The role of the adaptor protein CbbO in the Rubisco activation system of chemoautotrophic bacteria

Liu, Di

2019

Liu, D. (2019). The role of the adaptor protein CbbO in the Rubisco activation system of chemoautotrophic bacteria. Doctoral thesis, Nanyang Technological University, Singapore.

<https://hdl.handle.net/10356/105488>

<https://doi.org/10.32657/10220/48727>

**NANYANG
TECHNOLOGICAL
UNIVERSITY**

SINGAPORE

**THE ROLE OF THE ADAPTOR PROTEIN CBBO
IN THE RUBISCO ACTIVATION SYSTEM OF
CHEMOAUTOTROPHIC BACTERIA**

LIU DI

SCHOOL OF BIOLOGICAL SCIENCES

2019

**THE ROLE OF THE ADAPTOR PROTEIN CBBO
IN THE RUBISCO ACTIVATION SYSTEM OF
CHEMOAUTOTROPHIC BACTERIA**

LIU DI

School of Biological Sciences

A thesis submitted to the Nanyang Technological
University in partial fulfilment of the requirement for the
degree of Doctor of Philosophy

2019

Statement of Originality

I hereby certify that the work embodied in this thesis is the result of original research done by me except where otherwise stated in this thesis. The thesis work has not been submitted for a degree or professional qualification to any other university or institution. I declare that this thesis is written by myself and is free of plagiarism and of sufficient grammatical clarity to be examined. I confirm that the investigations were conducted in accord with the ethics policies and integrity standards of Nanyang Technological University and that the research data are presented honestly and without prejudice.

24/05/2019

.....
Date

.....
Liu Di

Supervisor Declaration Statement

I have reviewed the content and presentation style of this thesis and declare it is free of plagiarism and of sufficient grammatical clarity to be examined. To the best of my knowledge, the research and writing are those of the candidate except as acknowledged in the Author Attribution Statement. I confirm that the investigations were conducted in accord with the ethics policies and integrity standards of Nanyang Technological University and that the research data are presented honestly and without prejudice.

24 January 2019

.....
Date

.....
Assoc. Prof. Oliver Mueller-Cajar

Authorship Attribution Statement

This thesis contains material from 1 paper(s) published in the following peer-reviewed journal(s) where I was the co-first author.

One figure was adapted from:

Liu, D., Ramya, R.C.S., Mueller-Cajar, O., 2017. Surveying the expanding prokaryotic Rubisco multiverse. FEMS Microbiology Letters 364, fnx156-fnx156.

The contributions of the co-authors are as follows:

- Assoc. Prof. Oliver Mueller-Cajar provided the initial project direction and edited the manuscript drafts.
- I prepared the manuscript drafts together with co-author Ramya RCS.
- I prepared the figures for the paper together with co-author Ramya RCS
- The manuscript was revised together with Ramya RCS and Assoc. Prof. Oliver Mueller-Cajar.

.....
Liu Di

24 January 2019

.....
Assoc. Prof. Oliver Mueller-Cajar

24 January 2019

ACKNOWLEDGEMENTS

I am grateful to my supervisor, Assoc. Prof. Oliver Mueller-Cajar whose expertise, guidance and support made it possible for me to learn and make progress in my research. His enthusiasm and passion for science is an inspiration to all of us. I am lucky to be a part of his team.

I extend my heartfelt appreciation for my thesis advisory committee; Prof. Gerhard Grüber and Asst. Prof. Wu Bin for their constructive input and discussions. I am also thankful for my collaborators; Prof. Dr. Florian Stengel of the University of Konstanz for MS analysis of our cross-linked samples as well as Asst. Prof. Shashi Bhushan for EM support.

It is a privilege to work with my mentor, Candace whose technical expertise and experience has contributed much to this project. Her fun-loving nature ensures that there is rarely a dull moment in the lab. I have learnt a lot from her, not only on how to do research but also to enjoy my work as a researcher.

I want to express my gratitude towards all members of OMC lab; Tobi, Nitin, Dev, Ramya, Zhi Jun, Zhen Guo, Jed, Jian Ann, Warren and Lynette for making the lab a happy place. Doing science would be a lot less enjoyable without these people and I will cherish the times we had. I owe special thanks to Ankita, Arvind and Satabdi, without whom, timely completion of this thesis would not have been possible. My students, Gina and Wei Jie, also played a part in helping with various aspects of this work.

Finally, I am grateful to my family and friends for their unconditional love and unwavering support throughout this PhD journey.

TABLE OF CONTENT

ACKNOWLEDGEMENTS.....	i
LIST OF FIGURES.....	viii
LIST OF TABLES.....	xi
ABBREVIATIONS.....	xii
ABSTRACT.....	xv
1 INTRODUCTION.....	1
1.1 Microbial Carbon Fixation.....	1
1.2 The Calvin-Benson-Bassham (CBB) Cycle.....	2
1.3 Ribulose 1,5-bisphosphate carboxylase/oxygenase (Rubisco).....	4
1.4 Structural diversity of Rubisco.....	5
1.5 Rubisco catalytic mechanism and regulation.....	9
1.6 Rubisco Activase (Rca).....	13
1.7 The CbbQO Rca system of chemoautotrophic bacteria.....	19
1.7.1 The AAA+ ATPase CbbQ.....	21
1.7.2 The adaptor protein CbbO.....	25
1.7.3 Putative mechanism of Rubisco activation by AfCbbQO.....	27
1.8 Aims and Scope of the Project.....	28
2 MATERIALS AND METHODS.....	31
2.1 Chemicals.....	31
2.2 Media.....	34
2.3 Molecular Biology.....	34
2.3.1 Polymerase Chain Reaction (PCR).....	34
2.3.2 pGemT Vector Cloning.....	35
2.3.3 General cloning strategy.....	36
2.3.4 Site-directed mutagenesis.....	36
2.4 Protein biochemistry.....	37
2.4.1 SDS and Native PAGE.....	37
2.4.2 Coomassie blue staining of SDS and native PAGE.....	38
2.4.3 Western Blotting and Immunodetection.....	38
2.4.4 Protein expression and purification.....	39
2.4.5 Protein expression test for FLAG-AfO2VWA.....	40
2.4.6 Purification of FLAG-AfO2VWA from the inclusion bodies.....	41

2.4.7	Refolding of FLAG-AfO2VWA.....	42
2.4.8	Circular dichroism	42
2.4.9	Analytical size-exclusion chromatography	42
2.5	Enzymatic assays	43
2.5.1	Formation of ECM/ECMC/ER.....	43
2.5.2	ATPase Assay	43
2.5.3	Rubisco Assay	43
2.6	Chemical cross-linking	44
2.6.1	Glutaraldehyde cross-linking.....	44
2.6.2	Dithiobis(succinimidyl propionate) (DSP) crosslinking	44
2.6.3	Cross-linking coupled to mass spectrometry (XL-MS)	45
2.7	Negative stain EM	45
2.7.1	Sample preparation	45
2.7.2	Grid preparation and data collection.....	45
2.7.3	Data processing.....	46
2.8	Cryo-EM	46
2.8.1	Sample preparation	46
2.8.2	Grid preparation and data collection.....	47
2.8.3	Data processing.....	47
2.8.4	Model building for AfLS.....	49
2.9	Bioinformatics	49
2.9.1	Bioinformatic characterization of CbbQ orthologs	49
2.9.2	Bioinformatic analysis of CbbO sequences.....	50
3	ANALYSIS OF THE HETERO-OLIGOMERIZATION OF THE	
	AAA+ ATPASE CBBQ AND THE ADAPTOR PROTEIN CBBO	51
3.1	Purification of the form II Rubisco activase Q2O2	51
3.2	Negative stain EM analysis of the form II Rubisco activase Q2O2	53
3.3	Chemical cross-linking coupled to mass spectrometry (XL-MS) analysis of CbbQO activase	56
3.4	Bioinformatic characterization of CbbQ orthologs	60
3.5	Bioinformatical analysis of CbbO orthologs.....	62
3.6	Biochemical analysis of progressive C-terminal truncations of CbbO from the closely related homologue from <i>H. neapolitanus</i> (Hn).....	68

3.6.1	Purification and analysis of C-terminal truncation mutants of HnCbbO	69
3.6.2	Functional analysis of C-terminal truncation mutants of HnCbbO	72
3.7	Discussion	74
4	INSIGHTS INTO THE CBBQO RUBISCO ACTIVATION MECHANISM FOR FORM II RUBISCO	78
4.1	Native PAGE analysis of the Rubisco-Rubisco activase complex formation	78
4.2	The Walker B mutant Q2O2 (FLAG-Q2 ^{WB} O2) forms complexes with stably inhibited form II Rubisco (HA-AfM ECMC)	79
4.2.1	Q2 ^{WB} O2 preferentially interacts with the inhibited AfM ECMC over the active ECM	82
4.2.2	The surface exposed Glu75 on the Rubisco large subunit (RbcL) is essential for facilitating interactions with Q2 ^{WB} O2	83
4.2.3	An intact RbcL C-tail is required for a high affinity interaction with Q2 ^{WB} O2	84
4.2.4	The adaptor protein CbbO is required to form complexes with the stably inhibited form II Rubisco, AfM ECMC	85
4.2.5	Q2O2 MIDAS motif mutants are able to form complexes with the stably inhibited form II Rubisco, AfM ECMC	86
4.2.6	The soluble chimeric CbbO mutant O1O2VWA interacts specifically with AfM	89
4.2.7	Test expression and purification of O2VWA domain from the inclusion bodies	92
4.2.8	The re-folded FLAG-O2VWA interacts specifically with AfM	96
4.3	Characterization of the AfM-Q2 ^{WB} O2 complex with analytical size exclusion chromatography (SEC)	98
4.4	Negative stain EM analysis of HA-ECMC-FLAG-Q2 ^{WB} O2 complexes	101
4.5	Cross-linking coupled to mass spectrometry analysis (XL-MS) of the HA-AfM-FLAG-Q2 ^{WB} O2 complex	106
4.6	DISCUSSION	109
5	INSIGHTS INTO THE CBBQO RUBISCO ACTIVATION MECHANISM FOR FORM I RUBISCO	113
5.1	Purification of the chimeric Rubisco activase Q3O1	113

5.2	The chimeric activase Q3O1 is a functional Rca	115
5.2.1	The functional chimeric activase Q3O1 forms complexes with stably inhibited form I Rubisco, AfLS ECMC.....	116
5.2.2	Conformational changes induced by ATP hydrolysis are required for interaction between AfLS and Q3O1	119
5.2.3	Q3O1 is more suitable for structural characterization than the WT Q1O1.....	120
5.2.4	An intact RbcL C-tail is required for a high affinity interaction with Q3O1	121
5.2.5	The surface exposed Asp82 on Rubisco large subunit (RbcL) is essential for facilitating interactions with Q3O1.....	123
5.2.6	The adaptor protein CbbO is required to form complexes with the stably inhibited form I Rubisco, AfLS ECMC.....	124
5.3	Characterization of the AfLS-Q3O1 complex with analytical size exclusion chromatography (SEC)	125
5.4	Isolation of AfLS ECMC-Q3O1 complex for negative-stain EM.....	128
5.5	Negative-stain EM analysis of isolated AfLS ECMC-Q3O1 complex	129
5.6	Discussion	133
6	CRYO-ELECTRON MICROSCOPY STUDY OF THE FORM I RUBISCO-ACTIVASE COMPLEX	136
6.1	Isolation of AfLS ECMC-Q3O1 complex for cryo-EM and screening.....	136
6.2	Data collection on the FEI Titan Krios microscope (Krios)	138
6.3	Particle picking and 2D classification.....	139
6.4	3D refinement and classification	140
6.5	Focused refinement of Q3O1	142
6.6	Cryo-EM structure of AfLS	145
6.7	Modelling of AfLS.....	148
6.8	Discussion	152
7	CONCLUSION AND OUTLOOK.....	156
8	APPENDIX.....	158
8.1	Cross-linking coupled to mass spectrometry (XL-MS) results	158
8.2	Sequence alignment of 50 CbbQ orthologs.....	203
8.3	Sequence alignment of 50 CbbO orthologs.....	209

9	REFERENCES	223
----------	-------------------------	------------

LIST OF FIGURES

Figure 1-1: The Calvin-Benson-Bassham (CBB) cycle.....	3
Figure 1-2: Structure of Rubisco.....	6
Figure 1-3: Rubisco reaction mechanism.....	11
Figure 1-4: Regulation of Rubisco activity and inhibition by sugar phosphates.....	13
Figure 1-5: The red-type prokaryotic <i>R. sphaeroides</i> Rca.	16
Figure 1-6: The green-type eukaryotic <i>N. tabacum</i> Rca.	18
Figure 1-7: Arrangement of the CO ₂ fixation operons encoding form I and form II Rubisco in <i>A. ferrooxidans</i>	21
Figure 1-8: The CbbQO-type Rca.....	23
Figure 1-9: Structure of the Integrin α L I domain in complex with domain 1 of Intercellular adhesion molecule-3.....	26
Figure 3-1: Purification of Q2O2.....	52
Figure 3-2: Negative-staining EM image analysis of Q2O2.	55
Figure 3-3:Cross-linking coupled to mass spectrometry analysis (XL-MS).	58
Figure 3-4: Analysis of CbbQ2 structure.	61
Figure 3-5: Phylogenetic tree of CbbO isoforms from chemoautotrophic bacteria....	63
Figure 3-6: ClustalW alignment of CbbO sequences.....	67
Figure 3-7: Schematic representation of HnCbbO truncation mutants.	69
Figure 3-8: Analysis of purified HnCbbO truncation mutants.	70
Figure 3-9: Purified HnCbbO Mutants.....	71
Figure 3-10: Analytical size exclusion chromatography analysis of. hetero-oligomerization between HnCbbQ and HnCbbO Mutants.....	73
Figure 3-11: Activity assay of HnCbbO mutants.....	74
Figure 4-1: Cross-linking of HA-AfM and FLAG-Q2 ^{WB} O2 with glutaraldehyde.....	81
Figure 4-2: Cross-linking of HA-AfM ECM and FLAG-Q2 ^{WB} O2 with glutaraldehyde.....	83
Figure 4-3: Cross-linking of AfM ^{E75A} ECMC and FLAG-Q2 ^{WB} O2 with glutaraldehyde.....	84
Figure 4-4: Cross-linking of AfM ^{ΔC1} ECMC and FLAG-Q2 ^{WB} O2 with glutaraldehyde.....	85
Figure 4-5: Cross-linking of HA-AfM ECMC and FLAG-Q2 ^{WB} with glutaraldehyde.....	86

Figure 4-6: Cross-linking of HA-AfM ECMC and Q2O2 ^{MIDAS} mutants with glutaraldehyde.....	88
Figure 4-7: Purification of O1O2VWA.....	90
Figure 4-8: Cross-linking of O1O2VWA with AfM and AfLS.....	91
Figure 4-9: Protein expression test for His ₆ -ubiquitin-AfO2VWA.	93
Figure 4-10: Purification of FLAG-O2VWA from the inclusion bodies.....	95
Figure 4-11: Cross-linking of FLAG-O2VWA with AfM and AfLS.	97
Figure 4-12: Analytical size exclusion chromatography analysis of interaction between HA-AfM ECMC and FLAG-Q2 ^{WB} O2.	100
Figure 4-13: Analytical size exclusion chromatography analysis of interaction between activated Rubisco (HA-AfM ECM) and AfM ^{E75A} ECMC with FLAG-Q2 ^{WB} O2.	101
Figure 4-14: Isolation and visualization of the HA-AfM ECMC-FLAG-Q2 ^{WB} O2 complex.	103
Figure 4-15: Isolation and visualization of the HA-AfM ECMC-FLAG-Q2 ^{WB} O2 complex cross-linked with glutaraldehyde.	106
Figure 4-16: Cross-linking coupled to mass spectrometry analysis (XL-MS).	108
Figure 5-1: Purification of Q3O1.....	114
Figure 5-2: Q3O1 activity assays.....	116
Figure 5-3: Cross-linking of AfLS and Q3O1 with glutaraldehyde.	119
Figure 5-4: Cross-linking of AfLS and Q3 ^{WB} O1 with glutaraldehyde.	120
Figure 5-5: Cross-linking of AfLS and Q1O1 with glutaraldehyde.	121
Figure 5-6: Cross-linking of AfLS Δ C1 and Q3O1 with glutaraldehyde.....	123
Figure 5-7: Cross-linking of AfLS ^{D82P} and Q3O1 with glutaraldehyde.....	124
Figure 5-8: Cross-linking of AfLS and Q3 with glutaraldehyde.	125
Figure 5-9: Analytical SEC analysis of interaction between AfLS ECMC and Q3O1.....	127
Figure 5-10: Isolation of the AfLS-Q3O1 complex with analytical SEC.	129
Figure 5-11: Negative stain EM image analysis of the AfLS-Q3O1 complex.....	131
Figure 6-1: Isolation of Rubisco-activase complexes and screening of sample for cryo-EM data collection.	137
Figure 6-2: Micrograph evaluation.	138

Figure 6-3: 2D classification of AfLS ECMC-Q3O1 complexes derived from RELION-3.	140
Figure 6-4: 3D classification of AfLS ECMC-Q3O1 complexes.	142
Figure 6-5: Independent refinement and 3D classification and of the chimeric activase Q3O1.	144
Figure 6-6: 2D classification of AfLS ECMC derived from RELION-3.	146
Figure 6-7: Cryo-EM and 3D reconstruction of AfLS at 2.9 Å.	148
Figure 6-8: Analysis of the AfLS model.	150
Figure 6-9: Structural alignment of AfLS to the closely related homologue HnLS.	152
Figure 6-10: Proposed mechanistic model for the activation of AfLS by CbbQO.	154

LIST OF TABLES

Table 2-1: General Chemicals	32
Table 2-2: General buffer	33
Table 2-3: LB Media	34
Table 2-4: PCR reaction condition.....	34
Table 2-5: List of primers used in this study	35
Table 2-6: Components of SDS and Native gels.	38
Table 6-1: Atomic model validation statistics for one subunit of AfLS.	149
Table 8-1: AfQ2 and AfO2 inter-links	166
Table 8-2: AfQ2 to AfO2 intra-links.....	171
Table 8-3: AfQ1 and AfO1 inter-links	174
Table 8-4: AfQ1 to AfO1 intra-links.....	178
Table 8-5: FLAGQ2 ^{WB} and AfO2 inter-links with FLAG-Q2 ^{WB} O2 to HA-AfM interlinks	184
Table 8-6: FLAGQ2 ^{WB} and AfO2 intra-links with FLAG-Q2 ^{WB} O2 to HA-AfM intra-links.....	202

ABBREVIATIONS

2PG	2-Phosphoglycolate
3PG	3-phosphoglycerate
AAA+	ATPase associated with various cellular activities
ADP	Adenosine diphosphate
Af	<i>Acidithiobacillus ferrooxidans</i>
AfLS	<i>Acidithiobacillus ferrooxidans</i> form I Rubisco
AfM	<i>Acidithiobacillus ferrooxidans</i> form II Rubisco
ATP	Adenosine Triphosphate
CA1P	2-Carboxy-arabinitol 1-phosphate
CABP	2-Carboxy-arabinitol-1,5-bisphosphate
CAM	Crassulacean acid metabolism
CBB	Calvin-Benson-Bassham
CCM	Carbon-concentrating mechanism
CD	Circular dichroism
CO ₂	Carbon dioxide
CPR	Candidate phyla radiation
DMF	Dimethylformamide
DSS	Disuccinimidyl suberate
DTT	Dithiothreitol
<i>E. coli</i>	<i>Escherichia coli</i>
ECL	Enhanced chemiluminescence
ECM	Activated Rubisco complex
ECMC	CABP-inhibited Rubisco (carbamylated)
ECMI	Inhibited Rubisco complex (carbamylated)
EDTA	Ethylenediaminetetraacetic acid
EM	Electron Microscopy
ER	RuBP-inhibited Rubisco
FSC	Fourier shell correlation
G3P	Glyceraldehyde-3-phosphate
GA	Glutaraldehyde
HA tag	Hemagglutinin tag

HCL	Hydrochloric acid
HEPES	N-(2-hydroxyethyl)piperacin-N'-2-ethanesulfonic acid
His ₆ Ub	Hexa-histidine tagged ubiquitin
HMW	High molecular weight
Hn	<i>Halothiobacillus neapolitanus</i>
HRP	Horseradish peroxidase
IEX	Ion-exchange chromatography
IMAC	Immobilized metal ion affinity chromatography
IPTG	Isopropyl- beta-D-thiogalactopyranoside
ITC	Isothermal titration calorimetry
LB	Lysogeny broth
MbR	<i>Methanococcoides burtonii</i> Rubisco
Mg ²⁺	Magnesium ion
MIDAS	Metal ion-dependent adhesion site
NADPH	Nicotinamide adenine dinucleotide phosphate
NTA	Nitrilotriacetic acid
O ₂	Oxygen
OD	Optical Density
PAGE	Polyacrylamide gel electrophoresis
PDB	Protein Data Bank
PDBP	D-glycero-2,3-pentodiulose-1,5-bisphosphate
PEP	Phosphoenolpyruvate
PMSF	Phenylmethylsulfonylfluoride
RAD	Rubisco assembly domain
RbcL	Rubisco large subunit
RbcS	Rubisco small subunit
Rca	Rubisco activase (form I Rubisco- green-type)
RLP	Rubisco-like proteins
RpM	<i>Rhodopseudomonas palustris</i> form II Rubisco
RrM	<i>Rhodospirillum rubrum</i> form II Rubisco
RT	Room temperature
Rubisco	Ribulose 1,5- bisphosphate carboxylase/oxygenase
RuBP	Ribulose-1, 5-bisphosphate

SDS	Sodium dodecyl sulfate
SEC	Size exclusion chromatography
TBS	Tris-buffered saline
TBST	Tris-buffered saline, 0.1% Tween 20
TEMED	Tetramethylethylenediamine
TIM-barrel	Triose isomerase-like α - β barrel protein fold
Tris	Tris(hydroxymethyl)aminomethane
VWA	von Willebrand factor type A
WB	Walker B
XL-MS	Cross-linking coupled to mass spectrometry
XUBP	Xylulose 1,5- biphosphate

ABSTRACT

Ribulose 1, 5-bisphosphate carboxylase/oxygenase (Rubisco) is responsible for almost all biological CO₂ assimilation, but has a tendency to form dead-end inhibition complexes with sugar phosphates, including its own substrate ribulose 1,5-bisphosphate (RuBP). AAA+ molecular chaperones Rubisco activases (Rca) remodel inhibited Rubisco to effectively release the inhibitors. Recently, a new class of Rca from the chemolithoautotrophic bacteria *Acidithiobacillus ferrooxidans* was characterized. This new Rubisco activation system is unique in requiring the participation of both the AAA+ motor CbbQ, and a VWA (von Willebrand factor type A) domain-containing CbbO as an adaptor to be functional. The CbbO protein is thus far poorly characterized and its role in Rubisco activation is unclear. In this work, biochemical conditions that allow the isolation of a stabilized Rubisco-CbbQO complex through chemical cross-linking were established. Cryo-electron microscopy revealed that, unlike other activases studied so far, the convex surface of the CbbQ hexamer is not involved in Rubisco remodeling. Instead, mechanical force for disruption of the active site is likely transmitted to Rubisco via the flexible adaptor CbbO featured on the concave face of CbbQ. Interaction of CbbO with the C-terminus of the Rubisco large subunit, which governs the accessibility of the active site would trigger a sequence of conformational changes that would ultimately result in liberation of the inhibitor.

1. INTRODUCTION

1.1 Microbial Carbon Fixation

Autotrophic carbon fixation sustains life on earth by assimilating inorganic carbon dioxide into organic biomass. This process allows an otherwise unavailable form of carbon to become accessible to other organisms, making it an integral part of ecosystems. Microorganisms contribute a significant fraction of reduced carbon on earth, utilizing both light and inorganic substrates as a source of energy to drive synthesis of organic compounds (Field et al., 1998; Badger and Bek, 2008; Reinfelder, 2011). Photoautotrophic microbes such as cyanobacteria and algae harness energy from the photons of sunlight and transform it into chemical energy in a process known as photosynthesis.

Photosynthesis encompasses two reaction stages: light-dependent and light-independent reactions. The light-dependent reactions produce the reducing equivalents nicotinamide adenine dinucleotide phosphate (NADPH) and adenosine triphosphate (ATP), which are needed for carbon fixation via the reductive pentose phosphate pathway, more commonly referred to as the Calvin-Benson-Bassham (CBB) cycle (Nelson, 2005). This cycle encompasses the light-independent reaction of photosynthesis and involves the synthesis of organic molecules from the fixation of carbon dioxide. Oxygenic photosynthesis, commonly associated with terrestrial plants removes carbon dioxide and reintroduces oxygen into the atmosphere. It sustains the oxidizing environment and is responsible for the majority of primary productivity on earth.

Chemoautotrophic microorganisms such as certain proteobacteria and archaea are not dependent on solar energy but gain chemical energy from oxidation of inorganic material such as reduced sulfur compounds and iron containing minerals. These organisms utilize the same CBB cycle in what is known as dark CO₂ fixation to drive primary production in vastly different and sometimes extreme habitats (Shively et al., 1998).

Although other carbon fixation pathways have been identified in some bacteria and archaea, including the reductive tricarboxylic acid (TCA) cycle, reductive acetyl-CoA

pathway and the hydroxypropionate pathway, the CBB cycle is the most predominant form of organic matter production used by most microorganisms and all plants (Berg, 2011; Fuchs, 2011). Consequently, 90% of the carbon fixed on earth has been ascribed to the CBB cycle, making it the most biologically significant pathway.

1.2 The Calvin-Benson-Bassham (CBB) Cycle

In the 1950s, Melvin Calvin, along with colleagues Andrew Benson and James Bassham elucidated the metabolic pathway through which CO₂ was assimilated into the organic molecules of the cells (Bassham et al., 1950). The CBB cycle consists of a series of enzymatic reactions and can be divided into three phases: carbon fixation, reduction and regeneration (Figure 1-1). During carbon fixation, the enzyme ribulose-1,5-bisphosphate carboxylase/oxygenase (Rubisco) catalyzes the addition of CO₂ to the acceptor molecule ribulose 1,5-bisphosphate (RuBP) to form 2 molecules of 3-phosphoglycerate (3PG). In the reduction phase, 3PG is reduced to glyceraldehyde 3-phosphate (G3P) by consumption of ATP and NADPH. While some G3P is shunted out of the CBB cycle to produce sugars, amino acids and fatty acids, most of it is recycled back into the pathway to regenerate the substrate RuBP, known as the regeneration phase. The net result of the activity of the CBB cycle is the formation of one molecule of G3P from three molecules of CO₂ at the expense of nine molecules of ATP and six of NADPH.

Figure 1-1: The Calvin-Benson-Bassham (CBB) cycle.

In many organisms, the CBB cycle is an essential process in the generation of organic molecules for metabolic processes. In addition to sustaining autotrophic growth, it has also been shown to have an alternative role during heterotrophic growth. In some non-photosynthetic bacteria the CBB cycle acts as an electron sink in recycling excess reduced cofactors. This ensures a sufficient amount of oxidized electron carriers to balance the redox state of the cell and maintain metabolic flux (McKinlay and Harwood, 2010).

With the advent of challenges such as global warming, rising CO₂ levels and food shortages, research in the field of carbon assimilation has been motivated by the prospect of improving photosynthetic efficiency, transforming greenhouse gases and

production of sustainable energy (Zhu et al., 2010). Due to the importance of the CBB cycle, the key enzyme Rubisco, responsible for the actual fixation of CO₂ has been the subject of intensive study.

1.3 Ribulose 1,5-bisphosphate carboxylase/oxygenase (Rubisco)

Rubisco can be encountered in most autotrophic organisms from all domains of life. (Andersson and Backlund, 2008). It catalyzes the first and often rate-limiting carbon fixation reaction in the CBB cycle, covalently attaching CO₂ to the five-carbon sugar RuBP which culminates in the production of two 3-PG molecules. Despite its role as the principal carboxylation enzyme, Rubisco is not primed for optimal performance and possesses a characteristically low catalytic rate of only ~3-10 CO₂ molecules per second (Hartman and Harpel, 1994). Not only does Rubisco have a low affinity for its substrate CO₂, it also has a tendency to confuse the carboxylation substrate CO₂ with O₂. This results in a competing oxygenase reaction that produces one molecule of 3-PG and one molecule of 2-phosphoglycolate (2-PG) (Lorimer et al., 1973). In plants, 2-PG is recycled back to 3-PG via photorespiration, an energy-consuming pathway that liberates previously fixed carbon as CO₂ (Bauwe, 2001; Andersson and Backlund, 2008). The promiscuity of Rubisco adversely affects the photosynthetic efficiency and reduces the growth rate of plants. Such deficiencies necessitate the investment of large amount of resources into expression of Rubisco, making it the most abundant protein on earth (Ellis, 1979).

Considering the energetic cost of increased biomass production, it is therefore not surprising that a myriad of organisms have evolved carbon-concentrating mechanisms (CCM) to increase the local concentration of CO₂ around the Rubisco active site. This strategy not only subverts the problem of Rubisco's high Michaelis-Menten constant for CO₂ and also avoids carbon loss due to the non-productive oxygenation of RuBP.

Specialized microcompartments composed of a polyhedral protein shell known as carboxysomes are found in all cyanobacteria and certain chemoautotrophic bacteria that accumulate inorganic bicarbonate via dedicated pumps (Moroney and Somanchi, 1999; Badger, 2003; Yeates et al., 2008; Rae et al., 2013). Carbonic anhydrases located within these compartments convert bicarbonate back to CO₂, increasing the substrate

concentration around the encapsulated Rubisco to minimize the proportion of unproductive oxygenation reactions.

The majority of unicellular algae and eukaryotic phytoplankton possess subcellular organelles known as the pyrenoid, localized in the chloroplast stroma (Meyer et al., 2012; Meyer et al., 2017). Unlike carboxysomes that are surrounded by a proteinaceous shell, pyrenoids are sometimes enclosed within starch sheaths. In the model algae *Chlamydomonas reinhardtii*, bicarbonate enters the pyrenoid matrix via thylakoid lumen tubules where carbonic anhydrases are present to convert the bicarbonate to CO₂ (Engel et al., 2015). Recent work has revealed that the pyrenoid matrix is densely-packed with Rubisco and has liquid-like properties that allow dynamic regulation during cell division (Freeman Rosenzweig et al., 2017; Wunder et al., 2018).

Land plants also employ carbon concentrating strategies known as C₄ and CAM (Crassulacean Acid Metabolism) photosynthesis which utilize spatial and temporal separation of carbon fixation from other CBB cycle reactions, respectively (Dodd et al., 2002; Sage, 2004). In both, CO₂ is carboxylated to form 4-carbon compounds such as oxaloacetate and malate in a pre-fixation step before final fixation by Rubisco.

1.4 Structural diversity of Rubisco

Despite low sequence similarity, the overall fold of the Rubisco large subunit (RbcL) monomer is well-conserved in all forms of Rubisco (Figure 1-2A). It is made up of a N-terminal α + β domain and a larger C-terminal β 8 α 8 triosephosphate isomerase (TIM)-barrel domain (Hauser et al., 2015). The active site is at the intra-dimer (RbcL₂) interface between the N-domain of one subunit and the C-domain of the adjacent subunit (Figure 1-2B). Consequently, the functional unit of Rubisco is an L₂ dimer with two active sites per dimer (Whitney et al., 2011).

Figure 1-2: Structure of Rubisco.

(A) Surface representation of different Rubisco forms. All forms are made up of the functional units of antiparallel dimers of large subunits shown in light and dark blue. Small subunits of form I Rubisco and the RAD of form II/III Rubisco are shown in wheat. PDB codes: Form IA_c: 1SVD; form II: 4LF2; Form II/III: 5MAC; Form III: 1GEH; Form IV: IYKW Scale bar, 100 Å. (B) Rubisco active site. Surface representation of the spinach L₈S₈ Rubisco (Protein Data Bank 8RUC). Structural details for one L-subunit of an L₂ dimer is shown. An active site within the α/β -barrel of the C-terminal domain is highlighted. Boxed area represents the conserved Rubisco active-site residues relative to carbamylated Lys-201 (K201x). These residues and the six-carbon reaction intermediate analog 2-carboxyarabinitol 1,5-bisphosphate (CABP) are shown as sticks.

The dominant form I Rubisco is the best studied and is present in organisms ranging from metabolically diverse bacteria to algae and terrestrial plants. It consists of large (L) (~55 kDa) and small (S) (~15 kDa) subunits in a hexadecameric (L₈S₈) quaternary structure (Spreitzer and Salvucci, 2002) (Figure 1-2A, B). Although the form I RbcL amino acid sequences are highly conserved at ~50% identity, the Rubisco small subunit (RbcS) sequences are more diverse with only 30–40% homology between different species (Hauser et al., 2015). Structurally, they consist of a common core of four-

stranded anti-parallel β -sheet with two helices on one side. The most striking variations occur at the loop between β strands A and B as well as the C-terminus, which is elongated in some enzymes. Although not strictly required for enzymatic function, the small subunits in the form I enzymes are important for maximal catalytic rates and specificity (Spreitzer, 2003).

The form I Rubisco can be further subdivided into “green” (form IA and IB) and “red” (form IC and ID) subclasses (Delwiche and Palmer, 1996; Tabita et al., 2008). The green-type enzymes encompass those found in plants, green algae and cyanobacteria while the red-type Rubisco are present mainly in non-green algae and phytoplankton (Andersson and Backlund, 2008). Proteobacteria possess both the green- and red-type enzymes (Badger and Bek, 2008). The green and red enzymes differ in several structural features, the most notable being an extension in the C-terminal domains of the small subunits in the red-type enzymes. The C-terminal extensions form a small barrel that protrudes into the solvent channel of the holoenzyme (Hansen et al., 1999; Joshi et al., 2015). It is possible that the structural variance between the two subclasses could explain the higher substrate specificity for CO₂ over O₂ (S_{c/o}) sometimes observed in red algal Rubiscos as compared to green-type enzymes (Whitney et al., 2001).

In prokaryotes the genes encoding for form I Rubisco, *cbbL* and *cbbS* are found within the *cbb* operon. In eukaryotes, the RbcL is always encoded by the chloroplast genome while the RbcS is encoded by the plastid genome only in non-green algae and nuclear encoded in plants and green algae (Hauser et al., 2015). Although the majority of studies thus far have focused on the form IB Rubisco in higher plants, the bulk of sequence variation can be found in the prokaryotic realm which are represented in all clades of Rubisco (Liu et al., 2017).

The simpler form II Rubisco encoded by *cbbM*, is composed of different number of RbcL dimeric pairs (L₂)_n and is widely distributed in proteobacteria and eukaryotic dinoflagellates. It was initially discovered in the phototrophic purple non-sulfur bacteria, *Rhodospirillum rubrum* and functions as a dimer (Figure 1-2A) (Badger and Bek, 2008; Satagopan et al., 2014). Since then, other form II enzymes that present as hexamers have also been described (Varaljay et al., 2016). Although form II Rubisco maintain conserved active site residues, they exhibit substantially different catalytic

properties, with a higher carboxylation rate that comes at a cost of lower specificity for CO₂ (Tabita et al., 2008). As a result, form II enzymes are believed to have arose from an anaerobic environment where maximum catalysis can be achieved without the need for high specificity.

The crystal structure of the hexameric form II Rubisco from *Rhodopseudomonas palustris* (RpM) in complex with the transition state analog 2-carboxyarabinitol 1,5-bisphosphate (CABP) revealed subtle differences to the arrangement of a similar set of active site residues in three-dimensional space. The unique placement of the C-terminus, residing at the apex of the enzyme away from the active site region, is in contrast to the C-terminus of form I Rubisco which lies at the equatorial region of the protein and functions to regulate access to the catalytic site (Duff et al., 2000; Satagopan et al., 2014). Consistently, the recently characterized CABP-bound hexameric form II Rubisco from GWS1B, an uncultivated bacterium of the *Gallionellaceae* family in β -proteobacteria was observed to share the same C-terminal conformation (Varaljay et al., 2016). It is conceivable that the conformational differences observed account for the differences in catalytic properties between two different forms of Rubiscos.

Many photoautotrophic and chemoautotrophic proteobacteria such *Rhodobacter sphaeroides*, *Hydrogenovibrio marinus* and *Acidithiobacillus ferrooxidans* have form II Rubisco acting in concert with form I enzymes within a single cell (Yoshizawa et al., 2004; Badger and Bek, 2008; Tsai et al., 2015; Toyoda et al., 2018). Therefore, it appears that maintaining multiple Rubisco forms within one organism is a common strategy among proteobacteria to maintain flexibility in dealing with dynamic environmental conditions, especially those concerning relative concentrations of the gases CO₂ and O₂.

Form III Rubisco, believed to be the ancestor to all forms of modern day Rubisco was found predominantly in anaerobic archaea which lack the CBB cycle (Tabita et al., 2007a). It features conserved active sites and catalyzes carboxylation in the nucleoside assimilation pathway to regenerate RuBP (Sprott et al., 1993; Sato et al., 2007). In terms of quaternary structure, form III enzymes generally exist as dimers but there are

interesting exceptions such as the pentamer of dimers (L₂)₅ from *Thermococcus kodakaraensis* (Figure 1-2A) (Maeda et al., 1999).

Studies into organisms that inhabit diverse and sometimes extreme environments reveal interesting Rubisco variants in both structure and function which have adapted to alternative evolutionary pressures different from those of model organisms studied so far. Rubisco from the thermophilic archaea *Methanococcoides burtonii* assembles into pentamers of dimers (L₂)₅ only when stimulated by the carboxylation substrate RuBP (Alonso et al., 2009) (Figure 1-2A). Additionally, the novel crystal structure of *M. burtonii* Rubisco (MbR) showed a distinct domain of 29 amino acids proximal to the C-terminus of the RbcL. Its location between adjacent large subunit (L₂) dimers is reminiscent of the Rubisco small subunits present in form I Rubisco (L₈S₈). Therefore, the unique structural appendage is proposed to have a similar role to the small subunits in concentrating the RbcL dimers and is termed the Rubisco Assembly Domain (RAD) (Gunn et al., 2017). The RAD is unique to MbR and is not present in other (L₂)₅ archaea Rubisco such as that from *T. kodakarensis*. Phylogenetic analysis places the MbR Rubisco in a sub-group of form III Rubisco called form II/III. Genes encoding form II/III Rubisco have also been discovered in the uncultivated bacteria of the candidate phyla radiation (CPR) and are believed to act in the same capacity of purine/pyrimidine metabolism analogous to the archaea (Wrighton et al., 2012; Wrighton et al., 2016).

Form IV Rubisco also known as the Rubisco-like protein (RLP) possess the highest diversity amongst Rubisco sequences (Tabita et al., 2007b; Ashida et al., 2008; Badger and Bek, 2008). The RLPs do not catalyze carboxylation due to substitutions in critical active site residues, instead, some have been shown to act in methionine salvage pathways (Ashida et al., 2008; Erb et al., 2012). RLP is found in organisms that do not utilize CO₂ as their main carbon source. The enzyme in the green sulfur phototrophic bacterium *Chlorobium tepidum* forms an antiparallel dimer of large subunits (Figure 1-2A) (Li et al., 2005b).

1.5 Rubisco catalytic mechanism and regulation

The Rubisco apoenzyme (E) requires carbamylation at an essential lysine residue (Lys201 of spinach Rubisco) at the catalytic site by a non-substrate CO₂, followed by

stabilization of the resulting carbamate by a Mg^{2+} ion to form a catalytically active ternary complex, Enzyme- CO_2 - Mg^{2+} (ECM) (Cleland et al., 1998) (Figure 1-3A).

The Rubisco reaction mechanism has been studied extensively and involves critical active site residues that are highly conserved. The reaction proceeds with the enolization of RuBP which results in an intermediate that can then potentially react with both CO_2 and O_2 producing a ketonic acid (Figure 1-3B). Subsequently, the ketone gets hydrated and cleaved to produce two molecules of 3PG in the case of CO_2 or one molecule of 3PG and one molecule of 2PG for O_2 (Andersson and Backlund, 2008).

The Rubisco activation process happens spontaneously when Rubisco exist in a solvent-accessible open state where the α/β -barrel Loop 6 is retracted away from the active site and the C-terminal end of Rubisco is disordered. Binding of the ligand RuBP triggers the closure of loop 6 over the active site (Duff et al., 2000) (Figure 1-3C). This movement brings Lys334 located at the tip of loop 6 into the active site and in close proximity to the ligand. Migration of Loop 6 is dependent upon a critical hinge residue Val331 situated in the N-terminal of the loop which allows for the rotational movement. Changes also take place in the N-terminal domain of the adjacent large subunit. The 60s loop (residues 60-66) which connects helix αB and strand βC of the N-terminal domain is brought into position over the active site and provides several residues to the catalytic site. Glu60 and Thr65 of the 60s loop makes hydrogen bonds with Lys334 of loop 6 and forms a major interface with the C-terminal α/β -barrel of the neighbouring dimer. Lastly, the flexible C-terminal tail of RbcL packs against the protracted loop 6 and stabilizes it via Asp473 which functions as a latch to facilitate the closed state of the active site (Andersson and Backlund, 2008). Conformational changes in these regions lead to occlusion of the active site from the solvent, after which carboxylation of RuBP can proceed.

Figure 1-3: Rubisco reaction mechanism.

(A) Activation of Rubisco by non-substrate CO_2 and Mg^{2+} binding. A conserved active site lysine (Lys-201 in spinach RbcL) must react with a non-substrate CO_2 to form a carbamate (EC), followed by the binding of a Mg^{2+} ion to form the catalytically competent holoenzyme ECM. (B) Rubisco reaction

mechanism. The reaction is initiated by abstraction of the C-3 proton from the five-carbon substrate to give a enediol intermediate. Stereospecific carboxylation of C-2 of the enediol followed by rapid hydration gives rises to a six-carbon intermediate which is deprotonated and cleaved to eventually yield two molecules of 3-PG, if carboxylated. In the oxygenation side reaction, the enediol intermediate reacts with oxygen to form peroxyketone which undergoes bond cleavage to produce one molecule of 3PGA and one molecule of 2PG. Figure adapted with modifications from (Mueller-Cajar, 2017) and (Cleland et al., 1998). (C)Structural details for one L-subunit of an L₂ dimer (*Spinacia oleracea*, PDB-ID: 8RUC) highlighting the C-terminus and loop 6 in red and yellow, respectively. In the boxed area, residues involved in closure of the active site are shown as sticks.

Rubisco is known to form inhibited complexes with a number of sugar phosphates, including its own substrate RuBP. Premature binding of RuBP to the non-carbamylated form of Rubisco result in a dead-end binary complex known as Enzyme-RuBP (ER) (Figure 1-4). Binding of RuBP to the inactive Rubisco induces a conformational change that prevents carbamylation by CO₂ and binding of Mg²⁺. These inactive Rubisco complexes can accumulate and inhibit carbon fixation during active photosynthesis (Mueller-Cajar et al., 2014). It is also the reason for the decline in Rubisco catalytic activity over time in vitro, a phenomenon known as ‘fallover’ (Edmondson et al., 1990; Zhu and Jensen, 1991). A plethora of other sugar phosphates also bind non-productively to the active sites of non-carbamylated Rubisco. They include catalytic ‘misfire’ products such as xylulose 1,5-bisphosphate (XuBP) formed by misprotonation of the initial enediolate intermediate (Pearce, 2006).

Additionally, other sugar-phosphates such as the night time inhibitor, carboxy-arabinitol-1-phosphate (CA1P) binds preferentially to the carbamylated form of Rubisco (ECM) and inhibits catalysis by obstructing the active site, resulting in the formation of stably, inhibited complex, ECMI (Figure 1-4) (Berry et al., 1987; Parry et al., 2008). The activity of Rubisco that has been inhibited by sugar phosphates can be restored by the action of the enzyme, Rubisco activase (Bhat et al., 2017a; Mueller-Cajar, 2017).

Figure 1-4: Regulation of Rubisco activity and inhibition by sugar phosphates.

In order to form the activated ternary complex ECM, Rubisco (E) must be carbamylated by a non-substrate CO₂ and bind Mg²⁺. Premature binding of the substrate RuBP, or the misfiring XUBP to the inactivated enzyme result in an inhibited form of the enzyme (EI). Binding of the night time inhibitor CA1P or another misfiring product, glycerol-2,3-bisphosphate (PDBP), to activated Rubisco similarly result in a stable, inhibited complex (ECMI). Inhibited complexes are activated by the ATP-dependent Rubisco activase (Rca) (Parry et al., 2008). Figure adapted from (Liu et al., 2017).

1.6 Rubisco Activase (Rca)

Rubisco activase (Rca) belongs to the AAA⁺ (ATPases Associated with various cellular Activities) superfamily. They are typically involved in ATP-driven unfolding, degradation and dissociation of macromolecules in diverse settings. A distinctive feature of members of the superfamily is a structurally conserved ATPase domain of ~250 amino acids which consists of a large N-terminal P-loop NTPase α/β subdomain and a smaller C-terminal α -domain (Ogura and Wilkinson, 2001). The large subdomain contributes the Walker-A, Walker-B and sensor I motifs that are crucially involved in nucleotide binding and hydrolysis by coordinating the β and γ phosphates of ATP and the water activating magnesium ion (Wendler et al., 2012). The adjacent AAA⁺ domain provides an arginine residue known as the arginine finger, responsible for ATP

hydrolysis and oligomer formation. The functional state of AAA+ proteins is generally a cyclic hexamer, with the nucleotide-binding pocket situated at the interface of neighboring subunits. Many AAA+ proteins also possess conserved pore loop sequences that line the axial channel of hexameric ring assembly and are involved in substrate translocation (Hanson and Whiteheart, 2005a; Martin et al., 2008; Puchades et al., 2017; de la Peña et al., 2018). In addition to the classical AAA+ module, many clades within the AAA+ superfamily are embellished with distinctive secondary structural elements that necessitate specific substrate recognition and manipulation (Iyer et al., 2004).

