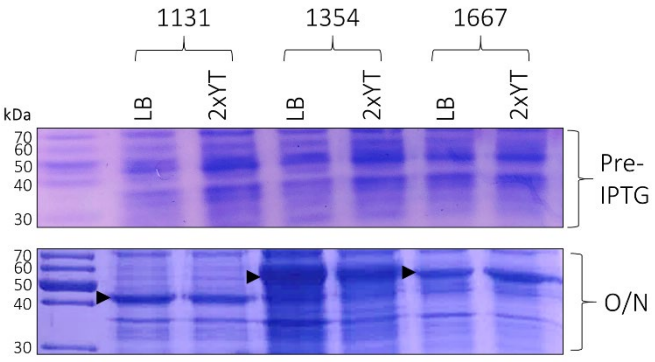


Figure 21 Test expressions of Nmar1131, Nmar1354 and Nmar1667 in LB and 2xYT. Samples before IPTG addition, and after overnight (O/N) incubation.

3.3.3. Test Expressions with 1mM $CuSO_4$

The effect of exogenous copper on target protein expression was tested with the addition of 1mM $CuSO_4$ to 2xYT growth media. The same protocol was repeated as above and clear expression bands were seen for Nmar1131 and Nmar1667 in both samples, where Nmar1354 was again less defined both with and without copper (Figure 22). However, these results were ambiguous, so future expressions further tested the addition of $CuSO_4$ to the growth media.

3.3.4. Large Scale Expression of Nmar1131

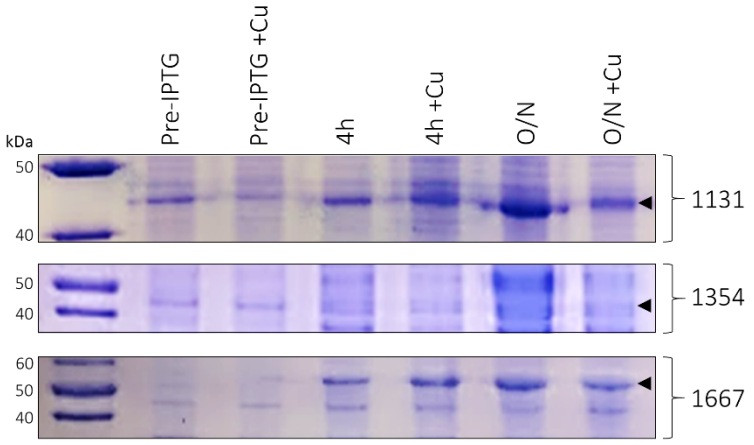


Figure 22 Test Expression of Nmar1131, Nmar1354 and Nmar1667 with 1mM $CuSO_4$ in 2xYT.

Nmar1131 gave the most promising results from the expression tests and was chosen to be taken forward for large scale expression to establish a basic protocol for purification. Nmar1131 was over-expressed in four 2L flasks with 750 ml of 2xYT broth, two of which had 1 mM of $CuSO_4$ and two without. Cells were harvested and loaded onto an IMAC column and 2 ml fractions collected by ÄKTA Prime. An unclear elution profile was seen, with negligible absorption values with and without copper (Figure 23A). However, fractions were still collected from under the peaks and ran on an SDS-PAGE. Nmar1131 grown with copper showed no bands at the expected molecular weight (figure not shown). However, one fraction from the copper-less media showed

a potential band at the correct weight for Nmar1131 (Figure 23B, indicated with an arrow) However, it was eventually decided this band was too high in MW, and the samples discarded.

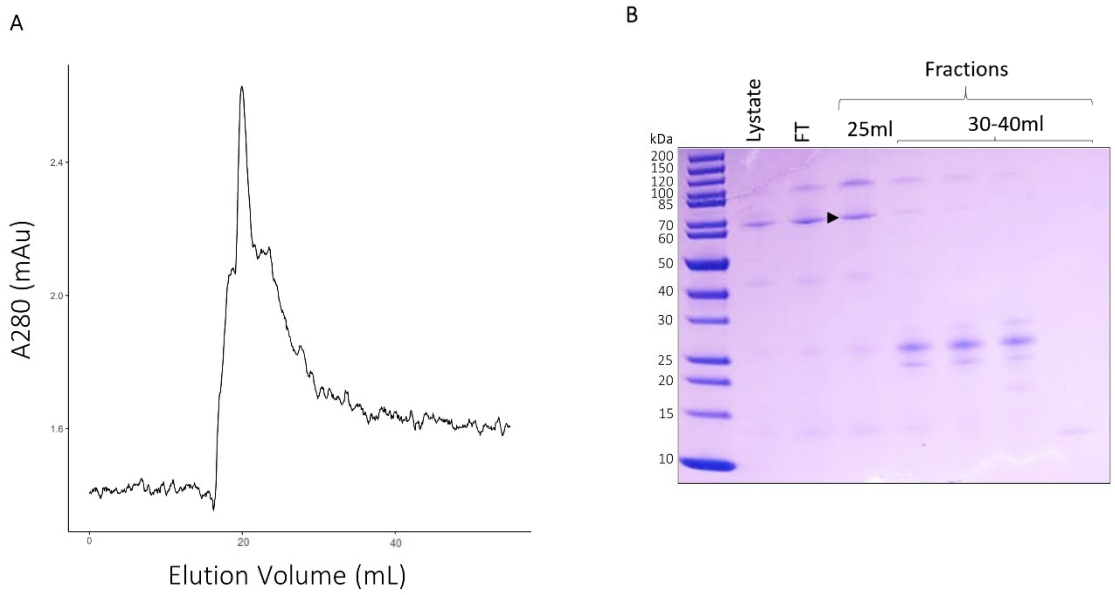


Figure 23 Elution profile (A) and SDS-PAGE of fractions from the first IMAC Purification Nmar1131 with no Copper 2xYT. Predicted band indicated by an arrowhead.

Large scale over-expression was also carried out with Nmar1667 (51.17 kDa) with no copper in 2xYT media. Overexpression showed a clear expression band (Figure 24A). This band was still seen after IMAC purification albeit very faintly at approximately 50kDa (Figure 24B). However, the fractions were pooled, concentrated and further purified on a G75 GF column, however no correct MW bands were seen (Figure 24C).

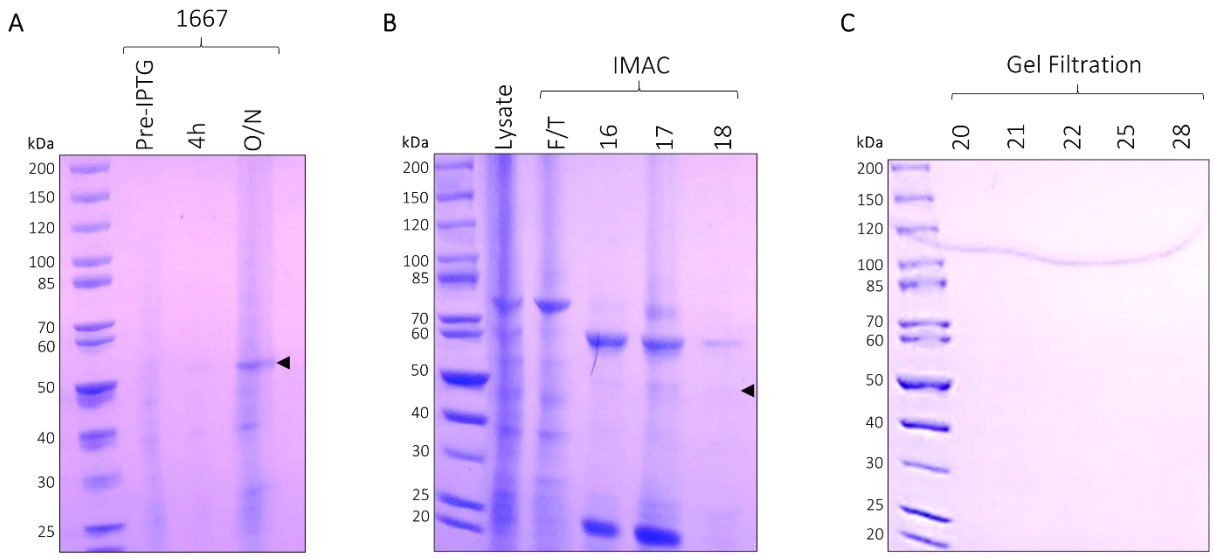


Figure 24 Nmar1667 in 2xYT with no copper. A) Expression sampes B) IMAC fractions and C) Gel Filtration fractions (G75).

3.3.5. Expressions with 0.1mM $CuSO_4$

As above, a 750ml expression of Nmar1667 and a test expression of Nmar1354 was carried out supplemented with a final concentration of 0.1 mM of $CuSO_4$ (Figure 25A). From the expression samples, the only clear bands were below the expected molecular weight for both proteins. From this series of unsuccessful experiments, a different cell line and media were attempted.

3.3.6. Test Expression in C43 (DE3)

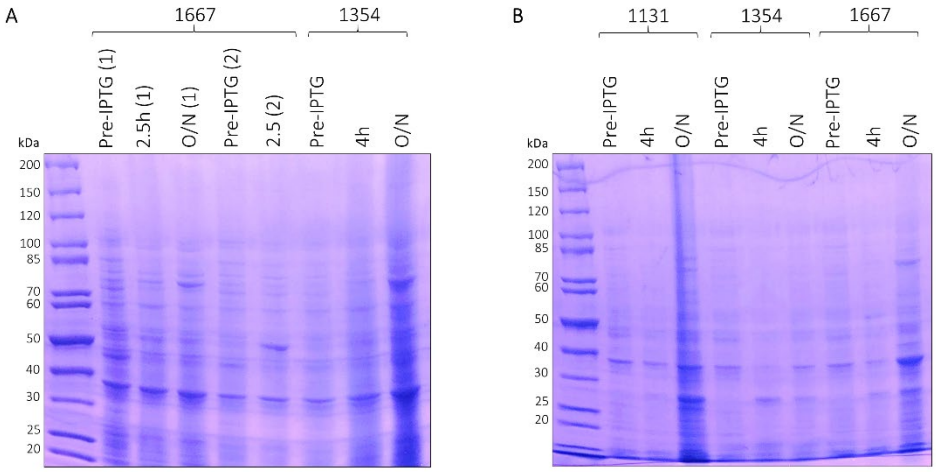


Figure 25 Overexpression A) in 2xYT with 0.1mM $CuSO_4$ B) in LB in C43 DE3 cell line.