Rubisco activases utilize energy from ATP hydrolysis to remove tightly bound, inhibitory sugar phosphates from the catalytic site of both carbamylated and non-carbamylated forms of Rubisco (Figure 1-4). Having a pivotal role in maintaining the activity of Rubisco, these ancillary proteins are therefore essential for maintenance of metabolic flux through the CBB cycle. Thus far, three Rca systems have been characterized. The green-lineage Rca initially discovered in *Arabidopsis thaliana*, remodels Rubisco from plants (Portis, 2003; Stotz et al., 2011; Shivhare and Mueller-Cajar, 2017) while the red-type Rca have been shown to function in α -proteobacteria and red lineage phytoplankton (Mueller-Cajar et al., 2011; Loganathan et al., 2016). More recently, hetero-oligomeric complexes CbbQ and CbbO were demonstrated to function as Rca in chemoautotrophic γ -proteobacteria (Sutter et al., 2015; Tsai et al., 2015). These molecular chaperones have low sequence homology among them and are regarded as products of convergent evolution.

1.6.1 Red-type Rca

The *cbbX* gene found downstream of Rubisco genes (*cbbL* and *cbbS*) in certain proteobacteria and red algae encodes the 35 kDa red-type Rca (Gibson and Tabita, 1997; Maier et al., 2000). In addition to the AAA+ core module, CbbX from the α -proteobacteria *Rhodobacter sphaeroides* (RsRca) contains a small helical extension at the N-terminus of AAA+ core (Figure 1-5A, B). RsRca requires the presence of ATP and allosteric binding of RuBP, the substrate of Rubisco to form an active hexameric ring (Figure 1-5C). In the absence of RuBP, RsRca arrange into high molecular weight fibrillar structures that are deficient in ATPase activity. Such allosteric regulation

ensures that RsRca takes on its functional form only when photosynthesis is active. The ATPase activity of RsRca can be further stimulated by the addition of inhibited Rubisco *in vitro* (Mueller-Cajar et al., 2011).

A cryo-electron microscopy (cryo-EM) structure of RsRca chemically cross-linked to its cognate form IC Rubisco (RsLS) revealed that the AAA⁺ protein interacts with Rubisco via a side-on approach, establishing contacts with both RbcL and RbcS (Figure 1-5D). The highly conserved top surface (convex face) of RsRca docks onto the edge of Rubisco and is positioned over the flexible C-terminus of RbcL (Bhat et al., 2017b). It is noteworthy that the C-terminal is extended in the red-lineage Rubisco of proteobacteria and red-algae by ~5-10 residues in comparison to the green-type RbcL. Several critical points of attachment needs to be broken before the C-terminal strand becomes detached from the body of the enzyme and is transiently threaded into the pore (~25 Å) of the RsRca hexamer, to promote destabilization of the Rubisco active site. The reversible threading of the C-terminal tail is facilitated by the classical pore loop sequence (aromatic-hydrophobic-glycine) also present in other AAA⁺ unfoldases such as ClpX (Sauer and Baker, 2011; Olivares et al., 2015). Interestingly, while ClpX progressively mediates the global unfolding of its substrate, the conformational modeling by RsRca is confined locally to the Rubisco active site (Bhat et al., 2017b). The liberated sugar phosphate inhibitor is then degraded by selective phosphatases to prevent it from rebinding (Bracher et al., 2015).

Figure 1-5: The red-type prokaryotic *R. sphaeroides* Rca.

(A) Schematic representation of the domain organization of RsRca. (B) Crystal structure of the monomer (PDB-ID: 3YSL) shown in ribbon representation. (C) Top and side view of the RsRca hexameric model fit into the electron microscope reconstruction (EMDB: EMD-1932; PDB-ID: 3ZUH) (D) Model of the mechanism of RsRca. Cryo-EM reconstruction of the cross-linked complex of the inhibited *R. sphaeroides* Rubisco (ECM-CABP) with RsRca hexamer bound. *R. sphaeroides* Rubisco (PDB-ID: 5NV3) and RsRca (PDB-ID: 3ZUH) are fitted into the electron density. Figure adapted with modifications from (Bhat et al., 2017a).

The eukaryotic Rca from the red algae *Cyanidioschyzon merolae* (CmRca) shares ~60% identity with RsRca. It forms a heterooligomeric complex composed of gene products from two copies of *cbbX* genes, in the nuclear and the plastid genomes at a 1:1 ratio. Although the alternately-arranged subunits function in a highly coordinated manner, they are not functionally equivalent with the nuclear-encoded subunit being more

catalytically dominant. In contrast to the prokaryotic system, CmRca presents as a hexamer even in the absence of RuBP and possesses a basal low ATP rate that is stimulated when RuBP is present (Loganathan et al., 2016). Despite the differences in composition and regulation, the mechanism of Rubisco repair appears to be conserved for both prokaryotic (form IC) and eukaryotic (form ID) systems in the red-lineage (Mueller-Cajar et al., 2011; Loganathan et al., 2016).

1.6.2 Green-type Rca

Although extensively studied, the mechanism of action of Rca in plants containing eukaryotic form IB Rubisco remains unclear. Structurally, green-type Rca resembles the p97/CDC48 ATPase in having an additional N-terminal domain, missing in the crystal structure of both NtRca and the Arabidopsis homologue (Hasse et al., 2015) (Figure 1-6A, B). N-terminal truncation or mutation of a highly conserved Tryptophan residue in this region preserved ATPase activity but abolished activase function (van de Loo and Salvucci, 1996; Stotz et al., 2011), implicating the N-domain in substrate interaction. The specificity helix, H9 of the α -helical subdomain is believed to interact with residues 89 and 94 (spinach numbering) of the Rubisco holoenzyme (Stotz et al., 2011) (Figure 1-6D). This interaction has been shown to be important for the green-type Rca to differentiate between solanaceous species (e.g. Tobacco) and non-solanaceous Rubisco (e.g. Spinach) (Li et al., 2005a; Wachter et al., 2013).

Although the canonical aromatic-hydrophobic-glycine motif is missing in pore loop 1, all three pore loops of NtRca which project into the central cavity have been shown to be involved in Rubisco repair (Stotz et al., 2011). Interestingly, NtRca has a wider pore (~ 36 Å) compared to the red-type RsRca (Figure 1-6C), therefore it is plausible that NtRca might engage a more complex secondary structural element of Rubisco for remodeling and not just a single peptide motif like the red-type rca.

Recently, the $\alpha 4$ - $\beta 4$ loop in *Oryza sativa* (Rice) Rca nucleotide-binding domain has become implicated as a Rubisco-interacting motif. This flexible loop, missing in the crystal structure of NtRca is located on the top surface of the hexamer in close proximity to helix H9 of the adjacent subunit, establishing the top surface as the primary Rubisco-interacting side (Shivhare and Mueller-Cajar, 2017) (Figure 1-6D).

Figure 1-6: The green-type eukaryotic *N. tabacum* Rca.

(A) Schematic representation of the domain organization of NtRca. (B) Crystal structure of the monomer (PDB-ID: 3T15) shown in ribbon representation. The specificity helix, H9 is highlighted in yellow. Missing parts of the structure are shown with dashed lines. (C) Top and side view of the NtRca hexameric model fit into the electron microscope reconstruction (EMDB: EMD-1940; PDB-ID: 3ZW6) (D) Model of the mechanism of NtRca. The interaction of the NtRca with inhibited Rubisco (PDB-ID: 1EJ7) requires the additional N-domain and the H9 specificity helix. Arg89 on RbcL is critical for interaction with NtRca. Figure adapted with modifications from (Bhat et al., 2017a).

Many plants (Arabidopsis, Spinach and rice) harbor two isoforms of Rca, the α and β isoforms as a result of alternative splicing of a single gene transcript. There are also examples (cotton and soybean) where the isoforms are encoded by two different genes. Sequences of plant Rca have indicated the presence of a C-terminal extension not resolved in crystal structures. This extension is essential in ensuring that plant Rca is

constitutively ATPase active (Portis et al., 2008). The eukaryotic green-type Rca is not known to be regulated by RuBP, and unlike many members of the AAA+ family, no stimulation of ATP activity is observed in the presence of its substrate. Instead, activity of plant Rca is governed by the intracellular ATP/ADP ratio and the redox state of two conserved cysteine residues on the C-terminal extension of the larger α isoform (Streusand and Portis, 1987; Portis et al., 2008). In limiting light conditions, the cysteine residues on the α isoform are oxidized to form a disulfide bond and enables ADP inhibition of Rca activity. During periods of illumination and active photosynthesis, thioredoxin-f reduces the disulfide bond and largely diminishes the inhibitory effects of ADP (Zhang and Portis, 1999; Zhang et al., 2002).

Rca thermolability limits the photosynthetic capacity of plants (Salvucci and Crafts-Brandner, 2004). Transforming a more thermostable Rca variant into *Arabidopsis* enabled these plants to grow at elevated temperatures with increased photosynthetic efficiency (Kurek et al., 2007; Kumar et al., 2009). Recent characterization of Rca from the CAM plant *Agave tequilana* revealed that it is ~ 10 °C more thermostable than Rca from *O. sativa* and has a higher enzymatic activity in activating inhibited *O. sativa* Rubisco (Shivhare and Mueller-Cajar, 2017), further reinforcing the potential of introducing thermostable Rca into host plants to increase rate of photosynthesis.

1.7 The CbbQO Rca system of chemoautotrophic bacteria

Chemoautotrophic proteobacteria often contain two poorly characterized open reading frames, *cbbQ* and *cbbO* directly downstream of Rubisco genes in multiple operons, reminiscent of the position of the *cbbX* gene in organisms encoding red-type Rubisco (Badger and Bek, 2008). The *cbbQ* and *cbbO* genes have been found to co-occur with Rubisco in several chemoautotrophic proteobacteria suggesting a functional relation among these gene products. CbbQ is a ~ 30 kDa protein belonging to the AAA+ superfamily while CbbO is a von Willebrand factor A (VWA) domain-containing adaptor protein of ~ 85 kDa. CbbQ shares $\sim 50\%$ identity with nirQ and norQ while the genetically-linked CbbO carries the VWA domain similar to norD. The *nor* operon encodes proteins involved in bacterial denitrification. In *Paracoccus denitrificans*, both norQ and norD are essential for the formation of functional cytochrome c dependent nitric oxide reductase suggesting that the actions of the two proteins are coupled (Kahle

et al., 2018). When CbbQ and CbbO of *Hydrogenovibrio marinus* were co-expressed with Rubisco in *E. coli*, more active purified Rubisco as well as a modest increase in Rubisco activity were observed, indicative of a role in post-translational regulation of Rubisco (Hayashi et al., 1997, 1999).

Recent work from our laboratory has shown that two sets of isoforms of CbbQ and CbbO recombinantly expressed in *E. coli*, form hetero-oligomers, AfQ1O1 and AfQ2O2 that function as specific Rubisco activases in the chemolithoautotroph *Acidithiobacillus ferrooxidans* (*A. ferrooxidans*, ATCC23270), a γ -proteobacterium found in low pH environments (Tsai et al., 2015). It was demonstrated that full ATPase activity and Rubisco activation require both CbbQ and CbbO to work in tandem.

The *cbbQ1O1* and *cbbQ2O2* genes are found directly downstream of Rubisco-encoding genes, *cbbLS-1* and *cbbM* genes respectively (Tsai et al., 2015) (Figure 1-7). The *cbbLS-1* genes encode the hexadecameric (L₈S₈) Rubisco AfLS, which belongs to form IAq, while *cbbM* encode AfM which classifies as form II, (L₂)_n. Although no high-resolution structure exists for AfM, biochemical data indicates that it functions as a hexamer, similar to the form II Rubisco in *Rhodopseudomonas palustris* (Satagopan et al., 2014). *A. ferrooxidans* also possesses an additional *cbb* operon where the third set of *cbbQ* and *cbbO* genes encoding Q3O3 are not found directly associated with the form I Rubisco gene. Instead, genes encoding structural proteins of the α -carboxysome are found between them. The Rubisco activase function for the carboxysomal associated Q3O3 has not been confirmed.

Figure 1-7: Arrangement of the CO₂ fixation operons encoding form I and form II Rubisco in *A. ferrooxidans*.

In the marine chemoautotroph *Hydrogenovibrio marinus*, three similar operons are expressed at varying CO₂ concentrations. Form II Rubisco is the main functioning enzyme at high CO₂ concentrations and form I is preferentially expressed at lower CO₂ concentrations, consistent with form II enzymes having lower specificity for CO₂. Carboxysomes enhance carbon fixation by concentrating Rubisco and its substrate CO₂ within a proteinaceous shell. Therefore, the third gene cluster is expressed when CO₂ is extremely limiting (Yoshizawa et al., 2004; Toyoda et al., 2018). The expression of three Rubisco isoenzymes is regulated by CbbR, a member of the LysR-family of transcriptional regulators (Dangel and Tabita, 2015; Toyoda et al., 2018). CbbR-mediated expression is influenced by the binding of RuBP, similar to the red-type RsRca. Such a controlled Rubisco expression system confers the organism ability to adapt rapidly to environmental changes in CO₂ concentrations. Notably, the expression of active Rubisco alone was inadequate to support optimal growth (Toyoda et al., 2018), suggesting the necessity of ancillary factors such as CbbQ and CbbO.

1.7.1 The AAA+ ATPase CbbQ

CbbQ belongs to the diverse but poorly characterized MoxR group of AAA⁺ ATPases, believed to function as molecular chaperones in bacteria and Archaea (Snider and

Houry, 2006; Wong and Houry, 2012). Members of this family frequently co-occur in close proximity to genes encoding proteins, which contain a von Willebrand factor A (VWA) domain, known for mediating protein-protein interactions (Xiong et al., 2002; Santelli et al., 2004) (Figure 1-8A). Based on limited data available there appears to be a tendency for MoxR AAA+ proteins to work in conjunction with VWA domain-containing proteins to assemble and maintain the activated state of its substrate (Snider and Houry, 2006). Well-known MoxR proteins include NirQ/NorQ associated with NorD which cooperate to facilitate non-heme Fe insertion to activate the nitric oxide reductase enzyme (Kahle et al., 2018). Another extensively characterized MoxR protein, RavA has been shown to interact with the inducible lysine decarboxylase, LdcI to prevent inhibition by the small molecule ppGpp (Snider et al., 2006; El Bakkouri et al., 2010).

Figure 1-8: The CbbQO-type Rca.

(A) Schematic representation of the domain organization of *A. ferrooxidans* CbbQ2 and CbbO2 (B) Crystal structure of the CbbQ2 monomer shown in ribbon representation. The helix-2-insert and pre-sensor 1 β hairpin are highlighted in yellow and orange, respectively. (C) Top and side view of the CbbQ2 hexameric model fit into the electron microscope reconstruction (EMDB: EMD-6477) (D) Structural details for the CbbQ2 active site. Critical residues involved in ATP hydrolysis are shown in sticks.

The *cbbQ* genes in *A. ferrooxidans* encode proteins of ~260 residues (~30 kDa) that

adopt a ring-shaped, hexameric assembly as visualized by negative-stain electron microscopy (Tsai et al., 2015). The crystal structure of AfQ2 has recently been solved in collaboration with Assoc Prof Gao Yonggui, by his Ph.D student, Ye Fu Zhou (Tsai and Ye et al., unpublished) using the molecular replacement model of *Halothiobacillus neapolitanus* carboxysomal CsoCbbQ (Sutter et al., 2015). Similar to other AAA+ motors, AfQ2 has an N-terminal P-loop NTPase domain consisting of a core of parallel β -sheets that are arranged in a $\beta 5$ – $\beta 1$ – $\beta 4$ – $\beta 3$ – $\beta 2$ sequence (Figure 1-8B). The structure of AfQ2 reveals β -hairpin insertions between $\alpha 3$ and $\beta 4$ as well as within $\alpha 2$, consistent with other members of the helix-2 insert clade of AAA+ proteins. The helix-2 β -hairpins project out of the AAA+ core module at an angle that juts conspicuously into the central cavity and fills the pore. As a result, the AfQ2 pore is only ~ 10 Å wide whereas the green and red-type activases have central pore diameters of ~ 36 Å and ~ 25 Å respectively (Mueller-Cajar et al., 2011; Stotz et al., 2011) (Figure 1-8C).

The N-terminal domain also contains the characteristic Walker A (GXXXGK[T/S]) and Walker B (hhhhDE, h for hydrophobic residue) motifs of AAA+ ATPases that are essential for nucleotide binding and hydrolysis (Figure 1-8D). Analysis of the active site reveals an ADP stabilized by a Mg^{2+} cation. The arginine finger is contributed by the C-terminal domain of the adjacent subunit. The C-terminus of AfQ2 is made up of five-helix bundles that fill in the gaps between the NTPase domains, presenting a hexagonal-shaped profile with a diameter of 107 Å (Figure 1-8C). The C-terminus is featured on the concave face of the hexamer, where regions of high sequence conservation is believed to be important for interaction with a substrate or other proteins (Sutter et al., 2015).

CbbO encoded downstream of CbbQ contains a VWA domain at its C-terminus. Biochemical characterization of the CbbQ homologue from *A. ferrooxidans* showed that in contrast to other Rca systems, its ATPase activity and activase function was dependent on the formation of a hetero-oligomeric complex with CbbO (Tsai et al., 2015). CbbQ alone or in complex with C-terminally truncated O2 (Q2O2 Δ C444, residues 1-444) was unable to activate Rubisco and had a low ATPase activity < 1 ATPmin⁻¹, whereas Q1O1 and Q2O2 had an ATPase activity of ~ 3 ATPmin⁻¹ (activities refer to protomer CbbQ). Association with inhibited Rubisco preferentially stimulated

the ATPase activity of Q1O1 and Q2O2 by 2- and 20-fold respectively.

1.7.2 The adaptor protein CbbO

The ~770 residue CbbO proteins consist of a poorly conserved N-terminus and a VWA domain with a perfect metal-ion-dependent adhesion site (MIDAS) at its C-terminus. The N-terminal region (residues 1-444) of CbbO is believed to interact with CbbQ forming a hetero-oligomer with a stoichiometry of CbbQ₆O₁ (Tsai et al., 2015). Thus, the subunit organization of CbbQO has been proposed to resemble the largest yeast protein, ribosome maturation protein Rea1 or Midasin, which presents as a hexameric head domain from which an elongated flexible tail carrying the VWA domain protrude (Barrio-Garcia et al., 2016; Chen et al., 2018; Sosnowski et al., 2018). However, the insolubility of CbbO isoforms in *E. coli* unless co-expressed with their respective CbbQ isoforms poses as a challenge in biochemically characterizing CbbO and understanding its role in this Rubisco activation system (Tsai et al., 2015). Further, because of its inherent flexibility, structural information for CbbO is also lacking.

The VWA domain is well studied and has been shown to be involved in a range of processes, including cell adhesion, transport and the complement system (Springer, 2006). A common feature seems to be involvement in multiprotein complexes. The structures of VWA domains in several proteins have been determined. Among these, the integrin I domains and the VWA domains of the von Willebrand factor are best characterized (Lee et al., 1995; Bienkowska et al., 1997; Celikel et al., 1998; Emsley et al., 2000; Hughes, 2001; Song et al., 2005).

In proteins whose structure has been determined to atomic resolution, the MIDAS-containing domain adopts a classic Rossmann-fold with five parallel and one short antiparallel β strands that form a central sheet surrounded by amphipathic α helices (Figure 1-9). The MIDAS motif is located in a crevice at the top of the sheet and comprises a highly conserved DxSxS sequence, where x is any amino acid, and/or noncontiguous T and D residues that together provide 5 coordination sites for a divalent cation (usually Mg²⁺). The carboxylate side chain of the ligand usually contributes the sixth co-ordination site (Haas and Plow, 1994; Xiong et al., 2002; Song et al., 2005).

Point mutations in the MIDAS motif often weaken or eliminate the ligand-binding capacity of these proteins (Michishita et al., 1993; Lee et al., 1995; Xiong et al., 2002). On the basis of known structures, the role of the MIDAS motif in ligand binding has been studied in detail. In the integrin I domains, it was shown that the MIDAS motif can exist in two distinct conformations with high and low affinities for ligand, respectively (Lee et al., 1995). The conformational changes of the MIDAS are often propagated to other parts of the proteins (Liddington, 2002; Shimaoka et al., 2002). It is possible that non-integrin VWA domains may be subjected to a similar regulation, where the binding of ligand is coupled to long-distance conformational changes in large multi-protein complexes.

Figure 1-9: Structure of the Integrin α_L I domain in complex with domain 1 of Interleukin-3.

Boxed area represent close-up of the MIDAS motif. The metal ion is shown as a green sphere. Coordinating side chains are shown in sticks. Figure adapted with modifications from (Song et al., 2005).

Preservation of the MIDAS motif is required for Rubisco-activase interactions as mutations of CbbOVWA domain MIDAS residues to alanine disrupted the activase function but not the ATPase activity of CbbQ (Tsai et al., 2015). This observation suggests that the CbbOVWA interaction with inhibited Rubisco is similar to those observed between α -integrin and their MIDAS-dependent ligands. In support of this, site-directed mutagenesis localized an essential acidic residue on the Rubisco large

subunit that is essential for activation. The highly conserved surface acidic residues, D82 and E75 on AfLS and AfM, respectively, are thus proposed to serve as the sixth coordination site of the MIDAS ion, supporting a model where CbbO functions as an essential substrate adaptor that binds to the carboxylate group of the inhibited Rubisco via its VWA domain (Tsai et al., 2015). Interestingly, the corresponding residue in the plant counterpart, P89 is also necessary for activation by green-type Rca (Larson et al., 1997).

1.7.3 Putative mechanism of Rubisco activation by AfCbbQO

The *A. ferrooxidans* RbcL C-terminus was determined to be important for activation, as deleting one or two residues ($\Delta C1$ and $\Delta C2$), or adding an alanine to the conserved C-terminal HK/R motif abolished activase function and ATPase stimulation (Tsai et al., 2015). However, unlike the red-type Rca which functions by threading of the C-terminal tail of RbcL through the pore of the hexamer (Mueller-Cajar et al., 2011; Bhat et al., 2017b), there is no evidence of the involvement of the CbbQ pore in Rubisco remodeling. Therefore, it is still unknown if CbbQ interacts with Rubisco directly.

Sequence analysis of CbbQ has indicated relatedness to dynein (Iyer et al., 2004), which employ long-distance structural changes as a central part of its mechanism (Kon et al., 2012). The hexameric AAA⁺ motor dynein consists of six distinct AAA protomers occurring in tandem on a single polypeptide chain. ATP-driven conformational changes of the AAA⁺ ring are used to drive the movement of the N-terminal linker domain, which delivers the ‘powerstroke’ required for moving along the microtubule (Kon et al., 2012; Roberts et al., 2013; Schmidt et al., 2015).

In this constellation, I hypothesized that CbbO act as an adaptor to its cognate ATPase motor CbbQ. The CbbQ₆O₁ complex will interact with the stalled Rubisco via the MIDAS motif of the VWA domain of CbbO, by engaging the identified acidic surface residue Glu75 (form II) or Asp82 (form I). Productive binding with inhibited Rubisco requires presence of the conserved RbcL C-terminus and stimulates the ATPase activity of CbbQO. The energy produced from the hydrolysis of ATP is transmitted vectorially from CbbQ to the VWA domain through the CbbO linker to disrupt the active site of Rubisco.

Structural characterization of the red and green-type Rca has shed some light on potential models of interaction between CbbQO and Rubisco. However, as expected of convergently-evolved systems, CbbQO is believed to utilize a different mechanism of Rubisco activation to its counterparts, with CbbQ requiring an adaptor in the form of CbbO to be functional. Further investigation of the molecular structure of the CbbQO activase and its interaction with Rubisco, is needed to assess the validity of the proposed model as well as improve our rudimentary understanding of how this activase promotes conformational changes in Rubisco.

1.8 Aims and Scope of the Project

Rubisco activases restore Rubisco catalytic competence by removing the tightly bound inhibitors from the catalytic sites in an ATP-dependent manner. Improving the activation of Rubisco has the potential to enhance Rubisco activity, making it the focus of much research efforts. Previous biochemical characterization of the hetero-oligomeric complex has set the stage for probing the detailed mechanism of Rubisco activation by the CbbQO system. However, the lack of a full molecular structure for the active CbbQO activase hampers progress in understanding how the two subunits associate with each other and in turn, with Rubisco.

A detailed dissection of these interactions will be advantageous in our understanding of Rubisco maintenance system in diverse organisms and also provide new insights into the poorly studied MoxR group of AAA+ ATPases and proteins containing the VWA domain. The aim of this study therefore, is to investigate in detail how CbbQ and CbbO associate in the proposed Q_6O_1 stoichiometry and in turn interact with its substrate Rubisco. A key goal of this work is to establish conditions to stabilize a Rubisco-CbbQO complex for structural analysis by electron microscopy.

A major obstacle that prevents us from learning more about Rubisco activation mechanistically is the highly transient nature of Rubisco-Rca interaction. Detailed analysis of Rubisco-activase complexes has been hindered by the dynamic nature of the interactions (Bhat et al., 2017b). Therefore, Chemical cross-linkers will be used to capture the interaction between Rubisco and CbbQO activase. Chemical cross-linking involves the formation of stable covalent bonds between proximal amino acid residues

in native protein structures with distance constraints. The cross-linked complexes will then be isolated and visualized with negative stain EM and cryo-EM to obtain information on how activase binds to Rubisco and specific interaction sites on the two proteins.

In recent years, single particle analysis by cryo-EM has enabled routine near-atomic resolution structure determination of proteins, from large complexes to small membrane proteins (Kühlbrandt, 2014; Bai et al., 2015b). The advent of direct electron detector technology and motion correction for beam-induced motion has revolutionized the field and at intermediate to high resolution, cryo-EM has the potential to visualize the arrangement of the CbbQO activase in complex with Rubisco and provide critical details on the mechanism of Rubisco activation.

From the recently published cross-linked structure of RsLS and red-type RsRca, a certain amount of conformational heterogeneity is to be expected from the CbbQO hetero-oligomer and by the varying attachment points of CbbQO on Rubisco (Bhat et al., 2017b). Single particle cryo-EM has emerged as a technique of choice as it is tolerant of structural heterogeneity to a certain extent. Images of particles of variable conformation and composition are averaged into more homogenous class averages of characteristic 2D views of the complex. The 2D projections can be combined to produce 3D reconstructions. Computational software such as RELION make use of the Bayesian approach to distinguish transitional or heterogenous structures that exist within the dataset during 3D classification (Scheres, 2012a; Scheres, 2012b). These algorithms allow dynamic conformational states of a macromolecular complex to be captured (Banerjee et al., 2016; Gates et al., 2017; Dong et al., 2019).

Parallel to cryo-EM, I aim to use chemical cross-linking coupled with mass spectrometry (XL-MS) to validate and refine models of the 3-D structure of the complex. XL-MS is a method that is increasingly being used to obtain insights into the structure and topology of large protein assemblies (Faini et al., 2016). It is often complementary to cryo-EM in determining high resolution structures, especially when resolution is not sufficient in cryo-EM models and side chains are not readily distinguishable (Rappsilber, 2011; Leitner et al., 2016). During XL-MS, proteins cross-linked in their native state are enzymatically digested into peptides and then identified

via tandem MS. This well-established method will enable identification of side chains that are involved in contact sites. This will facilitate docking of proteins into cryo-EM models and reveal important clues to the subunit organization, orientation of the CbbQO hetero-oligomer as well as the CbbQO-Rubisco complex. Importantly, the distance constraint imparted by the cross-linker can also reveal information about the conformational flexibility of the protein assembly.

Finally, emerging insights from different structural methods will be utilized to drive mutagenesis studies and *in vitro* biochemical assays in order to obtain a thorough understanding of this newly discovered Rubisco activase system. This will facilitate bioengineering efforts to increase the efficiency of carbon fixation.

2. MATERIALS AND METHODS

2.1 Chemicals

All the chemicals used in the current study were of at least analytical grade. Chemicals other than the ones listed here, were purchased from Sigma-Aldrich.

Reagent	Supplier
30% Acryl /Bis solution 37.5:1	BioRad
Acetic acid (Glacial)	Merck
Agarose (molecular grade)	Vivantis
Ammonium per sulphate (molecular grade)	Promega
Ampicillin	Gold Biotechnology
Chloramphenicol	Gold Biotechnology
Coomassie brilliant blue G-250	Applied Chemistry
DMF (N, N-Dimethylmethanamide)	Merck
DTT (Dithiothreitol), Cleland's reagent	Thermo Scientific
Ethanol 99.5-99.8%, absolute	Merck
Glycerol	Affymetrix
Glycine	Bio-Rad
IPTG (Isopropyl β -D-1-thiogalactopyranoside)	Gold Biotechnology
Methanol, GR, min 99.8%	Merck
Magnesium chloride hexahydrate	MP Biomedicals
HisPur™ Ni-NTA resin	Thermo Scientific
SDS, Sodium dodecyl sulphate	Hoefer
Sodium chloride	Merck

Sodium hydroxide, analysis grade pellets, 99%	Merck
Sodium phosphate, monobasic, 98%, ACS reagent grade	Affymetrix
TEMED (Tetramethylethylenediamine)	BioRad
Tris-HCl Base	BioRad
Tryptone	BD Difco
Yeast extract	Affymetrix

Table 2-1: General Chemicals

Buffers	Composition
5x Native PAGE loading buffer	62.5 mM Tris-HCl, pH 8.0, 40% (v/v) glycerol, 0.1% (w/v) Bromophenol Blue
Native PAGE running buffer	120 mM Tris-HCl, pH 8.0, 134 mM glycine
5x SDS PAGE loading buffer	225 mM Tris-HCl, pH 6.8, 50% (w/v) glycerol, 5% (w/v) SDS, 5% (w/v) β -mercaptoethanol, 0.01% (w/v) bromophenolblue
SDS PAGE running buffer	120 mM Tris-HCl, pH 8.0, 134 mM glycine, 0.1% (w/v) SDS
Coomassie Staining solution	0.1% (w/v) Serva Coomassie Blue R-250, 40% (v/v) ethanol, 8% (v/v) acetic acid
Coomassie Destaining solution	10 % (v/v) ethanol, 7% (v/v) acetic acid
TAE	40 mM Tris-acetate, pH 8.5, 2 mM EDTA
TBS	50 mM Tris-HCl, pH 8.0, 137 mM NaCl, 2.7 mM KCl
TBS-T	TBS, 0.1% (v/v) TWEEN 2.0
Western blot transfer buffer	120 mM Tris-HCl, pH 8.0, 134 mM glycine, 20% (v/v) methanol
Ponceau S staining solution	0.1% (w/v) Ponceau S, 5% acetic acid

Lysis buffer	50 mM Tris-HCl, pH 8.0, 50 mM NaCl, 10 mM imidazole
Wash buffer	50 mM Tris-HCl, pH 8.0, 50 mM NaCl, 25 mM imidazole,
Elution buffer	50 mM Tris-HCl, pH 8.0, 50 mM NaCl, 200 mM imidazole
Buffer A	20 mM Tris-HCl, pH 8.0, 50 mM NaCl
Buffer B	20 mM Tris-HCl, pH 8.0, 1 M mM NaCl
Cross-linking Buffer	20 mM HEPES-NaOH, pH 8.0, 50 mM NaCl
Inclusion body lysis buffer	40 mM Tris-HCl, pH 8.0, 0.25 M sucrose, 1% Triton X-100
Inclusion body wash buffer A	40 mM Tris-HCl, pH 8.0, 0.25 M sucrose, 1% (v/v) Triton X-100, 2 M Urea
Inclusion body wash buffer B	40 mM Tris-HCl, pH 8.0, 0.25 M sucrose, 10 mM EDTA
Denaturation Buffer	40 mM Tris-HCl, pH 8.0, 6 M Guanidine hydrochloride, 1 mM EDTA
Refolding Buffer	50 mM Tris-HCl, pH 8.0, 50 mM NaCl, 1 mM MgCl ₂
Circular Dichroism Buffer	20 mM K ₂ HPO ₄ /KH ₂ PO ₄ , pH 8.0
Assay Buffer	100 mM Tricine/NaOH, pH 8.0, 5 mM MgCl ₂

Table 2-2: General buffer

2.2 Media

Media were prepared with deionised, double distilled water and subsequently autoclaved.

LB Medium	10 g/l tryptone
	5 g/l yeast extract
	10 g/l NaCl
	pH 7.1 adjusted with NaOH

Table 2-3: LB Media

2.3 Molecular Biology

2.3.1 Polymerase Chain Reaction (PCR)

A high-fidelity polymerase with proofreading activities, *Pfu* polymerase (Thermo Scientific), was used for both DNA amplification and site-directed mutagenesis. Reaction conditions are listed below:

Step	Temperature	Time	cycles
Initial denaturation	95 °C	2 min	1
Denaturation	95 °C	20 sec	35
Annealing	(T_m -5) °C	30 sec	
Elongation	72 °C	1 min /kb	
Elongation	72 °C	10 min	1

Table 2-4: PCR reaction condition

T_m refer to annealing temperature of primers

A list of primers used in this study is given below:

Name	Sequence (5'→ 3')
5SacIIHnCbbO	AACCGCGGTATGAACCCAGCGAC
3HindIIHnCbbO1-429STOP	TTAAGCTTTCATGCATTGCTCTTGC
3HindIIHnCbbO1-471STOP	TTAAGCTTCTAAAGTTCGGTGGCAA
3HindIIHnCbbO1-589STOP	TTAAGCTTACGTCGACGAATGCCAAT
AfO2S575AS577A_for	CTGCTGGACTTGGCCGAAGCACTGA
AfO2S575AS577A_rev	TCAGTGCTTCGGCCAAGTCCAGCAG
AfO2S575A_for (Tsai et al., 2015)	CTTGCTGGACTTGGCCGAATCACTG
AfO2S575A_rev (Tsai et al., 2015)	CAGTGATTCGGCCAAGTCCAGCAAG
AfO2D684A_for (Tsai et al., 2015)	AATGGTACTCACCGCCGGCCGGCCC
AfO2D684A_rev (Tsai et al., 2015)	GGGCCGGCCGGCGGTGAGTACCATT
AfQ3E107Q_For	TGCTATCTCGACCAAGTGGTGAAG
AfQ3E107Q_Rev	CTTCCACCACTTGGTCGAGATAGCA

Table 2-5: List of primers used in this study

2.3.2 pGemT Vector Cloning

The PCR products were polyadenylated at both the 5' and 3' ends to introduce a deoxyadenosine (dA) tail which facilitated ligation into pGem T vector (TV; Promega, USA). Polyadenylation was done by adding 10 U of *Taq* polymerase and 2 mM dATP to the amplified DNA and incubated at 72 °C for 30 min. The polyadenylated PCR product was mixed with 5x DNA loading dye prior to loading into 1% TAE-agarose gel (50 mL) supplemented with 5 µL of SYBR Safe DNA Gel Stain (Invitrogen, Singapore). GeneRuler 1 kb DNA Ladder (Thermo Scientific) was loaded to serve as marker. Agarose gel electrophoresis was performed in TAE buffer at 120 V for 20-30 min. DNA fragments were purified from the gel using the Wizard SV Gel and PCR Clean-Up system (Promega, USA) according to the manufacturer's instructions. Ligation of the polyadenylated PCR product into TV was carried out according to manufacturer's instruction. The ligation mixture was transformed into DH5α competent cells and plated on ampicillin (200 µg/ml ampicillin) LB plate supplemented with X-gal (5-bromo-4-

chloro-3-indolyl- β -D-galactopyranoside). X-gal enables blue white screening of TV and only white colonies containing insert was inoculated and grown overnight in a shaking incubator at 37 °C, 220 rpm. The plasmids were purified using Wizard Plus SV Miniprep DNA Purification System (Promega, USA) and their concentrations were measured using NanoDrop 2000 Spectrophotometer (Thermo-scientific, Singapore) at 260 nm. Cloning products were verified by the DNA sequencing service from 1st base (Singapore).

2.3.3 General cloning strategy

After verifying the protein-coding sequences with sequencing, 1-2 μ g of TV was digested using 20-60 U of restriction enzymes *SacII* and *HindIII* at 37 °C for at least 2 hours. The vectors and inserts were separated with 1% agarose gel electrophoresis and the respective DNA bands were extracted for purification by the Wizard Plus SV gel clean DNA Purification System. Purified DNA fragments were ligated into the pHUE vector using 200 U of T4 Ligase at 4 °C overnight. Ligation mixtures were directly transformed into DH5 α .

Cloning of PCR products into the pHUE vector allows cleavage of His₆-ubiquitin fusions at the native N-terminus (Catanzariti et al., 2004) and produced plasmids: *pHueUbHnCbbO(1-429)*, *pHueUbHnCbbO(1-471)* and *pHueUbHnCbbO(1-589)*. These plasmids were then digested and positioned into the pBAD33 vector with *XbaI* and *HindIII* enzymes to create *pBAD33UbHnCbbO(1-429)*, *pBAD33UbHnCbbO(1-471)*, and *pBAD33UbHnCbbO(1-589)*.

2.3.4 Site-directed mutagenesis

Site-directed mutagenesis was used to introduce point mutations in existing plasmids. It was done using the QuikChange protocol (Stratagene) according to manufacturer's instructions. Primers used were designed to have 12 complementary nucleotides to the plasmid at the 3' and 5' end with the centre of the primer containing the desired nucleotide changes. The forward and reverse primers were reverse complementary to each other and the list of primers used for mutagenesis are listed in Table 2-4. *DpnI* treatment was used to remove the original plasmids, 10 U of *DpnI* was added to the

mutagenesis PCR product followed by incubation at 37 °C for one hour. The treated PCR product was subsequently transformed into DH5 α , colonies were inoculated and purified plasmids sequenced to verify for the mutations.

2.4 Protein biochemistry

2.4.1 SDS and Native PAGE

Proteins were separated based on apparent molecular weight, under reducing, denaturing conditions by discontinuous SDS-PAGE was used. 12.5% SDS gels were prepared according to the gel composition listed above. Samples were mixed with 5x SDS-sample buffer, boiled for 5 min at 95 °C and centrifuged for 2 min at RT at 17,600 x g before loading onto the gel. Gel electrophoresis was carried out in Mini Protean (Bio-rad) electrophoresis chambers in the presence of electrophoresis migration buffer employing a starting constant voltage of 120 V followed by 160 V for the rest of the run.

Native gels were prepared according to the gel composition listed above which does not include SDS. Samples were mixed with 5x native loading dye which does not include SDS and the reducing agent β -mercaptoethanol. Samples were not boiled and the gel and tank buffer solutions were without SDS. An electric field of 160 V was applied across the native PAGE gel for 100 min. Both sample loading and run were conducted at 4 °C.

Solutions	SDS PAGE (4 gels)			Native PAGE (4 gels)	
	12.5% Resolving gel	6% Resolving gel	4% Stacking gel	6% Resolving gel	4% Stacking gel
Deionized water (ml)	4.22	7.11	4.27	7.24	4.30
1.5M Tris-HCl, pH 8.8 (ml)	3.33	3.33	-	3.33	-
0.5M Tris-HCl, pH 6.8 (ml)	-		1.75	-	1.75
10% (w/v) SDS (μl)	133.33	133.33	70.00	-	-
30% Acrylamide /Bis solution 37.5:1 (ml)	5.54	2.66	0.91	2.66	0.91
10% (w/v) Ammonium persulfate (μl)	66.67	66.67	35.00	66.67	35.00
TEMED (μl)	4	4	7	4	7

Table 2-6: Components of SDS and Native gels.

2.4.2 Coomassie blue staining of SDS and native PAGE

Coomassie blue staining was used for the visualization of protein bands in polyacrylamide gels. Protein bands were fixed and stained in the gel matrix by incubating the gels for 1 h with Coomassie blue staining solution. Background staining was removed by incubation with Coomassie destaining solution for 1-3 h.

2.4.3 Western Blotting and Immunodetection

Western blot followed by specific immunodetection was employed for specific identification of protein bands after SDS and native PAGE. Separated proteins were transferred onto a nitrocellulose membrane in a semi-dry western blot apparatus (BioRad) in transfer buffer at a current of 15 volts per gel for 15 min. The nitrocellulose membranes containing proteins were stained with Ponceau S to ensure efficient transfer, before being blocked for an hour at RT with 5% (w/v) skim milk powder in TBS. The membranes were incubated with primary antibodies (1/5,000) diluted in TBST overnight at 4 °C, with shaking. The various primary antibodies used included

monoclonal mouse anti-FLAG (Sigma Aldrich, F1804), mouse monoclonal anti-HA (Protein tech, 66006), rabbit polyclonal anti-AfM, anti-form I Rubisco and anti-O1O2VWA. The membranes were washed thrice with TBST for 10 min each and subsequently incubated with horseradish peroxidase (HRP)-conjugated secondary antibodies (1/1,000 dilution) for 1 hr at RT. The secondary antibodies used were goat anti-mouse (Thermo scientific, 32430) and goat anti-rabbit (Sigma-Aldrich, A6154). The membranes were washed thrice with 0.001% (v/v) Tween-20/Tris-buffered saline (TBST) as described above before signal detection. Signals from hrp-conjugated secondary antibodies were detected with chemiluminescence (ECL) detection reagents (Pierce) with either LAS-3000 Imaging system (Fuji) or film.

2.4.4 Protein expression and purification

All CbbQ and CbbO proteins used in this studied were purified according to the protocol established by Dr. Candace Tsai (Tsai et al., 2015). To obtain the hetero-oligomeric CbbQO, AfCbbQ and AfCbbO proteins were co-expressed in *E. coli* BL21 (DE3) cells and co-purified as His₆-ubiquitin fusion proteins. The cells harbouring both pHue*UbcbbQ* and pBAD33*UbcbbO* were grown in LB supplemented with respective antibiotics (200 µg/ml ampicillin for pHUE vector, 34 µg/ml chloramphenicol for pBAD33 vector) at 37 °C in an Innova44R incubator (New Brunswick Scientific) with moderate shaking (200 rpm). When the culture has reached an optimum optical density, the temperature of the incubator was lowered to 23 °C and final concentration of 0.5 mM isopropyl β-D-1-thiogalactopyranoside (IPTG) and 0.2% w/v L-arabinose was added to start the induction. The cells were harvested after 16 hours by centrifugation in an Avanti J-25 centrifuge with JA 10 rotor (Beckman Coulter) at 4,000 rpm for 20 min. The pellet was re-suspended in 40 ml of lysis buffer and 0.3 mg/ml lysozyme for 30 min on ice. Cells were disrupted by ultrasonication with a Sonicator (20 x 15 sec bursts, 70 W, 30 sec cooling on ice between bursts) after the addition of 500 mM Phenylmethylsulfonylfluoride (PMSF) and 1 mM Ethylenediaminetetraacetic acid (EDTA). The supernatant obtained from centrifugation in an Avanti J-25 with JA25.50 rotor (Beckman Coulter) (20,000 x g, 45 min, 4 °C) was applied to Ni²⁺-nitrilotriacetic acid resin (Ni-NTA) (Thermo Scientific) equilibrated with purification lysis buffer. The His₆Ub-fusion protein was washed with the purification wash buffer to prevent non-

specific binding, and eluted using elution buffer. The ubiquitin moiety was cleaved overnight at 4 °C using the deubiquitylating enzyme Usp2 to yield the native N-terminus (Baker et al., 2005). Significant precipitation was observed during purification for some CbbQO mutants, the precipitated proteins were removed at this stage by centrifugation and the soluble fraction was applied to a Source 15Q anion exchange column (GE Healthcare) pre-equilibrated with buffer A, and eluted with a linear salt gradient to 0.5 M NaCl. A flow rate of 1 ml/min was used. Fractions were analysed using 12% SDS-PAGE.

Typically, protein purification of CbbQO-type activases will yield 1:2 ratio of CbbO (~80 kDa) to CbbQ (~30 kDa) when analysed with SDS-PAGE, assuming equal binding of Coomassie dye (Tsai et al., 2015). During purification of hetero-oligomers, fractions containing both CbbQ and CbbO in the correct ratio were combined and concentrated to 5 mL using Amicon Ultra-15 Centrifugal Filter Units (50 kDa cut-off, Merck, USA). The combined fraction was applied to a Superdex 200 16/600 size-exclusion column (GE Healthcare, USA) and proteins were eluted with buffer A. Flow rate of 1 ml/min was used. Fractions were analysed with 12% SDS-PAGE and those containing purest protein fractions (>95% pure as judged by SDS-PAGE) were pooled and concentrated to ~ 3-4 mg/ml before being flash-frozen in liquid nitrogen. All CbbQ2 and CbbO2 mutants were purified with the addition of 5 mM MgCl₂ and 5% glycerol in the purification buffers used in Ni-NTA agarose and anion exchange columns.