OverExpress™ C43 (DE3) (sigma-aldrich) is another strain for over-expression, derived from the BL21 (DE3) cell line and chosen for its resistance to toxic recombinant proteins. Nmar1131, Nmar1354 and Nmar1667 were transformed into C43 competent cells, and expression carried out as above, however, these experiments showed no more promising results (Figure 25B).

3.3.7. Cloning into pET26b Vector

After a series of unsuccessful attempts to express and purify these proteins in pET28a, which would express the proteins in the cytoplasm, *Nmar1131*, *Nmar1354* and *Nmar1667* were cloned into a new expression vector, pET26b. In contrast to pET28a, pET26b is a periplasmic expression vector which contains an N-terminal pelB signal peptide which exports the unfolded protein to the periplasm. This new expression system was chosen as all the target proteins are predicted to be periplasmic in their native organism.

NcoI and HindIII restriction sites were used to clone the target genes into pET26b. However, Nmar1354 and Nmar1667 contained an internal NcoI restriction site in their genes. Before

cloning could take place, this restriction site had to be closed. This was done by creating a silent mutation to, close the site but maintaining the same amino acid identity at that codon; P196 (CCC -> CCT) and H154 (CAT -> CAC), respectively. Although these sites were no longer using the optimised codon for these amino acids, this was not predicted to negatively affect expression. The success of these mutations was confirmed by sequencing from the T7 terminator primer, after DnpI digestion and transformation into DH5α cells (another *E. coli* cell line, with maximised transformation efficiency). Sequencing data corroborated that *Nmar1354* was successfully mutated (Figure 26) however *Nmar1667* did not produce transformants.

NcoI and HindIII restriction sites were already present in the pET26b vector. Primers to amplify these genes from the pET28a vector were designed to contain NcoI and HindIII sites at the 5' and 3' ends of the target genes, to enable ligation of the products into the pET26b vector. PCR was performed twice per gene to produce a large amount of mutated DNA for ligation. 5µl of PCR product was ran on a 1% agarose gel to confirm the success of the PCR reaction. The remainder was digested with NcoI and HindIII to produce sticky ends and ran on a 1% agarose gel to clean the restriction digest product. The cleaned product was then extracted by band excision and a Nucleospin®gel and PCR clean up kit. This last step was also performed on the miniprepped pET26b backbone to produce sticky ends, these products were then ligated to insert the target gene into the new backbone.

The success of this ligation was tested by running the resulting ligated product on an agarose gel. A single band at ~6,500bp (Figure 28) suggested a successful ligation. Although the ladder is not fully resolved, making the size of ligation band difficult to estimate. It is clear that this ligation band represents a longer DNA segment that the gene fragment (~1,000bp) and backbone (~5,500bp) alone. Even so, the ligation product was sent for sequencing for more robust confirmation. The reference sequences produced earlier were cloned in-silico into the pET26b plasmid and aligned with the sequencing results to confirm successful cloning (Figure 27).

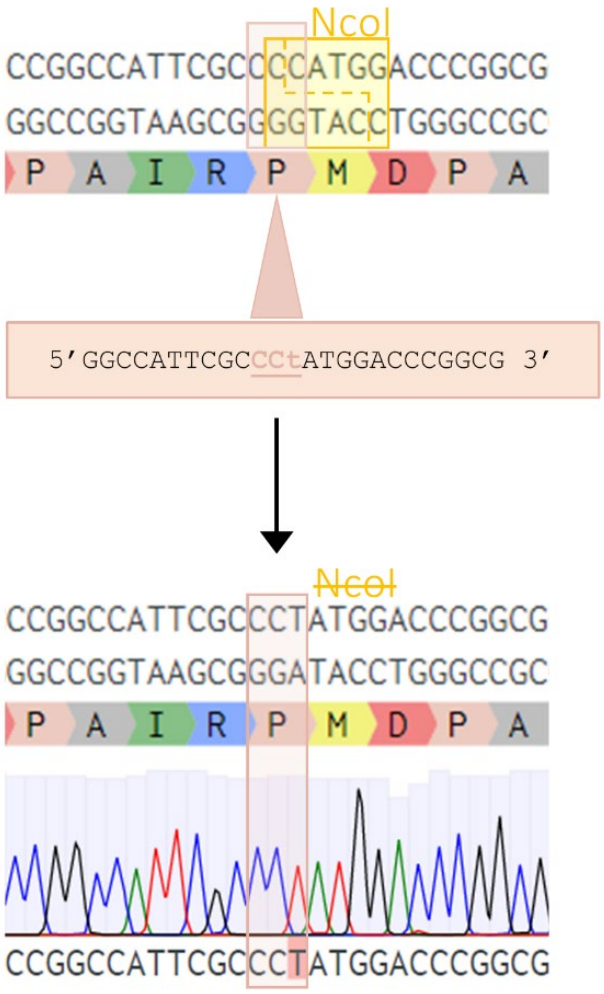


Figure 26 Site Directed Mutation of Nmar1354 to remove NcoI restriction site for cloning into pET26b vector. Sequencing results are aligned under the desired sequence to show successful mutation.

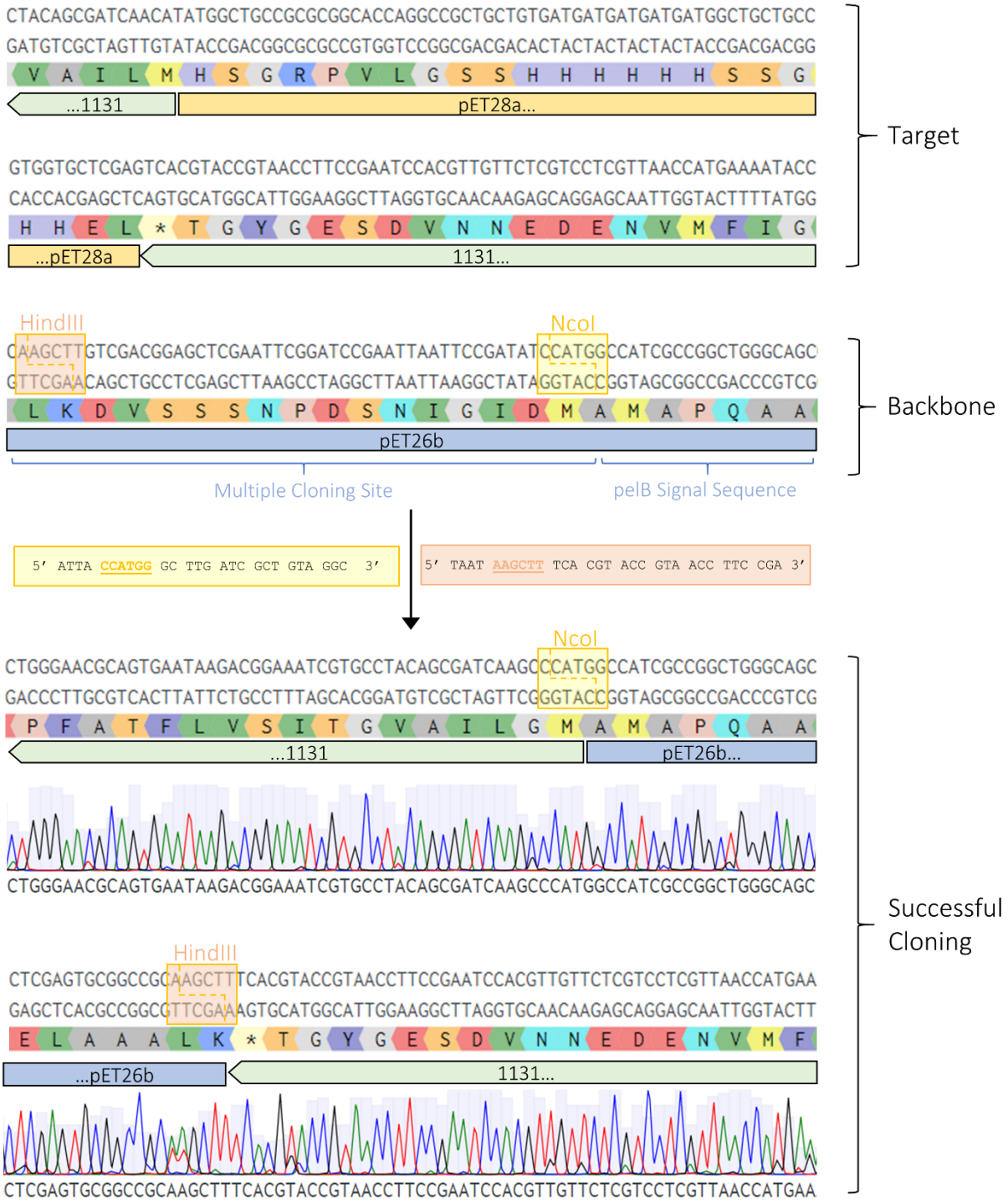


Figure 27 Cloning of Nmar1131 into pET26b vector. Beginning and end of Nmar1131 gene in pET28a shown with 6x-His tag. pelB signal sequence and multiple cloning site of pET26b shown along with HindIII and NcoI sites highlighted. Primers used to add the HindIII and NcoI restriction sites shown in orange and yellow boxes, respectively. Restriction sequence of each is underlined. Sequencing results were aligned with in-silico cloned template, showing successful cloning. An additional "GC" is added after the NcoI restriction sequence, to restore the correct open reading frame of the gene, adding an additional glycine residue.

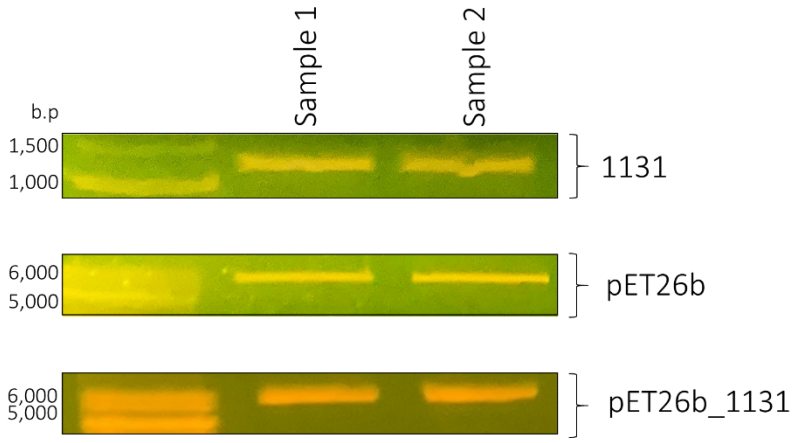


Figure 28 1131 and pET26b restriction digest with the subsequent ligation 1131_pET26b.

3.3.8. Overexpression of Nmar1131 in pET26b

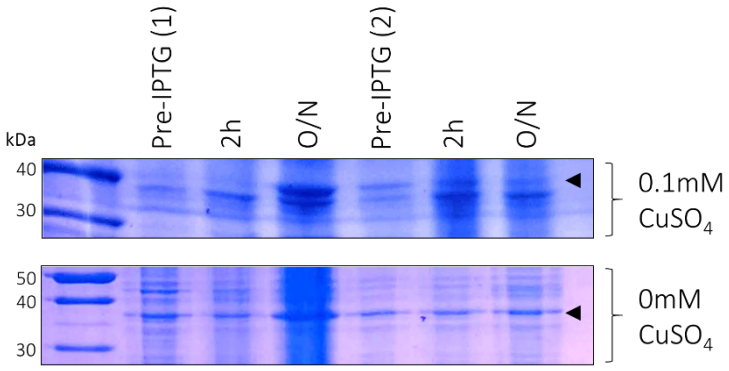


Figure 29 Expression test Nmar1131 pET26b in LB. Arrowhead denotes potential Nmar1131 expression band.

Nmar1131 was over-expressed as previously in LB cultures. One culture was supplemented with 0.1mM $CuSO_4$ and the other contained no additional copper in the growth media. These cultures were seen to grow slower than previously, with one flask taking five hours to grow to an OD_{600} of 0.8, compared to roughly 2-3 hours as seen using the pET28a construct. Clear expression bands

were seen, however, these were at slightly lower MWs than the expected 40.7kDa (Figure 29). As the presence of copper seemed to inhibit protein expression, the cultures with no additional copper in the growth medium were carried forward to Ion Exchange chromatography (IE). IE replaced IMAC as the initial purification step as cloning into pET26b removed the 6x His tag. Likewise, as pET26b tagged the target proteins for exportation to the periplasm, a sucrose shock was used to harvest these proteins, as a whole cell lysis was not required.

The IE elution profile showed two promising peaks, eluting at roughly 50% and 70% Buffer B with absorptions around 1,200 mAu (Figure 30A). Fractions collected from under these peaks showed that Peak 1 potentially contained the target protein, with a faint band around 40kDa. However, due to the COVID-19 pandemic and subsequent lab closure from March 20th, this sample was not taken any further.

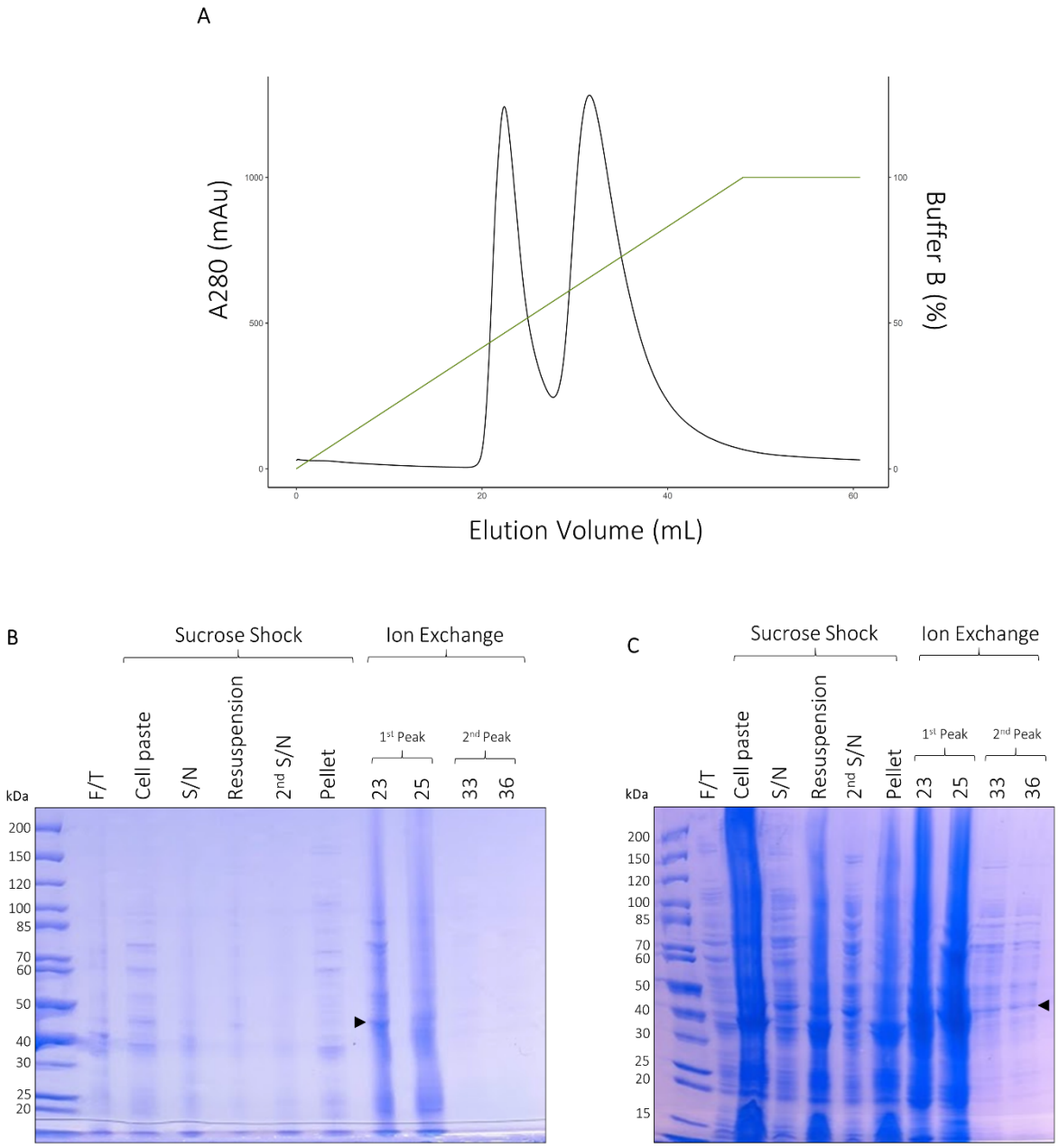


Figure 30 First Nmar1131 pET26b Ion Exchange. Gel repeated (C) due to underloading in B.

3.3.9. Second Over-Expression of Nmar1131 in pET26b

The above protocol was repeated, and slow growth also observed with a five-hour incubation needed to achieve an OD_{600} of 0.6. Three 750ml flasks of LB yielded a combined pellet of 7.7g. The samples collected during expression showed a robust expression band at around 40kDa (Figure 31). The calculated predicted weight of Nmar1131 expressed from pET26b is 40.7kDa

SDS-PAGE analysis of the sucrose shock samples showed the presence of a defined band just below 40kDa in the first supernatant and to a lesser degree in the second supernatant (Figure 32C). From this, the two supernatants were pooled and dialysed overnight into IE Buffer and loaded onto an IE the following day.

As previously, from IE chromatography two well defined peaks were seen (Figure 32A), with a max absorption of 283.0 mAu) at 30.7ml. Samples from these were ran on an SDS-PAGE. A band at 40kDa was seen from the first peak, albeit at a slightly higher molecular weight than those seen from the sucrose shock (Figure 32C). Regardless of this, all fractions from under the first peak were pooled and concentrated to 2 ml for GF.

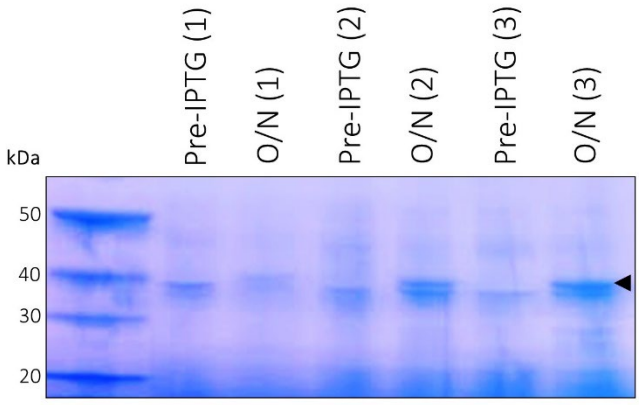


Figure 31 Expression test Nmar1131_pET26b. Showing pre-IPTG and post IPTG overnight (O/N) samples.

GF chromatography also produced two distinct peaks. (Figure 32B). The major peak had a small shoulder and was at a max absorption of 31.4 mAu at 67.2ml. The second peak had an absorption of only roughly 5 mAu. When ran on an SDS-PAGE samples from the major peak produced a band just above 40kDa. Although this is a potential Nmar1131 band, the sample was still contaminated. However, due to time restraints this was not purified any further.

The GF fractions from the major peak were pooled and concentrated to 2ml, then a 5 times dilution was made, and scanned on a spectrophotometer from 200-800nm. This scan showed a strong peak at 280nm confirming the presence of protein in the sample. A concentration of 13.7 µM of undiluted protein was calculated using the predicted extinction coefficient (42,290 $M^{-1}cm^{-1}$).