2.4.5 Protein expression test for FLAG-AfO2VWA

E. coli BL21 cells carrying the pHUE vector (Catanzariti et al., 2004) for expression of His₆Ubiquitin-FLAG-AfO2VWA were grown in LB supplemented with 200 µg/mL of Ampicillin at 37 °C in an Innova44 incubator (New Brunswick Scientific) at moderate shaking (220 rpm) overnight. Five millilitre LB media supplemented with 200 µg/mL of Ampicillin was inoculated with 50 µl of the overnight culture and grown at 37 °C to an optical density of ~0.6 at a wavelength of 600 nm (OD₆₀₀). The temperature of the incubator was lowered to 23 °C and final concentration of 0.5 mM IPTG was added to start the induction. Respective controls were left uninduced under identical conditions. For co-expression of chaperonin GroEL/ES to enhance solubility, IPTG was added an hour later after addition of 0.2% w/v L-arabinose to the culture. Cells were pelleted down

after 16 hours and re-suspended in 1 ml buffer A supplemented with 10 mM imidazole, incubated at 4 °C for 30 min and lysed on ice by sonication. 80 µl of the lysate were added to 20 µl 5x SDS sample buffer for analysis of total protein expression. 100 µl of the lysate were clarified by centrifugation (16,000 x g, 15 min and 4 °C) and 80 µl of the supernatant were added to 20 µl 5x SDS sample buffer for analysis of soluble protein expression. The residual supernatant was discarded and the pellet was washed with 200 µl deionized water and spun down again (16,000 x g, 15 min and 4 °C). The pellet was re-dissolved in 125 µl 1x SDS sample buffer for analysis of insoluble protein expression. Samples were heated at 95 °C for 5 min and analysed by 12.5% SDS-PAGE.

2.4.6 Purification of FLAG-AfO2VWA from the inclusion bodies

E. coli BL21 cells carrying the pHUE plasmid for expression of His₆Ubiquitin-FLAG-AfO2VWA were grown in 1L of LB media supplemented with 200 µg/mL Ampicillin and grown to an OD₆₀₀ of 0.5 at 37 °C. The temperature of the incubator was lowered to 23 °C and final concentration of 0.5 mM IPTG was added to start the induction. The cells were harvested after 16 hours by centrifugation in an Avanti J-25 centrifuge with JA 10 rotor (Beckman Coulter) at 4,000 rpm for 20 min. The pellet was re-suspended in 40 ml of inclusion body lysis buffer and 0.3 mg/ml lysozyme for 30 min on ice. Cells were disrupted by ultrasonication with a Sonicator (20 x 15 sec bursts, 70 W, 30 sec cooling on ice between bursts) after the addition of 500 mM PMSF and 1 mM EDTA. The supernatant obtained from centrifugation in an Avanti J-25 with JA25.50 rotor (Beckman Coulter) (20,000 x g, 30 min) was discarded. The pellet was transferred to a glass homogenizer and pressure was applied manually to re-suspend pellet in 30 ml of inclusion body wash buffer A. The re-suspended pellet was centrifuged in an Avanti J-25 with JA25.50 rotor (Beckman Coulter) (20,000 x g, 30 min). The re-suspension and centrifugation step were repeated again with inclusion body wash buffer B. The final supernatant was discarded and the purified inclusion bodies were dissolved in 20 ml of denaturation buffer. His₆Ubiquitin-FLAG-AfO2VWA was quantified by their absorbance at 280 nm by Nanodrop and the concentration was calculated using the extinction coefficient calculated with the program, ProtParam. Final concentration of the protein is calculated using extinction coefficient and molecular weight.

$$C \text{ (mg/ml)} = (A_{280}/\epsilon) \times \text{Molecular Weight}$$

2.4.7 Refolding of FLAG-AfO2VWA

Sample was diluted to 1 mg/ml in 10 ml of denaturation buffer and dialyzed against 5 L refolding buffer for 16 h at 4 °C. Sample was centrifuged (4,000 x g, 10 min) to remove precipitates formed. The ubiquitin moiety was cleaved overnight at 4 °C using the deubiquitylating enzyme Usp2 to yield the FLAG-tag at the N-terminus of the CbbO2VWA domain. The refolded protein was loaded onto the Superdex 200 16/600 size exclusion column for further purification. Runs were performed at RT at a flow rate of 1 ml/min. Five millilitre fractions were collected and analysed with SDS or native PAGE gels. The purest protein fractions (>95% pure as judged by SDS-PAGE) were pooled, supplemented with 5% v/v glycerol and flash-frozen in liquid nitrogen. Protein aliquots are stored at -80 °C.

2.4.8 Circular dichroism

A Chirascan Circular Dichroism Spectrometer (Applied Photophysics) using a Quartz cuvette with 0.1 cm pathlength was used to record spectra between 190 and 260 nm in 0.5 nm steps with an averaging time of 2.0 s/step at 25 °C. Protein concentration was 0.15 mg/ml in CD buffer (20 mM K₂HPO₄/KH₂PO₄, pH 8.0). The measured ellipticity, θ , in millidegrees (mdeg) was converted to the mean molar ellipticity per residue, $[\theta]$, in degrees $\times$ cm²/dmol by using the equation $[\theta] = \theta \times 100 / (c \times l \times N)$, where c is the molar concentration of the sample in mM, l is the pathlength in cm, and N is the number of amino acids. The recorded ellipticity values were the average of three determinations for each sample. Secondary structure content was determined using the K2D2 algorithm (Perez-Iratxeta and Andrade-Navarro, 2008).

2.4.9 Analytical size-exclusion chromatography

Analytical size-exclusion chromatography of proteins was performed on a Superdex 200 increase 3.2/30 GL analytical size exclusion column using an ÄKTAmicro chromatography system (both GE Healthcare). Fifty microlitres of protein-containing sample was injected into the column pre-equilibrated with either buffer A or the cross-linking buffer. Runs were performed at RT at a flow rate of 0.075 ml/min. Fifty

microlitre fractions were collected and analysed with SDS or native PAGE gels. The column was calibrated with globular proteins of known molecular weight (ovalbumin 43 kDa, conalbumin 75 kDa, aldolase 158 kDa, ferritin 440 kDa, thyroglobulin 669 kDa), provided by the calibration kit (GE Healthcare).

2.5 Enzymatic assays

2.5.1 Formation of ECM/ECMC/ER

The Rubisco active ternary complex ECM was obtained by incubating Rubisco (100 μ M monomer) in buffer A supplemented with 40 mM NaHCO₃ and 10 mM MgCl₂ for 10 min at 25 °C, while ECMC (20–100 μ M active sites depending on desired final concentration) was prepared by further addition of 400 μ M (fourfold molar excess) of 2-Carboxyarabinitol-1,5-bisphosphate (CABP) to the ECM complex for 10 min. The ER complex was obtained by incubating rubisco in buffer A with 4 mM EDTA for 10 min and then adding 0.8 mM RuBP.

2.5.2 ATPase Assay

ATPase activity was measured spectrophotometrically using a coupled enzymatic assay monitoring the oxidation of reduced nicotinamide adenine dinucleotide (NADH) (Kreuzer and Jongeneel, 1983; Tsai et al., 2015). The reactions were carried out in 100 μ l reactions at 25 °C with the assay buffer, in the presence of 2 mM phosphoenolpyruvate, 0.5 mM NADH, 2 mM ATP, 20 U/ml pyruvate kinase, 30 U/ml lactate dehydrogenase and 0.27 mM CbbQO oligomer. ECM, ER and ECMC (3 μ M active sites) were added as indicated. The oxidation of NADH, which is directly coupled to the generation of ADP, was followed by the reduction in the absorbance at 340 nm on a UV-1800 spectrophotometer (Shimadzu).

2.5.3 Rubisco Assay

Rubisco activity assay was performed as previously described in (Kubien et al., 2011; Tsai et al., 2015). Measurements were carried out in 100 μ l reaction volume with assay

buffer in the presence of 3 μ l coupling enzymes mixture (creatine P-kinase, glyceraldehyde-3-P dehydrogenase, 3-phosphoglycerate kinase and triose-P isomerase/glycerol-3-P dehydrogenase), 20 mM NaHCO₃, 0.5 mM NADH, 2 mM ATP, 10 mM creatine-P, 1 mM RuBP, 5-13% v/v of the crowding agent polyethylene glycol 3,350 (PEG) and a final concentration of 0.3 μ M Rubisco active sites. CbbQO Rca (0.27 μ M oligomer) was added as indicated. The oxidation of NADH, which is directly coupled to the generation of 3PG, was followed by the reduction in the absorbance at 340 nm on the UV-1800 spectrophotometer (Shimadzu).

2.6 Chemical cross-linking

2.6.1 Glutaraldehyde cross-linking

Protein concentrations in stock solutions were measured using the Bradford dye binding assay (BioRad) using bovine serum albumin as a standard. As Tris buffers contain primary amines that interfere with cross-linking reactions, purified proteins that were in Tris buffer were buffer exchanged to cross-linking buffer using Bio-Spin6 chromatography column (Bio-Rad) according to manufacturer's instructions. Proteins were cross-linked in the presence or absence of 5 mM ATP and 5 mM MgCl₂. A final concentration of 0.125% v/v of glutaraldehyde was used for crosslinking for 10 min at RT. The crosslinking reaction was quenched by addition of 1 M Tris-HCl, pH 8.0 to a final concentration of 50 mM. Samples were analysed by SDS and Native PAGE or subjected to analytical size-exclusion chromatography as described above.

2.6.2 Dithiobis(succinimidyl propionate) (DSP) crosslinking

Stock solutions of 25 mM DSP (5 mg dissolved in 542 μ l of DMF) were made immediately before use. 1 mM of DSP cross-linker was added to the protein samples in the cross-linking buffer for 30 min at room temperature. The crosslinking reaction was stopped by the addition of 1 M Tris-HCl, pH 8.0 to a final concentration of 50 mM. To cleave DSP, the reducing agent DTT was added to the samples to a final concentration of 50 mM, incubated at 37 °C for 30 min and analysed with SDS PAGE.

2.6.3 Cross-linking coupled to mass spectrometry (XL-MS)

To investigate the subunit organization of the CbbQO hetero-oligomer, 100 µl of sample containing 1 mg/ml of purified Q1O1 or Q2O2 was directly cross-linked by the addition of 1.5 mM isotope-labelled chemical cross-linker, d₀/d₁₂ (DSS) for 30 min at 37 °C with shaking at 330 rpm. Reactions were quenched by the addition of 50 mM NH₄HCO₃ for 10 min at 37 °C.

To understand more about the interaction between Rubisco and the CbbQO Rubisco activase, purified HA-AfM ECMC (13 µM monomer) was incubated with FLAG-Q2^{WB}O2 (104 µM monomer) in 500 µl of cross-linking for 30 min at RT. Equimolar of d₀/d₁₂ DSS cross-linker to lysine residues in the protein sample (assuming 50 nmol of lysines in 100 µg of protein) were added. The sample was loaded onto a Superdex 200 increase 10/300 gel filtration column to enrich for the cross-linked complexes. Fractions collected were separated on 6% Native Gel to determine which fractions contained cross-linked complexes. Appropriate fractions were pooled. All samples were dried in a heated vacuum before being sent to Prof. Dr. Florian Stengel (University of Konstanz) for mass spectrometric analysis. Cross-linked peptides were identified with xQuest (Rinner et al., 2008) and xProphet (Walzthoeni et al., 2012) programs.

2.7 Negative stain EM

2.7.1 Sample preparation

Samples for negative stain EM were loaded onto an analytical gel filtration column and the fraction corresponding to the centre of the major peak or the HMW peak containing Rubisco-activase complexes were diluted to a final concentration of 0.025-0.05 mg/ml in either buffer A or the cross-linking buffer. Samples were centrifuged at 4 °C and 16,000x g for 10 min, in order to sediment large aggregates, and immediately used for EM grid preparation.

2.7.2 Grid preparation and data collection

Commercial 200-mesh Cu-Rh grids coated with 6 nm carbon (Electron Microscopy

Sciences) were glow discharged for 1 min after which 4 μ l of the protein sample was applied to the grid. After 1 min incubation time, the sample was blotted from the grid using Whatman filter paper. 2% uranyl acetate was immediately added to the grid and incubated for 1 min. The excess stain was carefully wicked off from the grid by the touching the edge of the filter paper. The grid was further air dried for at least 10 minutes before loading into the microscope. Negative stain EM grids used in this study were stained by Dr. Tobias Wunder from our lab. The images were captured with a FEI Technai T12 microscope equipped with a 4K Eagle CCD camera at 49,000x magnification and electron dose of 20-25 e-/Å², resulting in a pixel size of 2.11 Å per pixel. Defocus range of -1.0 to -1.5 μ m was used.

2.7.3 Data processing

For the Q2O2 dataset, a total of 49 micrographs were collected. Particles were picked automatically with the RELION 3 auto-pick function. A total of 32,879 particles were used for reference-free unbiased 2D classifications after which 15,386 particles from selected 2D classes were subjected to 3D reconstructions. The highest quality model was used for further refinement to obtain a 3D EM map of Q2O2 at 16 Å resolution. Data processing was done with the help of Asst. Prof. Shashi Bhushan, SBS NTU.

An AfLS ECMC-Q3O1 dataset containing 87 micrographs and 8,332 particles was processed by the software EMAN2. Particles were picked semi-automatically with the ‘Swarm’ mode in EMAN2 together with manual inspection. Unbiased 2D classifications yield classes of mostly one or two Rca bound to Rubisco. Well-defined classes of one Rca bound to Rubisco (3234 particles) were selected for 3D reconstructions. The highest quality model was selected for further refinements to obtain a final 3D EM map of 21.1 Å.

2.8 Cryo-EM

2.8.1 Sample preparation

The AfLS ECMC-Q3O1 complex was prepared by incubating 16 μ M Rubisco (monomer) to 64 μ M Q3O1 (monomer) in the presence of 5 mM MgCl₂ and 5 mM ATP

for 2 min at RT. Cross-linking was initiated with the addition of 0.0125% glutaraldehyde and the reaction was allowed to proceed for 10 min at RT before quenching with 1 M Tris-HCl, pH 8.0 to a final concentration of 50 mM. Samples were loaded onto an analytical gel filtration column to enrich for cross-linked complexes and eluted proteins were analysed with 6% native PAGE. Fractions determined to contain complexes in the form of HMW bands were stained immediately for negative stain EM to verify the quality of samples. As the size of the form I Rubisco-activase complexes permitted visual identification without alignment, negative stain EM was a good method to quickly assess the suitability of samples for high resolution cryo-EM.

2.8.2 Grid preparation and data collection

Using a FEI Vitrobot, 4 μ l of crosslinked AfLS ECMC-Q3O1 complexes (1.4 mg/ml) prepared as above were applied on glow discharged Quantifoil R2/2 holey carbon grids with 2 nm carbon support film for 30 s at 4 °C and 100% humidity before blotting for 2 s and vitrification. The grids were plunged into liquid ethane and were carefully transferred to a grid box in the liquid nitrogen compartment. Once prepared, the grids were continuously maintained at liquid nitrogen temperature. Cryo-EM images were acquired under low-dose conditions at 300 kV on the in-house Titan Krios transmission electron microscope equipped with a Gatan K2 detector. Micrographs were collected at a nominal magnification of 130,000 with defocus ranging from -500 to 3,000 nm and sampled at 1.1 Å per pixel at the specimen level. 40 frames were recorded over 1.25 s giving an accumulated dose of 50 electrons/Å².

2.8.3 Data processing

All image processing and 3D reconstruction were performed using RELION-3 (Scheres, 2012a; Zivanov et al., 2018) with the help of Asst. Prof Shashi Bhushan of SBS, NTU. Micrographs with strong astigmatism, drift or broken ice were discarded, leaving a total of 4,943 good micrographs for whole-image drift correction using Motioncorr (Li et al., 2013). Contrast transfer functions were determined by CTFFIND4 (Rohou and Grigorieff, 2015). Approximately 2,000 manually selected particles were 2D-classified and a 2D class average of AfLS ECMC with Q3O1 bound was used as template for semi-automated particle selection (Scheres, 2015) resulting in an initial dataset of

1,103,387 particles. 2D classification with 4X decimated particles generated classes featuring both unbound Rubisco and Rubisco-activase complexes. Well-defined classes of complexes featuring one activase bound to Rubisco. (77,240 particles) were selected for 3D refinement. The negative-stain EM structure of the AfLS ECMC-Q3O1 complex was filtered to 40 Å resolution as a reference. A soft mask with 320 Å diameter was applied. This refinement generated an initial reconstruction of the complex. The initial reconstruction was filtered to 20 Å resolution and applied as a mask for subsequent 3D classification which yielded 5 classes. The majority class with 47% particle distribution resolved to ~11 Å was selected for subsequent 3D refinements and post-processing to produce a final map at 9.8 Å.

Q3O1 was subjected to focussed classifications using the particles selected from well-defined 2D class averages of one activase bound to Rubisco (77,240 particles). Signal subtraction was used to direct refinements and 3D classification of Q3O1 (Bai et al., 2015a). To this end, a low pass filtered reconstruction of the Rubisco-activase complex, with the density of Rubisco removed was used to produce a soft mask for signal subtraction as well as focused classification. The classification produced five classes with two dominant classes harboring 98% particle distribution (75,690 particles). Particles from the two majority classes were combined and further refined and subjected post-processing to yield the final 9.1 Å map.

In the case of the smaller initial dataset collected for the AfLS ECMC-Q3O1 complexes, 2,729 micrographs were used for data processing using a similar approach. Selected 2D classes from manual picking were used as templates for RELION autopicking. The initial particle extraction contained 647,661 particles which were used for 2D classifications. As most classes generated featured only unbound Rubisco particles, well-defined classes totaling 390,625 particles were selected for downstream processing. Three round of 3D refinements and classifications were conducted to yield 5 classes. Class 5 with 60% particle distribution (87,039 particles) resolved to ~4 Å was selected for further refinement and post-processing to produce the final 2.9 Å map. All refinements reported here used gold standard Fourier shell correlation (FSC) calculations and reported resolutions are based on the FSC = 0.143 criterion of mask-corrected FSC curves (Rosenthal and Henderson, 2003). All final maps were masked and sharpened using automatically determined negative B-factors.

2.8.4 Model building for AfLS

The sequences of AfRbcL and AfRbcS were manually built into the B-factor sharpened electron density with the program Coot (Emsley et al., 2010), using the crystal structure of the highly homologous *H. neapolitanus* Rubiso (HnLS) (PDB-ID: 1SVD) as template. An atomic model of one subunit of AfLS was initially generated with Chainsaw (Stein, 2008) and rigid-body fitted it into the density map by UCSF Chimera (Pettersen et al., 2004). Adjustments were made manually in Coot with Ramachandran restraints were applied throughout and secondary structure restraints applied whenever appropriate. During modelling in Coot, the model was refined with both the ‘Regularize Zone’ and ‘Real Space Refine Zone’ to obtain optimal fit into the EM density and maintain the correct protein geometry. Due to differences in conformation, RbcL residues 164-176 and the C-terminal residues (431-460) did not fit into the map and was built *de novo* with a poly-alanine model followed by subsequent amino acid assignment based on the sequence and guided by defining densities of bulky side chains. The density for N-terminus of AfRbcS contains density that extends beyond the model. Extensions were built manually into the density by using the tool ‘add terminal residue’. To further improve the fitting of AfLS into the map, the model was refined with *phenix.refine* (Afonine et al., 2012) using *phenix.real_space_refine*. After refining, the Ramachandran outliers, bad rotamers and geometry angles were corrected in coot. Model validation for the AfLS model was done with the Molprobtity server (Chen et al., 2010). Visualization of cryo-EM maps and models were performed using Chimera.

2.9 Bioinformatics

2.9.1 Bioinformatic characterization of CbbQ orthologs

Chemoautotrophs with CbbO isoforms encoded within the *cbb* operon gene clusters were selected by combining search results from non-redundant homology searches with the Basic Local Alignment Search Tool (BLAST) and the Uniprot and UniParc database. The isoform number (CbbQ1, Q2, or Q3) was determined by the location of the gene encoding isoforms from the National Centre for Biotechnology Information (NCBI) database. Protein sequences were aligned using the ClustalW program (Chenna et al., 2003). The percentage conservation was mapped onto the surface of CbbQ3 using

Chimera ‘Render by attribute’ function. Surface charge of CbbQ2 was calculated using the APBS plugin in PYMOL.

2.9.2 Bioinformatic analysis of CbbO sequences

Chemoautotrophs with CbbO isoforms encoded within the cbb operon gene clusters were selected by combining search results from non-redundant homology searches with the Basic Local Alignment Search Tool (BLAST) and the Uniprot and UniParc database. The isoform number (CbbO1, O2, or O3) was determined by the location of the gene encoding isoforms from the National Centre for Biotechnology Information (NCBI) database. Protein sequences were aligned using the ClustalW program (Chenna et al., 2003). Colour-coded figures of alignments were generated with the program Esript (Robert and Gouet, 2014). The phylogenetic tree was generated with the ClustalW alignment using the Geneious Pro software (Kearse et al., 2012).

3. ANALYSIS OF THE HETERO-OLIGOMERIZATION OF THE AAA+ ATPASE CBBQ AND THE ADAPTOR PROTEIN CBBO

The CbbQO-type Rca in the chemoautotrophic proteobacteria *A. ferrooxidans* is unique among activases in consisting of both a AAA+ motor CbbQ and a substrate adaptor CbbO. While the crystal structure of CbbQ is solved, CbbO remains enigmatic with only a VWA domain identified at its C-terminus. In the absence of structural information for CbbO, the molecular architecture and the mechanism by which the hetero-oligomeric CbbQO remodels its substrate Rubisco continues to be obscure. In this chapter, negative stain EM analysis of the structure of Q2O2 was done to build on the pioneering biochemical characterization. In parallel, cross-linking coupled to mass spectrometry (XL-MS) was used to validate our EM models and decipher more details on the assembly of this new activase. Bioinformatics analysis of CbbO across different classes of chemoautotrophic bacteria reveal conserved stretches in the primary sequence that could be responsible for subunit oligomerization and interactions with Rubisco. Finally, progressive truncations on the closely related homologue from *H. neapolitanus* identified key regulatory regions on CbbO important for association with CbbQ and regulating ATP activity.

3.1 Purification of the form II Rubisco activase Q2O2

Recent studies have identified and established Q2O2 as the Rca specific for activating the form II Rubisco in *A. ferrooxidans*, AfM under low CO₂ conditions (Tsai et al., 2015). As such, the groundwork has been laid to explore the structural arrangements and the mechanism of this new class of activase. For further studies on Q2O2, I purified the hetero-oligomer recombinantly from *E. coli* BL21 using the purification protocol previously optimized by Dr. Candace Tsai. CbbQ2 and CbbO2 form hetero-oligomers upon co-expression. The proteins were both expressed as a fusion protein with a His₆Ubiquitin (His₆Ub) tag at the N-terminus to promote solubility and enable initial purification by Ni-NTA beads (Figure 3-1A). After cleavage of the His₆Ub tag by the specific protease USP2, a significant amount of precipitation is usually observed. The soluble fraction is then subjected to ion-exchange chromatography (IEX) and size-exclusion chromatography (SEC) to enrich for hetero-oligomers. Typically, protein purification of CbbQO-type activases will yield 1:2 ratio of CbbO (~80 kDa) to CbbQ

(~30 kDa) band intensity when analysed using SDS-PAGE, assuming equal binding of Coomassie dye. After SEC, the fraction determined to contain the most hetero-oligomeric complexes was concentrated to yield ~ 10 mg of pure protein (Figure 3-1B).

Figure 3-1: Purification of Q2O2.

(A) Representative 12.5% SDS PAGE separation of proteins from IMAC Ni-NTA and various fractions from IEX Source15Q column. L: lysate, S: soluble, FT: flow through, W: wash, BC: before cleavage by USP2, AC: after cleavage. * Represents fractions that were pooled for SEC. (B) Size-exclusion profile of Q2O2 on a Superdex 16/600 column (left). Purified fractions on a 12.5% SDS PAGE with CbbQ2 and CbbO2 subunits indicated (right).

3.2 Negative stain EM analysis of the form II Rubisco activase Q2O2

Even though the crystal structure of AfQ2 is available (Tsai and Ye et al., unpublished), the lack of structural information on AfO2 hampers understanding of the structural organization of the AfCbbQO hetero-oligomeric complex. To address this gap in knowledge, I examined Q2O2 in the APO state by negative stain EM and single particle analysis.

Given that Q2O2 is an ensemble of a hexameric activase and an 80 kDa adaptor protein, hexamers were not identified from individual particles on the micrographs (Figure 3-2A). Two-dimensional (2D) class averages (32,879 particles) however yielded clear densities for some subunits of the hexameric CbbQ2, depending on their orientations (Figure 3-2B). Interestingly, in almost all selected classes, an additional density was identified in the centre of the hexamer which in some classes extend to the periphery of the ring. This was not observed previously in the negative stain study with CbbQ2 alone and must be attributed to CbbO2 (Tsai et al., 2015). Density for CbbO2 proximal to the rim of the hexamer is less defined than that in the central cavity or those of CbbQ2, suggesting structural flexibility and variations in position relative to the hexameric ring. These observations support the proposed model of a C-terminal VWA domain tethered to a flexible linker region.

E

Figure 3-2: Negative-staining EM image analysis of Q2O2.

(A) Representative micrographs of Q2O2 (0.05 mg ml⁻¹) in APO state. Scale bar, 50 nm. (B) Selected reference free 2D class averages of Q2O2 derived from negative stain images in RELION-3. Scale bar, 10 nm. (C) 3D reconstruction of Q2O2 (top) and 3D reconstruction of Q2O2 fit with the EM reconstruction of CbbQ2 (EMDB: EMD-6477, (Tsai et al., 2015)) showing additional density on the concave face of the AAA+ ring attributed to CbbO2 (centre). 3D reconstruction of Q2O2 fit with crystal structure of CbbQ2 (pink ribbon; Tsai and Ye et al., Unpublished) and VWA domain of Integrin α L (Song et al., 2005) (blue ribbon; PDB-ID: 1T0P) (bottom). (D) Angular distribution of the dataset. Each cylinder represents one view, and the height is scaled to the number of particles in that view. (E) Fourier shell correlation (FSC) curves of Q2O2 as determined by gold standard FSC procedure in RELION-3 (0.143 criterion). The final resolution was estimated to be 16 Å.

Particles of selected 2D classes (15,386 particles) were subsequently subjected to 3D reconstruction and a well-defined initial model was refined (no symmetry applied) to a final resolution of 16 Å (Figure 3-2C, D, E). The 3D map when fitted with the electron density of the CbbQ2 hexamer (EMDB: EMD-647; (Tsai et al., 2015)) revealed extra densities on the concave side, corresponding to CbbO2 (Figure 3-2C; centre). From the map, CbbO2 extends from rim of the hexamer to the central cavity, occluding the central pore. The adaptor makes contact with at least three subunits of CbbQ and as a result, these three subunits are not clearly defined in the model. The other three adjacent subunits are clearly visible. Importantly, the density in the central cavity of CbbQ is reminiscent of the VWA domain docking on the AAA+ ring of the yeast ribosome maturation factor Rea1 and Mdn1 (Midasin), suggesting a similar arrangement (Chen et al., 2018; Sosnowski et al., 2018).

CbbQ which belongs to the Helix-2-insert (H2I) clade of AAA+ proteins, includes a β -hairpin insertion within the $\alpha 2$ of the ATPase core as well as between $\alpha 3$ and $\beta 4$ (PS1- β H). Recent biochemical data from our lab implicates residues of the H2I and PS1- β H of CbbQ2 which project into the central cavity and fill the pore in regulating the ATP hydrolysis rate of the AAA+ motor (Tsai and Ye et al., Unpublished). This is consistent with our model where CbbO is featured on the concave face of the ring.

3.3 Chemical cross-linking coupled to mass spectrometry (XL-MS) analysis of CbbQO activase

Information about dynamics and interactions within the multi-subunit CbbQO activase is a prerequisite to understanding the mechanistic details of Rubisco activation. In this regard, knowledge about the subunit arrangement and organisation is still rather limited. XL-MS of peptides is a technique that can provide spatial amino acid distance constraints imposed by the length of the spacer of the cross-linking agent, to obtain low resolution (10-30 Å) structural information (Leitner et al., 2016). I aimed to use this approach to determine the molecular architecture of the CbbQO-type activases. To this end, the wild type (WT) Q2O2 and Q1O1 were cross-linked with the lysine-specific, isotope-labelled chemical cross-linker, d_0/d_{12} DSS. A low protein concentration of ~1 mg/ml was used to prevent non-specific cross-linking. The residues involved in association of the subunits were analysed by mass spectrometry in collaboration with Prof. Dr. Florian Stengel at the University of Konstanz.

The spacer arm length of DSS is 11.4 Å. Considering the length of flexible lysine side chains, the maximum span allowed by the cross-linking with DSS is approximately 24 Å (Young et al., 2000). Therefore, the distance between the cross-linked lysine pairs should be within 24 Å. Cross-linking data generated included: Intra-links (cross-links within one subunit) and inter-links (cross-links between two different subunits) (Appendix 8.1). Based on the hexameric crystal structure of CbbQ2, intra-links within the AAA+ motor generally satisfied the distance constraints of the DSS cross-linker. The extensive inter-subunit self-crosslinks observed for AfQ1 and AfQ2 (Figure 3-3A, B; red) are consistent with the observation that the CbbQ subunits are arranged in an

oligomeric assembly, whereas the lack of such inter-subunit self-links in CbbO (except one in AfO1) would mean that it exists in a monomeric form. From two independent experiments, a total of 55 distinct inter-linked lysine pairs (Id score > 20, See Appendix table 8-1) were identified for Q2O2 with many cross-links identified multiple times. In addition, 19 unique cross-links were detected for Q1O1 in a single experiment.

A

B

Figure 3-3: Cross-linking coupled to mass spectrometry analysis (XL-MS).

(A) Cross-link map for the Q2O2 hetero-oligomeric complex. (B) Cross-link map for the Q1O1 hetero-oligomeric complex. Green and orange cross-links represent intermolecular links. Red and purple connections represent inter- and intra-subunit self-cross-links, respectively. VWA domains are in light green. (C) Cross-linking sites detected in CbbQ2 are highlighted in the crystal structure shown in surface representation. Lysine residues cross-linked to CbbO2 are shown in cyan and orange for the concave and convex face, respectively. All cross-linked Lysine residues highlighted are labelled except K137, which is deep within the AfQ2 pore. The only lysine on CbbQ2 not cross-linked to CbbO2, K16 is shown in yellow.

The Q2O2 and Q1O1 inter-links were evaluated to determine how the hexameric CbbQ associates with the monomeric CbbO. The cross-linking experiment showed that both the N- and C-termini of CbbO are in proximity to CbbQ (Figure 3-3A, B). Notably, the C-terminal VWA domains of CbbO1 and O2 are shown to cross-link extensively to the concave surface of CbbQ, suggesting that they are in close proximity. This result is consistent with the 3D model from negative stain EM where the density occupying the central cavity is proposed to be the VWA domain of CbbO2.

To understand more about the interface between CbbQ and CbbO, the crystal structure of CbbQ2 was inspected to localize the cross-linked lysines. Although lysine residues on both the concave and convex face of the hexamer were implicated in cross-linking to CbbO2, 38 out of 55 unique cross-links between CbbQ2 to CbbO2 were located on

the concave face of the hexamer (Figure 3-3A, C). Notably, both the N- and C-terminal regions of CbbO2 were found to cross-link multiple times to CbbQ2 K215, located on the rim of the concave face (Figure 3-3C). Similarly, CbbO1 lysines were found to cross-link extensively to the corresponding residue in CbbQ1 (K211) (Figure 3-3B). Thus, it can be inferred that the N- and C- termini of CbbO are within proximity to each other at the concave face of CbbQ, which has been shown to be highly conserved (Sutter et al., 2015). The intra-molecular cross-links of CbbO1 and O2 also support the finding that CbbO may fold back upon itself, with the C-terminal VWA domain docking on the concave face of CbbQ (Figure 3-3A, B).

Out of the 17 cross-links identified to be on the convex side of the hexamer, 6 were identified to be cross-linked to lysines on the outer edge of the hexamer, K36 and K63 (Figure 3-3C). These lysines were generally cross-linked to the N-terminus of CbbO2 (Figure 3-3A), which based on our negative stain model is located near the rim of the CbbQ2 cup. It is possible that the flexibility and dynamic movements of the N-terminal linker of CbbO enables it to make contact with residues on the opposite face of the CbbQ. Another 6 lysine residues on CbbO2 were found to be proximal to K87 of the CbbQ2 helix-2-insert, while 5 were cross-linked to K138 of the CbbQ2 PS1- β H (Figure 3-3C). This is consistent with our biochemistry data which supports a model where the disordered loops of the helix-2-insert and PS1- β H is in contact with CbbO to regulate force generation for Rubisco activation (Tsai and Ye et al., Unpublished).

It is notable that 11 of 12 lysine residues that can be found in CbbQ2 are involved in cross-linking to CbbO2 (Figure 3-3A, C). The only lysine not detected is K16, located on the convex face of the hexamer. This highlights that the cross-links found on the convex side of the hexamer, which are confined to the outer edge of the cup and on mobile elements such as the helix-2-insert and PS1- β H, are specific. Out of 21 lysines on CbbO2, only 5 are not detected to be involved in cross-linking to CbbQ2, all of them are located on the C-terminus (Figure 3-3A) and could be inaccessible to solvent. Taken together, the XL-MS results are consistent with a model where both the N- and C-termini of CbbO are located on the concave surface of the cup. The flexibility of the CbbO linker could enable the N-terminus to come into contact with some residues on the outer edge of the convex side terminus (Figure 3-3C). Movements of the disordered loops on the helix-2-insert and PS1- β H could also allow residues facing the convex face

to interact with CbbO.

3.4 Bioinformatic characterization of CbbQ orthologs

In light of the negative stain EM and XL-MS results, a bioinformatic analysis of CbbQ orthologs was performed to gain further insights into the interacting surfaces between CbbQ and CbbO.

Fifty CbbQ orthologs were identified from chemoautotrophic bacteria that are encoded proximal to either the non-carboxysomal form IA^Q Rubisco (CbbQ1), form II Rubisco (CbbQ2) or the carboxysomal form IA^C Rubisco (CbbQ3) (Appendix 8.2). The *A. ferrooxidans* (ATCC23270) genome encodes all three classes of CbbQ. In order to dissect the binding sites for CbbO on CbbQ, the sequence conservation between 50 different CbbQ orthologs and the electrostatic surface charge was mapped onto the CbbQ2 crystal structure. I hypothesized that potential CbbO binding surfaces will involve conserved hydrophobic patches on the CbbQ2 surface.

Consistent with a model where the concave face associates with CbbO, higher conservation on the concave face of the hexamer as compared to the opposite convex face (Figure 3-4A) was observed. The high conservation is found mainly on the N-terminal AAA+ domain and accumulated towards the pore, while residues on the C-terminal α -helical bundle are more variable especially towards the edge of the hexameric cup. On the convex surface, conservation is maintained on the anti-parallel β -sheets of the helix-2-insert and the PS1- β H, corroborating our XL-MS data which suggest that these elements are involved in CbbO interaction. In addition, regions of high sequence conservation were also found at the interface between CbbQ subunits and at the active sites, consistent with an oligomeric assembly. Given that both our negative stain EM and XL-MS data suggest CbbO associates with the concave face of the hexamer, it is conceivable that the high sequence conservation observed on this face is important for interaction with the adaptor. The conserved concave face features mainly negatively-charged residues, with residues on the α -helical bundle and loops on the helix-2-insert lining the central pore having neutral surface potential (Figure 3-4B). Recent biochemical data have implicated conserved hydrophobic residues on the helix-

2-insert and the PS1- β H in contacting CbbO and regulating ATP activity (Tsai and Ye et al., Unpublished). Our XL-MS also suggest that these regions are proximal to CbbO. Thus, it does not seem unreasonable to presume, that similar to our proposed model, conserved regions on the helix-2-insert and the PS1- β H of the CbbQ hexamer interact with CbbO via hydrophobic contacts to transmit mechanical force to the substrate.

Figure 3-4: Analysis of CbbQ2 structure.

(A) Surface representation of CbbQ2 coloured based on sequence conservation of 50 CbbQ sequences from chemoautotrophic bacteria from most conserved (maroon) to most divergent (cyan) as implemented using the ConSurf server (Ashkenazy et al., 2010). (B) Surface charge of AfQ2 crystal structure was calculated using the APBS plugin (Baker et al., 2001) in Pymol and is plotted between ± 5 kT/e.

3.5 Bioinformatical analysis of CbbO orthologs

To understand more about the role of the uncharacterized VWA-containing protein, CbbO, corresponding CbbO orthologs from chemoautotrophs were used to analyse CbbO protein distribution and conservation. By integrating bioinformatics analysis with the structural data from XL-MS, I expected to find regions of conservation important for CbbO function, which could later be confirmed with mutagenesis studies. Similar to what was done for CbbQ, 50 CbbO orthologs from chemoautotrophic bacteria were included in the analysis (Appendix 8.3). Phylogenetically, these CbbO orthologues fall into three classes: CbbO1, CbbO2 or CbbO3. CbbO1 is found to be proximal to non-carboxysomal form IA^Q Rubisco, CbbO2 is found downstream of form II Rubisco and CbbO3 is found associated with the carboxysomal form IA^C Rubisco in the expanded α -carboxysome locus. The *A. ferrooxidans* (ATCC23270) genome encodes all three classes of CbbO.

The primary structures of CbbO1 and CbbO3 have higher sequence identity (60%) than individually to CbbO2 (32.4 and 32.9%). This is reflected in the phylogenetic tree generated from a ClustalW alignment of 50 CbbO amino acid sequences from 30 chemoautotrophic bacteria (Figure 3-5). CbbO2 proteins form a distinct clade separate from the ones associated with form I Rubisco (CbbO1 and O3). Interestingly, some CbbO1 proteins are more closely related to CbbO3 than other CbbO1 proteins of similar genetic contexts (i.e. form IA^Q Rubisco-associated).

Figure 3-5: Phylogenetic tree of CbbO isoforms from chemoautotrophic bacteria.

Sequences were aligned with ClustalW and the phylogenetic tree was constructed with the Neighbour-Joining method using the Geneious software. The *A. ferrooxidans* sequences are highlighted.

The ClustalW multiple sequence alignment revealed that the N-terminal linker of 50 CbbO sequences contain only 4.7% of identical sites, while the VWA domains are better conserved with 20.3% of sites being the same (Figure 3-6). This domain contains a full set of MIDAS sequence motifs consisting of hhhhDxSxS, followed by a conserved threonine after ~70 residues and an aspartate after ~30 residues, where h is any hydrophobic residue and x is any residue (Figure 3-6). The key residues of the MIDAS motif are absolutely conserved among all sequences. This was expected given its critical role in CbbO function. The MIDAS site is believed to attach ligands to CbbQO through

a mechanism involving the participation of a critical acidic residue on Rubisco. This is similar to that by which the MIDAS site in integrin I-domains appear to mediate attachment of collagen (Emsley et al., 2000; Luo et al., 2007).

Further inspection of the alignment shows that CbbO1 and O3 sequences are longer than CbbO2 sequences due to an extension of (~40 amino acids) in the linker region (AfO1 425-470) (Figure 3-6). Several residues within this extension are well conserved among the CbbO1 and CbbO3 sequences, suggesting that it is important for function. The extension in the linker of CbbO1 and O3 is likely due to the requirement for interaction with a structurally different form of Rubisco. In addition, ~120 residues preceding the hhhhDxSxS motif were also found to be well-conserved and include many lysines that cross-link to CbbQ as shown by our XL-MS results (Figure 3-6). This is a strong indication that the conservation of this region is linked to associations with the AAA+ motor.

1MTDHPDASLPRLAAYRKQDRCDFPRVGGVFSDCIAKASARLGPAQVSA.YV
1MKHLEKYAEINQAFGEGSQAILQSKCKEATRVFSPSPGLDQ.YL
1MAVELEKYQDINAEELGEHADEVILRASMEAEARVETPRGLEN.YL
.....leY.eId.ge.eI.....wqeA.rvf.P.Gldq.YI
consensus>50

Afo2_1B7J555| 53 DAARLCKRCGCEPVLIFLEWPDGAALDGTLEMVQVQFMQRTPNGNAIGGFLQS
 Afo1_1B7JA37| 43 DGAMLRLCGCGVLSSYLDAAPHAREGEDAVWELVGNNASKTSGAVEAVST
 Afo3_1B7JB16| 45 NGATSKSLGCTDIVSEISAPAVAHEICGDAVREMLAAKNYSKTSATVIAMFS
 consensus>50 #GA...LSLGGCeely.s1#_ap...ia.elG#dav.elm.av.kM.sktsg.vi.a.ss

113	LAPVSRALLSRLQLGHH	YLNTLRDRMMERTT	GSIHGHQTHPS
Af02_1B7J5E51			
Af01_1B7JA271	SPTAANRLADLELFKSYL	GMLDRLLAQVE	RAIRPMLENIDQLFGYLT
Af03_1B7J5B16	SEVAAARLGRDLEFNG	YLHLLDTLLAAGP	RGVRFPMDLHSLTLLGLT
consensus>10	sp.aa.r.l.dpelf.yl.mld.\$\$ag.p		

173 DGLRNWVDYGVRYNLHHPERKQDFELOSADSRVLRQRRHGTGLADVERKLDLYLRGLW
 Afo2 |B7J5E5| 151 GGLRRWALNGAHARTDFEACIAYFVPEALAMQQRKGTGVDDIQRRLMYRLRLW
 Afo1 |B7JA37| 153 GGLRRWALNGCAHKTFDCAKAFVPEESIGVLRQRRHGTGLDVRRLSYLRRLW
 Afo3 |B7JB16| 153 GGLRRWALNGCAH.tdf#Q.4frqsp#s.avLQ.ERKgrfvd#Rrl.\$YLRalw
 consensus>50

```

Af02_1B7J55| 233 QDSEVLVPVSTAFATLRTPQPYDALGRNLPDVLDD...LPG
Af01_1B7JA27| 211 GRDFELRPISCEFDREGYRPYIEGIFHLPDAYDD...WED
Af03_1B7JB16| 213 GRDFEMRPISCEFDREGYRPYIEGIFHLPDAYDDLVWQ...APSGSETH
consensus>50 grdf#rptSGredregyPy#g..ihLPdayDD...g

```

Af02_1B7J5E51	I	G	A	L	D	H	R	A	I	L	A	H	M	V	G	H	R	R	W	S	T	P	O	I	A	N	W	S	P	F	Q	R	L	A	V	E	F	E	D	A	R	I	D	L	L	R	T	Y	P	G	I	R	L		
Af01_1B7JA27	I	D	G	M	T	R	Y	R	A	V	H	A	A	H	V	A	Y	M	R	K	A	I	S	A	E	A	L	T	P	E	Q	M	A	V	I	G	L	E	D	A	R	V	E	A	L	A	M	R	A	S	P	G	I	R	A
Af03_1B7JB16	I	G	I	E	L	L	Y	R	A	S	A	H	A	C	H	O	V	I	T	Q	E	D	S	T	Q	A	L	L	I	G	L	I	E	D	A	R	V	E	A	L	A	M	R	A	S	P	G	I	R						
consensus>50	I	g	l	d	.	Y	R	A	S	A	H	A	.	H	.	Y	.	t	.	q	.	i	.	n	.	l	s	p	.	Q	.	a	.	g	l	f	e	d	a	r	.	#	t	a	m	r	.	@	p	g	i	r	.		

Af02 |B7J5E5| 331 FLALHPKPGEDACDPETT...SCLHRLAMLSRALIDPEHG...YRNPVLHDFVDR.
 Af01 |B7JA27| 310 WLPFHIAVAGP...VASMGHLFSRLARAIIDPDYV...DDNAWVQCR.
 Af03 |B7JB16| 321 KLPIHTATPA...SGNEASLMARLALIDPDYR...DDHPWVAMC.
 consensus>50 wlpIhta.....l.hllarlarAlIDP#y...ddnnpwv.dg...

Afo2_1B7J5E5|
 Afo1_1B7JA27|
 Afo3_1B7JB16|
 consensus>50

Af02_1B7J5E1 437 ESGDEEEAFDAEQRKAPDL.....DDAQEINVVLSEFFRDEEDMG
 Af01_1B7JA27 408 EIDWEHETILPG.....AAQRKYVNVNMVNVDTELAG
 Af03_1B7JB16 419 DIRADGGIAGVITQQRKYVNVNMVNLDVPGAG
 #1.e.e.e.i.....agvrkyvvmemndagddaqevnlsefed
 consensus>50

Af02_1B7J5E1		LPRRYPEDYQS	S	SRPDWVSLYEGLHASAP	AATDQLLEKHA
Af01_1B7JA27		SYNQMEKEPVS	D	PFFHYPEWDYQIQARPLMATVLERRRLGDAVVREVLKYK	
Af03_1B7JB16		SLNQEGREPP	P	DPFHYPEWDYQMLERPEWCIVLEKKA	KSGTLETIEHVAKKH
consensus>50		s.nq.e.g.e.p.	d	p.fhyepwdy.q.q.rpdw	t.v.l.k.h.k

Af02_1B7J5E1 504 **A****L****K****H****L****K****R****L****D****L****L****K****P****Q****K****V****R****I****Y****Q****E****G****S****E****L****D****L****V****A****L****R****S****W****I****D****F****K****S****G****S****T****P****D****P****R****I****N****M****S****H****R****T****A****G**
 Af01_1B7JA27 519 **P****V****S****R****L****R****L****I****E****A****L****Q****P****G****V****R****L****R****K****Q****E****D****E****I****D****V****E****A****A****V****R****A****L****V****D****I****R****M****G****L****A****P****D****T****R****I****H****M****R****S****L****R****K****V**
 Af03_1B7JB16 530 **P****I****G****L****R****K****E****L****L****I****Q****P****G****V****R****L****R****K****Q****E****D****E****I****D****L****N****A****V****R****A****I****E****M****E****M****G****E****Q****P****D****P****R****I****M****N****R****N****R****K****V**
 consensus>50 pv..rLkylI#alqPqGvqRlRkQE#GdeIdL#aAvral!#.rmg.FDPdRI.Mr..rkv

Figure 3-6: ClustalW alignment of CbbO sequences.

50 CbbO sequences from chemoautotrophic bacteria were aligned with ClustalW, colour-coded figures of alignments were generated with the program Esript. Only the *A. ferrooxidans* sequences are shown. In the alignment, similar residues are shown in red and identical residues in white using bold lettering on red background. Blue frames indicate homologous regions. The consensus sequence is shown at the bottom.

3.6 Biochemical analysis of progressive C-terminal truncations of CbbO from the closely related homologue from *H. neapolitanus* (Hn)

Unlike CbbQ, minimal structural information is known about CbbO except that it contains a VWA domain at the C-terminal end. I aimed to understand the contribution of different regions of CbbO in interaction and regulation of CbbQ. Previous studies have shown that the *A. ferrooxidans* C-terminal truncation mutant Q2O2 Δ C444 can be successfully purified as a hetero-oligomer. However, it was highly unstable and prone to precipitation, rendering it unsuitable for extensive characterization. The closely related CbbQO homologue from *H. neapolitanus* (HnQO), which have also been reported to form a CbbQ-CbbO complex in 6:1 stoichiometry (Sutter et al., 2015) has the advantage of being more stable, with no precipitation observed during purification and is potentially more favourable for biochemical characterization. A member of the expanded α -carboxysome locus, HnCbbQ maintains ~70% identity with its counterparts in *A. ferrooxidans*, while HnCbbO is ~50% identical to AfCbbO orthologs. Therefore, progressive C-terminal truncation mutants of HnCbbO were made and the ability of the mutants to interact and stimulate the activity of HnCbbQ were investigated.

Given that AfQ2O2 Δ C444 was soluble and formed hetero-oligomeric complexes, HnCbbO was truncated at residue 429 which corresponded to residue 444 of AfCbbO2 based on sequence alignment of HnCbbO with AfCbbO2. Similar to all form I Rubisco associated CbbO sequences, HnCbbO also contains a unique linker sequence from residues 430 to 471 not present in the form II Rubisco associated AfCbbO2. It would be interesting to test if retaining this linker sequence might affect the protein's stability and interaction with HnCbbQ compared to truncation at residue 429. In addition, alignment of 50 CbbO orthologs revealed a stretch of relatively conserved sequences preceding the VWA domain (residue 589 onwards) (Figure 3-6). XL-MS experiments have shown that many lysine residues in this conserved region were cross-linked to CbbQ, an indication that this region of CbbO is important for complex formation with CbbQ. To investigate the importance of these regions, three progressive truncation mutants *HnCbbO* Δ C429, *HnCbbO* Δ C471 and *HnCbbO* Δ C589 were cloned into the pBAD33 expression system (Figure 3-7). The plasmids were subsequently transformed into competent BL21 *E. coli* harbouring pHue*HnCbbQ* and the proteins were co-expressed. HnQO Δ C429, HnQO Δ C471 and HnQO Δ C589 were purified similarly to the *A.*

ferrooxidans CbbQO proteins.

Figure 3-7: Schematic representation of HnCbbO truncation mutants.

The N-terminal HnO (1-429) is shown in blue. The linker region present only in form I Rubisco associated CbbO isoforms is shown in green. The stretch of relatively conserved sequences preceding the VWA domain is shown in yellow. The C-terminal VWA domain is shown in red.

3.6.1 Purification and analysis of C-terminal truncation mutants of HnCbbO

During purification it became evident that HnCbbQ (~30 kDa) eluted separately from the truncation mutants HnCbbOΔC429 (~45 kDa) and HnOΔC471 (~55 kDa), suggesting that they did not form hetero-oligomers. A representative native PAGE analysis show that no distinct band corresponding to the HnQ₆O₁ complex were seen in any of the SEC purification fractions for HnQOΔC429 and HnQOΔC471 (Figure 3-8A). Instead, the migration patterns suggest that these fractions contain mainly free HnCbbQ. High molecular weight smears in fraction 5 of HnQOΔC429 and fraction 4 of HnQOΔC471 indicate the presence of soluble aggregates. Distinct novel bands near the bottom of the gel in fraction 6 of HnQOΔC429 and fraction 5 of HnQOΔC471 likely represent free HnOΔC429 and HnOΔC471. When the SEC fractions were exposed to the analytical size exclusion chromatography, peaks corresponding to the HnCbbQO complexes (1.22 ml) were not observed (Figure 3-8B). Instead, the major peaks corresponded to free HnCbbQ hexamers (~1.36 ml). This would mean that the HnCbbO C-terminal truncation mutants (HnOΔC429 and HnOΔC471) lacked critical interaction domains and were unable to oligomerize with HnCbbQ. Unexpectedly, purification of

HnQO Δ C589 was unsuccessful as the resulting protein was highly unstable with a tendency to precipitate and was insufficient for downstream biochemical analysis (not shown). The insolubility of this protein could be due to certain exposed hydrophobic patches that are prone to aggregation.

Figure 3-8: Analysis of purified HnCbbO truncation mutants.

(A) 6% Native PAGE analysis of HnQO Δ C429 and HnQO Δ C471. (B) Analysis of HnQO Δ C429 and HnQO Δ C471 by analytical gel filtration.

Given that the C-terminal truncation mutants eluted separately from HnCbbQ during SEC, it was inferred that these mutants would be soluble and stable enough to be

purified individually, unlike their full-length counterpart which remains insoluble unless co-expressed with HnCbbQ. With this, all three C-terminal truncation mutants were successfully purified (Figure 3-9A). The Far-UV circular dichroism (CD) spectra of these proteins consisted of negative bands at 208 nm and 222 nm and a positive band at 193 nm, consistent with a well-folded predominantly α -helical protein (Figure 3-9B, C). The obtained spectrums for the mutants are also comparable with the well-folded proteins, HnCbbQ and HnCbbQO, suggesting that they adopt the native conformation.

Figure 3-9: Purified HnCbbO Mutants.

(A) Representative 12.5% SDS PAGE of all *H. neapolitanus* proteins purified. 5 µg of protein was loaded in each lane. (B) CD spectroscopy of *H. neapolitanus* protein (~ 0.15 mg/ml, 20 mM phosphate buffer, pH 8.0) is consistent with properly folded structures. Well-folded proteins HnQ and HnQO were measured in parallel as controls. The recorded ellipticity values were the average of three determinations for each sample. (C) Table showing estimated secondary structure content determined using the K2D2 algorithm (Perez-Iratxeta and Andrade-Navarro, 2008).

3.6.2 Functional analysis of C-terminal truncation mutants of HnCbbO

Next, the analytical SEC (Superdex200 PC3.2/30 on an ÄKTAmicro system) was used to determine if incubation of C-terminal truncation mutants with HnCbbQ induced hetero-oligomeric complex formation. HnCbbQ (180 kDa) migrated as a single monodisperse peak at 1.36 ml, corresponding to a molecular mass of approximately 200 kDa (Figure 3-10A, B, C). When mutants HnOΔC429 (48 kDa) and HnOΔC471 (52 kDa) were incubated with HnCbbQ, the size exclusion profile displayed two peaks, corresponding to HnCbbQ and mutant HnCbbO migrating separately through the column (Figure3-10A, B). In contrast, when HnOΔC589 (66 kDa) was incubated with HnCbbQ, there was a shift in the size exclusion curve towards the higher molecular weight fractions with the major peak at 1.31ml, corresponding to complexes of ~260 kDa (Figure3-10C). Further, the peak consisting of free HnOΔC589 also reduced in size, suggesting a shift in the equilibrium towards hetero-oligomeric complex formation.

Figure 3-10: Analytical size exclusion chromatography analysis of hetero-oligomerization between HnCbbQ and HnCbbO Mutants.

(A) Size exclusion profile of HnCbbQ and HnCbbO(1-429), (B) HnCbbQ and HnCbbO(1-471), (C) HnCbbQ and HnCbbO(1-589). Elution peaks of globular molecular weight standards are indicated.

It was previously shown that the ATPase activity of CbbQ is stimulated when complexed with CbbO (Tsai et al., 2015). The finding that HnOΔC589 was able to oligomerize with HnCbbQ indicated that it might possibly induce conformation changes in HnCbbQ to increase the basal ATPase activity. In order to determine if any of the HnCbbO truncation mutants were able to stimulate the ATP activity of HnCbbQ, the truncation mutants were incubated with HnCbbQ and their effect on the AAA+ motor were analysed using a coupled ATPase activity assay monitoring the oxidation of

NADH. While HnCbbQ alone has a low ATP rate of $< 1 \text{ min}^{-1}$ protomer CbbQ⁻¹, the WT HnQO hydrolysed ATP with a rate of $\sim 4 \text{ min}^{-1}$ protomer CbbQ⁻¹ (Figure 3-11). These values are comparable to previously published ATPase rate of *A. ferrooxidans* Q1O1 and Q2O2 enzymes (Tsai et al., 2015). When mutants HnOΔC429 and HnOΔC471 were incubated with HnCbbQ, no noticeable change in ATP activity was observed and the ATPase rate remained at the basal level. In contrast, when HnOΔC589 was used, ATPase activity was stimulated to a rate that is higher than the WT at $\sim 6 \text{ min}^{-1}$ suggesting that HnO(1-589) harboured essential motifs necessary for hetero-oligomerization and stimulating the ATP activity of the AAA+ motor CbbQ.

Figure 3-11: Activity assay of HnCbbO mutants.

ATPase activity assays of HnCbbO were measured after incubation of 1.6 μM HnCbbQ with 3.2 μM HnCbbO mutants for 30 min. Error bars indicate the mean and s.d. of at least three independent experiments.

3.7 Discussion

In this chapter, the form II Rubisco associated activase Q2O2 from *A. ferrooxidans* was visualized with negative stain EM to understand more about the architecture of this newly discovered class of activase. Two-dimensional class averages of the Q2O2 hetero-oligomers revealed a density in the central cavity which is not present in the

previous study of the CbbQ2 hexamer alone (Figure 3-2). The 3D model at 16 Å supported an arrangement where the N-terminus of CbbO stems from the rim of the CbbQ cup and extends towards the central cavity with the C-terminal VWA domain, well-known for mediating protein-protein interactions, plugging the pore.

The recently published cryo-EM structures of the ribosome maturation factors Mdn1 (Midasin) and Rea1 resemble the structure of CbbQO with a hexameric head domain from which an elongated flexible tail carrying the MIDAS domain protrude (Chen et al., 2018; Sosnowski et al., 2018). In the reported structures, the linker stems from the outer edge of the AAA+ ring and the C-terminal VWA domain docks onto the AAA+ ring in a nucleotide state-specific manner. Given that CbbQ has been classified under the same evolutionary lineage group as Rea1 and Mdn1 (Iyer et al., 2004), it is conceivable the molecular architecture is conserved among this class of molecular chaperones. In this regard, it is conceivable that the flexibility of the CbbO linker allows movement of the C-terminal MIDAS domain, with the favoured position being on the central cavity of the hexameric CbbQ2.

In parallel to negative stain EM, XL-MS was used to determine the subunit organisation of the CbbQO activase. Integrating information from the XL-MS results with available structural data of CbbQ2, revealed that the concave face of the hexamer is extensively cross-linked to both the N- and C-termini of CbbO, in line with the notion that CbbO is featured on the concave side of the cup (Figure 3-3). Some lysine residues on the convex side near the outer rim of hexamer is also implicated in associating with CbbO. It could be interpreted that the flexibility of the N-terminal CbbO linker allows the adaptor to become proximal to residues on both the concave and convex face of the hexamer and enables the connecting MIDAS domain at the C-terminus to reside near the central pore of CbbQ.

Although the current model for green and red-type Rca function involves substrate-engagement of the disc-shaped hexamer using its convex surface, no evidence for an involvement of the CbbQ pore has been obtained so far. In contrast, the pore of CbbQ is occluded by the helix-2-insert, and pore-loop 1 is not accessible from the convex surface. Recent biochemistry data from our lab reveal that mutation of L85 on the Helix-2-insert (H2I), which faces the concave face uncouples ATPase and Rca activities of

Q2O2 (Tsai and Ye et al., Unpublished). This suggests that movement of the H2I is likely involved in coupling ATP hydrolysis to Rubisco remodelling via the interaction with CbbO. In support of this, lysine residues facing away from the pore such as K87 on the H2I were cross-linked to CbbO multiple times (Figure 3-3). This would only be possible if dynamic movements of these elements allowed access to the adaptor residing on the concave face. Interestingly, the H2I of AA2L in Rea1 also serves as an ATPase activity regulator and through nucleotide-induced conformational changes enable the VWA domain to dock on to the AAA+ ring (Sosnowski et al., 2018). The related AAA+ molecular motor dynein also utilizes the H2I of AAA2 to regulate dynamic movement of the dynein linker (Kon et al., 2012). On the basis of these findings, it is conceivable that the H2I from at least one CbbQ subunit is needed to be in contact with CbbO to trigger the sequence of conformational changes that would lead to mechanical force transmission from the AAA+ motor CbbQ to CbbO for Rubisco activation.

Other than the C-terminal VWA domain, the N-terminal linker region of CbbO is poorly conserved (Figure 3-6). Bioinformatic analysis was used to clarify the role of CbbO in Rubisco activation. The multiple sequence alignment of CbbO orthologs revealed that the CbbO1 and O3 sequences contain an extension of ~ 40 amino acids in the linker not present in CbbO2. It is plausible that the variant length of the linker in different isoforms is an adaptation to accommodate the conformational differences of two forms of Rubisco. Indeed, when known Rubisco structures were compared, it was found that the C-terminus of form I Rubisco resides in an equatorial region of the protein while the C-terminus of form II Rubisco lies at the top of the enzyme (Liu et al., 2017). Therefore, it is possible that the length of the linker is directly relevant to interaction with Rubisco and the activase function. In addition to this, it was found that ~120 residues prior to the VWA domain is generally well conserved, with all lysine residues in that region implicated in cross-linking to CbbQ (Figure 3-6). Truncating these residues are expected to destabilize association with CbbQ and abrogate Rubisco activation.

To validate this prediction, three progressive truncation mutants of the closely related homologue *H. neapolitanus* CbbO, HnCbbO Δ C429, HnCbbO Δ C471 and HnCbbO Δ C589 were purified. The homologue was used as it was deemed more stable during purification and mutants were more likely to be soluble upon expression. While both HnCbbO Δ C429 and HnCbbO Δ C471 were unable to oligomerize with HnCbbQ,

HnCbbO Δ C589 which harboured the well-conserved region preceding the VWA domain was shown to retain binding affinity for HnCbbQ (Figure 3-10) and was able to stimulate the ATPase rate of HnCbbQ (Figure 3-11), demonstrating that this region is an essential CbbQ-CbbO interaction domain of CbbO.

Taken together, the results in this chapter support the hypothesis that the structural ensemble of the CbbQO-type Rubisco activases resemble that of other Helix-2-insert clade of AAA+ proteins such as Mdn1 where the adaptor CbbO extends from the AAA+ ring and folds back upon itself to enable the tethered C-terminal VWA domain to dock on the AAA+ ring (Chen et al., 2018). This raises the question if the mechanism of action is similar as well. Given that Pore-loop 1 mutants of CbbQO were functionally equivalent to the WT (Tsai and Ye et al., Unpublished), a pore-loop threading mechanism as utilized by other activases and AAA+ unfoldases seem unlikely. Instead, our findings so far are consistent with motor force transmission from the AAA+ ring to the CbbO adaptor protein residing on the concave face. Clearly, structures at higher resolutions and in complex with the substrate is required for validation of the proposed model. A detailed mechanistic understanding of Rubisco activation by the CbbQO-type of activases will by extension help elucidate the common principles of the poorly studied MoxR class of AAA+ motors.

4. INSIGHTS INTO THE CBBQO RUBISCO ACTIVATION MECHANISM FOR FORM II RUBISCO

With the recent discovery and characterization of the third convergently-evolved Rca system (Tsai et al., 2015), I attempted to capture the complex between Rubisco and the CbbQO-type activase. Structural and biochemical characterization of such a complex would allow more detailed insights into the mechanism of activation. In this chapter, it was established that the interaction between the form II Rubisco, AfM and the Walker B mutant form of its activase, Q2^{WB}O2 could be biochemically stabilized with chemical cross-linkers and confirmed using native PAGE and analytical size exclusion chromatography. I took advantage of the finding that complex formation can be visualized on the native gel and used this efficient method to obtain more mechanistic insights into the form II Rubisco activation system. In addition, the CbbO2 VWA domain was purified from inclusion bodies and was used to demonstrate that this domain is sufficient for forming complexes with AfM in a cross-linker dependent manner. Finally, negative stain electron microscopy (EM) and cross-linking coupled to mass spectrometry (XL-MS) provided more insights into interaction sites of the Rubisco-activase complex.

4.1 Native PAGE analysis of the Rubisco-Rubisco activase complex formation

Within the CbbQ AAA+ module, the sequence of the Walker B motif is highly conserved, as the acidic residues of the Walker B hhhhDE sequence are crucial for ATPase activity. The critical glutamate of this motif activates water for the hydrolysis reaction and mutation of this residue inhibits nucleotide hydrolysis but not binding (Hanson and Whiteheart, 2005b). Previous biochemical assays have shown that the Q2O2 Walker B mutant (Q2^{WB}O2) displayed no ATPase activity and was unable to activate inhibited Rubisco (Tsai et al., 2015). Stabilized substrate interactions have been achieved for some AAA+ proteins with mutations in the Walker B motif (Weibezahn et al., 2003; Graham et al., 2013). These proteins with the E to Q Walker B mutation have been used to create ‘substrate traps’ that bind but cannot release substrates.

Under non-denaturing conditions, polypeptides will retain their higher-order structure as well as the ability to interact with other polypeptides. Therefore, native PAGE

analysis was initially used to determine if the Walker B mutant (E111Q in Q2O2) can trap Rubisco as a stable complex. Previous biochemical assays have shown that the ATPase rate of the CbbQO activases were specifically stimulated by inhibited Rubisco (Tsai et al., 2015), an indication that the activases specifically recognize inhibitor-bound forms of Rubisco. Hence, the form II (AfM) carbamylated Rubisco were stably inhibited by the transition state analogue, 2-carboxyarabinitol 1,5-bisphosphate (CABP) to form ECMC. Material of lower mobility on the native gels, which migrate differently from the Rubisco and activase standards were expected to represent complexes formed between ECMC and the Walker B mutants.

Samples were treated with the chemical cross-linker glutaraldehyde (GA) to introduce cross-links that stabilize the complexes formed. Glutaraldehyde (~5-Å spacer) is a homo-bifunctional cross-linker that specifically reacts with primary amine groups (i.e., ε-amino groups of lysine residues) at around neutral pH, facilitating the formation of covalent bonds between proteins that are in close spatial proximity.

4.2 The Walker B mutant Q2O2 (FLAG-Q2^{WB}O2) forms complexes with stably inhibited form II Rubisco (HA-AfM ECMC)

A representative native PAGE analysis of the cross-linking between HA-tagged form II Rubisco, AfM ECMC and the FLAG-tagged Q2^{WB}O2 is shown in Figure 4-1A. HA-AfM ECMC migrated as a stack of defined bands (Figure 4-1A; lane 1). A faint higher band can also be seen in the same lane believed to be AfM in higher oligomeric state. FLAG-Q2^{WB}O2 migrated as an intense well-defined band (Figure 4-1A; lane 3), indicating conformational homogeneity. With the addition of the cross-linker, a downward shift to the migration pattern of the corresponding bands was observed, likely due to a change in the charge of the proteins after reacting with glutaraldehyde (Figure 4-1A; lane 2 and 4). In addition, since glutaraldehyde reacts non-specifically with any primary amine, cross-linked bands on native gels are expected to appear less focused than bands without cross-linker. It is interesting that the addition of Mg²⁺ and ATP to the sample induced some FLAG-tagged Q2^{WB} to dissociate from the hetero-oligomer FLAG-Q2^{WB}O2. This was seen as a lower molecular weight band migrating below the FLAG-Q2^{WB}O2 standard on the anti-FLAG immunoblot (Figure 4-1C; lanes 5 and 9).

This indicated that ATP binding results in a conformational change within the CbbQ hexamer that causes the association between Q2^{WB} and O2 to weaken.

Using the migration pattern of HA-ECMC and FLAG-Q2^{WB}O2 individually as controls, interaction between the two proteins was observed as high molecular weight (HMW) smears (Figure 4-1 A, B; Lane 7). The smearing is likely caused by the heterogeneous conformations of the complexes due to the transient and reversible nature of the interaction. The addition of cross-linker resulted in the formation of a well-defined band (denoted Complex*) migrating higher than both the HA-ECMC and HA-Q2^{WB}O2 standards (Figure 4-1 A; Lane 8 and 10). In addition, other HMW bands migrating above Complex* were also observed. The formation of these bands suggested that addition of the cross-linker restricted the complexes to a few distinct conformations, with the major species represented by the Complex* band. It is also possible that the HMW bands represented protein complexes of different subunit stoichiometry (i.e. one Rubisco with multiple activases). Both Complex* and the HMW forms represented complexes of HA-ECMC and FLAG-Q2^{WB}O2, as confirmed by immunoblotting with anti-HA and anti-FLAG antibodies (Figure 4-1 B; Lane 8 and 10). The presence of ATP did not increase the amount of complexes formed despite the requirement of ATP binding for substrate interaction reported for many AAA+ proteins (Hanson and Whiteheart, 2005b; Bhat et al., 2017b). The ability to assay protein-protein interactions between Rubisco and the CbbQO activase using cross-linking followed by native PAGE made it possible to test multiple combinations of variants that had earlier been shown to be biochemically impaired in Rubisco activation. These will be described in the following sections.

A

B

Figure 4-1: Cross-linking of HA-AfM and FLAG-Q2^{WBO2} with glutaraldehyde.

(A) 6% native gel analysis of complex formation between HA-AfM and FLAG-Q2^{WB}O2. (B) Immunoblot of native-PAGE using anti-HA (Panel 1) and anti-FLAG (Panel 2). 5 μ M inhibited Rubisco (monomer) and 20 μ M activase (monomer) were cross-linked in the presence or absence of 5 mM ATP and 5 mM MgCl₂. A final concentration of 0.125% v/v of glutaraldehyde was used for crosslinking for 10 min at RT. The crosslinking reaction was quenched by addition of 1 M Tris-HCl, pH 8.0 to a final concentration of 50 mM.

4.2.1 Q2^{WB}O2 preferentially interacts with the inhibited AfM ECMC over the active ECM

Previous biochemical assays show that unlike the inhibited form ECMC, the active Rubisco ECM did not stimulate the ATPase activity of Q2O2 suggesting a preference for the inhibited form (Tsai et al., 2015). Therefore, the carbamylated active form of Rubisco (ECM) was tested for complex formation with FLAG-Q2^{WB}O2. The migration of proteins on the native gel depends on many factors, including size, shape, and native charge. HA-AfM ECM was observed to migrate as a distinct band instead of stacks of bands like HA-AfM ECMC, suggestive of a different conformation (Figure 4-2; lane 1). No high molecular weight smearing indicative of an interaction was observed (lane 7). Addition of glutaraldehyde induced the formation of a markedly smaller amount of Complex* (lanes 8 and 10) as compared to when HA-AfM ECMC was used (Figure 4-1A; lanes 8 and 10). Also, no HMW bands could be observed. This suggests that Q2O2 activase is able to differentiate between the two different conformations of the enzyme and preferentially interacts with the stably inhibited HA-AfM ECMC over the active HA-AfM ECM.

Figure 4-2: Cross-linking of HA-AfM ECM and FLAG-Q2^{WB}O2 with glutaraldehyde.

6% native PAGE analysis of complex formation between HA-AfM ECM and FLAG-Q2^{WB}O2. Cross-linking was performed as described in Figure 4.1.

4.2.2 The surface exposed Glu75 on the Rubisco large subunit (RbcL) is essential for facilitating interactions with Q2^{WB}O2

To probe for the mechanism of activation, cross-linking assays were done with inhibited AfM mutants (E75A), previously established to be incapable of both stimulating the ATPase activity of Q2O2 and being activated by the activase (Tsai et al., 2015). It was hypothesized that similar to the integrin I domain, the sixth coordination site for the Mg²⁺ of the MIDAS motif on CbbO is provided by a carboxylate side-chain of an acidic residue on the surface of its substrate, Rubisco. Native gel analysis of AfM^{E75A} ECMC with FLAG-Q2^{WB}O2 did not show smears indicative of an interaction (Figure 4-3; lane 5). Although the addition of cross-linker produced the Complex* band (Figure 4-3; lanes 6 and 8), it was not of similar intensity compared to that formed by WT HA-AfM ECMC and FLAG-Q2^{WB}O2 complexes (Figure 4-1A; lanes 8 and 10). No HMW bands migrating above the Complex* were observed. This confirms our hypothesis that the solvent exposed E75 on RbcL is essential to mediate the interaction with the activase.

Figure 4-3: Cross-linking of AfM^{E75A} ECMC and FLAG-Q2^{WB}O2 with glutaraldehyde.

6% native PAGE analysis of complex formation between AfM^{E75A} ECMC and FLAG-Q2^{WB}O2. Cross-linking was performed as described in Figure 4-1.

4.2.3 An intact RbcL C-tail is required for a high affinity interaction with Q2^{WB}O2

Previous studies have shown that the CbbQO system exhibits very stringent requirements for the presence and position of the RbcL C-terminal HK/R motif to perform its activase function (Tsai et al., 2015). The red-type Rubisco activase CbbX functions by transiently pulling the C-terminal peptide of the RbcL into the hexameric pore, resulting in the release of the sugar phosphates (Mueller-Cajar et al., 2011; Bhat et al., 2017b). Many other AAA+ proteins such as ClpX/ ClpA have also been shown to interact with their substrates by recognition of a specific peptide motif (Gottesman et al., 1998; Siddiqui et al., 2004; Olivares et al., 2015). In order to assess the importance of the C-terminus of RbcL for interaction with the Rubisco activase, the proteins AfM^{ΔC1} ECMC and FLAG-Q2^{WB}O2 were assessed for their ability to form complexes using native PAGE. Similar to the AfM^{E75A} ECMC mutant, no HMW weight smears indicative

of interaction were observed (Figure 4-4; lane 5). Complex* bands induced by glutaraldehyde cross-linking were also not as intense as those of WT HA-AfM ECMC and FLAG-Q2^{WB}O2 (Figure 4-4; lanes 6 and 8). No HMW bands migrating above the Complex* were observed. The deletion of one residue from the C-tail of Rubisco significantly reduced the affinity for interaction with the activase, consistent with previous biochemical assays suggesting that an intact C-tail is needed for activation.

Figure 4-4: Cross-linking of AfM^{AC1} ECMC and FLAG-Q2^{WB}O2 with glutaraldehyde.

6% native PAGE analysis of complex formation between AfM^{AC1} ECMC and FLAG-Q2^{WB}O2. Cross-linking was performed as described in Figure 4-1.

4.2.4 The adaptor protein CbbO is required to form complexes with the stably inhibited form II Rubisco, AfM ECMC

Although CbbQ is unable to remodel Rubisco without the adaptor protein CbbO, it is still unclear if the AAA+ motor in is direct contact with Rubisco (Tsai et al., 2015). Therefore, native gel analysis was used to investigate binding of AfQ2^{WB} to HA-AfM ECMC. In contrast to the hetero-oligomer FLAG-Q2^{WB}O2, the hexameric FLAG-

AfQ2^{WB} migrated as a diffuse smear on the native gel indicative of conformational heterogeneity (Figure 4-5; lane 3). Addition of Mg²⁺ and ATP appeared to facilitate stabilization of FLAG-AfQ2^{WB} into a structured hexameric form, which migrated as a distinct band (lane 7). However, no HMW smears or bands representing potential complexes between the FLAG-AfQ2^{WB} hexamer and HA-AfM ECMC were observed (lane 8). This highlighted the requirement of the adaptor protein CbbO for a stable interaction with the Rubisco substrate.

Figure 4-5: Cross-linking of HA-AfM ECMC and FLAG-Q2^{WB} with glutaraldehyde.

6% native PAGE analysis of complex formation between HA-AfM ECMC and FLAG-Q2^{WB}. 5 μ M inhibited Rubisco (monomer) and 10 μ g of FLAG-Q2^{WB} was used in the cross-linking reaction as described in Figure 4-1.

4.2.5 Q2O2 MIDAS motif mutants are able to form complexes with the stably inhibited form II Rubisco, AfM ECMC

Previous biochemical assays showed that all but one Q2O2 MIDAS motif mutants (Q2O2^{T656A}) were ATPase active but unable to activate Rubisco (Tsai et al., 2015). In

addition, their ATPase activities were no longer stimulated by inhibited AfM, indicating that they could no longer sense their substrate. I next investigated if Q2O2 MIDAS motif mutants were able to maintain the interaction with stably inhibited HA-AfM ECMC. Interestingly, all five MIDAS motif mutants were capable of interaction with AfM ECMC, indicated by distinct HMW bands after the cross-linker was added (Figure 4-6B; lanes 4, 6, 8, 10, 12). This is similar to what was observed on the native gel for FLAG-Q2^{WB}O2 with HA-AfM ECMC (Figure 4-1A lanes 8 and 10).

I deduced that single amino acid substitutions in the MIDAS motif were not enough to induce a detectable effect on substrate binding. As the MIDAS cation (probably Mg²⁺) is coordinated by 6 ligands (Figure 4-6A), it is possible that single mutations may not have abolished metal cation binding due to the presence of the other 5 ligands. Therefore, several double MIDAS mutants were purified and investigated them for their ability to interact with HA-AfM ECMC. Unexpectedly, all double MIDAS mutants retained the ability to interact with their substrate and Rubisco-activase complexes in the form of HMW bands were observed (Figure 4-6C). This was interpreted to mean that additional points of contact between Rubisco and the activase allowed mutants to interact with Rubisco, albeit at a lower affinity that prevented it from performing its function.

Figure 4-6: Cross-linking of HA-AfM ECMC and Q2O2^{MIDAS} mutants with glutaraldehyde.

(A) Diagram depicting the putative cis and trans sites of the MIDAS interaction. The numbers indicate the position of the residues within the AfO2 amino acid sequence. (B) 6% native PAGE analysis of complex formation between HA-AfM ECMC and Q2O2^{MIDAS} mutants. (C) 6% native PAGE analysis of complex formation between HA-AfM ECMC and Q2O2^{MIDAS} double mutants. Cross-linking was performed as described in Figure 4-1.

4.2.6 The soluble chimeric CbbO mutant O1O2VWA interacts specifically with AfM

Although the WT CbbO is insoluble when purified without CbbQ, the mutant chimeric protein O1O2VWA encoding the fusion product of the N-terminal linker (1-455) from CbbO1 and VWA domain (445-759) from CbbO2 is fully soluble. The CbbO mutant forms a hetero-oligomeric complex with CbbQ2 when co-expressed but the mutant activase is non-functional in activating inhibited Rubisco (Tsai and Ye et al., Unpublished). Given that no complex formation were detected between Rubisco and the AAA+ motor CbbQ alone, it would be interesting to determine if the adaptor CbbO is able to achieve stable interactions with Rubisco without CbbQ. Therefore the soluble chimeric CbbO mutant, O1O2VWA was purified as described in material and methods (Figure 4-7A, B).

Figure 4-7: Purification of O1O2VWA.

(A) Representative 12.5% SDS PAGE separation of proteins from IMAC Ni-NTA and various fractions from IEX Source15Q column. L: lysate, S: soluble, FT: flow through, W: wash, BC: before cleavage by USP2, AC: after cleavage. (B) Size-exclusion profile of O1O2VWA on a Superdex 200 16/600 column (left). Purified fraction on a 12.5% SDS PAGE.

As the only soluble full length CbbO available in the lab, the chimeric mutant O1O2VWA was mixed with AfM and used for Native PAGE analysis. In the anti-AfM blot, the inhibited AfM (ECMC) migrated differently on the native gel as compared to the inactive (E) or active (ECM) AfM, reflecting differences in conformation (Figure 4-8; panel 1; lanes 3-4, 7-8, 11-12). The appearance of an additional higher band above the AfM band when O1O2VWA is added in the presence of glutaraldehyde is suggestive of complex formation (Figure 4-8; panel 1; lanes 6, 10, 14). However the anti-O1O2VWA blot showed a heavy smear throughout the lanes that contain O1O2VWA protein and was not useful in confirming the presence of O1O2VWA in these bands of lower mobility formation (Figure 4-8; panel 2). It is possible that the flexibility of the adaptor protein resulted in heterogenous conformations visualized as smears on the western blot.

Figure 4-8: Cross-linking of O1O2VWA with AfM and AfLS.

(A) Immunoblot of 6% native PAGE analysis of complex formation between O1O2VWA and AfM using anti-AfM (Panel 1) and anti-O1O2VWA (Panel 2). Boxes represent suspected complexes. (B) Immunoblot of 6% native PAGE analysis of complex formation between O1O2VWA and AfLS using

anti-form I Rubisco (Panel 1) and anti-O1O2VWA (Panel 2). 5 μ M inhibited Rubisco (monomer) and 10 μ g of O1O2VWA was used in the cross-linking reaction as described in Figure 4-1.

When AfLS was used for the same Native PAGE analysis with O1O2VWA, no lower mobility bands above the AfLS bands were observed for the active (ECM) and the inhibited (ECMC) AfLS in the anti-Rubisco blot formation (Figure 4-8B; panel 1; lanes 10 and 14). Unexpectedly, the non-carbamylated, inactive (E) AfLS migrated as a higher molecular weight smear when the chimeric CbbO was added formation (Figure 4-8B; panel 1; lane 6). The anti-O1O2VWA blot also confirmed the presence of O1O2VWA in these smears and indicated that they were a manifestation of the interaction between the flexible adaptor protein and Rubisco (Figure 4-8B; panel 2; lane 6). Since this interaction was only observed for the inactive AfLS, this was interpreted as a non-specific interaction.

4.2.7 Test expression and purification of O2VWA domain from the inclusion bodies

Given that the full length mutant CbbO, O1O2VWA was capable of forming complexes with AfLS, I wanted to functionally characterize the essential VWA domain of CbbO2. Although CbbO2 is insoluble on its own, the solubility of the VWA domain has not been tested. The integrin α I domains had been reported to fold independently into its native conformation and could be purified as a soluble protein. For this reason, several of the α I domains have been studied in detail (Shimaoka et al., 2002; Song et al., 2005; Luo et al., 2007). Therefore, the FLAG-tagged O2VWA domain (561-759) was heterologously expressed in *E. coli* and its solubility was investigated with protein-expression tests.

The Flag-tagged His₆-ubiquitin-AfO2VWA protein in the pHUE vector was expressed in *E. coli* BL21 to check for the expression of soluble protein. This was done in the presence and absence of chaperonin GroEL/ES, which mediates the folding of proteins in prokaryotes by isolating it in a hydrophilic chamber (Hartl and Hayer-Hartl, 2002). Co-expression of GroEL/ES has been found to improve the yield of folded proteins, increasing protein solubility (de Marco et al., 2007). The pHUE expression vector

contains the T7 promoter and protein expression was induced by IPTG. The GroEL/ES genes were under the regulation of BAD promoter in the pBAD33 vector and gene expression of GroEL/ES was induced by arabinose (Guzman et al., 1995).

Samples from the total, soluble and pellet fraction of the un-induced and induced culture were denatured and loaded onto the 12.5% SDS gel (Figure 4-9). The gel image showed that the Flag-tagged His₆-ubiquitin-AfbO2VWA fusion protein migrated to ~34 kDa and was insoluble. The leakiness of the T7 promoter resulted in expression of the protein even without induction. Upon induction with IPTG alone or with arabinose, higher amount of O2VWA protein was observed in the total and pellet fraction. No soluble O2VWA expression was induced with IPTG. In the presence of GroEL/ES, a faint band corresponding to O2VWA can be observed in the 'S' fraction. This likely represented unfolded O2VWA proteins that were encapsulated in the GroEL/ES cavity.

Figure 4-9: Protein expression test for His₆-ubiquitin-AfO2VWA.

12.5% SDS PAGE separation of Total (T), soluble (S) and Pellet (P) fractions. +: Induced, -: un-induced.

In order to acquire enough protein for cross-linking, the FLAG-O2VWA protein was purified from the inclusion bodies of *E. coli* cells (Figure 4-10A, B). Inclusion bodies are aggregated proteins of higher density found in both the cytoplasmic and periplasmic space of *E. coli* during high level of foreign protein expression (Jurgen et al., 2010). The

expressed protein can be isolated and recovered in denatured form (Rodriguez-Carmona et al., 2010). The higher density inclusion bodies were separated from the soluble protein fraction by high-speed centrifugation after cell disruption. The pellet was extensively washed with buffers A and B, containing detergent and low concentration of denaturing agent. After washing, the FLAG-O2VWA protein was solubilized by the use of 6 M guanidine hydrochloride (Denaturation buffer). The protein was refolded to its native state after the removal of the denaturing agent by dialysis. The protein was further purified by size-exclusion chromatography with the Superdex 200 16/600 column, which produced a single symmetrical peak indicating homogeneity (Figure 4-10B). It was observed from the 12.5% SDS gel that the ubiquitin tag was able to be cleaved from the fusion protein by the deubiquitinating enzyme USP2, resulting in a band corresponding to the FLAG-tagged AfO2VWA protein (~23 kDa). CD was then used to assess that the protein was properly refolded after its denaturation. The obtained spectrum consists of negative bands at 208 nm and 222 nm and a positive band at 193 nm, consistent with a well-folded α -helical protein (Figure 4-10C, D).

Figure 4-10: Purification of FLAG-O2VWA from the inclusion bodies.

(A) 12.5% SDS PAGE separation of proteins from various stages of purification. L: lysate, S: soluble, A: wash buffer A, B: wash buffer B, D: Denaturation buffer, BC: Before USP2 mediated cleavage of Ubiquitin, AC: After USP2 cleavage. P: pellet. Protein was cleaved for 3h (left) and 16 h (right) with different results. (B) Size-exclusion profile of FLAG-O2VWA on a Superdex 200 16/600 column (left). Purified fraction on a 12.5% SDS PAGE (right). (C) CD spectroscopy of FLAG-O2VWA (0.15 mg/ml, 20 mM phosphate buffer, pH 8.0) is consistent with properly folded structures. Well-folded proteins (O1O2VWA and lysozyme) were measured in parallel as controls. The recorded ellipticity values were the average of three determinations for each sample. (D) Table showing estimated secondary structure content determined using the K2D2 algorithm (Perez-Iratxeta and Andrade-Navarro, 2008).

4.2.8 The re-folded FLAG-O2VWA interacts specifically with AfM

The purified FLAG-O2VWA fusion was subsequently used for interaction studies with AfM. In the anti-AfM blot, the presence of smears above the AfM band in the presence of the cross-linker suggested the formation of a complex between AfM and FLAG-O2VWA (Figure 4-11A; panel 1; lanes 6, 10, 14). The anti-O1O2VWA blot also confirmed the presence of FLAG-O2VWA in these higher smears and indicated that there were O2VWA bound to the AfM that had slower mobility (Figure 4-11A; panel 2; lanes 6, 10, 14). It was inferred that parts of the protein could be flexible, resulting in heterogenous conformations visualized as smears on the western blot.

The complex formation between FLAG-O2VWA and AfM was cross-linker dependent as no higher band was seen above AfM without glutaraldehyde (Figure 4-11A; panel 1; lanes 5, 9, 13). This is different from the interaction between FLAG-Q2^{WB}O2 and HA-ECMC where higher molecular weight smears can be observed even without the addition of cross-linker (Figure 4-1A). In addition, FLAG-O2VWA was shown to interact with AfM E, ECM and ECMC without discrimination, whereas FLAG-Q2^{WB}O2 only interacts with HA-AfM ECMC indicating that the VWA domain alone does not exhibit selectivity for the inhibited Rubisco.

When the form I Rubisco AfLS was mixed with FLAG-O2VWA, there was no indication of complex formation in both the anti-Rubisco and the anti-O1O2VWA blot (Figure 4-11B; panels 1 and 2; lanes 6, 10, 14), suggesting that the interaction between AfM and FLAG-AfO2VWA is specific.

Figure 4-11: Cross-linking of FLAG-O2VWA with AfM and AfLS.

(A) Immunoblot of 6% native PAGE analysis of complex formation between FLAG-O2VWA and AfM using anti-AfM (Panel 1) and anti-O1O2VWA (Panel 2). Boxes represent suspected complexes. (B) Immunoblot of 6% native PAGE analysis of complex formation between FLAG-O2VWA and AfLS using anti- form I Rubisco (Panel 1) and anti-O1O2VWA (Panel 2). 5 μ M inhibited Rubisco (monomer) and 10 μ g of FLAG-O2VWA was used in the cross-linking reaction as described in Figure 4-1.

4.3 Characterization of the AfM-Q2^{WB}O2 complex with analytical size exclusion chromatography (SEC)

The finding that it is possible to visualize the HA-AfM and FLAG-Q2^{WB}O2 complex on the native gel prompted me to characterize the complex via analytical size exclusion chromatography. I aimed to efficiently isolate the Rubisco-Rubisco activase complex in a biochemically stable form for structural characterization. Analysis of the binding of FLAG-Q2^{WB}O2 to HA-ECMC was carried out using analytical SEC (Superdex 200 PC3.2/30 on an ÄKTAmicro system). Fifty microliter fractions were collected from 0.95 ml elution volume and analysed using SDS-PAGE. HA-ECMC eluted as a single peak at 1.21 ml, corresponding to a molecular mass of approximately 260 kDa (Figure 4-12A). The 12.5% SDS-PAGE showed that HA-AfM ECMC could be found in fraction 5, 6 and 7. All the bands corresponded to the RbcL which migrated to the same position corresponding to ~52 kDa (Figure 4-12C). AfM is believed to occupy a well-defined hexamer in solution, resembling the form II Rubisco from *Rhodospseudomonas palustris* (RpM) (Satagopan et al., 2014). The FLAG-Q2^{WB}O2 mutant also migrated as a single peak at 1.19 ml with an apparent molecular mass of ~286 kDa representing the Q₆O₁ hetero-oligomer (Figure 4-12A). On the 12.5% SDS gel, FLAG-Q2^{WB}O2 can be seen as 2 bands corresponding to the ~90 kDa AfO2 and the ~30 kDa FLAG-AfQ2^{WB}. The estimated molecular weight from the analytical SEC corresponded closely to the published values of 313 kDa and 285 kDa for AfM and Q2O2 respectively (Tsai et al., 2015). The elution volumes of the two proteins were also previously determined not to be concentration dependent.

A 1:1 ratio of 4 μ M HA-ECMC (monomer) and 4 μ M FLAG-Q2^{WB}O2 (monomer) resulted in a small shift in the size exclusion curve (Figure 4-12B). Both proteins were also shown to elute in earlier fractions on the 12.5% SDS gel (Figure 4-12C). This effect became more pronounced when a 1:4 ratio of 4 μ M HA-ECMC to 16 μ M FLAG-Q2^{WB}O2 was used. The elution of proteins in earlier fractions indicated a shift in the equilibrium from free to complex formation. Upon incubation of 4 μ M HA-ECMC with large excess of FLAG-Q2^{WB}O2 at 32 μ M (1:8) and 64 μ M (1:16), an additional peak of higher molecular weight was obtained (Figure 4-12B). The new peak corresponded to fractions 3 and 4 on the SDS gel and was shown to contain both HA-ECMC and FLAG-Q2^{WB}O2 (Figure 4-12C). Notably, there was a dramatic shift of almost all HA-ECMC

proteins to fractions 3 and 4, a strong indication of complex formation. The excess unbound FLAG-Q2^{WB}O2 eluted at the expected volume and corresponded to the second peak (Figure 4-12B). No protein was detected by absorbance at 280 nm to be in the void volume (0.89 ml). Taken together, the data presented here demonstrates that HA-ECMC and FLAG-Q2^{WB}O2 exist in a concentration-dependent equilibrium between free and complex *in vitro*. It was inferred from the overlay of the size exclusion curves that using a molar ratio of 1:8 Rubisco to activase, resulted in maximum complex formation as indicated by the higher A₂₈₀ absorbance for the peak of higher molecular weight (Figure 4-12B).

Figure 4-12: Analytical size exclusion chromatography analysis of interaction between HA-AfM ECMC and FLAG-Q2^{WBO2}.

(A) Size exclusion profiles of HA-AfM ECMC (left) and FLAG-Q2^{WBO2} (right) monomers. Elution peaks of molecular weight standards are indicated. (B) Overlay of size exclusion profiles of HA-AfM ECMC and FLAG-Q2^{WBO2} at varying molar ratio. (C) 12.5% SDS PAGE analysis of size exclusion fractions. 50 μ l fractions were collected from 0.95 ml elution volume. Boxes represent major peaks.