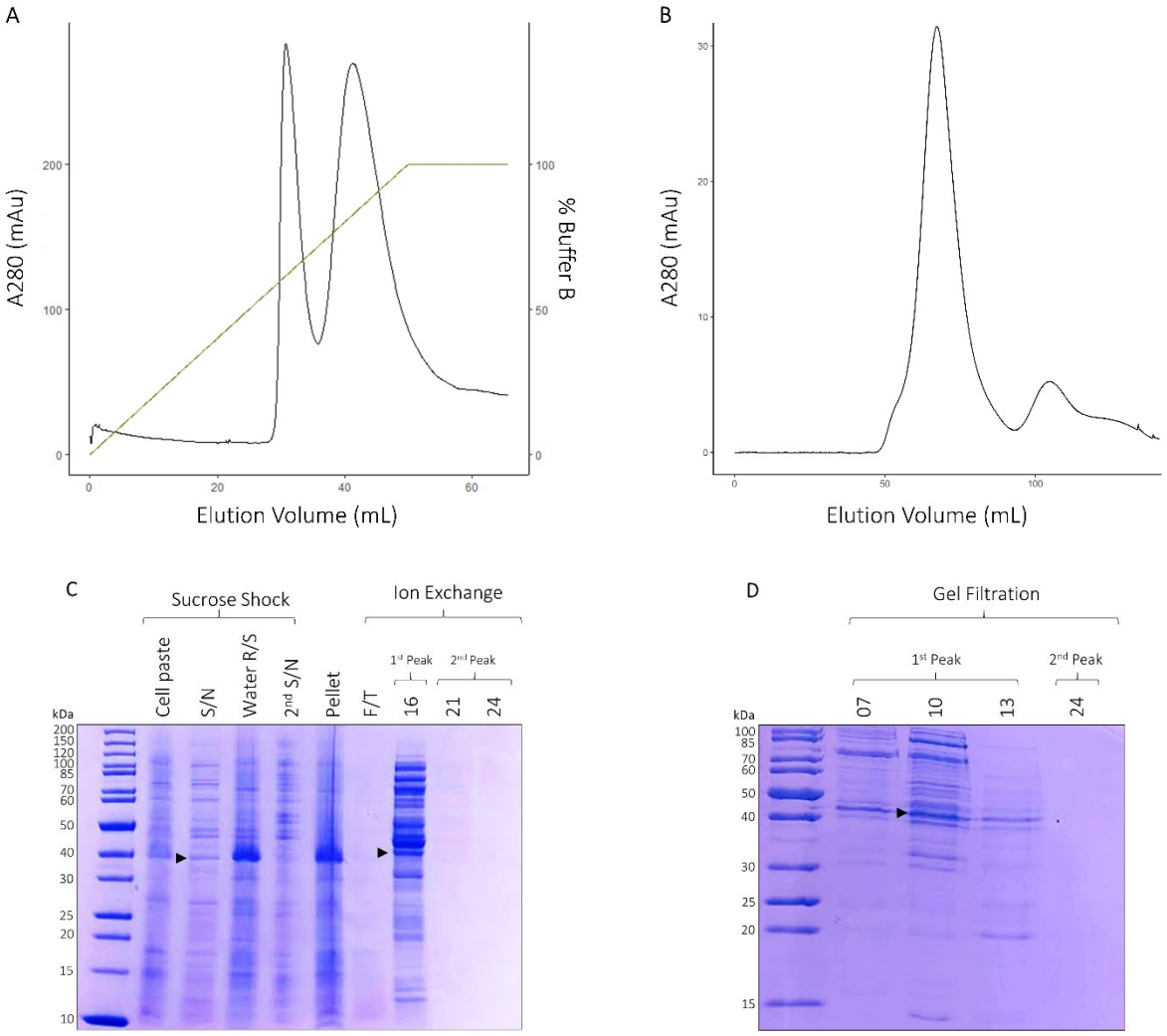


Figure 32 Purification of Nmar1131. Elution Profiles from purification of Nmar113: A) Ion Exchange step, ran from 0-100% Buffer B (green). B) Gel filtration. SDS-PAGE analysis of purification fractions and sucrose shock samples.

3.3.10. UV Vis Spectroscopy

The purified Nmar113 was tested for the presence of bound copper using electronic absorbance spectroscopy. A spectrum of the as purified sample was taken, and revealed an absorbance peak at 280 nm attributed to arise from the protein. Between 350-800 nm no absorbance peaks were observed. In this region absorbance bands arising from a Type 1 copper site (600 nm) in the oxidised (Cu(II)) form would have been expected. Alternatively, it could be that copper is not present in the protein, so a titration using a stock $CuSO_4$ solution was carried out (Figure 34). No spectral changes in the 350-800 nm region were observed, but it was noted that the absorbance band at 280 nm increased with increasing additions of $CuSO_4$ (Figure 34).

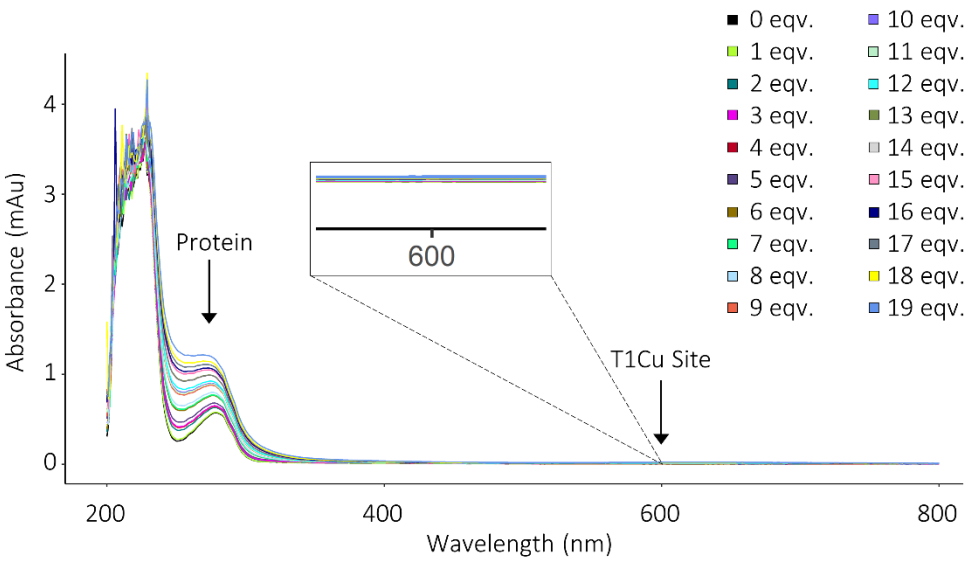


Figure 34 UV Vis Spectroscopy. Protein peak at 280nm and expected T1Cu site position at 600nm labelled

3.4. Discussion

Three previously uncharacterised proteins from *N. maritimus* were selected due to their likely role in the ammonia oxidation pathway of AOA. Previous bioinformatics studies have predicted these proteins to belong to a family of copper-binding cupredoxin-like proteins (Walker *et al.*, 2010). More specifically bioinformatics studies in this project have predicted Nmar1131 and Nmar1354 to be Multicopper Oxidases (MCO), and Nmar1667 a Nitrite Reductase (CuNiR). Where bioinformatics predictions for structure and function have been suggested for many proteins in this pathway, almost all remain uncharacterised *in-vitro*. These MCO and CuNiRs could play a vital role in the Hydroxylamine oxidation mechanism and. As structure and function are closely linked, the purification and subsequent structural identification through methods, such as X-ray crystallography is of paramount importance for understanding how this pathway functions in AOA, and for the wider role AOA play in global nitrogen cycling.

As archaea are notoriously slow growing it is difficult to attain a large enough biomass for biochemical studies. Therefore, isolation from the native organism would be extremely difficult to carry out, especially in a limited time frame. Therefore, this chapter has explored the possibility of setting-up a recombinant expression system for three proteins Nmar1131, Nmar1354 and Nmar1667, predicted to be two MCOs and a CuNiR, respectively. Out of these three proteins Nmar1131 was recombinantly overexpressed in *E. coli* for purification and further characterisation.

Bioinformatics studies predicted Nmar1131 to be periplasmic due to the presence of a Sec signal peptide on its N-terminus, thus the unsuccessful expression of Nmar1131 in the pET28a vector was possibly due the use of pET28a as a cytoplasmic expression vector. Although periplasmic proteins can be expressed in the cytoplasm, the closer an expression system is to the native system, the greater chance of successful expression. By cloning Nmar1131 into pET26b, a vector

which adds an N-terminal pelB signal peptide, the unfolded Nmar1131 would be exported to the periplasm for folding, where it is less likely to aggregate, and thus increase the likelihood of successful expression.

Furthermore, the periplasmic localisation of Nmar1131 negates the need to rupture the cells for harvesting. Instead, a sucrose shock is used to selectively release the periplasmic proteins. The lower concentration of proteins in the periplasm is also advantageous as the protein of interest is of higher relative concentration than it would be in the cytoplasm and have fewer contaminants.

A disadvantage is that, upon cloning into the pET26b vector, the N-terminal 6xHis-tag was lost, rendering IMAC unsuitable for this purification. Instead, IE Chromatography was used as the initial purification step. This method relies on the pH of the buffers creating an environment where the protein of interest is opposingly charged to the column used. The theoretical pI of Nmar1131 is 4.5 so an anionic column was used.

3.4.1. Nmar1131_pET26b Over-Expression

Samples were taken at different stages of the sucrose shock, allowing the location of the protein of interest to be predicted, indicating which fraction should be used for the next step in purification. Typically, periplasmic proteins would be in the first soluble fraction, although could be found in any fraction depending on the efficiency of this technique.

An SDS-PAGE confirmed the presence of soluble protein just below 40kDa in the sucrose shock supernatants (Figure 32C) although this was at a slightly lower MW than the expression band seen from the expression samples (Figure 31).

An intense band at the correct molecular weight is also seen in the pellet of the sucrose shock. This could either suggest low periplasmic transport efficiency, or more likely that the protein has aggregated. Regardless, both supernatants were pooled to ensure the highest possible yield and dialysed overnight into IE buffer.

IE chromatography showed two high absorbance peaks on the elution profile indicating the presence of protein in the sample. Fractions from under both peaks were selected and run on an SDS-PAGE to determine the presence of the target protein. Fraction 16 from the first peak showed a strong band at 40 kDa, thus peak 1 fractions were pooled and taken forward the GF.

Significantly fewer bands were observed after GF compared to the selected fraction (16) in IE. Another purification step would be required to produce a pure sample, but due to time restraints no further purification could take place. Without a pure sample, confident identification of the protein could not take place. However, the characteristics of MCO proteins, gave a potential crude test for identifying if copper proteins are present in the final sample, as discussed below.

3.4.2. Nmar1131 pET26b Characterisation

As Nmar1131 is a predicted MCO, which contains a Type I copper site, the addition of $CuSO_4$ should in theory produce a blue sample and a peak in the visible region of the absorbance spectrum at 600nm if the T1Cu site is loaded with Cu(II). However, up to 19 equivalents of Cu(II) were added to the sample but no changes were seen in the 600nm region.

It is noted that, in a type I copper site the coordinated, Cu(I) is spectroscopically silent in the visible region, so even if the T1Cu site is saturated with reduced copper, no 600nm signal would be produced (Yamaguchi *et al.*, 2004). This could be overcome by the addition of an oxidant such as potassium ferricyanide to the solution. Furthermore, if the T1Cu site is buried in the structure of the protein, titrating copper may not be sufficient for reconstituting this site. However, other

similar copper proteins have been reconstituted through this method, such as CueO, a MCO protein from E.coli (Musci et al., 2004).

Interestingly, an increase in 280nm signal was shown with increasing $CuSO_4$ concentration. This 280nm absorbance increase could be due to ligand to metal charge transfer (LMCT) between the cross-linked cysteine residues and potential Cu(I) bound to the T1Cu site. (Tarasava, et al. 2016), potentially showing the presence of a copper-binding protein in these samples. However, this could also be due to protein unfolding, as the absorption given by these readings is primarily produced by exposed tryptophan, tyrosine, and cysteine residues. If $CuSO_4$ additions caused the protein to unfold, these would be more exposed potentially producing a larger signal.