In order to investigate the specificity of the interaction observed between FLAG-Q2^{WBO2} and HA-ECMC by analytical size exclusion chromatography, the experiment was repeated with the activated form II Rubisco, HA-AfM ECM and the AfM^{E75A} ECMC mutant. Both proteins did not form stable complexes with FLAG-Q2^{WBO2} as previously analysed by native-PAGE (Figure 4-2 and 4-3). When 4 μ M of HA-ECM or HA-AfM^{E75A} ECMC protein was incubated with 32 μ M of FLAG-Q2^{WBO2}, only one peak that corresponded to both the unbound rubisco and activase was observed on the size exclusion curve (Figure 4-13A). The proteins also did not shift to earlier fractions, indicating a lack of affinity for complex formation (Figure 4-13B). This is agreeable with previous native-PAGE studies and confirmed that the interaction observed between HA-AfM ECMC and FLAG-Q2^{WBO2} is specific.

Figure 4-13: Analytical size exclusion chromatography analysis of interaction between activated Rubisco (HA-AfM ECM) and AfM^{E75A} ECMC with FLAG-Q2^{WB}O2.

(A) Size exclusion profiles of HA-AfM ECM (left) and AfM^{E75A} ECMC (right) with FLAG-Q2^{WB}O2. Elution peaks of globular molecular weight standards are indicated. (B) 12.5% SDS PAGE analysis of size exclusion fractions. 50 μ l fractions were collected from 0.95 ml elution volume. Boxes represent major peaks.

4.4 Negative stain EM analysis of HA-ECMC-FLAG-Q2^{WB}O2 complexes

With the aim of structurally characterizing the interaction between HA-ECMC and FLAG-Q2^{WB}O2, I analysed the complexes with negative stain EM. The two proteins were incubated to encourage complex formation and fractionated with analytical size exclusion chromatography. The fraction containing complexes, determined by both SDS and Native PAGE with western blotting were imaged using negative stain EM techniques.

The size-exclusion profile of the sample without cross-linker showed an early shoulder undifferentiated from the main peak (Figure 4-14A). The elution volume of the shoulder at 1.08 ml corresponded to the complex between FLAG-Q2^{WB}O2 and HA-ECMC, similar to what was obtained in previous SEC studies with lower concentrations of protein. The western blot detection of fraction 3 for both anti-AfM and anti-FLAG showed HMW smears on the native gel corresponding to the interaction between Rubisco and its activase (Figure 4-14B, panel 1; lane 3). However, free HA-ECMC and free FLAG-Q2^{WB}O2 could also be observed in lane 3 suggesting that some of the complexes may have dissociated. The unbound HA-ECMC and FLAG-Q2^{WB}O2 that did not participate in the interaction were seen to elute at later fractions (Figure 4-14A). Electron micrographs of fraction 3 at 0.1 mg/ml show numerous hetero-dimeric associations among the molecules, suspected to be complexes of one AfM and one FLAG-Q2^{WB}O2 (Figure 4-14C, left panel). These asymmetrical complexes were conformationally heterogeneous which explained the smearing observed on the native gel. Many of the complexes likely fell apart after a five times dilution of the sample, resulting in the visualization of single particles (Figure 4-14C, right panel).

A

B

C

Figure 4-14: Isolation and visualization of the HA-AfM ECMC-FLAG-Q2^{WB}O2 complex.

(A) Size-exclusion chromatography of the HA-AfM ECMC-FLAG-Q2^{WB}O2 complex. 50 μ l fractions were collected from 0.95 ml elution volume. (B) Corresponding size exclusion fractions separated by 12.5% SDS (Panel 1) and 6% native (Panels 2, 3, 4) gels. Western blot detection of 6% native gel with anti-AfM (Panel 3) and anti-FLAG (Panel 4) antibodies. Boxed fractions represent sample used for EM

analysis. (C) Raw electron micrograph of fraction 3 from SEC (0.1 mg/ml; left) and (0.02 mg/ml; right). Scale bar, 50 nm.

In order to stabilize the complexes to prevent dissociation, chemical cross-linking with glutaraldehyde was carried out. Analytical size exclusion chromatography of the cross-linked sample resulted in a high- and a low-molecular-mass peak (Figure 4-15A). The 6% SDS gel confirmed that proteins were cross-linked as majority of bands were seen to migrate above 200 kDa (Figure 4-15B, Panel 1). Low mobility bands on the native gel found in fractions 1 and 2 indicated presence of cross-linked complexes between Rubisco and activase. Unbound HA-AfM and FLAG-Q2^{WB}O2 eluted mostly in fractions 5 and 6. These could also be observed to migrate above 200 kDa on the SDS PAGE, suggesting that the monomeric subunits were cross-linked to each other. Some proteins can be seen to be retarded in the stacking gel and is speculated to be large non-specifically cross-linked complexes. Native PAGE show that Fraction 1 consisted mostly of HMW forms while fraction 2 contains both Complex* and the HMW forms. Western blot analysis confirmed the presence of HA-AfM and FLAG-Q2^{WB}O2 in both populations (Figure 4-15B, Panel 3 and 4). Therefore, both samples were analyzed with EM.

Electron micrographs of both fractions show protein complexes of two or three proteins (Figure 4-15C), consistent with multiple HMW bands observed on the native gel. This possibly represented complexes of different subunit stoichiometry (i.e. multiple activases interacting with one Rubisco). Due to similar molecular weight of HA-AfM and FLAG-Q2^{WB}O2, it was impossible to confirm from visual inspection of the micrographs if these were Rubisco-activase complexes. The heterogeneity of the complexes on the grid also complicated efforts to obtain informational 2D average classes. Therefore, no attempts were made to obtain a three-dimensional reconstruction of the glutaraldehyde cross-linked HA-AfM-FLAG-Q2^{WB}O2 complex.

A **B**

Figure 4-15: Isolation and visualization of the HA-AfM ECMC-FLAG-Q2^{WB}O2 complex cross-linked with glutaraldehyde.

(A) Size-exclusion chromatography of the HA-ECMC-FLAG-Q2^{WB}O2 complex. 50 µl fractions were collected from 0.95 ml elution volume. (B) Corresponding size exclusion fractions separated by 12.5% SDS (Panel 1) and 6% native (Panels 2, 3, 4) gels. Western blot detection of 6% native gel with anti-AfM (Panel 3) and anti-FLAG (Panel 4) antibodies. Boxed fractions represent sample used for EM analysis. (C) Raw electron micrographs of fraction 1 (0.1 mg/ml; left) and 2 (0.1 mg/ml; right). Scale bar, 50 nm. Suspected complexes of AfM and Q2^{WB}O2 are shown.

4.5 Cross-linking coupled to mass spectrometry analysis (XL-MS) of the HA-AfM-FLAG-Q2^{WB}O2 complex

Previously, XL-MS was used to gain more understanding on the molecular architecture of the CbbQO-type activases. Here, I aimed to use this approach to elucidate the interaction interface between AfM and Q2^{WB}O2. HA-AfM ECMC was incubated with FLAG-Q2^{WB}O2 to encourage complex formation. Once formed, the complexes were cross-linked with d₀/d₁₂ DSS and enriched by SEC. Fractions corresponding to the high molecular mass peak, believed to contain complexes were pooled. As per previous XL-MS experiments, the residues involved in association of the subunits were analysed by mass spectrometry in collaboration with Prof. Dr. Florian Stengel at the University of Konstanz (Appendix table 8-5, 8-6).

A total of 5 inter-links were obtained from the cross-linking of the two molecules, out of which 4 were connections between CbbO2 and AfM (Figure 4-16A), a strong indication of interaction between the adaptor protein and Rubisco. Surprisingly, no crosslinks supporting the expected interaction between the CbbO2 VWA domain to AfM were found. Instead, our data suggests that Rubisco is in proximity to regions of the CbbO2 N-terminal linker (1-560). Both the N- and C-termini of AfM were cross-linked to CbbO2, supporting its function as an adaptor protein. Among the 4 cross-links found on AfM, 2 were from lysine residues found on the N-terminus (K13), while 2 were from the C-terminus (K446 and K454) (Figure 4-16A, B). A single cross-link between HA-AfM (K206) and FLAG-Q2^{WB} hexamer (K138) was found. K138 is featured on the PS1-βH of the AAA+ motor and faces the convex face of the hexameric cup. However, based on previous cross-links between CbbQ2 and CbbO2 as well as biochemistry assays, it was determined that dynamic movements of the PS1-βH may enable it to interact with CbbO on the concave face. Thus, it is conceivable that AfM interacts with the concave face of CbbQ via CbbO. More CbbQ-Rubisco crosslinks would have to be found in future experiments to strengthen our confidence in this interpretation.

In addition to the inter-links between HA-AfM and FLAG-Q2^{WB}O2, 22 inter-links between FLAG-Q2^{WB} and CbbO2 were also identified (Figure 4-16A). Among the 22 cross-links, 18 were previously encountered in our XL-MS experiment for Q2O2, suggesting that in the presence of its substrate the conformation of Q2O2 remains largely the same, with both N- and C-termini featured on the concave face of CbbQ. Out of the 4 unique cross-links in this dataset, 2 supported the previously established structure with lysine residues on the concave face of CbbQ cross-linked to the C-terminus of CbbO2. Two other cross-links were between the FLAG-tag of FLAG-Q2^{WB} and CbbO2.

Taken together, these initial XL-MS results are consistent with a model both the N- and C-termini of CbbO are believed to be located at the concave surface of the ring. When Rubisco is present, the linker domain could provide the flexibility required for the C-terminus to interact with Rubisco and bring it to proximity with the concave face of the hexamer (Figure 4-16C). It is conceivable that CbbO2 makes multiple contacts with

AfM and possibly recruit it to the concave face of CbbQ where activation may take place.

Figure 4-16: Cross-linking coupled to mass spectrometry analysis (XL-MS).

Cross-link maps for the FLAG-Q2^{WB}O2 and HA-AfM ECMC. Green cross-links represent intermolecular links. Red and purple connections represent inter-subunit self-links and inter-subunit or intra-subunit self-cross-links, respectively. VWA domains are in light green. (B) Cross-linking sites detected on the surface of HA-AfM modeled with the crystal structure of the hexameric *Rhodospseudomonas palustris* (Satagopan et al., 2014). The RbcL subunits are shown in yellow and gray. One dimer is shown in cartoon representation under the transparent molecular surface of the hexamer. Both the side (left) and top (right) view are shown. Lysines cross-linked to CbbO2 and FLAG-CbbQ2^{WB} are shown in magenta and blue, respectively. Lysines highlighted are labeled with the corresponding residue of AfM in brackets. (C) Proposed model of CbbQO binding to form II Rubisco. CbbQO is modeled with the crystal structure of CbbQ2 (pink ribbon; Tsai and Ye et al., Unpublished) and the VWA domain of the integrin alpha-L VWA domain (blue ribbon, PDB-ID: 1T0P) (Song et al., 2005) superimposed on the negative stain EM reconstruction of Q2O2. CbbQO interacts with inhibited Rubisco via the VWA domain of its adaptor protein CbbO, engaging the exposed acidic residue Glu75 (green) on the RbcL subunit of the *R. palustris* Rubisco. Lysine residues cross-linked to CbbO2 are shown in magenta.

4.6 DISCUSSION

Rubisco activase is required for the ATP-dependent maintenance of Rubisco activity. A major challenge now is to deduce the mechanistic details of this new activase system. In this chapter, it was demonstrated that the interaction between the form II Rubisco, AfM and the Walker B mutant form of its activase, Q2^{WB}O2 can be biochemically stabilized with chemical cross-linkers and confirmed using native PAGE (Figure 4-1) and analytical size exclusion chromatography (Figure 4-12).

It was observed that Q2^{WB}O2 interacts specifically with the stably inhibited form of Rubisco, ECMC (Figure 4-2), with several conserved elements on the RbcL, including the surface acidic E75 (Figure 4-3) and the C-terminus (Figure 4-4) being critical for the interaction to take place. This indicates a mechanism where the initial recognition of Rubisco is mediated by these domains. Previous biochemical efforts support these findings as mutations of these regions abrogated stimulation of the ATPase activity of Q2O2 and Rubisco activation (Tsai et al., 2015).

The CbbQO activase differs from the Red- and green-type Rca systems in that it requires the two-part coordination of the AAA+ motor CbbQ and the adaptor CbbO to be functional. Cross-linking followed by native gel analysis did not show evidence of

interactions between the Q2^{WB} hexamer and AfM ECMC (Figure 4-5), emphasizing the importance of the adaptor CbbO for a stabilized and high affinity binding between the two proteins. Strikingly, the soluble chimeric adaptor O1O2VWA is able to form complexes with AfM even in the absence of the motor CbbQ (Figure 4-8). Therefore, remodelling of Rubisco likely involves a primary interaction with CbbO which recruits the substrate to the hetero-oligomeric Rca. Given that Q2-O1O2VWA hetero-oligomer is non-functional in activating AfM (Tsai et al., Unpublished), important interactions between the N-terminal linker (1-444) of CbbO2 and AfM, responsible for positioning the substrate during activation is likely to be missing. It remains to be seen if CbbQ is in direct contact with AfM.

Counter to our hypothesis that the C-terminal MIDAS motif is critical in mediating the interaction, neither single nor double MIDAS motif mutations of CbbO2 abrogated interactions with AfM (Figure 4-6), pointing to additional contacts between the two proteins. This is supported by our XL-MS data which indicate that the N-terminal linker of CbbO is also involved in binding Rubisco (Figure 4-16). Native gel analysis indicated that the FLAG-O2VWA domain is able to form complexes with AfM in a cross-linker dependent manner but was unable to interact with the form I Rubisco AfLS (Figure 4-11). This reflects that the interaction between the FLAG-O2VWA and AfM is specific. However, it was observed that the FLAG-O2VWA domain interacted indiscriminately with all the different forms of AfM, inactive, active and inhibited (Figure 4-11A), an indication that additional contacts between the activase and AfM are responsible for the specificity towards the inhibited Rubisco ECMC.

A form II Rubisco-activase complex has not been reported and this study represents the first structural characterization of the AfM-Q2^{WB}O2 complex. Analytical size exclusion chromatography of the native proteins showed that complexes eluted in earlier fractions (Figure 4-14). When visualized under the EM, they were seen as a mixture of hetero-dimeric associations and single particles. This indicated a tendency for the complexes to fall apart without the stabilization effect of cross-linkers. Glutaraldehyde cross-linking produced a heterogeneous mixture of complexes that comprised of different number of particles (Figure 4-15). The heterogeneity of particles on the grid together with the similarity in molecular weight of the two proteins made it difficult for downstream analysis. To resolve these difficulties, the interactions studies would be

extended to the form I Rubisco, AfLS because of its larger size (~500 kDa) and characteristic cuboid shape.

XL-MS was used to determine how the activase binds to Rubisco. Several cross-links between the CbbO linker and AfM suggest that CbbO may make multiple points of contact with both the N- and C- termini of Rubisco to secure its hold on the enzyme (Figure 4-16A). Although no cross-links were found between the O2VWA domain and Rubisco, earlier native PAGE and analytical SEC studies showed that the conserved acidic residue on RbcL (E75 for AfM) is essential for interaction, indicating that the VWA domain must participate in the association with Rubisco. In addition, the cross-links between FLAG-AfQ2^{WB} and CbbO2 supported our findings from the Q2O2 XL-MS dataset with the concave face of CbbQ serving as an intrinsic hub of interactions between CbbQ and CbbO (Figure 3-3). It is possible that the flexibility of the N-terminal linker enables the connecting VWA domain at the C-terminus to selectively recruit differently spaced and oriented Rubisco substrate and recruit it to the concave face of the hexamer.

The red-type RsRca has a conserved pore-loop, which engages the C-terminus of Rubisco and threads it through the central pore to allow release of the inhibitor from the active site (Mueller-Cajar et al., 2011; Bhat et al., 2017b). However, biochemical characterization of the CbbQ2 pore loop mutants show that they were still capable of activating Rubisco (Tsai and Ye et al., Unpublished), suggesting that a pore-loop threading mechanism might not apply to this system. This raises the question of how the critical C-terminus of *A. ferrooxidans* Rubisco is involved in the activation process. Based on the XL-MS data, the C-terminus of Rubisco interacts with the N-terminal linker of CbbO. Hence, the nucleotide-dependent conformational changes within CbbQ may apply tension force to the bound Rubisco through the long and flexible linker domain of CbbO, thereby disrupting the Rubisco active site. This is similar to how the ribosome maturation factor Rea1, believed to be the largest yeast protein at 560 kDa, removes the ribosome biogenesis factor Rsa4 from pre-60S ribosomal subunits in the nucleoplasm to allow the export of the preribosome into the cytoplasm (Barrio-Garcia et al., 2016). Rea1 possess six tandem AAA+ modules connected to one VWA domain at its C-terminus via a long flexible linker, resembling the architecture of CbbQO. The Rea1 VWA domain binds the substrate Rsa4 via its MIDAS motif, whereas the AAA

moiety is bound to the Rix1-subcomplex. Movement of the motor domain during ATP hydrolysis is transferred to the linker to generate the force required to disrupt the complex, leading to the release of these factors from the pre-60S particle (Sosnowski et al., 2018).

In conclusion, the data presented in this study led to a preliminary model for the mechanism of action for the CbbQO activase and provided the framework for further characterization of the form I CbbQO Rca system. The integrative approach of using both structural and biochemical methods will contribute to a more detailed understanding of Rubisco activation in chemoautotrophs.

5. INSIGHTS INTO THE CBBQO RUBISCO ACTIVATION MECHANISM FOR FORM I RUBISCO

In the previous chapter, characterization of interactions between the form II Rubisco, AfM and the Walker B mutant form of its activase was described. However, the size and heterogeneity of complexes during negative stain EM made data processing difficult. In this chapter, the interaction studies were extended to the form I Rubisco, AfLS and the functional chimeric activase Q3O1. The mechanistic details of the form I CbbQO Rca system was investigated using Native PAGE and revealed many similarities to the form II. The AfLS-Q3O1 complexes were successfully isolated with analytical SEC and examined with negative stain EM. Single particle analysis of negative stain EM images produced a 3D reconstruction of the complex, with one Q3O1 bound to the side of the cuboidal Rubisco at 21 Å.

5.1 Purification of the chimeric Rubisco activase Q3O1

Previous biochemical experiments by Dr Candace Tsai in the lab showed that the chimeric CbbQO-type activase, Q3O1 is functional in activating the form I Rubisco AfLS. Additionally, preliminary experiments also suggested that Q3O1 is less recalcitrant than the WT Q1O1 with no precipitation observed during purification, making it a good potential candidate for structural studies. Therefore, with the aim of investigating the form I Rubisco-activase interaction, the chimeric activase, Q3O1 was purified.

CbbQ3 is encoded in the carboxysomal operon while CbbO1 can be found directly downstream of the form I Rubisco encoding genes, *cbbLS-1*. The proteins CbbQ3 and CbbO1 form a “chimeric” hetero-oligomer Q3O1 when co-expressed, similar to the WT activases Q1O1 and Q2O2. Purification methods for Q3O1 are largely the same as for the WT proteins. Both CbbQ3 and CbbO1 were recombinantly expressed in *E. coli* BL21 as a fusion protein with a His₆Ubiquitin (His₆Ub) tag at the N-terminus to increase protein solubility and facilitate capture with Nickel-nitrilotriacetic acid (Ni-NTA), immobilized metal ion affinity chromatography (IMAC). The His₆-ubiquitin tag was subsequently cleaved by the specific protease, USP2 (Figure 5-1A). Unlike the WT activases such as Q1O1 and Q2O2 which had a tendency to precipitate at this stage, no

precipitation was observed before or after cleavage of the His₆Ub tag for the chimeric activase Q3O1. Eluted proteins were then used for ion-exchange chromatography (IEX) and size-exclusion chromatography (SEC) to further enrich for pure hetero-oligomers (Figure 5-1B). Typically, protein purification of CbbQO-type activases will yield 1:2 ratio of CbbO (~80 kDa) to CbbQ (~30 kDa) band intensity when analysed with SDS-PAGE. The fraction determined to contain mostly hetero-oligomeric complexes was concentrated to be used for subsequent studies. A typical yield of protein from 1 L culture of *E. coli* BL21 was found to be ~10 mg of pure protein.

Figure 5-1: Purification of Q3O1.

(A) Representative 12.5% SDS PAGE separation of proteins from IMAC Ni-NTA and various fractions from IEX Source15Q column. L: lysate, S: soluble, FT: flow through, W: wash, BC: before cleavage by USP2, AC: after cleavage. (B) Size-exclusion profile of Q3O1 on a Superdex 200 16/600 column (left). Purified fractions on a 12.5% SDS PAGE with CbbQ3 and CbbO1 subunits indicated (right).

5.2 The chimeric activase Q3O1 is a functional Rca

In order to confirm that Q3O1 is a functional Rca, a continuous enzyme-coupled spectrophotometric assay that couples the oxidation of NADH to the formation of 3-phosphoglycerate (3PG) was performed to assess Q3O1 for its ability to activate inhibited Rubisco. The Rubisco assay was conducted using conditions previously established by Dr. Candace Tsai and included the presence of 5% v/v of the crowding agent polyethylene glycol 3,350 (PEG) to facilitate a more pronounced effect of activase activity.

Three forms of Rubisco were generated to be used for the spectrophotometric assays: the fully activated carbamylated Rubisco complex (ECM), the apoenzyme inhibited by its substrate RuBP (ER) and the activated enzyme bound to the tight-binding transition state analogue 2-carboxyarabinitol 1,5-bisphosphate (CABP), (ECMC). While ECM displayed linear activity, ER and ECMC complexes were inactive as expected for the inhibited enzyme (Figure 5-2A). Rubisco activity increased when ER and ECMC was assayed in the presence of Q3O1, consistent with active removal of inhibitors from the rubisco active site and demonstrating that Q3O1 is a functional Rca.

While investigating the ATPase activity of Q3O1, it was shown that that Q3O1 has a basal ATPase rate of ~ 3 /min/protomer (CbbQ), similar to the WT Q1O1 (Tsai et al., 2015) (Figure 5-2B). An increase in the ATPase activity of Q3O1 was observed in the presence of AfLS, with a doubling of the ATPase rate when ECMC was added. Taken together, the evidence suggest that Q3O1 is stimulated by interaction with its substrate, the stably inhibited AfLS and is capable of rescuing it from inhibition.

Figure 5-2: Q3O1 activity assays.

(A) Rubisco activity assays of activated (ECM) and inactivated (ER and ECMC) AfLS (0.3 μM active sites, 20 mM NaHCO_3 and 5% PEG) in the absence and presence of Q3O1 (0.27 μM oligomer). (B) ATPase activity assays of 0.27 μM Q3O1 complex was measured in the absence and presence of the indicated Rubisco complex (3 μM active sites). Error bars indicate the mean and s.d. of at least three independent experiments.

5.2.1 The functional chimeric activase Q3O1 forms complexes with stably inhibited form I Rubisco, AfLS ECMC

Having established that the chimeric activase Q3O1 is a functional Rubisco activase, it must interact with Rubisco to remove tight-binding inhibitors. Experiments described

in the preceding chapters have shown that cross-linked form II Rubisco-activase complexes were able to be visualized on native PAGE as distinct high molecular weight (HMW) bands. In order to assess if complexes of the inhibited form I Rubisco, AfLS ECMC with the chimeric mutant activase Q3O1 could be stabilized using glutaraldehyde, the cross-linked products were similarly analysed by native PAGE after incubation. Cross-linked Rubisco-activase complexes would have lower mobility and present different migration patterns from individual protein standards. As seen in figure 5-3A lanes 1 and 2, AfLS ECMC resolved as a distinct band, consistent with a conformationally homogenous population. In lane 3 and 4, the activase Q3O1 consists of a population of free CbbQ3 hexamers seen as a faster migrating, less intense band below the main Q3O1 hetero-oligomer band. The free CbbQ3 hexamers remain from purification steps described above, which were unable to be completely separated from the hetero-oligomer population. Addition of the crosslinker glutaraldehyde caused a downward shift in the migration pattern due to charge modifications as a result of lysine residues reacting with the cross-linker (Figure 5-3A, lanes 2 and 4).

The appearance of a faint HMW band was observed when both Rubisco and activase were added together in lanes 7, different from the HMW smears observed for the form II system (Figure 4-1). These bands became more intense in the presence of ATP (Figure 5-3A; lane 9). Addition of the cross-linker in lane 10 resulted in the appearance of several distinct HMW bands, reminiscent of that observed for AfM and Q2^{WB}O2 (Figure 4-1). These bands could be seen to migrate above the Rubisco and activase proteins that ran in isolation and possibly correspond to one Rubisco with different numbers of activases bound. Notably, it was observed that the bands corresponding to free Rubisco and activase became less intense suggesting that some of the proteins have shifted into the HMW bands. It is interesting to note that in this system, the form I Rubisco-activase complex formation is dependent upon the presence of ATP while HMW complexes of form II Rubisco-activase complexes could be observed even in the absence of ATP, an indication that regulation vary between the isoforms.

These Rubisco-activase complexes were not observed in the presence of ADP or the non-hydrolysable analog ATPγS, where only very faint HMW bands could be seen (Figure 5-3B, C). The absence of distinct HMW bands suggest that ATP binding and

conformational changes during ATP hydrolysis is required to generate stable complexes, and further supports the functional relevance of the HMW complexes.

Figure 5-3: Cross-linking of AfLS and Q3O1 with glutaraldehyde.

(A) 6% native PAGE analysis of complex formation between AfLS ECMC and Q3O1 in the presence of ATP. (B) In the presence of ADP. (C) In the presence of ATP γ S. 5 μ M inhibited Rubisco (monomer) and 20 μ M activase (monomer) were cross-linked in the presence or absence of 5 mM nucleotide and 5 mM MgCl₂. A final concentration of 0.125% v/v of glutaraldehyde was used for crosslinking for 10 min at RT. The crosslinking reaction was quenched by addition of 1 M Tris-HCl, pH 8.0 to a final concentration of 50 mM.

5.2.2 Conformational changes induced by ATP hydrolysis are required for interaction between AfLS and Q3O1

In order to confirm the importance of ATP hydrolysis in interaction of Q3O1 with Rubisco, a Walker B motif mutant Q3^{WB}O1 (E107Q), where ATP hydrolysis ability is abolished, was constructed. A representative of the native PAGE analysis of cross-linking between AfLS ECMC and Q3^{WB}O1 is shown in figure 5-4. The hetero-oligomer Q3^{WB}O1 resolved as a distinct band, with a population of faster migrating free CbbQ hexamers. Addition of MgCl₂ and ATP caused the dissociation of the hexamer CbbQ3^{WB} from the hetero-oligomeric Q3^{WB}O1 (Figure 5-4; lanes 5 and 9), as was previously seen for the Q2^{WB}O2 (Figure 4-1). This could be observed as a lower migrating band below the population of Q3^{WB}O1. Previous immunoblot experiments have confirmed the identity of the hexamer within this lower molecular weight band

(Figure 4-1B). The reason for this dissociation is still unidentified. When Rubisco and Q3^{WB}O1 was cross-linked in the presence of ATP, only very faint bands could be seen, as compared to the distinct HMW bands observed with the WT Q3O1 (Figure 5-3). This is in support of our interpretation that conformational changes induced by ATP hydrolysis are obligatory for a strong interaction between Q3O1 and AfLS.

Figure 5-4: Cross-linking of AfLS and Q3^{WB}O1 with glutaraldehyde.

(A) 6% native PAGE analysis of complex formation between AfLS ECMC and Q3^{WB}O1. Cross-linking was performed as described in Figure 5-3.

5.2.3 Q3O1 is more suitable for structural characterization than the WT Q1O1

With the goal of achieving a form I Rubisco-activase complex, I looked at several activases capable of activating inhibited AfLS. The WT Q1O1 activase encoded downstream of form I Rubisco AfLS is a functional Rca capable of removing tight-binding inhibitors from Rubisco (Tsai et al., 2015). After incubation with AfLS ECMC in the presence of ATP and cross-linker, very diffuse HMW smears corresponding to the size of Rubisco-activase complexes were detected (Figure 5-5; lane 10). This is unlike the intense bands seen for Rubisco-Q3O1 complexes (Figure 5-3; lane 10) suggesting that cross-linking did not succeed in stabilizing the transient Rubisco-Q1O1

complexes. It is plausible that Q1O1 has a lower affinity for its substrate (AfLS ECMC) compared to the chimeric Q3O1. The biochemical properties of the two proteins differ as Q1O1 is prone to precipitation and has a tendency to elute in the void volume of gel filtration column after a few freeze-thaw cycles, unlike Q3O1 which remains stable and soluble. Taken together, this data suggests that Rubisco-Q3O1 complexes will be more suitable for downstream structural studies and hereafter all further experiments on form I Rubisco-activase complexes were conducted with AfLS and the functionally active chimeric activase, Q3O1.

Figure 5-5: Cross-linking of AfLS and Q1O1 with glutaraldehyde.

(A) 6% native PAGE analysis of complex formation between AfLS ECMC and Q1O1. Cross-linking was performed as described in Figure 5-3.

5.2.4 An intact RbcL C-tail is required for a high affinity interaction with Q3O1

To obtain more mechanistic insights into the form I Rubisco activation system, I took advantage of the finding that complex formation can be visualized on the native gel and used this efficient method to investigate the importance of certain conserved elements on the Rubisco LSU for complex formation. The red-type RsRca remodels Rubisco by

transient threading the C-terminus through the pore which leads to destabilization of the catalytic site and liberation of the inhibitor (Mueller-Cajar et al., 2011; Bhat et al., 2017b). Truncation of two to four amino acids from the Red-type Rubisco RsLS renders it non-activable by RsRca. The C-terminus of RsLS covers the catalytic site and proper attachment of the C-terminus to the body of the enzyme is a prerequisite for engagement with RsRca.

Alignment of Rubisco primary sequences associated with CbbQO revealed an unique HK/R motif at the C-terminus (Tsai et al., 2015). Previous biochemical experiments have determined that both presence and position of this motif is strictly required for activase stimulation and function. Deleting one or two amino acids from the C-terminal end abolished both ATPase stimulation and activase function, suggesting that the C-terminus is crucially involved in the substrate recognition and activation process similar to the Red-type Rubisco activation system.

Results from the previous chapter has shown that the form II Rubisco mutant AfM^{ΔC1}, with one amino acid deletion from the C-terminus do not form HMW complexes with Q2^{WB}O2 activase (Figure 4-4). To investigate if the same phenomenon can be observed for the form I Rubisco activation system, native PAGE analysis was used to assess for Rubisco-activase complexes between the mutant form I Rubisco, AfLS^{ΔC1} with one amino acid deletion from the C-terminus and Q3O1 activase. AfLS^{ΔC1} migrated as a single distinct band (Figure 5-6, lane 1), similar to the WT AfLS. Incubation with Q3O1 with or without the presence of ATP and the cross-linker produced no HMW bands indicative of complex formation (Figure 5-6, lanes 9 and 10). This is consistent with the previously published biochemical data (Tsai et al., 2015) and points to a role for an intact C-tail in substrate recognition and engagement for Rca.

Figure 5-6: Cross-linking of AfLS^{ΔC1} and Q3O1 with glutaraldehyde.

(A) 6% native PAGE analysis of complex formation between AfLS^{ΔC1} ECMC and Q3O1. Cross-linking was performed as described in Figure 5-3.

5.2.5 The surface exposed Asp82 on Rubisco large subunit (RbcL) is essential for facilitating interactions with Q3O1

The surfaced exposed acidic residue D82 on the LSU of AfLS was determined to be critical for activation and stimulation of CbbQO ATPase activity (Tsai et al., 2015). This acidic residue is believed to be the sixth coordination site for the MIDAS motif on the adaptor CbbO and essential for initial contact between Rubisco and activase. It has previously been found that when the corresponding residue E75 on the form II Rubisco AfM was mutated to alanine, the mutant did not form stabilized HMW complexes with the Q2^{WB}O2 as analysed with native PAGE (Figure 4-3). To find out if the form I Rubisco activation system maintains a similar mechanism, the mutant AfLS^{D82P} ECMC was incubated with Q3O1 and assessed for complex formation. AfLS^{D82P} ECMC migrated as a single distinct band (Figure 5-7; lane 1), similar to the WT. After mixing with Q3O1, faint bands running higher than the proteins standards could be observed

(Figure 5-7; lane 10). However, it is far less intense than those formed by the WT, suggesting a reduction in affinity and further corroborating evidence of the importance of this acidic residue in interaction with the activase.

Figure 5-7: Cross-linking of AfLS^{D82P} and Q3O1 with glutaraldehyde.

(A) 6% native PAGE analysis of complex formation between AfLSD82P ECMC and Q3O1. Cross-linking was performed as described in Figure 5-3.

5.2.6 The adaptor protein CbbO is required to form complexes with the stably inhibited form I Rubisco, AfLS ECMC

Previous biochemical data supported a requirement of the hetero-oligomer in activating Rubisco as the CbbQ motor alone was unable to rescue Rubisco inhibition (Tsai et al., 2015). When analysed with native PAGE, it was discovered that the form II Rubisco AfM did not form complexes with the Q2^{WB} hexamer alone, pointing to the importance of the adaptor CbbO in engaging the substrate. It would be interesting to investigate if this requirement is also necessary in the form I Rubisco activation system. Therefore, native gel analysis was used to study the interaction of the Q3 hexamer with AfLS ECMC. Similar to the Q2^{WB} hexamer, Q3 hexamer migrated as a smear on the native gel (Figure 5-8), indicating that conformational heterogeneity exists within the hexamer population. In the presence of ATP, migration pattern of Q3 remained as a smear unlike

the Q2^{WB} hexamers which migrated as a distinct band (Figure 4-5). As Q3 hexamers are capable of hydrolysing ATP, this could indicate a mixed nucleotide state with conformational differences within the subunits which manifests as a smear on the native gel. Incubation of Q3 with AfLS ECMC did not produce any indication of complex formation as the migration pattern remains largely similar to when both proteins were run in isolation (Figure 5-8). These results were interpreted to mean either a lack of direct interaction between the hexameric motor and Rubisco or a low-affinity and transient interaction that cannot be captured with cross-linking. This experiment demonstrated the importance of the adaptor CbbO for a stable interaction with Rubisco.

Figure 5-8: Cross-linking of AfLS and Q3 with glutaraldehyde.

(A) 6% native PAGE analysis of complex formation between AfLS ECMC and Q3. 5 μ M inhibited Rubisco (monomer) and 10 μ g of CbbQ3 was used in the cross-linking reaction as described in Figure 5-3.

5.3 Characterization of the AfLS-Q3O1 complex with analytical size exclusion chromatography (SEC)

Given that preliminary native PAGE experiments facilitated visualization of complex formation between AfLS ECMC and Q3O1 via cross-linking, I wanted to determine if the AfLS ECMC-Q3O1 complexes could be isolated in a biochemically stable form

using analytical size exclusion chromatography (Superdex 200 PC3.2/30 on an ÄKTAmicro system). The size-exclusion profile showed that AfLS ECMC eluted as a monodisperse peak at 1.17 ml, corresponding to a molecular mass of approximately 500 kDa (Figure 5-9A). Fifty microliter fractions were collected from 0.90 ml elution volume. The corresponding 12.5% SDS-PAGE showed that the RbcL (52 kDa) of AfLS could be found in fractions 4 to 9 (Figure 5-9B). Due to the high mobility of the SSU (13 kDa), it was not able to be resolved on the gel and assumed to have co-eluted with the LSU. Q3O1 eluted with a major peak at 1.28 ml which corresponds to the Q₆O₁ hetero-oligomer at ~ 300 kDa and a minor shoulder of lower molecular weight (Figure 5-9A). This is caused by the contamination of Q3O1 with a population of free Q3 hexamers that could not be separated during purification. SDS-PAGE showed that fractions 5 to 7 contained the hetero-oligomeric Q3O1, with representative bands at 30 and 88 kDa for CbbQ and CbbO respectively while fractions 8 and 9 consisted mainly of CbbQ (Figure 5-9B).

Figure 5-9: Analytical SEC analysis of interaction between AfLS ECMC and Q3O1.

(A) SEC profiles of AfLS ECMC, Q3O1 and the cross-linked sample containing both proteins. Elution peaks of molecular weight standards are indicated. (B) 12.5% SDS PAGE analysis of size exclusion fractions. 50 μ l fractions were collected from 0.90 ml elution volume. Boxes represent major peaks.

I incubated 8 μ M AfLS ECMC (monomer) with an excess of FLAG-Q2^{WB}O2 at 32 μ M (monomer) (~1:5 molar ratio Rubisco to activase) in the presence of ATP before cross-linking with the cleavable cross-linker dithiobis(succinimidyl propionate) (DSP). The cross-linking with DSP can be reversed to allow analysis of the components of each fraction with SDS-PAGE. The size-exclusion curve of AfLS ECMC and Q3O1 treated with the cleavable cross-linker, DSP in the presence of Mg^{2+} and ATP caused the formation of a new higher molecular weight peak at 1.05 ml (Figure 5-9A). The components of the fractions corresponding to the new peak were analysed by reversing the cross-linking with the reducing agent DTT. This allowed us to resolve the SEC fractions on an SDS gel and confirmed the presence of CbbQ3, CbbO1 and the LSU of AfLS in fractions 2 to 4 (Figure 5-9B). Notably, there was a shift in the elution pattern

of both AfLS and Q3O1 to earlier fractions, suggesting the formation of HMW Rubisco-activase complexes as seen on the native gel earlier (Figure 5-3). The free Rubisco and activase that did not participate in the complex formation eluted in the two later peaks with minor shifts in elution volume. No protein was detected by absorbance at 280 nm to be in the void volume (0.89 ml).

5.4 Isolation of AfLS ECMC-Q3O1 complex for negative-stain EM

After establishing that the AfLS ECMC-Q3O1 complex could be isolated, the Rubisco-activase complexes were analysed with negative stain EM. To this end, the proteins were treated with the cross-linker glutaraldehyde in the presence of Mg^{2+} and ATP and used the analytical SEC to separate complexes from the unbound proteins. The size-exclusion curve of AfLS ECMC and Q3O1 treated with glutaraldehyde differed from the profile of proteins treated with DSP, with a higher A_{280} absorbance at 1.05 ml which was previously determined to consist of AfLS-Q3O1 complexes (Figure 5-10). Notably, the size of the peak corresponding to unbound proteins reduced in size suggesting a shift of more proteins towards the HMW fractions. This discrepancy in the SEC profile between the two cross-linkers could be attributable to the different lengths of the spacer arm. Since glutaraldehyde has a much shorter spacer at ($\sim 5\text{-}\text{\AA}$) as compared to DSP ($\sim 12\text{ }\text{\AA}$), it is considerably more aggressive and is likely to be more effective in cross-linking transient interactions between proteins.

Fifty microliter fractions of the AfLS ECMC-Q3O1 cross-linking reaction were collected from 0.95 ml elution volume and analysed by native PAGE. Fractions 1 to 7 collected from the cross-linked sample showed that the new peak contained the HMW bands observed in earlier experiments (Figure 5-10). Bands corresponding to unbound Rubisco and activase eluted in later fractions. The migration pattern of the collected fractions is similar to that of the form II CbbQO Rca system (Figure 4-15) and is also reminiscent of that observed for the red-type RsRca system (Bhat et al., 2017b).

Figure 5-10: Isolation of the AfLS-Q3O1 complex with analytical SEC.

Size exclusion profile of the AfLS ECMC-Q3O1 complex. 50 μ l fractions were collected from 0.95 ml elution volume (Top). Elution peaks of globular molecular weight standards are indicated. Corresponding size exclusion fraction fractions separated by 6% native gel (bottom). Boxed fraction was selected for negative stain EM analysis.

5.5 Negative-stain EM analysis of isolated AfLS ECMC-Q3O1 complex

In order to obtain more structural insights into the Rubisco-activase complexes, fraction 2 of the SEC (Figure 5-10) believed to contain the cross-linked complexes was used for initial screening with negative-stain EM. Electron micrographs of the fractions show a heterogenous population of protein complexes, suggesting that they have been successfully cross-linked (Figure 5-11A). Among the complexes, the textbook cuboidal

architecture of form I Rubisco could be easily identified to be associated with varying numbers of smaller proteins. This is consistent with multiple activases interacting with one Rubisco and is in accordance with a ladder of distinct HMW bands observed on the native PAGE. Dimensions of the cube-shaped particles of ~10 nm by ~10 nm matches the size of Rubisco (Hauser et al., 2015). The size of the partner protein at the longest axes ~4 nm is also similar to that of the CbbQ hexamer on its side (Tsai et al., 2015).

Our reference-free analysis of two-dimensional class averages (8,332 particles) revealed classes where mostly one or two activases were bound to the side of Rubisco (Figure 5-11B). Because the activases bound to Rubisco were represented by strong and clear densities, they are likely to be specific interactions. Interestingly, classes where up to three activases are bound were observed. This is consistent with class averages obtained from the red-type Rca cross-linked to RsLS, where up to three activases are bound to Rubisco in rare cases (Bhat et al., 2017b). This suggested that a full occupancy of up to four activases can bind to Rubisco at one time, as single particles with four activases associated with one Rubisco can be observed on micrographs (Figure 5-11B; right). Steric constraints may limit the maximum number of bound activases to four as particles with more than four activases bound was not observed in any micrographs or represented in 2D class averages. In some well-resolved classes, the four-fold symmetry (top view) of Rubisco could be clearly observed and positions of the RbcL dimers could be discerned. However, the density of activases bound to Rubisco were generally less defined and distinguishing internal features remained difficult. This indicated a greater degree of conformational heterogeneity of the activase despite being cross-linked to Rubisco and could be due to the continuum of flexibility exhibited by the adaptor protein, CbbO in attachment to Rubisco.

Figure 5-11: Negative stain EM image analysis of the AfLS-Q3O1 complex.

(A) Representative micrographs of complexes 0.05 mg/ml. Scale bar, 50 nm. (B) Selected reference free 2D class averages of complexes derived from negative stain images in EMAN2 (left). Selected particles with 4 Q3O1 activases bound to 1 Rubisco (right). Scale bar, 20 nm. (C) 3D reconstruction of complexes (top) and 3D reconstruction of complexes fit with the crystal structure of *H. neapolitanus* Rubisco (PDB-ID: 1SVD) and crystal structures of CbbQ2 (pink ribbon; Tsai and Ye et al., Unpublished) and the integrin

alpha-LVWA domain (blue ribbon; PDB-ID 1T0P) (Song et al., 2005).

Classes were manually inspected and well-defined classes with one activase bound to Rubisco (3234 particles) were selected for 3D reconstruction. A well-defined initial model was used for further refinements with no symmetry applied to obtain the final map. The final 21.1 Å structure included recognizable features of a form I Rubisco with an activase bound to the equatorial region of AfLS, at the interface of two RbcL. The crystal structure of the closely related *H. neapolitanus* Rubisco (PDB-ID: 1SVD) could be accommodated inside the density of AfLS. Despite the less refined density for Q3O1, it was possible to discern that the conserved concave face of the hexamer is oriented towards Rubisco, with the adaptor protein CbbO sandwiched between the AAA+ motor and the substrate (Figure 5-11C). Our model of Q2O2 from negative stain EM analysis placed the VWA domain in the central cavity of the AfQ2 cup (Figure 3-2). Biochemical evidence supported a model where the MIDAS motif of the VWA domain on CbbO interacts with AfLS at the surface exposed acidic residue D82. In accordance with this, additional density could be observed inside the AfQ3 ring making direct contact with Rubisco (Figure 5-11C).

Q3O1 is in contact with one of the Rubisco active site, at the interface of a RbcL dimer. This supports of a “side on” model where the activase mainly associates with the RbcL as opposed to a “top-on” model where the activase is believed to make substantial contacts with the RbcS (Wachter et al., 2013) (Bhat et al., 2017b). Current model of Rubisco activation by CbbQO involve the functionally critical VWA domain making primary contact with Rubisco followed by communication to CbbQ motor. However, low resolution of the electron density map did not permit the precise contact points between Rubisco and activase to be defined unambiguously. It was not possible to determine if the adaptor CbbO alone or both CbbQ and CbbO establishes contact with Rubisco. Full characterization of the detailed interaction must await higher resolution cryo-EM studies.

5.6 Discussion

A central question in understanding the mechanisms of the CbbQO-type activase is how the activase interacts with its substrate, Rubisco. To address this, the form I Rubisco-activase complex was isolated for structural analysis by negative stain EM. It was shown that the chimeric activase Q3O1 is a functional Rca and is capable of rescuing the form I Rubisco AfLS from inhibition (Figure 5-2). Compared to the WT Q1O1, Q3O1 forms a more stable hetero-oligomer that is not as prone to precipitation and therefore more suitable for downstream structural analysis (Figure 5-5). It was shown with native PAGE analysis that the AfLS-Q3O1 complex could be captured via cross-linking only in the presence of MgCl₂ and ATP, demonstrating that most of the observed crosslinking was specific (Figure 5-3). The transient nature of the Rubisco-activase interaction dictates that complexes can only be stabilized with the addition of the cross-linker and present as bands with lower electrophoretic mobility. The lack of distinct HMW bands with ADP, ATP_γs and the Walker B mutant activase Q3^{WB}O1 suggest that nucleotide-induced conformational changes during ATP hydrolysis is important for both substrate engagement and activation. This is in contrast to several other AAA⁺ proteins where ATP binding is often sufficient for substrate association and Walker B mutants have been used as substrate ‘traps’ (Weibezahn et al., 2003; Hanson and Whiteheart, 2005a). It is interesting to note that where the Q2^{WB}O2 was capable of interacting with the form II Rubisco AfM, only very faint HMW bands could be observed with Q3^{WB}O1 and AfLS ECMC suggesting different regulation between isoforms.

Examples of Rubisco-activase interaction studies are limited due to the dynamic character of the interaction between the proteins. A recent study of the red-type Rca cross-linked to the *R. sphaeroides* form I Rubisco reflected our observation that multiple activases can work on a Rubisco at any one time (Bhat et al., 2017b). Our 3D model suggests a configuration where Q3O1 is bound to the equatorial region of Rubisco where the catalytic sites are located, making extensive contacts with the Rubisco large subunits (Figure 5-11C). This differs slightly from the cryo-EM structure of the red-type Rubisco-activase complex where the hexameric activase is docked in a tilted position to a corner of the cube-shaped Rubisco. However, it was proposed that the position of Rca shifts after initial substrate engagement, and is likely to be positioned

closer to the equator of Rubisco. The overall architecture of the Rubisco-activase complex in the reported model is comparable with our density and confirms the relevance of our study.

The CbbQO activase is unique in hosting an adaptor protein for sensing its substrate and regulating the ATPase activity of the motor. Due to its inherent flexibility, structural information for CbbO is lacking. Previously by using negative stain EM and XL-MS, it was established that the C-terminal VWA domain of CbbO occupies the central cavity of the CbbQ cup, making extensive interactions with the concave face of the hexamer. Unlike the other activases studied thus far (Mueller-Cajar et al., 2011; Stotz et al., 2011; Loganathan et al., 2016; Bhat et al., 2017b; Shivhare and Mueller-Cajar, 2017), the convex face of the AAA+ motor CbbQ is not involved in Rubisco activation (Tsai and Ye et al., Unpublished). Instead, our model shows the concave face featuring CbbO associating with Rubisco. This is consistent with a remodelling strategy that involves mechanical force transmission from CbbQ to the adaptor protein CbbO instead of the conventional pore-loop type mechanism utilized by other activases. Recent biochemical data suggest that residue L85 on the Helix-2-insert (H2I) of CbbQ, facing the concave face, binds to CbbO and is critical for communication and functional coupling between the CbbQ motor and the adaptor (Tsai and Ye et al., Unpublished). In the AAA+ protein dynein, nucleotide changes in the AAA+ ring is linked to remodelling of the dynein linker involved in motor force generation (Schmidt et al., 2015). Given that both proteins belong to the Helix-2-insert clade of AAA+ proteins and taking into account the similarities in architecture, it is tempting to speculate that similar mechanisms may underlie Rubisco activation by the CbbQO-type activases.

From native PAGE analysis, the highly conserved surface acidic residue, D82 of the β C- β D loop on N-terminus of the Rubisco large subunit was found to be critical for the formation stable interactions with the activase. This led us to hypothesize that the carboxylate group of the inhibited Rubisco serves as the sixth coordination site of the MIDAS motif on the VWA domain of CbbO. In support of this, a direct interaction between the density occupying the central cavity of the CbbQ3 cup, believed to be the VWA domain and Rubisco can be seen in our map. Further, both the yeast ribosome maturation factors Rea1 and Mdn1 (Midasin) VWA domain is docked onto the AAA+

ring and is in close contact with the substrate (Chen et al., 2018; Sosnowski et al., 2018). Manual docking of the CbbQ2 crystal structure into the EM density positions the activase over the critical C-terminus of RbcL which governs the accessibility of the active site (Figure 5-11C). This is in line with our finding that an intact RbcL C-terminus is essential for association with the activase and a deletion of even one residue would greatly reduce the affinity for interaction (Figure 5-6). Overall, our model is in close agreement with the previously generated biochemistry data which points to involvement of the CbbO C-terminal VWA domain in engaging Rubisco and RbcL C-tail manipulation as part of the mechanism to reactivate Rubisco.

In this chapter, a low-resolution model of the CbbQO-type activase interacting with its substrate Rubisco reveals the binding position of activase relative to Rubisco and hinted at the possible conformation of the activase when interacting with its substrate. The resolution of the negative stain model was insufficient to accurately identify the interacting motifs and more structural details are expected to be revealed by moving forward with cryo-EM.

6. CRYO-ELECTRON MICROSCOPY STUDY OF THE FORM I RUBISCO-ACTIVASE COMPLEX

In order to develop a detailed picture of the interaction of the CbbQO-type activase with Rubisco, single particle cryo-EM was pursued. The cryo-EM structure of the functional chimeric Rca, Q3O1 from *A. ferrooxidans* in complex with the cognate inhibited form I Rubisco, AfLS was resolved to a global resolution of 9.8 Å. The structure of the form I Rubisco inhibited by the transition state analogue, 2-carboxyarabinitol 1,5-bisphosphate (CABP) was solved at a resolution of ~2.9 Å. Together, the structural studies provided insights into how CbbQO bind to Rubisco and lead to a better mechanistic understanding of Rubisco remodeling by the CbbQO-type Rca.

6.1 Isolation of AfLS ECMC-Q3O1 complex for cryo-EM and screening

For cryo-EM studies, 16 µM of AfLS (monomer) was mixed with 64 µM of Q3O1 (monomer) (~1:5 molar ratio Rubisco to activase) in the presence of 5 mM MgCl₂ and 5 mM ATP before addition of the chemical cross-linker glutaraldehyde. To enrich for the Rubisco-activase complexes, the cross-linking reaction was loaded onto the analytical gel filtration and analysed eluted fractions with native PAGE (Figure 6-1A). The presence of high molecular weight bands in the earlier fractions was similar to what was observed previously and suggested that cross-linking was successful (Figure 5-10).

After capturing the AfLS ECMC-Q3O1 complexes with analytical gel filtration, the sample was diluted to a suitable concentration and both negative stain EM and Arctica cryo-EM was used for assessment of the suitability of the sample for data collection. As the size of the form I Rubisco (~530 kDa) was large enough to enable visual identification on micrographs, negative stain EM provided an efficient method for assessment of the quality of the sample. Inspection of the negative stain micrographs showed a heterogenous mixture of Rubisco-activase complexes that were similar to what was obtained previously and suggest that AfLS ECMC-Q3O1 complexes were successfully captured (Figure 6-1B). Screening of vitrified grids on the FEI Arctica cryo-EM operated at 200 kV provided assurances that the particle distribution was optimum for data collection and the protein complexes were embedded in a sufficiently thin layer of vitreous ice that would enable high contrast between the sample and the

surrounding ice layer (Figure 6-1B).

Figure 6-1: Isolation of Rubisco-activase complexes and screening of sample for cryo-EM data collection.

(A) Size exclusion profile of the AfLS ECMC-Q3O1 complex. 50 μ l fractions were collected from 0.90 ml elution volume (left). Elution peaks of globular molecular weight standards are indicated. Corresponding size exclusion fraction fractions separated by 6% native gel (right). Boxed fraction was selected for screening with negative stain EM (0.05 mg/ml) and Arctica cryo-EM (1.4 mg/ml). (B) Representative micrographs of the selected fraction imaged with negative stain EM (left) and Arctica cryo-EM (right). Scale bar; 50 nm.

6.2 Data collection on the FEI Titan Krios microscope (Krios)

After validating the protein concentration and particle distribution on the grid, the in-house 300-kV FEI Titan Krios microscope (Krios) was used to collect images of the AfLS ECMC-Q3O1 complexes. Representative cryo-EM micrographs of the complexes collected at two different defocus values is shown in figure 6-2. Particles appeared to be equally distributed in an even layer of vitreous ice and image contrasts were determined to be sufficient for particle picking and alignment. Although minor ice crystal contaminations were observed, no obvious protein aggregation was found. The quality of the micrographs reflected good sample preparation, which was reproducible throughout the study. The power spectrum of the micrograph also displayed clearly identifiable circular Thon rings that stretched to high spatial frequencies, indicating that high resolution information was retained and that astigmatism and drift were absent (Figure 6-2).

Figure 6-2: Micrograph evaluation.

Representative micrographs collected at the defocus values of 0.6 μM (top) and 2.5 μM (bottom) with their respective power spectrum shown on the right.

6.3 Particle picking and 2D classification

Micrographs were collected over a wide range of defocus values. After beam-induced motion correction and CTF estimation, images with obvious astigmatism and drift were discarded, leaving 4943 good micrographs for image processing using the software RELION-3 (Zivanov et al., 2018) (Scheres, 2012a). Manual picking of ~ 2,000 particles produced initial 2D class averages featuring AfLS ECMC-Q3O1 complexes as well as unbound Rubisco. Classes featuring cross-linked complexes were selected and used as templates for the RELION auto-picking which generated 1,103,387 particles from the entire dataset.

Inspection of the 2D class averages shown in figure 6-3A revealed that free Rubisco is present in many classes, suggesting that a large percentage of Rubisco remained unbound within the sample. Classes with different number of activases bound to the edge of the cuboidal Rubisco can also be observed, similar to what was obtained for previous negative stain EM experiments (Figure 5-11B). However, majority of these classes featuring complexes only had low resolution information, presumably due to particle heterogeneity and poor alignment.

In classes where both Rubisco and activase were represented by strong and clear densities, good contrast enabled the four-fold symmetry of Rubisco (top) as well as the RbcL dimers (side) to be clearly identified from the multi-angle views of Rubisco (Figure 6-3A). The visibility of secondary structures suggested that high resolution information was retained and the dataset contained potential for reliable 3D reconstructions. However, structural details of Q3O1 were less defined. This was also encountered in the previous negative stain EM experiment and indicated that the class averages represented a distribution of positions of the flexibly attached activase relative to Rubisco. Although Q3O1 was not well-resolved, an intense density within the CbbQ cup, believed to be the VWA domain of CbbO1, could be seen to be making direct contact with Rubisco (Figure 6-3A). This corroborated our negative stain model where the VWA domain of CbbO featured on the concave face of the CbbQ AAA+ ring interacts with Rubisco (Figure 5-11). Due to the low quality of classes that featured Rubisco-activase complexes, only 4 classes (77240 particles) that included well-defined complexes with one activase bound to Rubisco and both proteins having equal intensity

were selected for further 3D refinement (Figure 6-3B).

Figure 6-3: 2D classification of AfLS ECMC-Q3O1 complexes derived from RELION-3.

(A) Initial classification of 1,103,387 particles from RELION auto-picking into 150 classes. (B) Selected classes with one activase bound to Rubisco (77240 particles) for downstream refinement and 3D classification. An intense density within the CbbQ cup, believed to be the VWA domain of CbbO1, can be seen to be making direct contact with Rubisco.

6.4 3D refinement and classification

After discarding the ‘bad’ particles with 2D classification, 3D refinement was performed. The negative stain model of the AfLS ECMC-Q3O1 complex (Figure 5-11C) was filtered to 40 Å and used as a reference. A soft mask of 320 Å was applied to both the particles and reference to reduce background noise and increase sensitivity for alignment. The refinement converged by the end of the 20th iteration.

Three dimensional classification was performed to further resolve structural heterogeneity within the dataset and the remaining ‘good’ particles were sorted into 5 classes. Classes 1, 2 and 5 supported identifiable features of Rubisco bound to one Q3O1

activase. Class 1 represented the major class at 47% particle distribution and was resolved to the $\sim 11\text{\AA}$, the highest among all classes. Class 3 resembled 2 activases and harbored only 9% of particle distribution. The 4th class was indistinguishable and contained only less than 1% of all particles. As such, the first class consisting of $\sim 36,000$ particles was selected for further refinements and post-processing to a global resolution of $9.76\text{\AA}$ estimated from the 0.143 FSC criterion (Figure 6-4B,C).

The final model resembled our negative stain EM where the activase is docked near the equatorial region of Rubisco, at the interface of two RbcL dimers (Figure 6-4B). Q3O1 appeared to be slightly tilted with the central density, believed to be the CbbO1VWA domain making direct contact with the RbcL. The dimensions of Q3O1 at $110\text{\AA} \times 90\text{\AA}$ is similar to that of the unbound Q2O2 negative stain 3D reconstruction, suggesting that there are no large conformational changes when CbbQO binds to the substrate. Compared to Rubisco, for which the RbcL dimers could be distinguished, the density corresponding to Q3O1 was poorly defined, indicating structural flexibility and variations in positions relative to Rubisco. Taken together, the 3D reconstruction of the AfLS ECMC-Q3O1 complex supports our interpretation that the adaptor CbbO located on the highly conserved concave face of the CbbQ hexamer makes contact with Rubisco via a side-on approach.

Figure 6-4: 3D classification of AfLS ECMC-Q3O1 complexes.

(A) Results from 3D classification. Particle distribution for each class is reported. (B) The post-processed map of the AfLS ECMC-Q3O1 complex at 9.8 Å. Q3O1 is docked at the equatorial region of Rubisco, at the RbcL dimer interface. (C) Mask-corrected Fourier shell correlation (FSC) curves of the AfLS ECMC-Q3O1 complexes as determined by gold standard FSC procedure (0.143 criterion) in RELION-3. All maps were masked and sharpened using automatically determined negative B-factors.

6.5 Focused refinement of Q3O1

The fall-off in map quality for the activase observed in the reconstruction of the Rubisco-activase complex is likely due to the structural heterogeneity caused by varying attachment points of Q3O1 on Rubisco. This was also observed in the cryo-EM structure of the red-type Rca system (Bhat et al., 2017b). To mitigate this, focused refinement of Q3O1 was initiated with the signal subtraction function in RELION-3. This approach refines Q3O1 separately against particle images where signal from Rubisco has been subtracted (Bai et al., 2015a). Using the particles selected from well-defined 2D class averages of one activase bound to Rubisco (77,240 particles), Q3O1 was refined independently with partial signal subtraction. At this stage, classification into five classes yielded 2 majority classes that showed good density for Q3O1 (Figure 6-5A). These two classes were combined and selected for further refinements to give rise to a final map at 9.1 Å estimated from the 0.143 FSC criterion (Figure 6-5B, C).

Docking of the CbbQ2 crystal structure into the model of Q3O1 resulted in unoccupied density in the periphery and also showed CbbQ2 protruding from the map in several

places (Figure 6-5B). The limited resolution of Rca is attributed to the heterogeneity that remains within the dataset, owing to movements of the subunits of the AAA+ ring relative to each other as well as the flexibility of the CbbO-Rubisco junction. The convex face of Q3O1 was poorly resolved and density for the Helix-2-insert (H2I) was missing (Figure 6-5B). It should be noted that the CbbQ crystal structure from *H. neapolitanus* is also missing the H2I, consistent with the dynamic nature of these mobile elements (Sutter et al., 2015). While CbbO linker was not readily distinguishable from the density, the map revealed a similar organization for Q3O1 compared to previous negative stain models, where the central cavity of the AAA+ ring is filled and the pore is occluded from the concave face of the hexamer (Figure 3-2C).

The 3D reconstruction of the AfLS ECMC-Q3O1 complex suggests that the density on the central cavity of Rubisco interacts with Rubisco and appears to be the only attachment point (Figure 6-4B). Given that previous biochemical studies have pointed to the importance of the CbbO VWA domain in interacting with Rubisco, the central density on CbbQ hexamer was assigned to be the CbbO1VWA domain and parts of the unresolved CbbO1 linker. Together, these data indicate that during Rubisco activation, the conformation of CbbQO involves docking of the CbbO VWA domain onto the AAA+ ring, comparable to that of the yeast ribosome maturation factors Mdn1 and Real (Chen et al., 2018; Sosnowski et al., 2018).

Figure 6-5: Independent refinement and 3D classification and of the chimeric activase Q3O1.

(A) Results from 3D classification. Only two majority classes are shown. Particle distribution for each class is reported. (B) The post-processed map of the Q3O1 complex at 9.1 Å. The density of Q3O1 fit with the crystal structures of CbbQ2 (pink ribbon; Tsai and Ye et al., Unpublished) and the integrin alpha-LVWA domain (blue ribbon; PDB-ID 1T0P) (Song et al., 2005). Density for the Helix-2-Insert is missing. (C) Mask-corrected Fourier shell correlation (FSC) curves of the Q3O1 model as determined by gold standard FSC procedure (0.143 criterion) in RELION-3. All maps were masked and sharpened using automatically determined negative B-factors.

6.6 Cryo-EM structure of AfLS

Prior to the collection of the large dataset of AfLS ECMC-Q3O1 complexes, a smaller dataset of 2,729 micrographs was collected. The data was processed using a similar pipeline described above. In our initial dataset, we failed to detect significant numbers of AfLS-Q3O1 complexes despite having used the cross-linker glutaraldehyde. Two dimensional class averages from 647,661 particles show mainly unbound rubisco, with multi-directional views represented and secondary structures clearly distinguishable in some classes (Figure 6-6A). In order to obtain insights into the structure of the CABP-bound *A. ferrooxidans* form I Rubisco, 390,625 particles in well-defined classes containing free Rubisco were selected for further processing (Figure 6-6B).

Figure 6-6: 2D classification of AfLS ECMC derived from RELION-3.

(A) Initial classification of 647,661 particles from RELION-3 auto-picking into 150 classes. (B) Selected classes of well-defined Rubisco (390,625 particles) for downstream refinement and 3D classification.

Over several rounds of 3D classification and refinements the final EM map of AfLS was refined to 2.9 Å with 87,039 particles by applying D4 symmetry (Figure 6-7A, C). The local resolution of the map was calculated with Resmap (Kucukelbir et al., 2013) and showed that the resolution at the core of Rubisco was generally higher than the peripheral regions (Figure 6-7A). The N- and C-termini of the RbcL as well as the N-terminus of RbcS exhibited higher B-factors, indicating structural flexibility (Figure 6-7B). Consistent with the estimated high resolution, side chains could be clearly distinguished for most of the map, with exceptions of the termini regions of each chain. Therefore, modeling for the protein could be carried out with high confidence.

A

Slices Through ResMap Results

B

Figure 6-7: Cryo-EM and 3D reconstruction of AfLS at 2.9 Å.

(A) Local resolution estimation with Resmap for AfLS. Both solvent (top) and sliced (bottom) views are shown. (B) B factors of the refined AfLS D4-symmetric model. B factors from 60 to 120 Å² is shown as a gradient from blue to yellow. (C) Mask-corrected Fourier shell correlation (FSC) curves of the AfLS model as determined by gold standard FSC procedure (0.143 criterion) in RELION-3. All maps were masked and sharpened using automatically determined negative B-factors.

6.7 Modelling of AfLS

The sequences of AfLS were manually built into the B-factor sharpened electron density, using the crystal structure of the highly homologous *H. neapolitanus* form I Rubisco (HnLS) as a starting model. The structure encompasses residues 4-460 of the 473-residue RbcL and 7-115 of the 118-residue RbcS (Figure 6-8A).

The atomic model of AfLS was validated using the Molprobity server. The model validation and refinement statistics are listed in Table 6.1.

Validation (proteins)	
Molprobability score	1.55 (94 th percentile)
Clashscore, all atoms	4.33 (96 th percentile)
Poor rotamers (%)	0.00
Favoured rotamers (%)	95.83
Ramachandran plot	
Favoured (%)	95.11
Outliers (%)	0.00
Bad bonds (%)	0.00
Bad angles (%)	0.03

Table 6-1: Atomic model validation statistics for one subunit of AflS.

Amino acid side chains were mostly well-resolved, with most residues enclosed by the map and most of the map filled with atoms. Exceptions include the N- and C-termini of RbcL and residues 7-17 on RbcS. Inspection of the Rubisco active site revealed clear densities for the inhibitor CABP and the Mg^{2+} ion (Figure 6-8B). Lys194, predicted to be carbamylated was remodeled to KCX.

Figure 6-8: Analysis of the AfLS model.

(A) 3D reconstruction of AfLS ECMC at 2.9 Å resolution with D4-symmetry applied. Both top and side (left) views are shown. Model of the RbcL dimer docked into the map (right) with the ‘Fit in map’ function in Chimera. The EM reconstruction is shown as a transparent surface. RbcL and RbcS subunits are shown in blue and yellow, respectively. (B) Cryo-EM density of the active site pockets of AfLS, showing density for CABP and the carbamylated Lys194 as displayed in the program Coot (Emsley et al., 2010) (left). A close up of the active site of the AfLS model is shown (right). Residues involved in metal ion coordination are shown in sticks that are coloured by element. (D) Cryo-EM density at the C-terminus (left). Phe460 is the last residue that can fit into the map density. Densities for residues 461 to 473 are missing. Structural details for one L-subunit of an RbcL dimer highlighting the C-terminus and loop 6 in red and yellow, respectively (right). Residues involved in closure of the active site are shown as sticks coloured by element.

The large subunit of HnLS possesses ~90% identity with AfLS while the small subunit is 55% identical. However, it must be noted that the structure of HnLS enzyme represents an uncarbamylated, inactive state during which the C-tail of RbcL is observed to be disordered. In contrast, the C-terminus of other inhibitor or substrate-bound

Rubisco structures have been shown to be anchored to the body of the enzyme, maintaining its rigidity and covering the active site. Due to these differences in conformation, certain regions of our starting model did not fit into the density of the map. Therefore, RbcL residues 164-176 and the C-terminal residues (431-460) was built *de novo* using a poly-alanine model followed by subsequent amino acid assignment based on the sequence and guided by defining densities of bulky side chains.

As the map density is only available until Phe460 for RbcL, 13 residues of the C-terminus are excluded from the model including the highly conserved Asp465, responsible for anchoring the C-terminus to the body of the enzyme when substrate is bound (Figure 6-8C). A close-up of the C-terminal region shows that similar to other closed Rubisco structures loop 6 is protracted over the active site during catalysis, excluding it from the solvent (Andersson and Backlund, 2008). Lys327 and Leu328 located on the tip of loop 6, reach into the active site and maintain close proximity to the ligand. This suggests that the missing C-tail residues must be folded over loop 6 to maintain it in a closed conformation.

An alignment of the AfLS with the homologue HnLS show that the structure is generally similar with an overall root-mean-square deviation (RMSD) of 0.617, suggesting that glutaraldehyde cross-linking did not significantly modify the structure (Figure 6-9). Major differences in the structure of the two Rubisco lie in the RbcL C-terminus and loop 6. Upon binding of the substrate, the flexible loop 6 slides over the active site and is pinned down by the C-terminus (Duff et al., 2000). While the C-tail of the inhibited AfLS is closed and lie proximal to the catalytic centre, the C-tail of HnLS is disordered and points away from the body of the enzyme. Correspondingly, loop 6 of HnLS is flexible and is positioned further away from the catalytic site, consistent with other Rubisco enzymes in the apo state.

Figure 6-9: Structural alignment of AfLS to the closely related homologue HnLS.

Model of one monomer of AfLS was aligned to that of HnLS. Ribbon representation of AfLS and HnLS is shown in shades of blue and pink, respectively. Major differences lie in the conformation of the RbcL C-terminus and Loop 6. The boxed area includes a close-up highlighting the different conformations of the two structures. The C-terminus of AfLS RbcL extends towards the active site while that of HnLS is pointing away. Loop 6 of AfLS is closed over the catalytic site while that of HnLS is positioned further away.

6.8 Discussion

In this chapter, the cryo-EM structure of the *A. ferrooxidans* form I Rubisco in complex with the chimeric activase Q3O1 at 9.8 Å (Figure 6-4B) was shown. Although resolution of the model did not permit detailed analysis of interaction sites on both proteins, the positioning of the activase relative to Rubisco was determined to be at the RbcL dimer interface, proximal to the active site. Q3O1 appears to be slightly more tilted as compared to the previous negative stain model, but remains near the equatorial region of Rubisco. This is in contrast to the red-type RsRca which docks near the top surface, at a corner of Rubisco cube (Bhat et al., 2017b), and highlights that the activase likely goes through a variety of orientations relative to Rubisco from initial engagement to activation. Access to the active site is governed by the C-terminus of Rubisco which lies

near the equator of the holoenzyme. Therefore it is possible that during initial recognition of Rubisco, Q3O1 is positioned close to the equator and makes contact with the critical C-tail. Q3O1 relocates after activation as seen for the red-type Rca system. This would be consistent with a continuous spectrum of structural heterogeneity exhibited by the Rca and is responsible for the low resolution of the model of the Rubisco-activase complex.

Independent refinements of the activase produced a 9.1 Å model of the chimeric Rca, Q3O1 (Figure 6-5). Previous biochemical data implicated the Helix-2-insert (H2I) in force transmission from the AAA+ motor to the CbbO adaptor and is believed to be a mobile element of the hexamer (Tsai and Ye et al., Unpublished). Correspondingly, the H2I densities are missing from the model and the AAA+ protein presents like a ring instead of cup-shaped as seen in the previous negative stain models. The central pore of the hexamer is closed by the density of the VWA domain, and suggests that a pore-loop threading mechanism is not feasible. Since direct contact between the central density and Rubisco was observed, the force generated from the AAA+ ATPase for inhibitor release must be transmitted via the adaptor, CbbO. At the current resolution it was not possible to resolve the CbbO linker and distinguish each subunit of the hexamer. As previously observed, the structure of the unbound CbbQO appears similar to the conformation the activase when it is docked onto Rubisco. Hence, moving forward with cryo-EM to determine the structure of the Rca alone is expected to shed light on the intricate details of the CbbO-Rubisco junction.

Processing of an initial smaller dataset collected on the in-house Krios microscope produced a high resolution model of the inhibited Rubisco, AfLS at 2.9 Å (Figure 6-7). Although high-resolution structures of Rubisco exist for different species, proteobacterial Rubisco which are represented in all clades remains sparsely sampled (Liu et al., 2017). Recent work performed in the lab on directed evolution using a Rubisco-dependent *E. coli* system produced several AfLS variants with increased rates of inhibitor release as compared to the WT (Guo et al., Unpublished). The structure of AfLS will be instrumental in structural analysis of the mutations that confer the different phenotypes observed. At the time of submission of this thesis the AfLS model has not been submitted to PDB and will be subjected to further refinements and validation.

The structure of the inhibited AfLS is generally similar to that of the closely related homologue from *H. neapolitanus* in the apo state, with exceptions of the C-terminus and loop 6 (Figure 6-9). AfLS assumes a closed conformation with loop 6 is stretch out over the active site and the C-tail is observed to be extending towards the active site to pin down loop 6. In contrast, these elements are disordered and flexible in HnLS and points away from the catalytic centre. Such differences in the open versus closed state of the holoenzyme have been noted in other Rubisco structures (Andersson and Backlund, 2008) and points to the validity of our model.

In conclusion, the structures presented in this chapter led us to conclude that the CbbQO-type Rca docks onto the equatorial region of Rubisco via a side-on approach and initiates engagement using the CbbO VWA domain (Figure 6-10). Unlike the other activases studied so far, the convex surface of the CbbQ hexamer is not involved in Rubisco remodeling. Instead, mechanical force for disruption of the active site is likely transmitted via the intrinsically flexible CbbO linker to Rubisco. Interaction of CbbO with the C-terminus of RbcL, which governs the accessibility of the active site would trigger a sequence of conformational changes that would ultimately result in liberation of the inhibitor. It is possible that the release of the inhibitory sugar phosphate might be coupled to repositioning of CbbQO activase to the top surface of Rubisco which will represent the end-state. Clearly, the proposed model will be greatly strengthened with the acquirement of a high resolution model for the activase and detailed mutational analysis. An integrative approach of using both structural and biochemical methods will contribute to a more detailed understanding of Rubisco activation in chemoautotrophs.

Figure 6-10: Proposed mechanistic model for the activation of AfLS by CbbQO.

Sequences of AfLS is superimposed on the EM reconstruction of AfLS map shown as a transparent surface. CbbQO is modelled with the crystal structure of CbbQ2 (pink ribbon; Tsai and Ye et al., Unpublished) and the VWA domain of the integrin alpha-L VWA domain (blue ribbon, PDB-ID: 1T0P)

(Song et al., 2005) superimposed on the cryo-EM reconstruction of Q3O1. CbbQO interacts with inhibited Rubisco via the VWA domain of its adaptor protein CbbO. Mechanical force for active site disruption is transmitted from the cbbQ hexamer to Rubisco through CbbO. Interaction of CbbO with the C-terminus of RbcL, which governs the accessibility of the active site is required to trigger conformational changes that would release the bound sugar phosphate inhibitor.

7. CONCLUSION AND OUTLOOK

In this work, a preliminary model for the mechanism of action for the CbbQO activase was presented and provides the framework for further characterization of this new Rca system. The structure of CbbQO alone and cross-linked to Rubisco suggest utilization of the adaptor protein CbbO in engaging the substrate. The force generated by the AAA+ motor CbbQ is transmitted vectorially to the substrate through the VWA-domain containing CbbO featured on the concave face of the hexamer. This reflects a common theme among the poorly studied MoxR class of AAA+ ATPases in enzyme modulation in conjunction with VWA-containing proteins. The coupling of ATP hydrolysis to molecular remodeling events is expected to be a commonality for this class of proteins. Further biochemical and structural characterization of the CbbQO-type Rubisco activase is likely to reveal more fascinating facets of this novel class of molecular chaperone.

8. APPENDIX

8.1 Cross-linking coupled to mass spectrometry (XL-MS) results

Id	Protein1	Protein2	XLType	AbsPos1	AbsPos2	Mr	Mz	z	Id-Score^a
LKGHGLDEGISTR-KLDLYLR-a2-b1	AfQ2	AfO2	inter-protein xl	215	223	2439.347	488.877	5	39.76
WLLDKDGTR-KLDLYLR-a5-b1	AfQ2	AfO2	inter-protein xl	87	223	2160.191	541.056	4	39.57
KRLPVMVK-KLDLYLR-a1-b1	AfQ2	AfO2	inter-protein xl	36	223	2027.233	406.454	5	37.93
DLKQSTK-KLDLYLR-a3-b1	AfQ2	AfO2	inter-protein xl	166	223	1876.063	470.024	4	37.19
SWIDFKSGSTPDPR-WLLDKDGTR-a6-b5	AfO2	AfQ2	inter-protein xl	546	87	2832.407	567.489	5	33.97
LKGHGLDEGISTR-KADKK-a2-b1	AfQ2	AfO2	inter-protein xl	215	673	2108.158	528.047	4	33.82

RLKGHGLDEGISTR-QKDFSLQSADSR-a3-b2	AfQ2	AfO2	inter-protein xl	215	194	3203.632	641.734	5	31.68
LPEKLPELFIALTR-LKGHGLDEGISTR-a4-b2	AfO2	AfQ2	inter-protein xl	749	215	3158.768	632.761	5	31.63
LPEKLPELFIALTR-RLKGHGLDEGISTR-a4-b3	AfO2	AfQ2	inter-protein xl	749	215	3314.884	663.985	5	31.46
LKGHGLDEGISTR-LLDLLKPQDKVR-a2-b10	AfQ2	AfO2	inter-protein xl	215	521	2956.67	740.175	4	29.31
LPEKLPELFIALTR-RTLPLDKK-a4-b7	AfO2	AfQ2	inter-protein xl	749	137	2746.653	550.338	5	25.85
LKGHGLDEGISTR-HAALAKHLK-a2-b6	AfQ2	AfO2	inter-protein xl	215	507	2507.396	502.487	5	25.26
QSDQFAKVHFADTEISYRDDNR-LKGHGLDEGISTR-a7-b2	AfO2	AfQ2	inter-protein xl	415	215	4161.013	694.51	6	21.37
RLKGHGLDEGISTR-HAALAKHLKR-a3-b9	AfQ2	AfO2	inter-protein xl	215	510	2819.593	564.926	5	20.45
LLDLLKPQDKVR-DLKQSTK-a6-b3	AfO2	AfQ2	inter-protein xl	517	166	2393.394	599.356	4	20.34

LKGHGLDEGISTR-KLDLYLR-a2-b1	AfQ2	AfO2	inter-protein xl	215	223	2439.339	814.121	3	42.16
TLFLALHPKPGEDACDPETTSCL R-LKGHGLDEGISTR-a9-b2	AfO2	AfQ2	inter-protein xl	337	215	4247.083	1062.779	4	41.81
VGEVFSDCIAKASAR- LKGHGLDEGISTR-a11-b2	AfO2	AfQ2	inter-protein xl	38	215	3128.579	626.724	5	41.06
LGKPLVTVACNEDMTAADLVG R-VGEVFSDCIAKASAR-a3-b11	AfQ2	AfO2	inter-protein xl	63	38	4076.028	1020.015	4	40.96
LKGHGLDEGISTR-ALCKLGR-a2-b4	AfQ2	AfO2	inter-protein xl	215	60	2336.255	779.759	3	40.66
LKGHGLDEGISTR- HAALAKHLK-a2-b6	AfQ2	AfO2	inter-protein xl	215	507	2507.386	627.854	4	40.59
RLKGHGLDEGISTR-ALCKLGR- a3-b4	AfQ2	AfO2	inter-protein xl	215	60	2492.354	624.096	4	40.43
LPEKLPELFIALTR-KGELIR-a4-b1	AfO2	AfQ2	inter-protein xl	749	138	2491.483	623.879	4	39.37
LKGHGLDEGISTR- LLDLLKPQDK-a2-b6	AfQ2	AfO2	inter-protein xl	215	517	2701.492	541.306	5	38.43

LPEKLPELFIALTR- LKGHGLDEGISTR-a4-b2	AfO2	AfQ2	inter-protein xl	749	215	3158.761	632.760	5	38.06
DLKQSTK-KLDLYLR-a3-b1	AfQ2	AfO2	inter-protein xl	166	223	1876.058	470.022	4	37.99
LPEKLPELFIALTR-DLKQSTK-a4- b3	AfO2	AfQ2	inter-protein xl	749	166	2595.478	649.877	4	37.94
LGKPLVTVACNEDMTAADLVG R-KQLDCR-a3-b1	AfQ2	AfO2	inter-protein xl	63	19	3285.637	822.417	4	37.89
LKGHGLDEGISTR- LLDLLKPQDKVR-a2-b10	AfQ2	AfO2	inter-protein xl	215	521	2956.664	740.174	4	37.34
VGEVFSDCIAKASAR- DLKQSTK-a11-b3	AfO2	AfQ2	inter-protein xl	38	166	2565.304	642.334	4	37.18
LKGHGLDEGISTR- QKDFFSLSADSAR-a2-b2	AfQ2	AfO2	inter-protein xl	215	194	3047.523	610.512	5	36.94
LKGHGLDEGISTR-KQLDCR-a2- b1	AfQ2	AfO2	inter-protein xl	215	19	2338.198	585.557	4	36.87
LPEKLPELFIALTR- LPVMVKGPTGCGK-a4-b6	AfO2	AfQ2	inter-protein xl	749	43	3119.737	780.942	4	36.74

VGEVFSDCIAKASAR- WLLDKDGTR-a11-b5	AfO2	AfQ2	inter-protein xl	38	87	2849.425	713.364	4	36.25
SWIDFKSGSTPDPR- WLLDKDGTR-a6-b5	AfO2	AfQ2	inter-protein xl	546	87	2832.398	709.107	4	36.20
LPVMVKGPTGCGK- QKDFFSLSQASDR-a6-b2	AfQ2	AfO2	inter-protein xl	43	194	3008.501	753.133	4	35.87
QKDFFSLSQASDR-GPTGCGKSR- a2-b7	AfO2	AfQ2	inter-protein xl	194	50	2584.226	647.064	4	35.24
KAPDLETQGLPPR- LKGHGLDEGISTR-a1-b2	AfO2	AfQ2	inter-protein xl	451	215	2940.557	981.193	3	34.73
KLMLVLTDGRPSDQDVQDDR- LKGHGLDEGISTR-a1-b2	AfO2	AfQ2	inter-protein xl	677	215	3790.938	948.742	4	34.13
ALCKLGR-KGELIR-a4-b1	AfO2	AfQ2	inter-protein xl	60	138	1668.963	418.249	4	34.07
LGKPLVTVACNEDMTAADLVG R-ALCKLGR-a3-b4	AfQ2	AfO2	inter-protein xl	63	60	3283.718	821.937	4	33.92
RLKGHGLDEGISTR- QKDFFSLSQASDR-a3-b2	AfQ2	AfO2	inter-protein xl	215	194	3203.620	641.732	5	33.67

WLLDKDGTR-KLDLYLR-a5-b1	AfQ2	AfO2	inter-protein xl	87	223	2160.186	541.054	4	33.18
LLDLLKPQDKVR-DLKQSTK-a6-b3	AfO2	AfQ2	inter-protein xl	517	166	2393.379	479.684	5	33.01
VGEVFSDCIAKASAR-GPTGCGKSR-a11-b7	AfO2	AfQ2	inter-protein xl	38	50	2665.284	889.436	3	32.95
ALCKLGR-TLPLDKK-a4-b6	AfO2	AfQ2	inter-protein xl	60	137	1768.020	443.013	4	32.94
GPTGCGKSR-KLDLYLR-a7-b1	AfQ2	AfO2	inter-protein xl	50	223	1976.045	659.690	3	32.88
LPVMVKGPTGCGK-KLDLYLR-a6-b1	AfQ2	AfO2	inter-protein xl	43	223	2400.320	601.088	4	32.82
ALCKLGR-DLKQSTK-a4-b3	AfO2	AfQ2	inter-protein xl	60	166	1772.974	444.251	4	32.77
KLDLYLR-QSTKQR-a1-b4	AfO2	AfQ2	inter-protein xl	223	170	1804.015	452.012	4	32.58
RLKGHGLDEGISTR-KLDLYLR-a3-b1	AfQ2	AfO2	inter-protein xl	215	223	2595.438	520.095	5	32.40

QKDFSLQSADSR- WLLDKDGTR-a2-b5	AfO2	AfQ2	inter-protein xl	194	87	2768.367	693.099	4	32.16
GPTGCGKSR-KQLDCR-a7-b1	AfQ2	AfO2	inter-protein xl	50	19	1874.902	625.975	3	32.07
LPVMVKGPTGCGK-ALCKLGR- a6-b4	AfQ2	AfO2	inter-protein xl	43	60	2297.236	575.317	4	31.63
LGKPLVTVACNEDMTAADLVG R-KLDLYLR-a3-b1	AfQ2	AfO2	inter-protein xl	63	223	3386.781	847.703	4	31.44
QSDQFAKVHFADTEISYR- LKGHGLDEGISTR-a7-b2	AfO2	AfQ2	inter-protein xl	415	215	3660.806	733.169	5	31.13
VGEVFSDCIAKASAR- LPVMVKGPTGCGK-a11-b6	AfO2	AfQ2	inter-protein xl	38	43	3089.568	773.400	4	31.06
LGKPLVTVACNEDMTAADLVG R-QKDFSLQSADSR-a3-b2	AfQ2	AfO2	inter-protein xl	63	194	3994.961	999.748	4	30.99
GPTGCGKSR-ALCKLGR-a7-b4	AfQ2	AfO2	inter-protein xl	50	60	1872.959	625.328	3	30.59
LLDLLKPQDKVR-DLKQSTK- a10-b3	AfO2	AfQ2	inter-protein xl	521	166	2393.379	599.353	4	30.48

VGGEVFSDCIAKASAR-KGELIR-a11-b1	AfO2	AfQ2	inter-protein xl	38	138	2461.294	616.331	4	30.10
AHPDFQLVISYNPGYQSLMKDL K-LLDLLKPQDKVR-a20-b10	AfQ2	AfO2	inter-protein xl	163	521	4238.271	848.662	5	29.74
VGGEVFSDCIAKASAR-QSTKQR-a11-b4	AfO2	AfQ2	inter-protein xl	38	170	2493.256	624.322	4	29.09
LKGHGLDEGISTR- LLDLLKPQDKVR-a2-b6	AfQ2	AfO2	inter-protein xl	215	517	2956.660	493.784	6	28.88
QKDFFSLSQASDR-KGELIR-a2-b1	AfO2	AfQ2	inter-protein xl	194	138	2380.233	596.066	4	28.18
QKDFFSLSQASDR-QSTKQR-a2-b4	AfO2	AfQ2	inter-protein xl	194	170	2412.197	604.057	4	27.69
QSDQFAKVHFADTEISYR- KGELIR-a7-b1	AfO2	AfQ2	inter-protein xl	415	138	2993.512	599.710	5	27.19
WLLDKDGTR-KQLDCR-a5-b1	AfQ2	AfO2	inter-protein xl	87	19	2059.045	515.769	4	26.15
SWIDFKSGSTPDPR- LKGHGLDEGISTR-a6-b2	AfO2	AfQ2	inter-protein xl	546	215	3111.552	623.318	5	25.53

LPEKLPELFIALTR- WLLDKDGTR-a4-b5	AfO2	AfQ2	inter-protein xl	749	87	2879.608	720.910	4	25.22
------------------------------------	------	------	---------------------	-----	----	----------	---------	---	-------

Table 8-1: AfQ2 and AfO2 inter-links

^aId score indicates the quality of cross-link identifications. It incorporates the quality of the measured MS spectrum, the quality of the reference spectrum and the likelihood for a certain match to be random. The higher the Id score, the higher is the likelihood that a certain MS spectrum is correctly matched to a certain cross-link, meaning that this interaction is most likely true. For this work, all cross-links with Id scores above 20 are listed.