3.4.3. Environmental Distribution of AOA and AOB

Positive results from the archaeal and bacterial AmoA PCR experiments showed the presence of both AOA and AOB in marine and agricultural soils, both occurring together (Soil sample 3 and Hythe) and separately (Soil samples 1 and 2). However, where the positive control for the bacterial PCR reaction presents at the expected length (~600bp), the archaeal positive control smeared and no clear band at the correct length (~700bp). From this, all archaeal PCR reactions must be repeated in order to confirm the reliability of these results.

3.4.4. Conclusions

In summary, Nmar1131 was successfully over-expressed in *E.coli*, however, further purification steps should be used to achieve a pure culture and mass spectrometry could be used confirm the identity of this protein. Presence of a correct molecular weight band in the insoluble fraction of the sucrose shock could also suggest this protein has aggregated during expression. If future experiments experience difficulties in over-expressing appropriate yields for further analytical experiments, the expression conditions and protocols should be revised. Furthermore, the addition of solubility-enhancing fusion peptides to the target gene could be considered.

Once a pure sample is achieved, future work should focus on crystallising Nmar1131 so X-ray crystallography can be performed, providing experimentally derived structural data. Attempts to reconstitute the copper sites were also unsuccessful. Future work should repeat this protocol in the presence of an oxidising agent such as potassium ferricyanide to ensure any bound copper is in the Cu(II) state so that the Type 1 copper site can be spectroscopically observed. Likewise, techniques such as inductively coupled plasma mass spectrometry (ICP-MS) and electron paramagnetic resonance (EPR) spectroscopy should be utilised to determine the identity of the bound metals more confidently.

Nmar1354 was successfully mutated to remove the NcoI site, however cloning was not successful in part due to the incorrect sequencing used to order the original primers. New primers were ordered to the correct Nmar1354 sequence, and these should be used in future experiments for successful cloning.

Nmar1667 was not successfully mutated. Site directed mutagenesis should be reattempted at a range of annealing temperatures in order for this to be successfully cloned into pET26b for periplasmic expression in future work.

Furthermore, ammonia oxidising archaea and ammonia oxidising bacteria were found in both marine sediment and agricultural soils. Unfortunately, due to time restraints the efficiency of Fresh Water Media and Synthetic Crenarchaea media for the selection of AOA was not established. Future work should continue to culture these environmental samples and perform qPCR on samples taken over the course of growth to determine how the identity of the organisms in these cultures changes over time.

3.5. Supplementary

3.5.1. Environmental Samples

Two different medias were made for culturing the environmental samples, as these were collected from marine and agricultural soils. Fresh Water Medium (FWM) was used for the agricultural soil samples. For 1l of FWM 1g NaCl, 0.4g $MgCl_2.6H_2O$, 0.2g KH_2PO_4 and 0.5g KCl were autoclaved. Separate solutions of Trace elements (100mM HCl, 2.1g $FeSO_4.7H_2O$, 30mg H_3BO_3, 100mg $MnCl_2.6H_2O$, 190mg $CoCl_2.6H_2O$, 24mg $NiCl_2.6H_2O$, 2mg $CuCl_2.2H_2O$, 144mg $ZnSO_4.7H_2O$, 36mg $Na_2MoO_4.2H_2O$ per litre), 1M $NaHCO_3$, 1M FeNaEDTA and 1M Na-Pyruvate were filter sterilised and 1ml Trace Elements, 1ml FeNaEDTA solution, 2ml $NaHCO_3$, 10ml 1M HEPES buffer (Sigma), 50µl of Na-Pyruvate, together with Kanamycin, Streptomycin and Ampicillin all to 50µg/ml added to the autoclaved mixture under aseptic conditions.

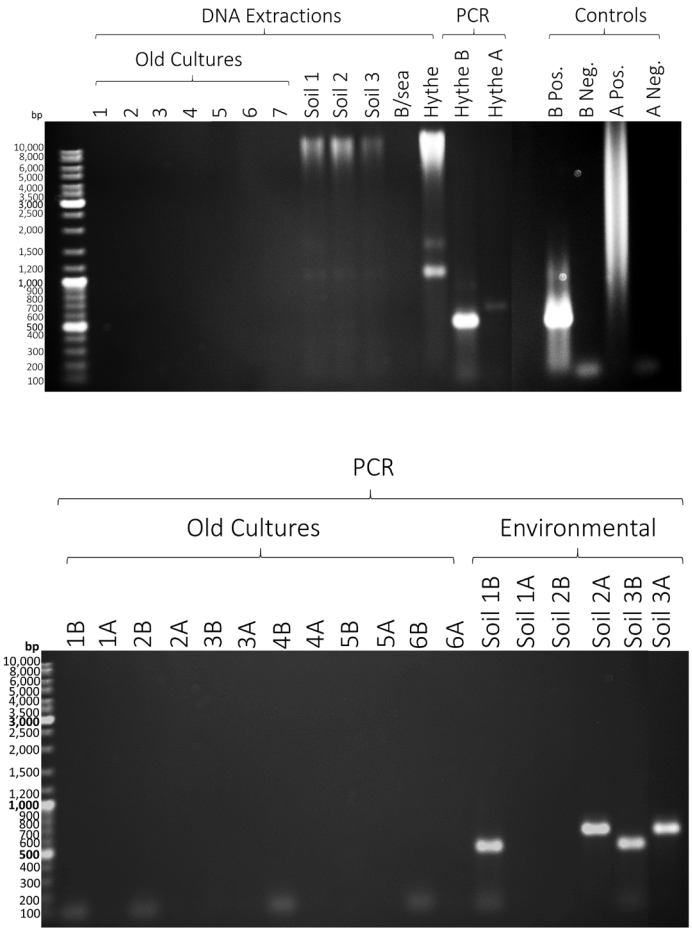
For the Brightlingsea (51.805409, 1.010030) and Hythe (51.879352, 0.929307) samples the suggested *Nitrosopumilus* medium (NM) from ATCC culture collection was used. NM was made by adding; 26 g NaCl, 5 g $MgSO_4 \cdot 7H_2O$, 5 g $MgCl_2 \cdot 6H_2O$, 1.5 g $CaCl_2 \cdot 2H_2O$, 0.1 g KBr per litre of water and autoclaving. Separately, stock solutions of 4g/L KH_2PO_4, 1M NH_4Cl, and 100ml Vitamin solution (10mM, pH7.1 Sodium phosphate buffer, 4mg 4-Aminobenzoic Acid, 1mg D-biotin, 10mg Nicotinic Acid, 5mg Calcium D-pantothenate, 15mg Pyridoxin dihydrochloride) were filter sterilised and 3 ml $NaHCO_3$ (1 M), 5 ml KH_2PO_4 (0.4 g/l), 0.45ml FeNaEDTA solution, 1 ml trace element solution, 0.2 ml NH_4Cl (1 M), 10ml HEPES, and 10ml Vitamin solution added per litre under aseptic conditions.

Each culture was prepared by adding 20ml of the respective media into sterile 50ml universal sample bottles. 0.5g of the environmental sample was added and the cultures incubated at 21°C without agitation. Every two weeks, 500µl of culture was used to inoculate 20ml of fresh media.

Fresh antibiotics were also added with every subculture, and the inoculate passed through a 4µm filter to remove fungal contaminants.

PCR was performed on the DNA extractions from the environmental sample cultures to determine the presence of AOA or AOB in the samples. The primers used in this PCR are specific to either the archaeal (crenamoA-23F (5' ATGGTCTGGCTWAGACG-3') and crenamoA-616R (5' GCCATCCATCTGTATGTCCA) (Tourna *et al.*, 2008) or bacterial (amoA-1F (5' GGGGTTTCTACTGGTGGT) and amoA-2R (5' CCCCTCKGSAAAGCCTTCTTC) (Rotthauwe et al., 1997) AmoA subunit gene. 12.5µl 2x AppTAQ RedMix (Appleton®), 1µl of each primer, 1µl of DNA template and 10µl ddH_2O were mixed to a final volume of 25µl and the target DNA amplified using the following protocol: 10 min 95°C, (45s at 95°C, 30s 55°C and 30 s 72°C) x35 and 7 min at 72°C performed in a Veriti 96-Well Thermal Cycler (Applied Biosystems). 25µ of the PCR products were run on a 1% w/v agarose gel with 1Kb GeneRuler (Thermo Scientific™) ladder to determine the success of the reaction. DNA extraction samples were also run on an agarose gel, and were prepared by mixing 15µl with 6x DNA loading dye (Thermo Scientific™)

3.5.2. *AmoA* Amplification in Environmental Samples



Supplementary 1 DNA extractions previous student's cultures and environmental samples from agricultural soils, and marine soils from Hythe and Brightlingsea (B/sea).. PCR results of amplification of environmental samples with bacterial (B) AmoA and archaeal (A) primers. The positive (Pos.) and negative (Neg.) controls for each of these PCR reactions are also shown.

Before DNA extraction, all samples were grown for a minimum of 6 weeks in either FWM (soil samples) or SC media (Hythe and Brightlingsea samples) to allow a suitable level of archaeal biomass to accumulate. Each of these growth medias has been designed to select for ammonia oxidising archaea (AOA), with the ultimate aim of producing a pure culture of AOA. Every 2 weeks, the cultures were tested with nitrite test strips (Quantofix®), when the presence of nitrite

was detected, and therefore ammonia oxidisers, 500 µl of each culture was inoculated into fresh culture to encourage further growth and AOA selection. When the DNA extractions and PCR products of old cultures inherited from a previous student were ran on a 1% agarose gel, no clear bands were seen. Likewise, with the DNA extraction from the environmental Brightlingsea culture. However, clear bands were seen for all three agricultural soil samples and the Hythe sample. Both the archaeal and bacterial AmoA primers produced clear bands for the Hythe PCR products. Likewise, for the third soil sample. Whereas only positive results were seen for the bacterial and archaeal primers in Soil 1 and Soil 2, respectively (Supp 1).

References

Abraham, Z. H. L., Lowe, D. J. and Smith, B. E. (1993) 'Purification and characterization of the dissimilatory nitrite reductase from Alcaligenes xylosoxidans subsp. Xylosoxidans (N.C.I.M.B. 11015): Evidence for the presence of both type 1 and type 2 copper centres', *Biochemical Journal*. Portland Press Ltd, 295(2), pp. 587–593. doi: 10.1042/bj2950587.

Almagro Armenteros, J. J. *et al.* (2019) 'SignalP 5.0 improves signal peptide predictions using deep neural networks', *Nature Biotechnology*. Nature Publishing Group, 37(4), pp. 420–423. doi: 10.1038/s41587-019-0036-z.

Anthony, C. (1982) *The Biochemistry of Methylotrophs*, *Academic Press*.

Arp, D. J., Sayavedra-Soto, L. A. and Hommes, N. G. (2002) 'Molecular biology and biochemistry of ammonia oxidation by Nitrosomonas europaea', *Archives of Microbiology*, pp. 250–255. doi: 10.1007/s00203-002-0452-0.

Baker, D. B. and Richards, R. P. (2002) 'Phosphorus budgets and riverine phosphorus export in northwestern Ohio watersheds', *Journal of Environmental Quality*, 31(1), pp. 96–108. doi: 10.2134/jeq2002.9600.

Bateman, A. *et al.* (2021) 'UniProt: The universal protein knowledgebase in 2021', *Nucleic Acids Research*. Oxford University Press, 49(D1), pp. D480–D489. doi: 10.1093/nar/gkaa1100.

Beaumont, H. J. E. *et al.* (2002) 'Nitrite reductase of Nitrosomonas europaea is not essential for production of gaseous nitrogen oxides and confers tolerance to nitrite', *Journal of Bacteriology*. American Society for Microbiology (ASM), 184(9), pp. 2557–2560. doi: 10.1128/JB.184.9.2557-2560.2002.

Beaumont, Hubertus J. E. *et al.* (2004) 'Expression of nitrite reductase in Nitrosomonas europaea involves NsrR, a novel nitrite-sensitive transcription repressor', *Molecular Microbiology*. John Wiley & Sons, Ltd, 54(1), pp. 148–158. doi: 10.1111/j.1365-2958.2004.04248.x.

Beaumont, Hubertus J.E. *et al.* (2004) 'Nitrosomonas europaea expresses a nitric oxide reductase during nitrification', *Journal of Bacteriology*. American Society for Microbiology Journals, 186(13), pp. 4417–4421. doi: 10.1128/JB.186.13.4417-4421.2004.

Beaumont, H. J. E. *et al.* (2005) 'Novel nirK cluster genes in Nitrosomonas europaea are required for NirK-dependent tolerance to nitrite', *Journal of Bacteriology*. American Society for Microbiology (ASM), 187(19), pp. 6849–6851. doi: 10.1128/JB.187.19.6849-6851.2005.

Bédard, C. and Knowles, R. (1989) 'Physiology, biochemistry, and specific inhibitors of CH4, NH4+, and CO oxidation by methanotrophs and nitrifiers.', *Microbiological reviews*. American Society for Microbiology (ASM), pp. 68–84. doi: 10.1128/mmbr.53.1.68-84.1989.

Beman, J. M. *et al.* (2007) 'Distribution and diversity of archaeal ammonia monooxygenase genes associated with corals', *Applied and Environmental Microbiology*. American Society for Microbiology (ASM), 73(17), pp. 5642–5647. doi: 10.1128/AEM.00461-07.

Benkert, P., Biasini, M. and Schwede, T. (2011) 'Toward the estimation of the absolute quality of individual protein structure models', *Bioinformatics*. Oxford Academic, 27(3), pp. 343–350. doi: 10.1093/bioinformatics/btq662.

Bock, E. *et al.* (1995) 'Nitrogen loss caused by denitrifying Nitrosomonas cells using ammonium or hydrogen as electron donors and nitrite as electron acceptor', *Archives of Microbiology*. Springer-Verlag, 163(1), pp. 16–20. doi: 10.1007/BF00262198.

Bollmann, A., Bullerjahn, G. S. and McKay, R. M. (2014) 'Abundance and Diversity of Ammonia-Oxidizing Archaea and Bacteria in Sediments of Trophic End Members of the Laurentian Great Lakes, Erie and Superior', *PLoS ONE*. Edited by J. L. Balcazar, 9(5), p. e97068. doi: 10.1371/journal.pone.0097068.

Bornhorst, J. A. and Falke, J. J. (2000) 'Purification of proteins using polyhistidine affinity tags', *Methods in Enzymology*. Academic Press Inc., pp. 245–254. doi: 10.1016/s0076-6879(00)26058-8.

Boulanger, M. J. and Murphy, M. E. P. (2002) 'Crystal structure of the soluble domain of the major anaerobically induced outer membrane protein (AniA) from pathogenic Neisseria: A new class of copper-containing nitrite reductases', *Journal of Molecular Biology*. Academic Press, 315(5), pp. 1111–1127. doi: 10.1006/jmbi.2001.5251.

Cantera, J. J. L. and Stein, L. Y. (2007) 'Molecular diversity of nitrite reductase genes (nirK) in nitrifying bacteria', *Environmental Microbiology*, 9(3), pp. 765–776. doi: 10.1111/j.1462-2920.2006.01198.x.

Caranto, J. D. and Lancaster, K. M. (2017) 'Nitric oxide is an obligate bacterial nitrification intermediate produced by hydroxylamine oxidoreductase', *Proceedings of the National Academy of Sciences of the United States of America*. National Academy of Sciences, 114(31), pp. 8217–8222. doi: 10.1073/pnas.1704504114.

Caranto, J. D., Vilbert, A. C. and Lancaster, K. M. (2016) 'Nitrosomonas europaea cytochrome P460 is a direct link between nitrification and nitrous oxide emission', *Proceedings of the National Academy of Sciences of the United States of America*. National Academy of Sciences, 113(51), pp. 14704–14709. doi: 10.1073/pnas.1611051113.

Cavalli, A. *et al.* (2007) 'Protein structure determination from NMR chemical shifts', *Proceedings of the National Academy of Sciences of the United States of America*. National Academy of Sciences, 104(23), pp. 9615–9620. doi: 10.1073/pnas.0610313104.

Clark, K. *et al.* (2016) 'GenBank', *Nucleic Acids Research*. Oxford University Press, 44(D1), pp. D67–D72. doi: 10.1093/nar/gkv1276.

Coyne, M. S. *et al.* (1989) 'Immunological identification and distribution of dissimilatory heme cd1 and nonheme copper nitrite reductases in denitrifying bacteria', *Applied and Environmental Microbiology*. American Society for Microbiology (ASM), 55(11), pp. 2924–2931. doi: 10.1128/aem.55.11.2924-2931.1989.

Crooks, G. E. *et al.* (2004) 'WebLogo: A Sequence Logo Generator'. doi: 10.1101/gr.849004.

Denk, T. R. A. *et al.* (2017) 'The nitrogen cycle: A review of isotope effects and isotope modeling approaches', *Soil Biology and Biochemistry*. Elsevier Ltd, pp. 121–137. doi: 10.1016/j.soilbio.2016.11.015.

Dodd, F. E. *et al.* (1998) 'X-ray structure of a blue-copper nitrite reductase in two crystal forms. The nature of the copper sites, mode of substrate binding and recognition by redox partner', *Journal of Molecular Biology*. Academic Press, 282(2), pp. 369–382. doi: 10.1006/jmbi.1998.2007.

Dundee, L. and Hopkins, D. W. (2001) 'Different sensitivities to oxygen of nitrous oxide production by Nitrosomonas europaea and Nitrosolobus multiformis', *Soil Biology and Biochemistry*. Pergamon, 33(11), pp. 1563–1565. doi: 10.1016/S0038-0717(01)00059-1.

Eggert, C. *et al.* (1998) 'Molecular analysis of a laccase gene from the white rot fungus Pycnoporus cinnabarinus', *Applied and Environmental Microbiology*. American Society for Microbiology, 64(5), pp. 1766–1772. doi: 10.1128/aem.64.5.1766-1772.1998.

Ellis, M. J. *et al.* (2007) 'Genomic analysis reveals widespread occurrence of new classes of copper nitrite reductases', *Journal of Biological Inorganic Chemistry*, 12(8), pp. 1119–1127. doi: 10.1007/s00775-007-0282-2.

Galloway, J. N. *et al.* (2003) 'The Nitrogen Cascade', *BioScience*. Oxford University Press (OUP), 53(4), p. 341. doi: 10.1641/0006-3568(2003)053[0341:tnc]2.0.co;2.

Galloway, J. N. *et al.* (2013) 'A Chronology of Human Understanding of the Nitrogen Cycle', *Philisophical Transactions of the Royal Society*, 368(1621). doi: 10.1098/rstb.2013.0120.

Galloway, J. N. and Cowling, E. B. (2002) 'Reactive nitrogen and the world: 200 Years of change', in *Ambio*. Royal Swedish Academy of Sciences, pp. 64–71. doi: 10.1579/0044-7447-31.2.64.

Gasteiger, E. *et al.* (2005) *Protein Analysis Tools on the ExPASy Server 571 571 From: The Proteomics Protocols Handbook Protein Identification and Analysis Tools on the ExPASy Server*.

Ghosh, S. *et al.* (2009) 'Spectroscopic and computational studies of nitrite reductase: Proton induced electron transfer and backbonding contributions to reactivity', *Journal of the American Chemical Society*. American Chemical Society, 131(1), pp. 277–288. doi: 10.1021/ja806873e.

Glockner, A. B., Jüngst, A. and Zumft, W. G. (1993) 'Copper-containing nitrite reductase from Pseudomonas aureofaciens is functional in a mutationally cytochrome cd1-free background (NirS-) of Pseudomonas stutzeri', *Archives of Microbiology*. Springer-Verlag, 160(1), pp. 18–26. doi: 10.1007/BF00258141.

Godden, J. *et al.* (1991) 'The 2.3 angstrom X-ray structure of nitrite reductase from Achromobacter cycloclastes', *Science*, 253(5018), pp. 438–442. doi: 10.1126/science.1862344.

Hagel, L. (2011) 'Gel Filtration: Size Exclusion Chromatography', in *Protein Purification: Principles, High Resolution Methods, and Applications: Third Edition*. Third. Wiley Blackwell, pp. 51–91. doi: 10.1002/9780470939932.ch3.

Hatzenpichler, R. (2012) 'Diversity, physiology, and niche differentiation of ammonia-oxidizing archaea', *Applied and Environmental Microbiology*. American Society for Microbiology (ASM), 78(21), pp. 7501–7510. doi: 10.1128/AEM.01960-12.

Hollocher, T. C., Tate, M. E. and Nicholas, D. J. (1981) 'Oxidation of ammonia by Nitrosomonas europaea. Definite 18O-tracer evidence that hydroxylamine formation involves a monooxygenase.', *Journal of Biological Chemistry*, 256(21), pp. 10834–10836.