Id	Protein1	Protein2	XLType	AbsPos1	AbsPos2	Mr	Mz	z	Id-Score
RLKGHGLDEGISTR-KGELIR-a3-b1	AfQ2	AfQ2	intra-protein xl	215	138	2390.338	479.075	5	43.74
QKDFSLQSAISR-KLDLYLR-a2-b1	AfO2	AfO2	intra-protein xl	194	223	2585.339	647.343	4	43.67
LKGHGLDEGISTR-LKGHGLDEGISTR-a2-b2	AfQ2	AfQ2	intra-protein xl	215	215	2901.531	581.314	5	38.83
RLKGHGLDEGISTR-RLKGHGLDEGISTR-a3-b3	AfQ2	AfQ2	intra-protein xl	215	215	3213.731	536.63	6	38.51
KRLPVMVK-KRLPVMVK-a1-b1	AfQ2	AfQ2	intra-protein xl	36	36	2077.298	416.467	5	36.69
WLLDKDGTR-WLLDKDGTR-a5-b5	AfQ2	AfQ2	intra-protein xl	87	87	2343.22	586.813	4	36.6
LKGHGLDEGISTR-WLLDKDGTR-a2-b5	AfQ2	AfQ2	intra-protein xl	215	87	2622.375	525.483	5	36.17
WLLDKDGTRWQDGPLTVAAR-WLLDKDGTR-a5-b5	AfQ2	AfQ2	intra-protein xl	87	87	3537.836	708.575	5	35.04

WLLDKDGTR-KGELIR-a5-b1	AfQ2	AfQ2	intra-protein xl	87	138	1955.082	489.778	4	34.61
WLLDKDGTR-KRLPVMVK-a5-b1	AfQ2	AfQ2	intra-protein xl	87	36	2210.258	443.06	5	33.26
KRLPVMVK-KGELIR-a1-b1	AfQ2	AfQ2	intra-protein xl	36	138	1822.121	456.538	4	31.77
LKGHGLDEGISTR-DLKQSTK-a2-b3	AfQ2	AfQ2	intra-protein xl	215	166	2338.251	468.658	5	31.7
TLPLDKKGELIR-WLLDKDGTR-a6-b5	AfQ2	AfQ2	intra-protein xl	137	87	2622.474	656.626	4	31.49
RLKGHGLDEGISTR-LKGHGLDEGISTR-a3-b2	AfQ2	AfQ2	intra-protein xl	215	215	3057.638	612.535	5	31
LKGHGLDEGISTR-QSTKQR-a2-b4	AfQ2	AfQ2	intra-protein xl	215	170	2266.201	454.248	5	30.17
RLKGHGLDEGISTR-TLPLDKKGELIR-a3-b6	AfQ2	AfQ2	intra-protein xl	215	137	3057.724	612.553	5	29.73
LLDLLKPQDKVR-KADKK-a6-b1	AfO2	AfO2	intra-protein xl	517	673	2163.298	433.667	5	29.12

TLPLDKKGELIR-TLPLDKKGELIR-a6-b6	AfQ2	AfQ2	intra-protein xl	137	137	2901.728	726.44	4	28.82
WLLDKDGTR-TLPLDKK-a5-b6	AfQ2	AfQ2	intra-protein xl	87	137	2054.14	514.543	4	28.25
RLKGHGLDEGISTR-DLKQSTKQR-a3-b7	AfQ2	AfQ2	intra-protein xl	215	170	2778.506	556.709	5	27.97
TLPLDKKGELIR-WLLDKDGTR-a7-b5	AfQ2	AfQ2	intra-protein xl	138	87	2622.472	525.502	5	27.91
QKDFFSLSADSR-KADKK-a2-b1	AfO2	AfO2	intra-protein xl	194	673	2254.162	564.548	4	27.12
LKGHGLDEGISTR-KGELIR-a2-b1	AfQ2	AfQ2	intra-protein xl	215	138	2234.237	447.855	5	27.07
AHPDFQLVISYNPGYQSLMKDLK-DLKQSTK-a20-b3	AfQ2	AfQ2	intra-protein xl	163	166	3619.861	724.98	5	26.07
KAPDLETQGLPPR-QKDFFSLSADSR-a1-b2	AfO2	AfO2	intra-protein xl	451	194	3086.561	1029.861	3	25.7
AHPDFQLVISYNPGYQSLMKDLK-LKGHGLDEGISTR-a20-b2	AfQ2	AfQ2	intra-protein xl	163	215	4183.14	1046.793	4	25.49

TLPLDKKGELIR-TLPLDKKGELIR-a6-b7	AfQ2	AfQ2	intra-protein xl	137	138	2901.723	581.352	5	24.78
RLKGHGLDEGISTR-WLLDKDGTR-a3-b5	AfQ2	AfQ2	intra-protein xl	215	87	2778.475	464.087	6	24.7
LKGHGLDEGISTR-RTLPLDKK-a2-b7	AfQ2	AfQ2	intra-protein xl	215	137	2489.394	623.356	4	24.63
KRLPVMVK-DLKQSTK-a1-b3	AfQ2	AfQ2	intra-protein xl	36	166	1926.133	482.541	4	24.2
WLLDKDGTR-RTLPLDKK-a5-b7	AfQ2	AfQ2	intra-protein xl	87	137	2210.243	553.568	4	23.89
TLPLDKKGELIR-KGELIR-a6-b1	AfQ2	AfQ2	intra-protein xl	137	138	2234.334	559.591	4	23.78
TLPLDKKGELIR-TLPLDKK-a7-b6	AfQ2	AfQ2	intra-protein xl	138	137	2333.391	584.356	4	22.85
LLDLLKPQDKVR-KADKK-a10-b1	AfO2	AfO2	intra-protein xl	521	673	2163.299	722.107	3	22.39
LKGHGLDEGISTR-TLPLDKKGELIR-a2-b6	AfQ2	AfQ2	intra-protein xl	215	137	2901.626	484.612	6	22.33

RTLPLDKK-TLPLDKK-a7-b6	AfQ2	AfQ2	intra-protein xl	137	137	1921.159	481.298	4	22.12
LGKPLVTVACNEDMTAADLVGR-RLPVMVKGPTGCGK-a3-b7	AfQ2	AfQ2	intra-protein xl	63	43	3966.057	662.017	6	21.37
RLKGHGLDEGISTR-QSTKQR-a3-b4	AfQ2	AfQ2	intra-protein xl	215	170	2422.3	404.724	6	20.74
LLDLLKPQDKVR-KADKK-a6-b4	AfO2	AfO2	intra-protein xl	517	676	2163.298	541.832	4	20.38
TLPLDKKGELIR-TLPLDKK-a6-b6	AfQ2	AfQ2	intra-protein xl	137	137	2333.392	584.356	4	20.32

Table 8-2: AfQ2 to AfO2 intra-links

Id	Protein1	Protein2	XLType	AbsPos1	AbsPos2	Mr	Mz	z	Id-Score
nNLKGHGLDEGISTR-IAGMTGKFSTR-a3-b7	AfQ1	AfO1	inter-protein xl	211	665	2801.45	934.825	3	41.97
NLKGHGLDEGISTR-KGTLFVDIQR-a3-b1	AfQ1	AfO1	inter-protein xl	211	191	2809.509	703.385	4	36.87
FKDFDEAYSEEAMGR-NLKGHGLDEGISTR-a2-b3	AfO1	AfQ1	inter-protein xl	645	211	3427.598	1143.54	3	36.82
NLKGHGLDEGISTR-EVLEKYKPVSSR-a3-b5	AfQ1	AfO1	inter-protein xl	211	516	3067.628	1023.55	3	36.77
KLLLVITDGEPADIDVRDPQYLR-NLKGHGLDEGISTR-a1-b3	AfO1	AfQ1	inter-protein xl	689	211	4272.287	1069.079	4	35.13
KQEDGDEIDVEAAVR-KAPYYR-a1-b1	AfO1	AfQ1	inter-protein xl	541	12	2607.277	652.827	4	35.12
NLKGHGLDEGISTR-YKPVSSR-a3-b2	AfQ1	AfO1	inter-protein xl	211	518	2469.294	494.867	5	34.4
NLKGHGLDEGISTR-KGTLFVDIQR-a3-b1	AfQ1	AfO1	inter-protein xl	211	191	2965.61	594.13	5	33.32

LRKQEDGDEIDVEAAVR-NLKGHGLDEGISTR-a3-b3	AfO1	AfQ1	inter-protein xl	541	211	3575.82	596.978	6	29.85
NLKGHGLDEGISTR-EVLEKYKPVSSR-a3-b7	AfQ1	AfO1	inter-protein xl	211	518	3067.63	614.534	5	29.39
DLGVLVMDLSESTNDKVR-NLKGHGLDEGISTR-a17-b3	AfO1	AfQ1	inter-protein xl	596	211	3736.927	748.393	5	29.15
SLGKGDGLVSSYLDAAPHIAR-NLKGHGLDEGISTR-a4-b3	AfO1	AfQ1	inter-protein xl	53	211	3759.946	752.997	5	28.11
SLGKGDGLVSSYLDAAPHIAR-EVADKLVSIAER-a4-b5	AfO1	AfQ1	inter-protein xl	53	199	3592.912	719.59	5	26.67
KQEDGDEIDVEAAVR-VLPLEKK-a1-b6	AfO1	AfQ1	inter-protein xl	541	133	2636.39	879.804	3	26.29
FKDFDEAYSEEAMGR-KAPYYR-a2-b1	AfO1	AfQ1	inter-protein xl	645	12	2728.257	683.072	4	25.96
IAGMTGKFSTR-KAPYYR-a7-b1	AfO1	AfQ1	inter-protein xl	665	12	2102.096	526.532	4	25.45
IAGMTGKFSTR-VLPLEKK-a7-b6	AfO1	AfQ1	inter-protein xl	665	133	2131.206	533.809	4	23.21

SLGKG DGLVSSYLDAAPHIAR-KAPYYR-a4-b1	AfO1	AfQ1	inter-protein xl	53	12	3060.599	613.128	5	22.73
SPEALAMLQQRKGTLFVDIQR-IKKAPYYR-a13-b2	AfO1	AfQ1	inter-protein xl	191	11	3705.021	742.012	5	22.62
ARNLKGHGLDEGISTR-KGTLFVDIQR-a5-b1	AfQ1	AfO1	inter-protein xl	211	191	3192.749	639.558	5	21.79
VRDLGVLVLMDLSESTNDKVR-MLVYAGSLINKGVEAK-a19-b11	AfO1	AfQ1	inter-protein xl	596	233	4188.257	838.659	5	21.17
KQEDGDEIDVEAAVR-NLKGHGLDEGISTR-a1-b3	AfO1	AfQ1	inter-protein xl	541	211	3306.631	827.666	4	20.37

Table 8-3: AfQ1 and AfO1 inter-links

Id	Protein1	Protein2	XLType	AbsPos1	AbsPos2	Mr	Mz	z	Id-Score
FKDFDEAYSEEAMGR-IAGMTGKFSTR-a2-b7	AfO1	AfO1	intra-protein xl	645	665	3099.431	1034.151	3	46.05
NLKGHGLDEGISTR-DLKQSTK-a3-b3	AfQ1	AfQ1	intra-protein xl	211	162	2452.288	818.437	3	42.79
SLGKGDGLVSSYLDAAPHIAR-KGTLFVDIQR-a4-b1	AfO1	AfO1	intra-protein xl	53	191	3439.834	860.966	4	42.03
NLKGHGLDEGISTR-NLKGHGLDEGISTR-a3-b3	AfQ1	AfQ1	intra-protein xl	211	211	3129.62	626.932	5	40.71
MLVYAGSLINKGVEAK-KAPYYR-a11-b1	AfQ1	AfQ1	intra-protein xl	233	12	2626.416	657.612	4	36.84
LRKQEDGDEIDVEAAVR-YKPVSSR-a3-b2	AfO1	AfO1	intra-protein xl	541	518	2915.495	584.107	5	36.5
NLKGHGLDEGISTR-KAPYYR-a3-b1	AfQ1	AfQ1	intra-protein xl	211	12	2430.266	487.061	5	35.77
FKDFDEAYSEEAMGR-KGTLFVDIQR-a2-b1	AfO1	AfO1	intra-protein xl	645	191	3107.494	1036.839	3	35.74

KQEDGDEIDVEAAVR-YKPVSSR-a1-b2	AfO1	AfO1	intra-protein xl	541	518	2646.314	883.112	3	35.64
ARNLKGHGLDEGISTR-NLKGHGLDEGISTR-a5-b3	AfQ1	AfQ1	intra-protein xl	211	211	3356.749	840.195	4	35.03
ARNLKGHGLDEGISTR-DLKQSTK-a5-b3	AfQ1	AfQ1	intra-protein xl	211	162	2679.427	670.865	4	34.27
MLVYAGSLINKGVEAK-NLKGHGLDEGISTR-a11-b3	AfQ1	AfQ1	intra-protein xl	233	211	3325.777	666.163	5	33.58
IAGMTGKFSTR-KGTLFVDIQR-a7-b1	AfO1	AfO1	intra-protein xl	665	191	2481.34	621.343	4	33.2
LRKQEDGDEIDVEAAVR-IAGMTGKFSTR-a3-b7	AfO1	AfO1	intra-protein xl	541	665	3247.648	812.92	4	31.54
NLKGHGLDEGISTR-VLPLEKK-a3-b6	AfQ1	AfQ1	intra-protein xl	211	133	2459.374	492.883	5	31.11
SLGKGDGLVSSYLDAAPHIAR-KGTLFVDIQR-a4-b1	AfO1	AfO1	intra-protein xl	53	191	3595.945	899.994	4	30.9
VLPLEKK-VLPLEKK-a6-b6	AfQ1	AfQ1	intra-protein xl	133	133	1789.132	597.385	3	30.08

IAGMTGKFSTR-IAGMTGKFSTR-a7-b7	AfO1	AfO1	intra-protein xl	665	665	2473.28	619.328	4	28.46
MLVYAGSLINKGVEAK-DLKQSTK-a11-b3	AfQ1	AfQ1	intra-protein xl	233	162	2648.442	663.118	4	26.83
NLKGHGLDEGISTR-QSTKQR-a3-b4	AfQ1	AfQ1	intra-protein xl	211	166	2380.245	477.057	5	26.57
NLKGHGLDEGISTR-IKKAPYYR-a3-b3	AfQ1	AfQ1	intra-protein xl	211	12	2671.444	668.869	4	25.98
KLLLVITDGEPADIDVRDPQYLR-YKPVSSR-a1-b2	AfO1	AfO1	intra-protein xl	689	518	3611.959	903.998	4	24.5
LRKQEDGDEIDVEAAVR-FKDFDEAYSEEAMGR-a3-b2	AfO1	AfO1	intra-protein xl	541	645	3873.793	969.456	4	23.68
DLGVLVMDLSESTNDKVR-IAGMTGKFSTR-a17-b7	AfO1	AfO1	intra-protein xl	596	665	3408.76	853.198	4	23.65
EVLEKYKPVSSR-IAGMTGKFSTR-a7-b7	AfO1	AfO1	intra-protein xl	518	665	2739.464	685.874	4	22.49
IAGMTGKFSTR-KGTLFVDIQRR-a7-b1	AfO1	AfO1	intra-protein xl	665	191	2637.44	660.368	4	22.18

ARNLKGHGLDEGISTR- MNDQLTEYRIKK-a5-b11	AfQ1	AfQ1	intra- protein xl	211	11	3398.768	567.469	6	21.75
EVADKLVSIAER-QSTKQR-a5-b4	AfQ1	AfQ1	intra- protein xl	199	166	2213.199	554.308	4	20.78

Table 8-4: AfQ1 to AfO1 intra-links

Id	Protein1	Protein2	XLType	AbsPos1	AbsPos2	Mr	Mz	z	Id-Score
LKGHGLDEGISTR-KLDLYLR-a2-b1	FLAGAfQ2 E111Q	AfO2	inter-protein xl	229	223	2439.347	814.124	3	41.42
DLKQSTK-KLDLYLR-a3-b1	FLAGAfQ2 E111Q	AfO2	inter-protein xl	180	223	1876.064	470.024	4	40.41
SWIDFKSGSTPDPR-WLLDKDGTR-a6-b5	AfO2	FLAGAfQ2 E111Q	inter-protein xl	546	101	2832.408	567.489	5	38.67
LPEKLPELFIALTR-RLKGHGLDEGISTR-a4-b3	AfO2	FLAGAfQ2 E111Q	inter-protein xl	749	229	3314.87	553.486	6	36.05
LKGHGLDEGISTR-HAALAKHLK-a2-b6	FLAGAfQ2 E111Q	AfO2	inter-protein xl	229	507	2507.395	627.857	4	35.69
LPEKLPELFIALTR-LKGHGLDEGISTR-a4-b2	AfO2	FLAGAfQ2 E111Q	inter-protein xl	749	229	3158.767	632.761	5	34.7
LKGHGLDEGISTR-KADKK-a2-b1	FLAGAfQ2 E111Q	AfO2	inter-protein xl	229	673	2108.159	528.048	4	34.14
RLKGHGLDEGISTR-KLDLYLR-a3-b1	FLAGAfQ2 E111Q	AfO2	inter-protein xl	229	223	2595.449	649.87	4	34.09

WLLDKDGTR-KLDLYLR-a5-b1	FLAGAfQ2 E111Q	AfO2	inter- protein xl	101	223	2160.196	541.057	4	32.16
LPEKLPELFIALTR-DLKQSTK-a4-b3	AfO2	FLAGAfQ2 E111Q	inter- protein xl	749	180	2595.489	649.88	4	31.42
AHPDFQLVISYNPGYQSLMKDL K-KADKK-a20-b1	FLAGAfQ2 E111Q	AfO2	inter- protein xl	177	673	3389.769	565.969	6	29.91
KRLPVMVK-KLDLYLR-a1-b1	FLAGAfQ2 E111Q	AfO2	inter- protein xl	50	223	2027.233	406.454	5	29.9
RLKGHGLDEGISTR- HAALAKHLKR-a3-b6	FLAGAfQ2 E111Q	AfO2	inter- protein xl	229	507	2819.594	564.927	5	27.73
LKGHGLDEGISTR- LLDLLKPQDKVR-a2-b10	FLAGAfQ2 E111Q	AfO2	inter- protein xl	229	521	2956.667	740.175	4	26.57
LLDLLKPQDK-EHKEFAR-a6-b3	AfO2	HAAfM	inter- protein xl	517	446	2235.22	559.813	4	24.75
LKGHGLDEGISTR- LLDLLKPQDKVR-a2-b6	FLAGAfQ2 E111Q	AfO2	inter- protein xl	229	517	2956.67	740.175	4	23.78
RLKGHGLDEGISTR- GFSEPWGDVVKGR-a3-b11	FLAGAfQ2 E111Q	AfO2	inter- protein xl	229	644	3108.603	778.159	4	21.73

KAPDLETQGLPPR-MDYKDDDDKAGR-a1-b4	AfO2	FLAGAfQ2 E111Q	inter-protein xl	451	4	2986.443	598.296	5	21.44
LKGHGLDEGISTR-KLDLYLR-a2-b1	FLAGAfQ2 E111Q	AfO2	inter-protein xl	229	223	2439.349	814.124	3	54.52
YADLSLKEEDLISGGK-ALCKLGR-a7-b4	HA-AfM	AfO2	inter-protein xl	28	60	2691.416	673.862	4	53.82
YADLSLKEEDLISGGK-KQLDCR-a7-b1	HA-AfM	AfO2	inter-protein xl	28	19	2693.354	674.346	4	52.5
VGEVFSDCIAKASAR-LKGHGLDEGISTR-a11-b2	AfO2	FLAGAfQ2 E111Q	inter-protein xl	38	229	3128.588	626.726	5	48.69
LKGHGLDEGISTR-HAALAKHLK-a2-b6	FLAGAfQ2 E111Q	AfO2	inter-protein xl	229	507	2507.401	627.858	4	47.88
RLKGHGLDEGISTR-HAALAKHLKR-a3-b6	FLAGAfQ2 E111Q	AfO2	inter-protein xl	229	507	2819.587	564.925	5	47.73
LKGHGLDEGISTR-ALCKLGR-a2-b4	FLAGAfQ2 E111Q	AfO2	inter-protein xl	229	60	2336.273	585.076	4	47.11
LKGHGLDEGISTR-QKDFFSLQSADSR-a2-b2	FLAGAfQ2 E111Q	AfO2	inter-protein xl	229	194	3047.526	762.889	4	45.65

LKGHGLDEGISTR-KQLDCR-a2-b1	FLAGAfQ2 E111Q	AfO2	inter-protein xl	229	19	2338.205	585.559	4	44.71
LGKPLVTVACNEDMTAADLVGR-KQLDCR-a3-b1	FLAGAfQ2 E111Q	AfO2	inter-protein xl	77	19	3285.652	822.421	4	41.37
SWIDFKSGSTPDPR-MDYKDDDDKAGR-a6-b4	AfO2	FLAGAfQ2 E111Q	inter-protein xl	546	4	3157.43	632.494	5	41.25
SWIDFKSGSTPDPR-WLLDKDGTR-a6-b5	AfO2	FLAGAfQ2 E111Q	inter-protein xl	546	101	2832.408	709.11	4	40.76
LPEKLPELFIALTR-LKGHGLDEGISTR-a4-b2	AfO2	FLAGAfQ2 E111Q	inter-protein xl	749	229	3158.77	790.7	4	38.91
LGKPLVTVACNEDMTAADLVGR-QKDFFSLQSADSR-a3-b2	FLAGAfQ2 E111Q	AfO2	inter-protein xl	77	194	3994.982	999.753	4	37.36
VGEVFSDCIAKASAR-RLKGHGLDEGISTR-a11-b3	AfO2	FLAGAfQ2 E111Q	inter-protein xl	38	229	3284.696	657.947	5	35.81
AHPDFQLVISYNPGYQSLMKDLK-KADKK-a20-b1	FLAGAfQ2 E111Q	AfO2	inter-protein xl	177	673	3389.767	678.961	5	35.45
LKGHGLDEGISTR-KADKK-a2-b1	FLAGAfQ2 E111Q	AfO2	inter-protein xl	229	673	2108.158	422.639	5	35.11

SWIDFKSGSTPDPR-MDYKDDDDDK-a6-b4	AfO2	FLAGAfQ2 E111Q	inter-protein xl	546	4	2873.269	719.325	4	34.1
RLPVMVKGPTGCGK-KLDLYLR-a7-b1	FLAGAfQ2 E111Q	AfO2	inter-protein xl	57	223	2556.428	640.115	4	33.96
RLLDLLKPQDK-DLKQSTK-a7-b3	AfO2	FLAGAfQ2 E111Q	inter-protein xl	517	180	2294.318	574.587	4	32.9
LLDLLKPQDKVR-DLKQSTK-a6-b3	AfO2	FLAGAfQ2 E111Q	inter-protein xl	517	180	2393.386	599.354	4	32.66
LKGHGLDEGISTR-LLDLLKPQDKVR-a2-b10	FLAGAfQ2 E111Q	AfO2	inter-protein xl	229	521	2956.672	740.176	4	32.29
RLKGHGLDEGISTR-QKDFSLQSADSR-a3-b2	FLAGAfQ2 E111Q	AfO2	inter-protein xl	229	194	3203.631	641.734	5	32.08
TLFLALHPKPGEDACDPETTSLR-LKGHGLDEGISTR-a9-b2	AfO2	FLAGAfQ2 E111Q	inter-protein xl	337	229	4247.096	850.427	5	32.06
RLKGHGLDEGISTR-KLDLYLR-a3-b1	FLAGAfQ2 E111Q	AfO2	inter-protein xl	229	223	2595.459	520.1	5	31.32
LGKPLVTVACNEDMTAADLVGR-ALCKLGR-a3-b4	FLAGAfQ2 E111Q	AfO2	inter-protein xl	77	60	3283.707	821.935	4	31

MDYKDDDDKAGR-KQLDCR-a9-b1	FLAGAfQ2 E111Q	AfO2	inter-protein xl	9	19	2384.072	597.026	4	30.17
LPEKLPELFIALTR-RLKGHGLDEGISTR-a4-b3	AfO2	FLAGAfQ2 E111Q	inter-protein xl	749	229	3314.867	663.981	5	28.6
AFESFPGDADKLYPGWR-QKDFFSLQSADSR-a11-b2	HA-AfM	AfO2	inter-protein xl	461	194	3620.717	906.187	4	28.1
LGKPLVTVACNEDMTAADLVGR-KLDLYLR-a3-b1	FLAGAfQ2 E111Q	AfO2	inter-protein xl	77	223	3386.801	1129.942	3	25.25
KAPDLETQGLPPR-EKLGVHR-a1-b2	AfO2	HA-AfM	inter-protein xl	451	469	2396.321	600.088	4	23
NDEPQGNQVFAPVKK-KGELIR-a14-b1	HA-AfM	FLAGAfQ2 E111Q	inter-protein xl	220	152	2522.342	631.593	4	21.52
QSDQFAKVHFADTEISYR-QSTKQR-a7-b4	AfO2	FLAGAfQ2 E111Q	inter-protein xl	415	184	3025.496	606.107	5	20.52

Table 8-5: FLAGQ2^{WB} and AfO2 inter-links with FLAG-Q2^{WB}O2 to HA-AfM interlinks

Id	Protein1	Protein2	XLType	AbsPos1	AbsPos2	Mr	Mz	z	Id-Score
LKGHGLDEGISTR-DLKQSTK-a2-b3	FLAGAfQ2 E111Q	FLAGAfQ2 E111Q	intra-protein xl	229	180	2338.249	780.424	3	41.1
ILGRPIKDGGYIAGTIKPKILGR PIKDGGYIAGTIKPK-a7-b7	HAAfM	HAAfM	intra-protein xl	170	170	4356.603	872.328	5	40.99
MDYKDDDDKAGR-KGELIR-a9-b1	FLAGAfQ2 E111Q	FLAGAfQ2 E111Q	intra-protein xl	9	152	2280.104	571.034	4	39.79
RLKGHGLDEGISTR-KGELIR-a3-b1	FLAGAfQ2 E111Q	FLAGAfQ2 E111Q	intra-protein xl	229	152	2390.335	479.075	5	38.45
LKGHGLDEGISTR- LKGHGLDEGISTR-a2-b2	FLAGAfQ2 E111Q	FLAGAfQ2 E111Q	intra-protein xl	229	229	2901.522	726.388	4	37.79
WLLDKDGTR-WLLDKDGTR-a5-b5	FLAGAfQ2 E111Q	FLAGAfQ2 E111Q	intra-protein xl	101	101	2343.235	782.086	3	37.53
GYTAFVLAKMSR-EHKEFAR-a9-b3	HAAfM	HAAfM	intra-protein xl	325	446	2396.229	480.254	5	36.41
WLLDKDGTR-KGELIR-a5-b1	FLAGAfQ2 E111Q	FLAGAfQ2 E111Q	intra-protein xl	101	152	1955.082	652.702	3	35.4

GYTAFVLAKMSR-EKLGVHR-a9-b2	HAAfM	HAAfM	intra-protein xl	325	469	2318.255	464.659	5	34.89
RLKGHGLDEGISTR-LKGHGLDEGISTR-a3-b2	FLAGAfQ2 E111Q	FLAGAfQ2 E111Q	intra-protein xl	229	229	3057.628	510.613	6	34.37
AGHGMITSPSAKR-AGHGMITSPSAKR-a12-b12	HAAfM	HAAfM	intra-protein xl	315	315	2761.419	461.244	6	33.83
LLDLLKPQDKVR-KADKK-a10-b1	AfO2	AfO2	intra-protein xl	521	673	2163.297	722.107	3	33.11
AGHGMITSPSAKR-HILVAYKMKPK-a12-b7	HAAfM	HAAfM	intra-protein xl	315	44	2776.524	556.313	5	32.9
MDYKDDDDKAGR-KGELIR-a4-b1	FLAGAfQ2 E111Q	FLAGAfQ2 E111Q	intra-protein xl	4	152	2280.106	571.034	4	32.47
TLPLDKKGELIR-TLPLDKKGELIR-a6-b6	FLAGAfQ2 E111Q	FLAGAfQ2 E111Q	intra-protein xl	151	151	2901.726	968.25	3	32.44
MDYKDDDDKAGR-WLLDKDGTR-a4-b5	FLAGAfQ2 E111Q	FLAGAfQ2 E111Q	intra-protein xl	4	101	2668.241	668.068	4	32.19
KRLPVMVK-KRLPVMVK-a1-b1	FLAGAfQ2 E111Q	FLAGAfQ2 E111Q	intra-protein xl	50	50	2077.296	416.467	5	32.15

WLLDKDGTRWQDGPLTVAAR- WLLDKDGTR-a5-b5	FLAGAfQ2 E111Q	FLAGAfQ2 E111Q	intra-protein xl	101	101	3537.836	708.575	5	32.04
KLDLYLR-KADKK-a1-b1	AfO2	AfO2	intra-protein xl	223	673	1645.973	412.501	4	31.69
MDYKDDDDKAGR- TLPLDKKGELIR-a4-b7	FLAGAfQ2 E111Q	FLAGAfQ2 E111Q	intra-protein xl	4	152	2947.495	590.507	5	31.44
LKGHGLDEGISTR- WLLDKDGTR-a2-b5	FLAGAfQ2 E111Q	FLAGAfQ2 E111Q	intra-protein xl	229	101	2622.375	525.483	5	31.02
RLKGHGLDEGISTR- TLPLDKKGELIR-a3-b7	FLAGAfQ2 E111Q	FLAGAfQ2 E111Q	intra-protein xl	229	152	3057.724	612.553	5	30.57
MDYKDDDDKAGR- TLPLDKKGELIR-a9-b7	FLAGAfQ2 E111Q	FLAGAfQ2 E111Q	intra-protein xl	9	152	2947.494	737.881	4	30.06
KRLPVMVK-KGELIR-a1-b1	FLAGAfQ2 E111Q	FLAGAfQ2 E111Q	intra-protein xl	50	152	1822.121	456.538	4	29.59
TLPLDKKGELIR- WLLDKDGTR-a7-b5	FLAGAfQ2 E111Q	FLAGAfQ2 E111Q	intra-protein xl	152	101	2622.472	525.502	5	29.17
TLPLDKKGELIR- WLLDKDGTR-a6-b5	FLAGAfQ2 E111Q	FLAGAfQ2 E111Q	intra-protein xl	151	101	2622.475	656.626	4	29.12

RLKGHGLDEGISTR- WLLDKDGTR-a3-b5	FLAGAfQ2 E111Q	FLAGAfQ2 E111Q	intra-protein xl	229	101	2778.474	556.703	5	28.95
KRLPVMVK-DLKQSTK-a1-b3	FLAGAfQ2 E111Q	FLAGAfQ2 E111Q	intra-protein xl	50	180	1926.132	482.541	4	28.88
LKGHGLDEGISTR- DLKQSTKQR-a2-b7	FLAGAfQ2 E111Q	FLAGAfQ2 E111Q	intra-protein xl	229	184	2622.407	656.61	4	28.62
AFESFPGDADKLYPGWR- EKLGVHR-a11-b2	HAAfM	HAAfM	intra-protein xl	461	469	2930.472	733.626	4	28.31
LKGHGLDEGISTR-KGELIR-a2- b1	FLAGAfQ2 E111Q	FLAGAfQ2 E111Q	intra-protein xl	229	152	2234.249	559.57	4	28.29
WLLDKDGTR-TLPLDKK-a5-b6	FLAGAfQ2 E111Q	FLAGAfQ2 E111Q	intra-protein xl	101	151	2054.139	514.542	4	27.21
TLPLDKKGELIR- TLPLDKKGELIR-a7-b7	FLAGAfQ2 E111Q	FLAGAfQ2 E111Q	intra-protein xl	152	152	2901.726	581.353	5	27.08
MDYKDDDDK-KGELIR-a4-b1	FLAGAfQ2 E111Q	FLAGAfQ2 E111Q	intra-protein xl	4	152	1995.945	666.323	3	26.97
QKDFFSLQSADSR-KLDLYLR- a2-b1	AfO2	AfO2	intra-protein xl	194	223	2585.349	862.791	3	26.83

GFSEPWGDVVKGR- HAALAKHLKR-a11-b9	AfO2	AfO2	intra-protein xl	644	510	2714.476	543.903	5	26.38
LKGHGLDEGISTR-TLPLDKK- a2-b6	FLAGAfQ2 E111Q	FLAGAfQ2 E111Q	intra-protein xl	229	151	2333.293	467.666	5	26.33
KGELIR-KGELIR-a1-b1	FLAGAfQ2 E111Q	FLAGAfQ2 E111Q	intra-protein xl	152	152	1566.944	523.322	3	25.94
AHPDFQLVISYNPGYQSLMKD LK-DLKQSTK-a20-b3	FLAGAfQ2 E111Q	FLAGAfQ2 E111Q	intra-protein xl	177	180	3619.861	724.98	5	25.26
LKGHGLDEGISTR- DLKQSTKQR-a2-b3	FLAGAfQ2 E111Q	FLAGAfQ2 E111Q	intra-protein xl	229	180	2622.406	525.489	5	25.23
TLPLDKKGELIR-DDDDKAGR- a7-b5	FLAGAfQ2 E111Q	FLAGAfQ2 E111Q	intra-protein xl	152	9	2410.268	603.575	4	24.59
RLKGHGLDEGISTR- TLPLDKKGELIR-a3-b6	FLAGAfQ2 E111Q	FLAGAfQ2 E111Q	intra-protein xl	229	151	3057.724	612.553	5	24.41
QKDFFSLQSADSR- KQLDCRFPR-a2-b1	AfO2	AfO2	intra-protein xl	194	19	2884.42	722.113	4	24.31
MDYKDDDDKAGR- MDYKDDDDKAGR-a4-b4	FLAGAfQ2 E111Q	FLAGAfQ2 E111Q	intra-protein xl	4	4	2993.262	599.66	5	24

LQGASGIHVGTMGYGKMEGG ADDR-AGHGMITSPSAKR-a16- b12	HAAfM	HAAfM	intra-protein xl	344	315	3855.839	964.968	4	23.99
TLPLDKKGELIR-KGELIR-a6-b1	FLAGAfQ2 E111Q	FLAGAfQ2 E111Q	intra-protein xl	151	152	2234.332	559.591	4	23.91
TLPLDKKGELIR-RTLPLDKK- a6-b7	FLAGAfQ2 E111Q	FLAGAfQ2 E111Q	intra-protein xl	151	151	2489.494	623.381	4	23
MDYKDDDDKAGR- TLPLDKKGELIR-a9-b6	FLAGAfQ2 E111Q	FLAGAfQ2 E111Q	intra-protein xl	9	151	2947.494	737.881	4	22.91
WLLDKDGTR-RTLPLDKK-a5- b7	FLAGAfQ2 E111Q	FLAGAfQ2 E111Q	intra-protein xl	101	151	2210.26	553.573	4	22.87
LLDLLKPQDKVR-KADKK-a6- b4	AfO2	AfO2	intra-protein xl	517	676	2163.297	541.832	4	22.26
TLPLDKKGELIR-DLKQSTK-a7- b3	FLAGAfQ2 E111Q	FLAGAfQ2 E111Q	intra-protein xl	152	180	2338.345	468.677	5	22.07
RLKGHGLDEGISTR- KRLPVMVK-a3-b1	FLAGAfQ2 E111Q	FLAGAfQ2 E111Q	intra-protein xl	229	50	2645.513	662.386	4	21.94
RLKGHGLDEGISTR-DLKQSTK- a3-b3	FLAGAfQ2 E111Q	FLAGAfQ2 E111Q	intra-protein xl	229	180	2494.345	499.877	5	21.93

QKDFSLQSAISR-KADKK-a2-b1	AfO2	AfO2	intra-protein xl	194	673	2254.155	752.393	3	21.88
TLPLDKKGELIR-TLPLDKK-a7-b6	FLAGAfQ2 E111Q	FLAGAfQ2 E111Q	intra-protein xl	152	151	2333.398	584.357	4	21.71
LGKPLVTVACNEDMTAADLV GR-LKGHGLDEGISTR-a3-b2	FLAGAfQ2 E111Q	FLAGAfQ2 E111Q	intra-protein xl	77	229	3848.977	963.252	4	60.19
LKGHGLDEGISTR-KGELIR-a2-b1	FLAGAfQ2 E111Q	FLAGAfQ2 E111Q	intra-protein xl	229	152	2234.237	559.567	4	58.5
RLKGHGLDEGISTR- LPVMVKGPTGCGK-a3-b6	FLAGAfQ2 E111Q	FLAGAfQ2 E111Q	intra-protein xl	229	57	3018.612	604.73	5	50.33
LKGHGLDEGISTR- LPVMVKGPTGCGK-a2-b6	FLAGAfQ2 E111Q	FLAGAfQ2 E111Q	intra-protein xl	229	57	2862.517	716.637	4	48.98
ALCKLGR-KQLDCR-a4-b1	AfO2	AfO2	intra-protein xl	60	19	1772.944	444.244	4	48.58
ILGRPIKDGGYIAGTIIKPK- ILGRPIKDGGYIAGTIIKPK-a7-b7	HA-AfM	HA-AfM	intra-protein xl	170	170	4356.61	872.33	5	48.05
MDYKDDDDK-KRLPVMVK-a4-b1	FLAGAfQ2 E111Q	FLAGAfQ2 E111Q	intra-protein xl	4	50	2251.13	451.234	5	47.81

VGEVFSDCIAKASAR-KQLDCR-a11-b1	AfO2	AfO2	intra-protein xl	38	19	2565.269	642.325	4	47.11
LGKPLVTVACNEDMTAADLV GR- LGKPLVTVACNEDMTAADLV GR-a3-b3	FLAGAfQ2 E111Q	FLAGAfQ2 E111Q	intra-protein xl	77	77	4796.422	1200.113	4	46.93
MDYKDDDDK-KGELIR-a4-b1	FLAGAfQ2 E111Q	FLAGAfQ2 E111Q	intra-protein xl	4	152	1995.955	499.997	4	46.15
LGKPLVTVACNEDMTAADLV GR-LPVMVKGPTGCGK-a3-b6	FLAGAfQ2 E111Q	FLAGAfQ2 E111Q	intra-protein xl	77	57	3809.956	953.497	4	46.13
LKGHGLDEGISTR- LKGHGLDEGISTR-a2-b2	FLAGAfQ2 E111Q	FLAGAfQ2 E111Q	intra-protein xl	229	229	2901.543	581.316	5	45.64
KTIPLVYDAMK- KTIPLVYDAMK-a1-b1	HA-AfM	HA-AfM	intra-protein xl	221	221	2693.479	674.378	4	45.42
LGKPLVTVACNEDMTAADLV GR-KGELIR-a3-b1	FLAGAfQ2 E111Q	FLAGAfQ2 E111Q	intra-protein xl	77	152	3181.683	796.429	4	45.37
MDYKDDDDK-WLLDKDGTR-a4-b5	FLAGAfQ2 E111Q	FLAGAfQ2 E111Q	intra-protein xl	4	101	2384.087	795.703	3	43.49

RLKGHGLDEGISTR- RLPVMVKGPTGCGK-a3-b7	FLAGAfQ2 E111Q	FLAGAfQ2 E111Q	intra-protein xl	229	57	3174.711	635.95	5	43.15
LGKPLVTVACNEDMTAADLV GR-DLKQSTK-a3-b3	FLAGAfQ2 E111Q	FLAGAfQ2 E111Q	intra-protein xl	77	180	3285.699	822.433	4	42.95
QKDFFSLQSADSR-KLDLYLR- a2-b1	AfO2	AfO2	intra-protein xl	194	223	2585.345	647.344	4	41.82
WLLDKDGTR-KGELIR-a5-b1	FLAGAfQ2 E111Q	FLAGAfQ2 E111Q	intra-protein xl	101	152	1955.079	652.701	3	41.55
LKGHGLDEGISTR-DLKQSTK- a2-b3	FLAGAfQ2 E111Q	FLAGAfQ2 E111Q	intra-protein xl	229	180	2338.253	780.425	3	41.33
WLLDKDGTR-TLPLDKK-a5-b6	FLAGAfQ2 E111Q	FLAGAfQ2 E111Q	intra-protein xl	101	151	2054.142	685.722	3	41.01
LGKPLVTVACNEDMTAADLV GR-QSTKQR-a3-b4	FLAGAfQ2 E111Q	FLAGAfQ2 E111Q	intra-protein xl	77	184	3213.66	804.423	4	40.82
DDDDKAGR-KGELIR-a5-b1	FLAGAfQ2 E111Q	FLAGAfQ2 E111Q	intra-protein xl	9	152	1742.88	581.968	3	40.69
LPVMVKGPTGCGK-KGELIR- a6-b1	FLAGAfQ2 E111Q	FLAGAfQ2 E111Q	intra-protein xl	57	152	2195.218	549.812	4	40.4

LKGHGLDEGISTR-MDYKDDDDK-a2-b4	FLAGAfQ2 E111Q	FLAGAfQ2 E111Q	intra-protein xl	229	4	2663.242	666.818	4	39.98
LKGHGLDEGISTR-RTLPLDKK-a2-b7	FLAGAfQ2 E111Q	FLAGAfQ2 E111Q	intra-protein xl	229	151	2489.405	623.359	4	39.7
RLPVMVKGPTGCGK-WLLDKDGTR-a7-b5	FLAGAfQ2 E111Q	FLAGAfQ2 E111Q	intra-protein xl	57	101	2739.463	685.874	4	39.43
AGHGMITSPSAKR-AGHGMITSPSAKR-a12-b12	HA-AfM	HA-AfM	intra-protein xl	315	315	2761.418	553.291	5	39.42
VGEVFSDCIAKASAR-ALCKLGR-a11-b4	AfO2	AfO2	intra-protein xl	38	60	2563.326	641.839	4	39.39
LKGHGLDEGISTR-WLLDKDGTR-a2-b5	FLAGAfQ2 E111Q	FLAGAfQ2 E111Q	intra-protein xl	229	101	2622.376	525.483	5	39.33
MDYKDDDDKAGR-KGELIR-a9-b1	FLAGAfQ2 E111Q	FLAGAfQ2 E111Q	intra-protein xl	9	152	2280.109	571.035	4	39.22
QKDFFSLQSADSR-ALCKLGR-a2-b4	AfO2	AfO2	intra-protein xl	194	60	2482.267	621.575	4	39.05
LPVMVKGPTGCGK-WLLDKDGTR-a6-b5	FLAGAfQ2 E111Q	FLAGAfQ2 E111Q	intra-protein xl	57	101	2583.353	862.126	3	38.95

MDYKDDDDKAGR-WLLDKDGTR-a9-b5	FLAGAfQ2 E111Q	FLAGAfQ2 E111Q	intra-protein xl	9	101	2668.246	534.657	5	38.62
ILGRPIKDGGYIAGTIKPK-QAYECWKAGADPIEYAR-a7-b7	HA-AfM	HA-AfM	intra-protein xl	170	433	4274.259	713.384	6	38.51
MDYKDDDDKAGR-MDYKDDDDK-a9-b4	FLAGAfQ2 E111Q	FLAGAfQ2 E111Q	intra-protein xl	9	4	2709.107	678.284	4	37.03
RLKGHGLDEGISTR-WLLDKDGTR-a3-b5	FLAGAfQ2 E111Q	FLAGAfQ2 E111Q	intra-protein xl	229	101	2778.476	556.703	5	36.9
WLLDKDGTR-WLLDKDGTR-a5-b5	FLAGAfQ2 E111Q	FLAGAfQ2 E111Q	intra-protein xl	101	101	2343.219	586.813	4	36.69
RLPVMVKGPTGCGK-DLKQSTK-a7-b3	FLAGAfQ2 E111Q	FLAGAfQ2 E111Q	intra-protein xl	57	180	2455.335	614.842	4	35.92
QSDQFAKVHFADTEISYR-VGEVFSDCIAKASAR-a7-b11	AfO2	AfO2	intra-protein xl	415	38	3887.891	778.586	5	35.73
KAPDLETQGLPPR-KQLDCR-a1-b1	AfO2	AfO2	intra-protein xl	451	19	2377.243	595.319	4	35.65
GFSEPWGDVVKGR-HAALAKHLK-a11-b6	AfO2	AfO2	intra-protein xl	644	507	2558.377	512.683	5	35.59

LPVMVKGPTGCGK-DLKQSTK-a6-b3	FLAGAfQ2 E111Q	FLAGAfQ2 E111Q	intra-protein xl	57	180	2299.23	575.815	4	35.37
MDYKDDDDKAGR-MDYKDDDDKAGR-a9-b9	FLAGAfQ2 E111Q	FLAGAfQ2 E111Q	intra-protein xl	9	9	2993.268	599.661	5	34.47
LKGHGLDEGISTR-KRLPVMVK-a2-b1	FLAGAfQ2 E111Q	FLAGAfQ2 E111Q	intra-protein xl	229	50	2489.423	498.892	5	34.17
HAALAKHLK-ALCKLGR-a6-b4	AfO2	AfO2	intra-protein xl	507	60	1942.131	486.541	4	33.77
RLKGHGLDEGISTR-KGELIR-a3-b1	FLAGAfQ2 E111Q	FLAGAfQ2 E111Q	intra-protein xl	229	152	2390.336	598.592	4	33.7
AHPDFQLVISYNPGYQSLMKD LK-LKGHGLDEGISTR-a20-b2	FLAGAfQ2 E111Q	FLAGAfQ2 E111Q	intra-protein xl	177	229	4183.141	837.636	5	33.36
MDYKDDDDKAGR-KGELIR-a4-b1	FLAGAfQ2 E111Q	FLAGAfQ2 E111Q	intra-protein xl	4	152	2280.106	457.029	5	33.21
AFESFPGDADKLYPGWR-EKLGVHR-a11-b2	HA-AfM	HA-AfM	intra-protein xl	461	469	2930.471	587.102	5	32.76
LPVMVKGPTGCGK-LPVMVKGPTGCGK-a6-b6	FLAGAfQ2 E111Q	FLAGAfQ2 E111Q	intra-protein xl	57	57	2823.491	706.88	4	32.48

GFSEPWGDVVKGR-ALCKLGR-a11-b4	AfO2	AfO2	intra-protein xl	644	60	2387.24	597.818	4	32.29
LGKPLVTVACNEDMTAADLV GR-MDYKDDDDKAGR-a3-b4	FLAGAfQ2 E111Q	FLAGAfQ2 E111Q	intra-protein xl	77	4	3894.844	779.977	5	32.11
AHPDFQLVISYNPGYQSLMKD LK-DLKQSTK-a20-b3	FLAGAfQ2 E111Q	FLAGAfQ2 E111Q	intra-protein xl	177	180	3619.866	724.981	5	31.93
GGMTATDSSILNQYLVGKEPF YQPQHDEVALFEAAYR- MDYKDDDDK-a18-b4	FLAGAfQ2 E111Q	FLAGAfQ2 E111Q	intra-protein xl	30	4	5426.498	1086.307	5	31.51
KTIPLVYDAMKR- KTIPLVYDAMK-a11-b1	HA-AfM	HA-AfM	intra-protein xl	231	221	2849.581	713.403	4	31.24
NDEPQGNQVFAPVKK- KTIPLVYDAMK-a14-b1	HA-AfM	HA-AfM	intra-protein xl	220	221	3085.617	772.412	4	30.98
KGELIR-KGELIR-a1-b1	FLAGAfQ2 E111Q	FLAGAfQ2 E111Q	intra-protein xl	152	152	1566.945	523.323	3	30.65
QSDQFAKVHFADTEISYR- QKDFFSLSADSAR-a7-b2	AfO2	AfO2	intra-protein xl	415	194	3806.825	762.373	5	29.75
LGKPLVTVACNEDMTAADLV GR-GPTGCGKSR-a3-b7	FLAGAfQ2 E111Q	FLAGAfQ2 E111Q	intra-protein xl	77	64	3385.681	847.428	4	29.75

LQGASGIHVGTMGYGKMEGG ADDR-DGGYIAGTHKPK-a16- b11	HA-AfM	HA-AfM	intra-protein xl	344	181	3875.919	969.988	4	29.57
GGMTATDSSILNQYLVGKEPF YQPQHDEVALFEAAYR- WLLDKDGTR-a18-b5	FLAGAfQ2 E111Q	FLAGAfQ2 E111Q	intra-protein xl	30	101	5385.632	1078.134	5	29.3
TLFLALHPKPGEDACDPETTSC LR-ALCKLGR-a9-b4	AfO2	AfO2	intra-protein xl	337	60	3681.831	921.466	4	29.29
LKGHGLDEGISTR-QSTKQR-a2- b4	FLAGAfQ2 E111Q	FLAGAfQ2 E111Q	intra-protein xl	229	184	2266.199	454.248	5	28.77
KTIPLVYDAMK- TIPLVYDAMKR-a1-b10	HA-AfM	HA-AfM	intra-protein xl	221	231	2721.481	681.378	4	28.65
LQGASGIHVGTMGYGKMEGG ADDR- AFESFPGDADKLYPGWR-a16- b11	HA-AfM	HA-AfM	intra-protein xl	344	461	4499.094	900.827	5	28.44
LGKPLVTVACNEDMTAADLV GR-RLPVMVKGPTGCGK-a3-b7	FLAGAfQ2 E111Q	FLAGAfQ2 E111Q	intra-protein xl	77	57	3966.052	794.218	5	28.43