Hooft, R. W. W., Sander, C. and Vriend, G. (1997) *Objectively judging the quality of a protein structure from a Ramachandran plot*, *CABIOS*.

Hooper, A. B. (1968) 'A nitrite-reducing enzyme from Nitrosomonas europaea Preliminary characterization with hydroxylamine as electron donor', *BBA - Bioenergetics*, 162(1), pp. 49–65. doi: 10.1016/0005-2728(68)90213-2.

Hooper, A. B. *et al.* (1997) 'Enzymology of the oxidation of ammonia to nitrite by bacteria', in *Antonie van Leeuwenhoek, International Journal of General and Molecular Microbiology*, pp. 59–67. doi: 10.1023/A:1000133919203.

Hooper, A. B. *et al.* (2004) 'The Oxidation of Ammonia as an Engergy Source in Bacteria', in.

Springer, Dordrecht, pp. 121–147. doi: 10.1007/978-1-4020-3163-2_6.

Horrell, S. *et al.* (2016) 'Serial crystallography captures enzyme catalysis in copper nitrite reductase at atomic resolution from one crystal', *IUCrJ*. International Union of Crystallography, 3(4), pp. 271–281. doi: 10.1107/S205225251600823X.

Horrell, S. *et al.* (2017) 'Recent structural insights into the function of copper nitrite reductases', *Metallomics*. Royal Society of Chemistry, pp. 1470–1482. doi: 10.1039/c7mt00146k.

Huerta-Cepas, J. *et al.* (2016) 'EGGNOG 4.5: A hierarchical orthology framework with improved functional annotations for eukaryotic, prokaryotic and viral sequences', *Nucleic Acids Research*. Oxford University Press, 44(D1), pp. D286–D293. doi: 10.1093/nar/gkv1248.

IPCC *et al.* (2007) *IPCC, 2007: Climate Change 2007 ; The Physical Basis. Contribution of Working Group I to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change*. Cambridge University Press, Cambridge, United Kingdom andNew York, NY, USA, 996pp.

Iverson, V. *et al.* (2012) 'Untangling genomes from metagenomes: Revealing an uncultured class of marine euryarchaeota', *Science*. American Association for the Advancement of Science, 335(6068), pp. 587–590. doi: 10.1126/science.1212665.

Jones, S. M. and Solomon, E. I. (2015) 'Electron transfer and reaction mechanism of laccases', *Cellular and Molecular Life Sciences*. Birkhauser Verlag AG, pp. 869–883. doi: 10.1007/s00018-014-1826-6.

Kakutani, T. *et al.* (1981) *A Blue Protein as an Inactivating Factor for Nitrite Reductase from Alcaligenes faecalis Strain S-6*, *J. Biochem*.

Kampschreur, M. J. *et al.* (2009) 'Nitrous oxide emission during wastewater treatment', *Water Research*. Elsevier Ltd, pp. 4093–4103. doi: 10.1016/j.watres.2009.03.001.

Karner, M. B., Delong, E. F. and Karl, D. M. (2001) 'Archaeal dominance in the mesopelagic zone of the Pacific Ocean', *Nature*, 409(6819), pp. 507–510. doi: 10.1038/35054051.

Kataoka, K. *et al.* (2000) 'Functional Analysis of Conserved Aspartate and Histidine Residues Located Around the Type 2 Copper Site of Copper-Containing Nitri Reductase', *Journal of Biochemistry*. Japanese Biochemical Society, 127(2), pp. 345–350. doi: 10.1093/oxfordjournals.jbchem.a022613.

Kelley, L. A. *et al.* (2015) 'The Phyre2 web portal for protein modeling, prediction and analysis', *Nature Protocols*. Nature Publishing Group, 10(6), pp. 845–858. doi: 10.1038/nprot.2015.053.

Kerou, M. *et al.* (2016) 'Proteomics and comparative genomics of Nitrososphaera viennensis reveal the core genome and adaptations of archaeal ammonia oxidizers', *Proceedings of the National Academy of Sciences of the United States of America*. National Academy of Sciences, 113(49), pp. E7937–E7946. doi: 10.1073/pnas.1601212113.

Kim, J. G. *et al.* (2016) 'Hydrogen peroxide detoxification is a key mechanism for growth of ammonia-oxidizing archaea', *Proceedings of the National Academy of Sciences of the United States of America*. National Academy of Sciences, 113(28), pp. 7888–7893. doi: 10.1073/pnas.1605501113.

Kobayashi, S. *et al.* (2018) 'Nitric oxide production from nitrite reduction and hydroxylamine oxidation by copper-containing dissimilatory nitrite reductase (Nirk) from the aerobic ammonia-oxidizing archaeon, nitrososphaera viennensis', *Microbes and Environments*. Japanese Society of Microbial Ecology, 33(4), pp. 428–434. doi: 10.1264/jsme2.ME18058.

Kondo, K., Yoshimatsu, K. and Fujiwara, T. (2012) 'Expression, and molecular and enzymatic characterization of Cu-containing nitrite reductase from a marine ammonia-oxidizing gammaproteobacterium, Nitrosococcus oceani', *Microbes and Environments*. Nakanishi Printing, 27(4), pp. 407–412. doi: 10.1264/jsme2.ME11310.

Könneke, M. *et al.* (2005) 'Isolation of an autotrophic ammonia-oxidizing marine archaeon', *Nature*. Nature Publishing Group, 437(7058), pp. 543–546. doi: 10.1038/nature03911.

Kosman, D. J. (2010) 'Multicopper oxidases: a workshop on copper coordination chemistry, electron transfer, and metallophysiology', *JBIC Journal of Biological Inorganic Chemistry*, 15(1), pp. 15–28. doi: 10.1007/s00775-009-0590-9.

Kowalchuk, G. A. and Stephen, J. R. (2001) 'Ammonia-Oxidizing Bacteria: A Model for Molecular Microbial Ecology', *Annual Review of Microbiology*. Annual Reviews, 55(1), pp. 485–529. doi: 10.1146/annurev.micro.55.1.485.

Kozlowski, J. A. *et al.* (2016) 'Pathways and key intermediates required for obligate aerobic ammonia-dependent chemolithotrophy in bacteria and Thaumarchaeota', *ISME Journal*. Nature Publishing Group, 10(8), pp. 1836–1845. doi: 10.1038/ismej.2016.2.

Kukimoto, M. *et al.* (1996) 'Studies on protein-protein interaction between copper-containing nitrite reductase and pseudoazurin from Alcaligenes faecalis S-6', *Journal of Biological Chemistry*. American Society for Biochemistry and Molecular Biology, 271(23), pp. 13680–13683. doi: 10.1074/jbc.271.23.13680.

Lawton, T. J. *et al.* (2013) 'Characterization of a nitrite reductase involved in nitrifier denitrification', *Journal of Biological Chemistry*. American Society for Biochemistry and Molecular Biology, 288(35), pp. 25575–25583. doi: 10.1074/jbc.M113.484543.

Lawton, T. J. *et al.* (2014) 'Structural conservation of the B subunit in the ammonia monooxygenase/particulate methane monooxygenase superfamily', *Proteins: Structure, Function and Bioinformatics*. John Wiley and Sons Inc., 82(9), pp. 2263–2267. doi: 10.1002/prot.24535.

Lawton, T. J. and Rosenzweig, A. C. (2011) 'Detection and characterization of a multicopper oxidase from nitrosomonas Europaea', in *Methods in Enzymology*. Academic Press Inc., pp. 423–433. doi: 10.1016/B978-0-12-386489-5.00017-8.

Lead, C.-O. *et al.* (2001) *TAR Climate Change 2001: The Scientific Basis. Chapter 4: Atmospheric Chemistry and Greenhouse Gases*.

Lüthy, R., Bowie, J. U. and Eisenberg, D. (1992) 'Assessment of protein models with three-dimensional profiles', *Nature*, 356(6364), pp. 83–85. doi: 10.1038/356083a0.

Majumdar, S. *et al.* (2014) 'Roles of Small Laccases from Streptomyces in Lignin Degradation'. doi: 10.1021/bi500285t.

Martens-Habbena, W. *et al.* (2009) 'Ammonia oxidation kinetics determine niche separation of nitrifying Archaea and Bacteria', *Nature*, 461(7266), pp. 976–979. doi: 10.1038/nature08465.

Martin, C. T. *et al.* (1981) *Reactions of Nitric Oxide with Tree and Fungal Laccaset*, *Biochemistry*.

Martins, L. O. *et al.* (2015) 'Laccases of prokaryotic origin: Enzymes at the interface of protein science and protein technology', *Cellular and Molecular Life Sciences*. Birkhauser Verlag AG, pp. 911–922. doi: 10.1007/s00018-014-1822-x.

Mulder, A. *et al.* (1995) 'Anaerobic ammonium oxidation discovered in a denitrifying fluidized bed reactor', *FEMS Microbiology Ecology*, 16(3), pp. 177–183. doi: 10.1016/0168-6496(94)00081-7.

Murata, K. and Wolf, M. (2018) 'Cryo-electron microscopy for structural analysis of dynamic biological macromolecules', *Biochimica et Biophysica Acta - General Subjects*. Elsevier B.V., pp. 324–334. doi: 10.1016/j.bbagen.2017.07.020.

Murphy, M. E. P., Lindley, P. E. and Adman, E. T. (1997) *Structural comparison of cupredoxin domains: Domain recycling to construct proteins with novel functions*, *Protein Science*. Cambridge University Press.

Musci, G. *et al.* (2004) 'Sequential reconstitution of copper sites in the multicopper oxidase CueO', *Article in JBIC Journal of Biological Inorganic Chemistry*. doi: 10.1007/s00775-003-0501-4.

Opperman, D. J. *et al.* (2019) 'A three-domain copper-nitrite reductase with a unique sensing loop', *IUCrJ*. International Union of Crystallography, 6(2), pp. 248–258. doi: 10.1107/S2052252519000241.

Osipov, E. *et al.* (2014) 'Effect of the L499M mutation of the ascomycetous Botrytis aclada laccase on redox potential and catalytic properties', *Acta Crystallographica Section D: Biological Crystallography*. International Union of Crystallography, 70(11), pp. 2913–2923. doi: 10.1107/S1399004714020380.

Otani, Y., Hasegawa, K. and Hanaki, K. (2004) *Comparison of aerobic denitrifying activity among three cultural species with various carbon sources*.

Plasencia, A. *et al.* (2013) 'Phylogenetic characterization and quantification of ammonia-oxidizing archaea and bacteria from Lake Kivu in a long-term microcosm incubation-oxidizing archaea and bacteria · ammonia monooxygenase alpha subunit (amoA) · Lake Kivu · microcosm · multi-color CARD-FISH', *InternatIonal MIcrobIology*, 16, pp. 177–189. doi: 10.2436/20.1501.01.192.

Qin, W. *et al.* (2017) 'Nitrosopumilus maritimus gen. nov., sp. nov., Nitrosopumilus cobalaminigenes sp. nov., Nitrosopumilus oxyclinae sp. nov., and Nitrosopumilus ureiphilus sp. nov., four marine ammoniaoxidizing archaea of the phylum thaumarchaeo', *International Journal of Systematic and Evolutionary Microbiology*. Microbiology Society, 67(12), pp. 5067–5079. doi: 10.1099/ijsem.0.002416.

Rees, A. P., Woodward, E. M. S. and Joint, I. (2006) 'Concentrations and uptake of nitrate and ammonium in the Atlantic Ocean between 60{ring operator} N and 50{ring operator} S', *Deep-Sea Research Part II: Topical Studies in Oceanography*, 53(14–16), pp. 1649–1665. doi: 10.1016/j.dsr2.2006.05.008.

Reiss, R. *et al.* (2013) 'Laccase versus Laccase-Like Multi-Copper Oxidase: A Comparative Study of Similar Enzymes with Diverse Substrate Spectra', *PLoS ONE*. Edited by C. M. Soares. Public Library of Science, 8(6), p. e65633. doi: 10.1371/journal.pone.0065633.

Rhodes, G. (2006) *Crystallography Made Crystal Clear*. Third, *Complementary Science* . Third. Academic Press. doi: 10.1016/B978-012587073-3/50012-X.

Rodionov, D. A. *et al.* (2005) 'Dissimilatory Metabolism of Nitrogen Oxides in Bacteria: Comparative Reconstruction of Transcriptional Networks', *PLoS Computational Biology*. Public Library of Science, 1(5), p. e55. doi: 10.1371/journal.pcbi.0010055.

Rotthauwe, J.-H., Witzel, K.-P. and Liesack, W. (1997) *The Ammonia Monooxygenase Structural*

Gene amoA as a Functional Marker: Molecular Fine-Scale Analysis of Natural Ammonia-Oxidizing Populations, *APPLIED AND ENVIRONMENTAL MICROBIOLOGY*.

Roy, A., Kucukural, A. and Zhang, Y. (2010) 'I-TASSER: A unified platform for automated protein structure and function prediction', *Nature Protocols*. Nat Protoc, 5(4), pp. 725–738. doi: 10.1038/nprot.2010.5.

Rütting, T. *et al.* (2011) 'Assessment of the importance of dissimilatory nitrate reduction to ammonium for the terrestrial nitrogen cycle', *Biogeosciences*, 8(7), pp. 1779–1791. doi: 10.5194/bg-8-1779-2011.

Selkirk, C. (2004) 'Ion-Exchange Chromatography', in *Protein Purification Protocols*. New Jersey: Humana Press, pp. 125–132. doi: 10.1385/1-59259-655-X:125.

Shraddha *et al.* (2011) 'Laccase: Microbial Sources, Production, Purification, and Potential Biotechnological Applications', *Enzyme Research*. Hindawi Limited, 2011, p. 11. doi: 10.4061/2011/217861.

Silva, C. S. *et al.* (2012) 'Crystal structure of the multicopper oxidase from the pathogenic bacterium Campylobacter jejuni CGUG11284: Characterization of a metallo-oxidase', *Metallomics*, 4(1), pp. 37–47. doi: 10.1039/c1mt00156f.

Singh, S. K. *et al.* (2004) 'Cuprous oxidase activity of CueO from Escherichia coli', *Journal of Bacteriology*. American Society for Microbiology (ASM), 186(22), pp. 7815–7817. doi: 10.1128/JB.186.22.7815-7817.2004.

Sirajuddin, S. and Rosenzweig, A. C. (2015) 'Enzymatic oxidation of methane', *Biochemistry*. American Chemical Society, 54(14), pp. 2283–2294. doi: 10.1021/acs.biochem.5b00198.

Skálová, T. *et al.* (2009) 'The Structure of the Small Laccase from Streptomyces coelicolor Reveals a Link between Laccases and Nitrite Reductases', *Journal of Molecular Biology*. Academic Press, 385(4), pp. 1165–1178. doi: 10.1016/j.jmb.2008.11.024.

Smyth, M. S. and Martin, J. H. J. (2000) 'X-Ray Crystallography', *Journal of Clinical Pathology - Molecular Pathology*. BMJ Publishing Group, pp. 8–14. doi: 10.1136/mp.53.1.8.

Soliman, M. and Eldyasti, A. (2018) 'Ammonia-Oxidizing Bacteria (AOB): opportunities and applications—a review', *Reviews in Environmental Science and Biotechnology*. Springer Netherlands, pp. 285–321. doi: 10.1007/s11157-018-9463-4.

Stoj, C. S. and Kosman, D. J. (2006) 'Copper Proteins: Oxidases', in *Encyclopedia of Inorganic Chemistry*. Chichester, UK: John Wiley & Sons, Ltd. doi: 10.1002/0470862106.ia055.

Strange, R. W. *et al.* (1999) 'Structural and kinetic evidence for an ordered mechanism of copper nitrite reductase', *Journal of Molecular Biology*. Academic Press, 287(5), pp. 1001–1009. doi: 10.1006/jmbi.1999.2648.

Szklarczyk, D. *et al.* (2019) 'STRING v11: Protein-protein association networks with increased coverage, supporting functional discovery in genome-wide experimental datasets', *Nucleic Acids Research*. Oxford University Press, 47(D1), pp. D607–D613. doi: 10.1093/nar/gky1131.

Tarasava, K., Loebus, J. and Freisinger, E. (2016) 'Localization and spectroscopic analysis of the Cu(I) binding site in wheat metallothionein Ec-1', *International Journal of Molecular Sciences*. MDPI AG, 17(3), p. 371. doi: 10.3390/ijms17030371.

Tolar, B. B. *et al.* (2017) 'Integrated structural biology and molecular ecology of N-cycling

enzymes from ammonia-oxidizing archaea', *Environmental Microbiology Reports*. Wiley-Blackwell, 9(5), pp. 484–491. doi: 10.1111/1758-2229.12567.

Torres, J. *et al.* (2002) 'Fast reduction of a copper center in laccase by nitric oxide and formation of a peroxide intermediate', *Journal of the American Chemical Society*. American Chemical Society , 124(6), pp. 963–967. doi: 10.1021/ja016107j.

Tourna, M. *et al.* (2008) 'Growth, activity and temperature responses of ammonia-oxidizing archaea and bacteria in soil microcosms', *Environmental Microbiology*, 10(5), pp. 1357–1364. doi: 10.1111/j.1462-2920.2007.01563.x.

Tourna, M. *et al.* (2011) 'Nitrososphaera viennensis, an ammonia oxidizing archaeon from soil.', *Proceedings of the National Academy of Sciences of the United States of America*. National Academy of Sciences, 108(20), pp. 8420–5. doi: 10.1073/pnas.1013488108.

Vajrala, N. *et al.* (2013) 'Hydroxylamine as an intermediate in ammonia oxidation by globally abundant marine archaea', *Proceedings of the National Academy of Sciences of the United States of America*, 110(3), pp. 1006–1011. doi: 10.1073/pnas.1214272110.

Walker, C. B. *et al.* (2010) 'Nitrosopumilus maritimus genome reveals unique mechanisms for nitrification and autotrophy in globally distributed marine crenarchaea', *Proceedings of the National Academy of Sciences of the United States of America*, 107(19), pp. 8818–8823. doi: 10.1073/pnas.0913533107.

Ward, B. B. (2011) 'Measurement and distribution of nitrification rates in the oceans', in *Methods in Enzymology*. Academic Press Inc., pp. 307–323. doi: 10.1016/B978-0-12-381294-0.00013-4.

Waterhouse, A. *et al.* (2018) 'SWISS-MODEL: homology modelling of protein structures and complexes', *Nucleic Acids Research*, 46. doi: 10.1093/nar/gky427.

Williams, C. J. *et al.* (2018) 'MolProbity: More and better reference data for improved all-atom structure validation', *Protein Science*. Blackwell Publishing Ltd, 27(1), pp. 293–315. doi: 10.1002/pro.3330.

Wunderlin, P. *et al.* (2012) 'Mechanisms of N2O production in biological wastewater treatment under nitrifying and denitrifying conditions', *Water Research*. Elsevier Ltd, 46(4), pp. 1027–1037. doi: 10.1016/j.watres.2011.11.080.

Yamaguchi, K. *et al.* (2004) 'Characterization of two type 1 Cu sites of Hyphomicrobium denitrificans nitrite reductase: A new class of copper-containing nitrite reductases', *Biochemistry*, 43(44), pp. 14180–14188. doi: 10.1021/bi0492657.

Yang, J. *et al.* (2015) 'The I-TASSER Suite: protein structure and function prediction', *Nature Methods*. Nature Publishing Group, 12(1), pp. 7–8. doi: 10.1038/nmeth.3213.

Zajicek, R. S. *et al.* (2009) 'D1 haem biogenesis - Assessing the roles of three nir gene products', *FEBS Journal*, 276(21), pp. 6399–6411. doi: 10.1111/j.1742-4658.2009.07354.x.

Zhang, L. *et al.* (2020) 'Engineering of Laccase CueO for Improved Electron Transfer in Bioelectrocatalysis by Semi-Rational Design', *Chemistry – A European Journal*. Wiley-VCH Verlag, 26(22), pp. 4974–4979. doi: 10.1002/chem.201905598.

Zhang, Y. (2008) 'I-TASSER server for protein 3D structure prediction', *BMC Bioinformatics*. BioMed Central, 9(1), pp. 1–8. doi: 10.1186/1471-2105-9-40.

Zhang, Y. and Skolnick, J. (2004) 'Scoring function for automated assessment of protein structure template quality', *Proteins: Structure, Function and Genetics*, 57(4), pp. 702–710. doi: 10.1002/prot.20264.

Zhang, Y. and Skolnick, J. (2005) 'TM-align: a protein structure alignment algorithm based on the TM-score', *Nucleic Acid Research* , 33(7), pp. 2303–2309.

Zumft, W. G. *et al.* (1988) 'Defects in cytochrome cd1-dependent nitrite respiration of transposon Tn 5-induced mutants from Pseudomonas stutzeri', *Archives of Microbiology*. Springer-Verlag, 149(6), pp. 492–498. doi: 10.1007/BF00446750.