MDYKDDDDK-MDYKDDDDK-a4-b4	FLAGAfQ2 E111Q	FLAGAfQ2 E111Q	intra-protein xl	4	4	2424.948	607.245	4	28.09
LPVMVKGPTGCGK-QSTKQR-a6-b4	FLAGAfQ2 E111Q	FLAGAfQ2 E111Q	intra-protein xl	57	184	2227.184	557.804	4	27.39
GPTGCGKSR-WLLDKDGTR-a7-b5	FLAGAfQ2 E111Q	FLAGAfQ2 E111Q	intra-protein xl	64	101	2159.074	720.699	3	27.13
WLLDKDGTR-KRLPVMVK-a5-b1	FLAGAfQ2 E111Q	FLAGAfQ2 E111Q	intra-protein xl	101	50	2210.26	553.573	4	27.06
QKDFFSLQSADSR-KADKK-a2-b1	AfO2	AfO2	intra-protein xl	194	673	2254.158	564.547	4	26.93
LGKPLVTVACNEDMTAADLV GR-MDYKDDDDKAGR-a3-b9	FLAGAfQ2 E111Q	FLAGAfQ2 E111Q	intra-protein xl	77	9	3894.842	974.718	4	26.64
QSDQFAKVHFADTEISYR-KQLDCR-a7-b1	AfO2	AfO2	intra-protein xl	415	19	3097.497	775.382	4	26.51
QAYECWKAGADPIEYAR-EHKEFAR-a7-b3	HA-AfM	HA-AfM	intra-protein xl	433	446	3080.45	771.12	4	26.49
LQGASGIHVGTMGYGKMEGG ADDR-AGHGMITSPSAKR-a16-b12	HA-AfM	HA-AfM	intra-protein xl	344	315	3855.846	772.177	5	26.4

QSDQFAKVHFADTEISYR- ALCKLGR-a7-b4	AfO2	AfO2	intra-protein xl	415	60	3095.548	620.117	5	25.48
LQGASGIHVGTMGYGKMEGG ADDRNIAYMIER- AFESFPGDADKLYPGWR-a16- b11	HA-AfM	HA-AfM	intra-protein xl	344	461	5489.589	1098.926	5	25.29
RLKGHGLDEGISTR- MDYKDDDDK-a3-b4	FLAGAfQ2 E111Q	FLAGAfQ2 E111Q	intra-protein xl	229	4	2819.344	564.877	5	24.98
LLDLLKPQDKVR-KADKK-a10- b1	AfO2	AfO2	intra-protein xl	521	673	2163.3	541.833	4	24.88
AHPDFQLVISYNPGYQSLMKD LK-KGELIR-a20-b1	FLAGAfQ2 E111Q	FLAGAfQ2 E111Q	intra-protein xl	177	152	3515.849	879.97	4	24.39
MDYKDDDDKAGR-DLKQSTK- a4-b3	FLAGAfQ2 E111Q	FLAGAfQ2 E111Q	intra-protein xl	4	180	2384.121	597.038	4	24.18
LQGASGIHVGTMGYGKMEGG ADDRNIAYMIER- AGHGMITSPSAKR-a16-b12	HA-AfM	HA-AfM	intra-protein xl	344	315	4846.342	970.276	5	23.84

AHPDFQLVISYNPGYQSLMKD LK- LGKPLVTVACNEDMTAADLV GR-a20-b3	FLAGAfQ2 E111Q	FLAGAfQ2 E111Q	intra-protein xl	177	77	5130.579	1027.124	5	23.46
RLKGHGLDEGISTR-DLKQSTK- a3-b3	FLAGAfQ2 E111Q	FLAGAfQ2 E111Q	intra-protein xl	229	180	2494.352	499.878	5	22.95
RLPVMVKGPTGCGK- MDYKDDDDK-a7-b4	FLAGAfQ2 E111Q	FLAGAfQ2 E111Q	intra-protein xl	57	4	2780.322	696.088	4	22.45
RLKGHGLDEGISTR- TLPLDKKGELIR-a3-b7	FLAGAfQ2 E111Q	FLAGAfQ2 E111Q	intra-protein xl	229	152	3057.717	765.437	4	22.31
AHPDFQLVISYNPGYQSLMKD LK-WLLDKDGTR-a20-b5	FLAGAfQ2 E111Q	FLAGAfQ2 E111Q	intra-protein xl	177	101	3903.982	977.003	4	21.63
DSCDGPVYHQEWYGMKPTTPI ISGGMNALR- DGGYIAGTIKPK-a16-b11	HA-AfM	HA-AfM	intra-protein xl	376	181	4849.361	970.88	5	21.48
LPVMVKGPTGCGK-TLPLDKK- a6-b6	FLAGAfQ2 E111Q	FLAGAfQ2 E111Q	intra-protein xl	57	151	2294.275	765.766	3	21.42
KAPDLETQGLPPR-ALCKLGR- a1-b4	AfO2	AfO2	intra-protein xl	451	60	2375.299	594.833	4	21.12

QSDQFAKVHFADTEISYR- KAPDLETQGLPPR-a7-b1	AfO2	AfO2	intra-protein xl	415	451	3699.855	740.979	5	20.58
MDYKDDDDKAGR-TLPLDKK- a4-b6	FLAGAfQ2 E111Q	FLAGAfQ2 E111Q	intra-protein xl	4	151	2379.169	595.8	4	20.08
NDEPQGNQVFAPVKK- TIPLVYDAMKR-a14-b10	HA-AfM	HA-AfM	intra-protein xl	220	231	3113.62	779.413	4	20.05

Table 8-6: FLAGQ2^{WB} and AfO2 intra-links with FLAG-Q2^{WB}O2 to HA-AfM intra-links

8.2 Sequence alignment of 50 CbbQ orthologs

```

AfQ2_|B7J5E4|
AfQ1_|B7JA26|
AfQ3|B7JB17|
Acidithiobacillus_ferriivorans_CbbQ3_|G0JL86|
Ferrovum_sp_CbbQ3_|A0A0R0LV7|
Thioalkalivibrio_sp_CbbQ3_|D3SG06|
Thioalkalivibrio_sulfidiphilus_CbbQ3_|B8GQN5|
Acidimicrobium_ferrooxidans_CbbQ3_|C7M202|
Acidithiobacillus_caldus_CbbQ3_|F9ZLQ7|
Halothiobacillus_neapolitanus_CbbQ3_|D0KZ75|
Nitrobacter_hamburgensis_CbbQ3_|Q1QFQ9|
Nitrosomonas_eutropha_CbbQ3_|Q0AHV0|
Thiomonas_intermedia_CbbQ3_|D5X343|
Acidithiobacillus_ferriivorans_CbbQ1_|G0JNK4|
Nitrosomonas_communis_CbbQ1_|A0A0F7KI56|
Nitrosomonas_europaea_CbbQ1_|Q82TG8|
Nitrosomonas_sp_CbbQ1_|F9ZF97|
Thiobacillus_denitrificans_CbbQ1_|Q3SFN1|
Thiobacillus_thioparus_CbbQ1_|WP_018507362.1|
Thiomonas_sp_CbbQ1_|A0A078BG06|
Hydrogenovibrio_marinus_CbbQ1_|Q75W43|
Sideroxydans_lithotrophicus_CbbQ1_|D5CQ61|
Sulfuricella_denitrificans_CbbQ1_|S6AM60|
Hydrogenovibrio_kuenenii_CbbQ1_|WP_024851894.1|
Hydrogenovibrio_halophilus_CbbQ1_|WP_019895627.1|
Thiomicrothabhdus_chilensis_CbbQ1_|WP_028486808.1|
Hydrogenovibrio_crunogenus_CbbQ1_|Q31IJ8|
Thiomicrothabhdus_sp_CbbQ1_|WP_029406683.1|
Mariprofundus_ferrooxydans_CbbQ1_|Q0EX20|
Thiothrix_nivea_CbbQ1_|I3BR11|
Acidithiobacillus_ferriivorans_CbbQ2_|G0JPS2|
Thiobacillus_denitrificans_CbbQ2_|Q3SFL6|
Thiobacillus_thioparus_CbbQ2_|WP_018507397.1|
Halothiobacillus_neapolitanus_CbbQ2_|D0KZQ7|
Thiomonas_arsenitoxydans_CbbQ2_|D6CTS9|
Thiomonas_intermedia_CbbQ2_|Q84918|
Sideroxydans_lithotrophicus_CbbQ2_|D5CTF9|
Sulfuricella_denitrificans_CbbQ2_|S6ABC2|
Gallionella_capsiferriiformans_CbbQ2_|D9SHP7|
Mariprofundus_ferrooxydans_CbbQ2_|Q0F2Z9|
Acidithiobacillus_caldus_CbbQ2_|F9ZR06|
Thiolapillus_brandeam_CbbQ2_|W0TLR2|
Hydrogenovibrio_marinus_CbbQ2_|Q9WXI1|
Hydrogenovibrio_kuenenii_CbbQ2_|WP_024851802.1|
Thiomicrothabhdus_sp_CbbQ2_|WP_029406689.1|
Hydrogenovibrio_halophilus_CbbQ2_|WP_019895619.1|
Thiomicrothabhdus_chilensis_CbbQ2_|WP_028486930.1|
Hydrogenovibrio_crunogenus_CbbQ2_|Q31IK4|
Thiothrix_nivea_CbbQ2_|I3BN91|
consensus>70

```

1MTATDSSILN*
1MND...QLTEYR*
1MSP...EIDQYLV*
1MSP...EIDPYLV*
1MS...VIAEYN*
1MQD...DLRDHL*
1MQD...ELRDYL*
1MVAP...QVAPYV*
1MTD...LASAYL*
1MTQN...ADQYRV*
1MALHAVTDTAGTAA...VLDAYR*
1MSLPATAVVAPTSS...SLDSYRV*
1 MRHEDYVRTYAPRQDARPARAGSVMPAAHRPAADLSAAPAS...GLDAYRV*
1MRD...QLAEYY*
1MSD...AIEQYLI*
1MSD...VIEQYF*
1MSK...IIDQYRV*
1MSN...IVDQYRI*
1MSN...IVDQYRI*
1MSS...IIDQYRI*
1MSEINALDPQ*
1MSD...IIQYRI*
1MSD...IIQYRI*
1MSEINALDPQ*
1MSEVNALDPS*
1MSEINALDPQ*
1MSEINALDPS*
1MSEVNALDPS*
1MSEM...TDQYRV*
1MSTVD...PNQYLV*
1MTATDSSILN*
1MTTN...KEQYKV*
1MTN...KDQYKV*
1MTD...KDQYKI*
1MSID...AQYRV*
1MSID...AQYRV*
1MDN...KEQYKV*
1MTA...KDQYKV*
1MTN...KNQYTV*
1MTTD...KEQYKV*
1MTD...LSEYRV*
1MSDID...IEQYKI*
1MD...ISQYKI*
1MD...ISQYKI*
1MD...ISQYKI*
1MD...ISQYKI*
1MD...ISQYKI*
1MD...ITQYKI*
1MTD...TNQYLI*
.....d.....qy!...ep%Y..

AfQ2_|B7J5E4|
 AfQ1_|B7JA26|
 AfQ3_|B7JB17|
 Acidithiobacillus_ferrivorans_CbbQ3_|G0JL86|
 Ferrovum_sp_CbbQ3_|A0A0R0LVR7|
 Thioalkalivibrio_sp_CbbQ3_|D3SG06|
 Thioalkalivibrio_sulfidophilus_CbbQ3_|B8GQN5|
 Acidimicrobium_ferrooxidans_CbbQ3_|C7M202|
 Acidithiobacillus_caldus_CbbQ3_|F9ZLQ7|
 Halothiobacillus_neapolitanus_CbbQ3_|D0KZ75|
 Nitrobacter_hamburgensis_CbbQ3_|Q1QFQ9|
 Nitrosomonas_eutropha_CbbQ3_|Q0AHV0|
 Thiomonas_intermedia_CbbQ3_|D5X343|
 Acidithiobacillus_ferrivorans_CbbQ1_|G0JNK4|
 Nitrosomonas_communis_CbbQ1_|A0A0F7KI56|
 Nitrosomonas_europaea_CbbQ1_|Q82TG8|
 Nitrosomonas_sp_CbbQ1_|F9ZF97|
 Thiobacillus_denitrificans_CbbQ1_|Q3SFN1|
 Thiobacillus_thioparus_CbbQ1_|WP_018507362.1|
 Thiomonas_sp_CbbQ1_|A0A078BG06|
 Hydrogenovibrio_marinus_CbbQ1_|Q75W43|
 Sideroxydans_lithotrophicus_CbbQ1_|D5CQ61|
 Sulfuricella_denitrificans_CbbQ1_|S6AM60|
 Hydrogenovibrio_kuenenii_CbbQ1_|WP_024851894.1|
 Hydrogenovibrio_halophilus_CbbQ1_|WP_019895627.1|
 Thiomicrothabodus_chilensis_CbbQ1_|WP_028486808.1|
 Hydrogenovibrio_crunogenus_CbbQ1_|Q31IJ8|
 Thiomicrothabodus_sp_CbbQ1_|WP_029406683.1|
 Mariprofundus_ferrooxydans_CbbQ1_|Q0EX20|
 Thiothrix_nivea_CbbQ1_|I3BR11|
 Acidithiobacillus_ferrivorans_CbbQ2_|G0JPS2|
 Thiobacillus_denitrificans_CbbQ2_|Q3SFL6|
 Thiobacillus_thioparus_CbbQ2_|WP_018507397.1|
 Halothiobacillus_neapolitanus_CbbQ2_|D0KZQ7|
 Thiomonas_arsenitoxydans_CbbQ2_|D6CTS9|
 Thiomonas_intermedia_CbbQ2_|O84918|
 Sideroxydans_lithotrophicus_CbbQ2_|D5CTF9|
 Sulfuricella_denitrificans_CbbQ2_|S6ABC2|
 Gallionella_capsiferriformans_CbbQ2_|D9SHP7|
 Mariprofundus_ferrooxydans_CbbQ2_|Q0F2Z9|
 Acidithiobacillus_caldus_CbbQ2_|F9ZR06|
 Thiola pillus_brandeum_CbbQ2_|W0TLR2|
 Hydrogenovibrio_marinus_CbbQ2_|Q9WX11|
 Hydrogenovibrio_kuenenii_CbbQ2_|WP_024851802.1|
 Thiomicrothabodus_sp_CbbQ2_|WP_029406689.1|
 Hydrogenovibrio_halophilus_CbbQ2_|WP_019895619.1|
 Thiomicrothabodus_chilensis_CbbQ2_|WP_028486930.1|
 Hydrogenovibrio_crunogenus_CbbQ2_|Q31IK4|
 Thiothrix_nivea_CbbQ2_|I3BN91|
 consensus>70

23 QHD E V A L F E A A Y R K R L P V M V K G P T G C G K S R F V E F M A W R L G K P L V T V A C N E D M T A A D L V
 19 T G D E T E L Y E A A Y S V R M P M M L K G P T G C G K T R F V E Y M A Y K L G K P L I T V A C N E D M T A S D L V
 19 V A D E V D L Y E A A Y S V R M P M M L K G P T G C G K T R F V E Y M A W K L G K P L I T V A C N E D M T A S D L V
 19 V A D E V E L Y E A A Y S V R M P M M L K G P T G C G K T R F V E Y M A W K L G K P L I T V A C N E D M T A S D L V
 18 V A D E I D L C E S D G S A R I P M I L K R L T G Y G R S R F V E Y M G W K L Q G P L I I L T C N E D M I P S N P V
 19 V A D E I E L Y E A A Y S V R M P M M L K G P T G C G K T R F V E H M A W K L G K P L I T V A C N E D M T A S D L V
 19 V A D E I E L Y E A A Y S V R M P M M L K G P T G C G K T R F V E Y M A W K L N K P L I T V A C N E D M T A S D L V
 20 V A N E I A L F E A A Y S V R M P V M L K G P T G S G K T R F V E Y M A W R L G K P L I T V A C N E D M T A S D L V
 19 V D D E I A L Y E A A Y A A R L P M M V K G P T G C G K S R F V E Y M A W R L K R P L I S V A C N E D M T A S D L V
 19 V Q D E I E L Y R A G Y E A R I P M M L K G P T G C G K S R F V E H M A W K L N R P L I T I A C N E D M T A S D L V
 30 V A D E V G L Y E A A Y S V R M P M M L K G P T G C G K T R F V E Y I A W K L G K P L V T V A C N E D M T A S D L V
 30 T G D E V A L Y E A A Y S V R M P M M L K G P T G C G K T R F V E Y M A W K L G K P L I T V A C N E D M T A S D L V
 57 L S D E V E L Y E A A Y S V R L P V M L K G P T G C G K T R F V E Y M A W K L G K P L I T V A C N E D M T A S D L V
 19 T G D E I E L Y E A A Y G V R M P M M L K G P T G C G K T R F V E Y M A Y K L G R P L I T V A C N E D M T A S D L V
 19 V A D E V T L Y E A A Y S A R I P M M L K G P T G C G K T R F V E F M A W K L K K P L I T V A C N E D M T A S D L V
 19 V A D E V K L Y E A A Y S V R M P M M L K G P T G C G K T R F V E Y M A W K L G K P L I T V A C N E D M T A S D L V
 19 V A D E V E L Y E A A Y S V R M P M M L K G P T G C G K T R F V E H M A W K L K K P L I T V A C N E D M T A S D L V
 19 V G D E V E L Y E A A Y S V R M P M M L K G P T G C G K T R F V E Y M A Y K L N K P L I T V A C N E D M T A S D L V
 19 V A D E V E L Y E A A Y S V R M P M M L K G P T G C G K T R F V E Y M A Y K L G K P L I T V A C N E D M T A S D L V
 19 V A D E V E L F E A A Y S V R M P M M L K G P T G C G K T R F V E S M A Y K L N K P L I T V A C N E D M T A S D L V
 23 V S D E V E L F E S A Y A S R M P M M L K G P T G C G K S R F V E Y M A H K L N K P L I T V A C N E D M T A S D L V
 19 V S D E V E L F E A A Y T V R M P M M L K G P T G C G K T R F V E H M A W K L G R P L I T V A C N E D M T A S D L V
 19 V S D E V E L Y E A A Y S V R M P M M L K G P T G C G K T R F V E H M A W K L G K P L I T V A C N E D M T A S D L V
 23 V S D E V E L F E S A Y A S R M P M M L K G P T G C G K S R F V E Y M A H K L N K P L I T V A C N E D M T A S D L V
 23 V A D E V D V F E S A Y A T R M P I M L K G P T G C G K S R F V E Y M A H K L N K P L I T V S C N E D M T A S D L V
 23 V A D E I E L F E S G Y D S R M P I M L K G P T G C G K S R F V E Y M A W K L G K P L I T V A C N E D M T A S D L V
 23 V A D E V E L F E S G Y A S R M P I M L K G P T G C G K S R F V E Y M A H K L G K P L I T V A C N E D M T A S D L V
 23 V A D E I E L F E S G Y A S R M P I M L K G P T G C G K S R F V E Y M A H K L N K P L I T V A C N E D M T A S D L V
 20 V G N E V E L Y N A A Y A R R L P M M L K G P T G C G K S R F V E Y M A W K L G R P L I T V A C N E D M T A S D L V
 20 V G D E I T L Y E A A Y N A R M P M M L K G P T G C G K S R F V E H M A W R L K R P L I T V A C N E D M T A S D L V
 23 Q H D E V A L F E A A Y R K R L P V M V K G P T G C G K S R F V E Y M A W R L G K A L I T V A C N E D M T A A D L V
 19 Q G R E V Q L Y E A A Y R N R L P V M V K G P T G C G K S R F V E Y M A W K L N K P L I T V A C N E D M T A S D L V
 18 Q K N E V A L Y E A A Y R N R L P V M V K G P T G C G K S R F V E Y M A W K L G K P L I T V A C N E D M T A S D L V
 18 Q T N E V A L Y E A A Y R N R L P V M V K G P T G C G K S R F V E Y M A W K L G K P L I T V A C N E D M T A S D L V
 19 Q G R E V A L F E A A Y R N R L P V M V K G P T G C G K S R F V E H M A W R L G K P L I T V A C N E D M T A S D L V
 19 Q G R E V A L F E A A Y R N R L P V M V K G P T G C G K S R F V E H M A W R L G K P L I T V A C N E D M T A S D L V
 18 Q R N E V A L Y E A A Y A A R L P V M V K G P T G C G K S R F V E Y M A W K L N K P L I T V A C N E D M T A S D L V
 18 Q T N E V T L Y E A A Y A A R L P V M I K G P T G C G K S R F V E Y M A W K L N K P L I T V A C N E D M T A S D L V
 18 Q G D E V A L Y E A A Y A A R L P V M V K G P T G C G K S R F V E Y M A W K L N K P L I T V A C N E D M T A S D L V
 19 Q G D E V E L Y E A A Y A A R L P V M V K G P T G C G K S R F V E H M A W K L G K P L I T V A C N E D M T A S D L V
 18 Q G H E V S L F A A A Y R R R L P V M V K G P T G C G K S R F V E Y M A W Q L D R P L I T V A C N E D M S A S D L V
 20 S G N E I D Q Y E A A Y A A R L P V M V K G P T G C G K S R F V E Y M A W K L G K P I I T V A C N E D M T A A D L V
 17 Q S N E V E L Y E A A Y A Q R L P V M V K G P T G C G K S R F V E H M A W K L G K P I I T V S C N E D I T A S D L V
 17 Q S N E I E L Y E A A Y D A R L P V M V K G P T G C G K S R F V E H M A W K L G K P I I T V S C N E D I T A S D L V
 17 Q S N E V E L Y E A A Y A A R L P V M V K G P T G C G K S R F V E H M A W K L G K P I I T V S C N E D I T A A D L V
 17 Q S N E I E L Y E A A Y A A R L P V M V K G P T G C G K S R F V E H M A W K L D K P I I T V S C N E D I T A A D L V
 17 Q S N E I E L Y E A A Y G A R L P V M V K G P T G C G K S R F V E H M A W K L G K P I I T V S C N E D I T A S D L V
 17 Q S N E I E L Y E A A Y S A R L P V M V K G P T G C G K S R F V E H M A W K L G K P I I T V S C N E D I T A S D L V
 18 T G R E V E L Y Q A A Y R S R L P V M V K G P T G C G K S R F V E H M A W K L G K P L I T V A C N E D M T A S D L V
 . . d E v . l y e a a y . . R l P . m . K g p T G c G k . R F ! E . m a w k L . k p l ! t v a C N E D m t a s # 1 v

AfQ2_|B7J5E4|
 AfQ1_|B7JA26|
 AfQ3|B7JB17|
 Acidithiobacillus_ferrivorans_CbbQ3_|G0JL86|
 Ferrovum_sp_CbbQ3_|A0A0R0LVR7|
 Thioalkalivibrio_sp_CbbQ3_|D3SG06|
 Thioalkalivibrio_sulfidophilus_CbbQ3_|B8GQN5|
 Acidimicrobium_ferrooxidans_CbbQ3_|C7M202|
 Acidithiobacillus_caldus_CbbQ3_|F9ZLQ7|
 Halothiobacillus_neapolitanus_CbbQ3_|D0KZ75|
 Nitrobacter_hamburgensis_CbbQ3_|Q1QFQ9|
 Nitrosomonas_eutropha_CbbQ3_|Q0AHV0|
 Thiomonas_intermedia_CbbQ3_|D5X343|
 Acidithiobacillus_ferrivorans_CbbQ1_|G0JNK4|
 Nitrosomonas_communis_CbbQ1_|A0A0F7KI56|
 Nitrosomonas_europaea_CbbQ1_|Q82TG8|
 Nitrosomonas_sp_CbbQ1_|F9ZF97|
 Thiobacillus_denitrificans_CbbQ1_|Q3SFN1|
 Thiobacillus_thioparus_CbbQ1_|WP_018507362.1|
 Thiomonas_sp_CbbQ1_|A0A078BG06|
 Hydrogenovibrio_marinus_CbbQ1_|Q75W43|
 Sideroxydans_lithotrophicus_CbbQ1_|D5CQ61|
 Sulfuricella_denitrificans_CbbQ1_|S6AM60|
 Hydrogenovibrio_kuenenii_CbbQ1_|WP_024851894.1|
 Hydrogenovibrio_halophilus_CbbQ1_|WP_019895627.1|
 Thiomicrobacter_chilensis_CbbQ1_|WP_028486808.1|
 Thiomicrobacter_crunogenus_CbbQ1_|Q31IJ8|
 Thiomicrobacter_sp_CbbQ1_|WP_029406683.1|
 Mariprofundus_ferrooxydans_CbbQ1_|Q0EX20|
 Thiothrix_nivea_CbbQ1_|I3BR11|
 Acidithiobacillus_ferrivorans_CbbQ2_|G0JPS2|
 Thiobacillus_denitrificans_CbbQ2_|Q3SFL6|
 Thiobacillus_thioparus_CbbQ2_|WP_018507397.1|
 Halothiobacillus_neapolitanus_CbbQ2_|D0KZQ7|
 Thiomonas_arsenitoxydans_CbbQ2_|D6CTS9|
 Thiomonas_intermedia_CbbQ2_|Q84918|
 Sideroxydans_lithotrophicus_CbbQ2_|D5CTF9|
 Sulfuricella_denitrificans_CbbQ2_|S6ABC2|
 Gallionella_capsiferriiformans_CbbQ2_|D9SHP7|
 Mariprofundus_ferrooxydans_CbbQ2_|Q0F2Z9|
 Acidithiobacillus_caldus_CbbQ2_|F9ZR06|
 Thiopapillus_brandeam_CbbQ2_|W0TLR2|
 Hydrogenovibrio_marinus_CbbQ2_|Q9WXI1|
 Hydrogenovibrio_kuenenii_CbbQ2_|WP_024851802.1|
 Thiomicrobacter_sp_CbbQ2_|WP_029406689.1|
 Hydrogenovibrio_halophilus_CbbQ2_|WP_019895619.1|
 Thiomicrobacter_chilensis_CbbQ2_|WP_028486930.1|
 Hydrogenovibrio_crunogenus_CbbQ2_|Q31IK4|
 Thiothrix_nivea_CbbQ2_|I3BN91|
 consensus>70

81 GRWLLDKDGTWQDGP LTV AARY GAICYLDEIVEARQDTTVV IHPLTDHRR TLPLDKK
 77 GRFLLDPPQGTWQDGP LTLAARL GAICYLDEVVEARQDTTVV IHPLTDNRRV LBLEKK
 77 GRFLLDASGTWQDGP LAIAARF GAICYLDEVVEARQDTTVV IHPLTDARRV LBLEKK
 77 GRFLLDASGTWQDGP LAIAARF GAICYLDEVVEARQDTTVV IHPLTDARRV LBLEKK
 76 GYLFDTSLSLTRFQDGP LTVVARL GEIFYRDEVAEASONTTFV IHPLINVRRL LBLEKK
 77 GRYLDDADGTWQDGP LTLAARH GAICYLDEIVEARQDTTVV IHPLTDNRRV LBLEKK
 77 GRFLLDANGTRWQDGP LAMAARH GAICYLDEVVEARQDTTVV IHPLTDNRRV LBLEKK
 78 GRFLLDAGGTWQDGP LTLAARY GALCYLDEIVEARQDTTVV IHPLTDARRV LBLEKK
 77 GRFLLDKDGTKWQDGP LTLAARI GAICYLDEVVEARQDTTVV IHPLTDHRR TLPLDKK
 77 GRYLDDKDGTRWQDGP LTV AARI GAICYLDEIVEARQDTTVV IHPLTDHRRV LBLEKK
 88 GRFLLDASGTWQDGP LTLAARF GGICYLDEVVEARQDTTVV IHSLTDARR TLBLEKK
 88 GRFLLDTSGTWQDGP LALAARY GAICYLDEVVEARQDTTVV IHPLTDTRR TLBLEKK
 115 GRYLDDANGTRWQDGP LAMAARL GAICYLDEVVEARQDTTVV IHPLTDARR TLBLEKK
 77 GRFLLDPPQGTWQDGP LTLAARL GAICYLDEVVEARQDTTVV IHPLTDNRR TLBLEKK
 77 GRFLLDAGGTWQDGP LTLAARY GAICYLDEVVEARQDTTVV IHPLTDNRR TLBLEKK
 77 GRFLLDAGGTWQDGP LTLAARY GAICYLDEVVEARQDTTVV IHPLTDNRR TLBLEKK
 77 GRFLLDANGTRWQDGP LTV AARY GAICYLDEVVEARQDTTVV IHPLTDNRRV LBLEKK
 77 GRFLLSAAGTEWQDGP LAIAARH GAICYLDEVVEARQDTTVV IHPLTDNRRV LBLEKK
 77 GRFLLSASGTWQDGP LAIAARH GAICYLDEVVEARQDTTVV IHPLTDNRRV LBLEKK
 77 GRFLLSATGTWQDGP LAIAARH GAICYLDEVVEARQDTTVV IHPLTDNRRV LBLEKK
 81 GRFLIDINGTKWQDGP LTV AARI GAICYLDEVVEARQDTTVV IHPLTDHRRV LBLEKK
 77 GRFLLDANGTRWQDGP LAIAARF GAICYLDEVVEARQDTTVV IHPLTDNRRV LBLEKK
 77 GRFLLDAGGTWQDGP LALGARF GAIVYLDEVVEARQDTTVV IHPLTDNRRV LBLEKK
 81 GRFLIDINGTKWQDGP LTV AARI GAICYLDEVVEARQDTTVV IHPLTDHRRV LBLEKK
 81 GRYLIDVNGTRWQDGP LTV AARI GAICYLDEVVEARQDTTVV IHPLTDHRRV LBLEKK
 81 GRYLIDINGTRWQDGP LTV AARI GAICYLDEVVEARQDTTVV IHPLTDHRRV LBLEKK
 81 GRFLIDINGTRWQDGP LTV AARI GAICYLDEVVEARQDTTVV IHPLTDHRRV LBLEKK
 81 GRFLIDINGTRWQDGP LTV AARI GAICYLDEVVEARQDTTVV IHPLTDHRR TLPLDKK
 78 GRFLLDKDGTKWQDGP LTLAARI GAICYLDEVVEARQDTTVV IHPLTDHRR TLPLDKK
 78 GRFLLDKDGTKWQDGP LTLAARI GAICYLDEVVEARQDTTVV IHPLTDHRR TLPLDKK
 81 GRWLLDKDGTWQDGP LTV AARY GAICYLDEIVEARQDTTVV IHPLTDHRR TLPLDKK
 77 GRYLLEANGTRWLDGP LTLAARI GAICYLDEVVEARQDTTVV IHPLTDHRR TLPLDKK
 76 GRYLLEANGTRWLDGP LTLAARI GAICYLDEIVEARQDTTVV IHPLTDHRR TLPLDKK
 76 GRYLLEANGTRWLDGP LTLAARI GAICYLDEIVEARQDTTVV IHPLTDHRR TLPLDKK
 77 GRYLLEGGGTWLDGP LTLAARI GAICYLDEIVEARQDTTVV IHPLTDHRR TLPLDKK
 77 GRYLLEGGGTWLDGP LTLAARI GAICYLDEIVEARQDTTVV IHPLTDHRR TLPLDKK
 76 GRYLDDAGGTWLDGP LTLAARI GAICYLDEVVEARQDTTVV IHPLTDHRR TLPLDKK
 76 GRYLLEANGTRWLDGP LTLAARI GAICYLDEIVEARQDTTVV IHPLTDHRR TLPLDKK
 76 GRYLDDANGTRWLDGP LTLAARI GAICYLDEIVEARQDTTVV IHPLTDHRR TLPLDKK
 76 GRYLDDAGGTWLDGP LTLAARI GAICYLDEVVEARQDTTVV IHSLTDHRR TLPLDKK
 78 GRYLDDADGTWLDGP LTLAARI GAICYLDEVVEARQDTTVV IHPLTDHRR TLPLDKK
 75 GRYLDDANGTRWLDGP LTLAARY GAICYLDEIVEARQDTTVV IHPLTDHRR TLPLDKK
 75 GRYLDDANGTRWLDGP LTLAARY GAICYLDEIVEARQDTTVV IHPLTDHRR TLPLDKK
 75 GRYLDDANGTRWLDGP LTLAARY GAICYLDEIVEARQDTTVV IHPLTDHRR TLPLDKK
 75 GRYLDDANGTRWLDGP LTLAARY GAICYLDEIVEARQDTTVV IHPLTDHRR TLPLDKK
 75 GRYLDDANGTRWLDGP LTLAARY GAICYLDEIVEARQDTTVV IHPLTDHRR TLPLDKK
 76 GRYLLEANGTRWLDGP LTLAARI GAICYLDEIVEARQDTTVV IHPLTDHRR TLPLDKK
 Grflld.ngTrw.DGpLtv.aAR.GaicylDE!vEARQ#TtvvIHplT#.RR.LpIdkK

AfQ2_|B7J5E4|
 AfQ1_|B7JA26|
 AfQ3_|B7JB17|
 Acidithiobacillus_ferriivorans_CbbQ3_|G0JL86|
 Ferrovum_sp_CbbQ3_|A0A0R0LVR7|
 Thioalkalivibrio_sp_CbbQ3_|D3SG06|
 Thioalkalivibrio_sulfidiphilus_CbbQ3_|B8GQN5|
 Acidimicrobium_ferrooxidans_CbbQ3_|C7M202|
 Acidithiobacillus_caldus_CbbQ3_|F9ZLQ7|
 Halothiobacillus_neapolitanus_CbbQ3_|D0KZ75|
 Nitrobacter_hamburgensis_CbbQ3_|Q1QFQ9|
 Nitrosomonas_eutropha_CbbQ3_|Q0AHV0|
 Thiomonas_intermedia_CbbQ3_|D5X343|
 Acidithiobacillus_ferriivorans_CbbQ1_|G0JNK4|
 Nitrosomonas_communis_CbbQ1_|A0A0F7KI56|
 Nitrosomonas_europaea_CbbQ1_|Q82TG8|
 Nitrosomonas_sp_CbbQ1_|F9ZF97|
 Thiobacillus_denitrificans_CbbQ1_|Q3SFN1|
 Thiobacillus_thioparus_CbbQ1_|WP_018507362.1|
 Thiomonas_sp_CbbQ1_|A0A078BG06|
 Hydrogenovibrio_marinus_CbbQ1_|Q75W43|
 Sideroxydans_lithotrophicus_CbbQ1_|D5CQ61|
 Sulfuricella_denitrificans_CbbQ1_|S6AM60|
 Hydrogenovibrio_kuenenii_CbbQ1_|WP_024851894.1|
 Hydrogenovibrio_halophilus_CbbQ1_|WP_019895627.1|
 Thiomicrobacter_chilensis_CbbQ1_|WP_028486808.1|
 Hydrogenovibrio_crunogenus_CbbQ1_|Q31IJ8|
 Thiomicrobacter_sp_CbbQ1_|WP_029406683.1|
 Mariprofundus_ferrooxydans_CbbQ1_|Q0EX20|
 Thiothrix_nivea_CbbQ1_|I3BR11|
 Acidithiobacillus_ferriivorans_CbbQ2_|G0JPS2|
 Thiobacillus_denitrificans_CbbQ2_|Q3SFL6|
 Thiobacillus_thioparus_CbbQ2_|WP_018507397.1|
 Halothiobacillus_neapolitanus_CbbQ2_|D0KZQ7|
 Thiomonas_arsenitoxydans_CbbQ2_|D6CTS9|
 Thiomonas_intermedia_CbbQ2_|O84918|
 Sideroxydans_lithotrophicus_CbbQ2_|D5CTF9|
 Sulfuricella_denitrificans_CbbQ2_|S6ABC2|
 Gallionella_capsiferriiformans_CbbQ2_|D9SHP7|
 Mariprofundus_ferrooxydans_CbbQ2_|Q0F2Z9|
 Acidithiobacillus_caldus_CbbQ2_|F9ZR06|
 Thiolaipillus_brandeum_CbbQ2_|W0TLR2|
 Hydrogenovibrio_marinus_CbbQ2_|Q9WXI1|
 Hydrogenovibrio_kuenenii_CbbQ2_|WP_024851802.1|
 Thiomicrobacter_sp_CbbQ2_|WP_029406689.1|
 Hydrogenovibrio_halophilus_CbbQ2_|WP_019895619.1|
 Thiomicrobacter_chilensis_CbbQ2_|WP_028486930.1|
 Hydrogenovibrio_crunogenus_CbbQ2_|Q31IK4|
 Thiothrix_nivea_CbbQ2_|I3BN91|
 consensus>70

139 GELI R AHPDFQLVISYNPGYQSLMKDLKQSTKQRF TGF EFD YPNAE LEAG ILVQETGV
 135 GELV Q AHPDFQLVISYNPGYQSLMKDLKQSTKQRF GAL DFN YPDHE IEAE IVSHETGV
 135 GELV Q AHPDFQLVISYNPGYQSLMKDLKQSTKQRF GAL DFN YPAHD IETE IVAHETGV
 135 GELV Q AHPDFQMVISYNPGYQSLMKDLKQSTKQRF GAL DFN YPAHA IETE IVAHETS
 134 GKLI H AHPDFQLVISYNPGYQSLMKDLKQSTKQRF CPL DFN YSSND IEVE TVEYDTKV
 135 GEMV Q AHPDFQVVISYNPGYQSLMKDLKQSTKQRF GAL DFN YPDHE IEAE IVAHETGV
 135 GELI Q AHPDFQVVISYNPGYQSLMKDLKQSTKQRF GAL DFN YPAHD IETE IVAHETGV
 136 GELV Q AHPDFQVVISYNPGYQSLMKDLKQSTKQRF GAL DFN YPAHA IETE IVAHETGV
 135 GELV H AHPDFQLVISYNPGYQSLMKDLKQSTKQRF GAL DFN YPDHE IEAE IVAHETGV
 135 GELV E AHPDFQLVISYNPGYQSLMKDLKQSTKQRF GAL DFN YPKPD IEAE IVSHETGV
 146 GEIV E AHPDFQVVISYNPGYQSLMKDLKQSTKQRF GAL DFN YPEHE IEAE IVVHETGV
 146 GEII E AHPDFQVVISYNPGYQSLMKDLKQSTKQRF GAL DFN YPEHE IEAE IVAHETGV
 173 GEII E AHPDFQVVISYNPGYQSLMKDLKQSTKQRF AAL DFN YPEHE IEAE IVAHETGV
 135 GELV H AHPDFQLVISYNPGYQSLMKDLKQSTKQRF GAL DFN YPEHA IETE IVSHETGV
 135 GELI N AHPDFQLVISYNPGYQSLMKDLKQSTKQRF GAL DFN YPDHD IETE IVSHETGV
 135 GELV G AHPDFQLVISYNPGYQSLMKDLKQSTKQRF GAL DFN YPDHD IETE IVSHETGV
 135 GELV H AHPDFQLVISYNPGYQSLMKDLKQSTKQRF GAL DFN YPSHD IETE IVAHETGV
 135 GELV N AHPDFQLVISYNPGYQSLMKDLKQSTKQRF GGL DFN YPEHA IETE IVAHETGV
 135 GELV H AHPDFQLVISYNPGYQSLMKDLKQSTKQRF GGL DFN YPEHA IETE IVTHETGV
 135 GELI H AHPDFQLVISYNPGYQSLMKDLKQSTKQRF GGL DFN YPEHA IETE IVVHETGV
 135 GELV E AHPDFQLVISYNPGYQSLMKDLKQSTKQRF GGF DFN YPAEQ IEAE IVAKEADV
 135 GELV H AHPDFQLVISYNPGYQSLMKDLKQSTKQRF GGL DFN YPEHD IETE IVVHETGV
 135 GELV H AHPDFQLVISYNPGYQSLMKDLKQSTKQRF GGL DFN YPDHA IETE IVAHETGV
 139 GELV E AHPDFQLVISYNPGYQSLMKDLKQSTKQRF GGF DFN YPAEQ IEAE IVAHETGV
 139 GELV E AHPDFQLVISYNPGYQSLMKDLKQSTKQRF GGF DFN YPEEA IEAE IVSKEAGI
 139 GELV E AHPDFQLVISYNPGYQSLMKDLKQSTKQRF GGF DFN YPEES IEAE IVAKEAGI
 139 GELV E AHPDFQLVISYNPGYQSLMKDLKQSTKQRF GGF DFN YPEES IEAE IVAKEAGI
 139 GELV E AHPDFQLVISYNPGYQSLMKDLKQSTKQRF GGF DFN YPEES IEAE IVAKEAGI
 136 GELI E AHPDFQLVISYNPGYQSLMKDLKQSTKQRF GAL DFN YPAET IEAE IVAKEAGI
 136 GELI E AHPDFQLVISYNPGYQSLMKDLKQSTKQRF GAL DFN YPAEN IEAE IVSKEAGI
 139 GELI S AHPDFQLVISYNPGYQSLMKDLKQSTKQRF TGF EFD YPNAE LESE ILVQETGV
 135 GELI E AHPDFQLVISYNPGYQSLMKDLKQSTKQRF AAF DFN YPDAE LETT ILARETGL
 134 GELI H AHPDFQLVISYNPGYQSLMKDLKQSTKQRF TAF DFN YPDAE LEAQ IVAKETGM
 134 GELI H AHPDFQLVISYNPGYQSLMKDLKQSTKQRF TAF DFN YPDAE LEAQ IVAKETGL
 135 GELI Q AHPDFQLVISYNPGYQSLMKDLKQSTKQRF TAF DFN YPEAG LETT ILCKESGL
 135 GELI Q AHPDFQLVISYNPGYQSLMKDLKQSTKQRF TAF DFN YPEAG LETT ILCKESGL
 134 GELI K AHPDFQLVISYNPGYQSLMKDLKQSTKQRF TAF EFD YPEASV EAGILSKEADI
 134 GELI E AHPDFQMVISYNPGYQSLMKDLKQSTKQRF TAF EFD YPDAQ LETS IVAKETGT
 134 GELI E AHPDFQLVISYNPGYQSLMKDLKQSTKQRF TGF EFD YPDAE LESE ILAKETGL
 135 GELI E AHPDFQLVISYNPGYQSLMKDLKQSTKQRF AAF EFD YPLPEV ETS IVAKETGI
 134 GELV R AHPDFQLVISYNPGYQSLMKDLKQSTKQRF AAF DFN YPPEL ETE ILVREAGV
 136 GEV V K AHPDFQLVISYNPGYQSLMKDLKQSTKQRF VGM DFN YPAD LEET IVAETGI
 133 GELI K AHPDFQLVISYNPGYQSLMKDLKQSTKQRF CAL DFN YAAPDVE AH ILQKEGQV
 133 GELI K AHPDFQLVISYNPGYQSLMKDLKQSTKQRF CAL DFN YAAPDVE AH ILQKEGQV
 133 GELI K AHPDFQLVISYNPGYQSLMKDLKQSTKQRF CGL DFN YADAET EAH ILQKEGQV
 133 GELI K AHPDFQLVISYNPGYQSLMKDLKQSTKQRF CAM DFN YASAE EAH ILQKEGQV
 133 GELI K AHPDFQLVISYNPGYQSLMKDLKQSTKQRF CAL DFN YAAPDVE AH ILQKEGQV
 133 GELI K AHPDFQLVISYNPGYQSLMKDLKQSTKQRF CAL DFN YALPEV EAH ILQKEGQV
 134 GELV Q AHPDFQLVISYNPGYQSLMKDLKQSTKQRF VAF DFN YPTEK EES ILQAEETGV

AfQ2_|B7J5E4|
 AfQ1_|B7JA26|
 AfQ3_|B7JB17|
 Acidithiobacillus_ferrovorum_CbbQ3_|G0JL86|
 Ferrovum_sp_CbbQ3_|A0A0R0LVR7|
 Thioalkalivibrio_sp_CbbQ3_|D3SG06|
 Thioalkalivibrio_sulfidophilus_CbbQ3_|B8GQ5|
 Acidimicrobium_ferrooxidans_CbbQ3_|C7M202|
 Acidithiobacillus_caldus_CbbQ3_|F9ZLQ7|
 Halothiobacillus_neapolitanus_CbbQ3_|D0KZ75|
 Nitrobacter_hamburgensis_CbbQ3_|Q1QFQ9|
 Nitrosomonas_eutropha_CbbQ3_|Q0AHV0|
 Thiomonas_intermedia_CbbQ3_|D5X343|
 Acidithiobacillus_ferrovorum_CbbQ1_|G0JNK4|
 Nitrosomonas_communis_CbbQ1_|A0A0F7KI56|
 Nitrosomonas_europaea_CbbQ1_|Q82TGS|
 Nitrosomonas_sp_CbbQ1_|F9ZF97|
 Thiobacillus_denitrificans_CbbQ1_|Q3SFN1|
 Thiobacillus_thioparus_CbbQ1_|WP_018507362.1|
 Thiomonas_sp_CbbQ1_|A0A078BG06|
 Hydrogenovibrio_marinus_CbbQ1_|Q75W43|
 Sideroxydans_lithotrophicus_CbbQ1_|D5CQ61|
 Sulfuricella_denitrificans_CbbQ1_|S6AM60|
 Hydrogenovibrio_kuenenii_CbbQ1_|WP_024851894.1|
 Hydrogenovibrio_halophilus_CbbQ1_|WP_019895627.1|
 Thiomicrobacter_chilensis_CbbQ1_|WP_028486808.1|
 Hydrogenovibrio_crunogenus_CbbQ1_|Q31IJ8|
 Thiomicrobacter_sp_CbbQ1_|WP_029406683.1|
 Mariprofundus_ferrooxydans_CbbQ1_|Q0EX20|
 Thiothrix_nivea_CbbQ1_|I3BR11|
 Acidithiobacillus_ferrovorum_CbbQ2_|G0JPS2|
 Thiobacillus_denitrificans_CbbQ2_|Q3SFL6|
 Thiobacillus_thioparus_CbbQ2_|WP_018507397.1|
 Halothiobacillus_neapolitanus_CbbQ2_|D0KZQ7|
 Thiomonas_arsenitoxydans_CbbQ2_|D6CTS9|
 Thiomonas_intermedia_CbbQ2_|Q84918|
 Sideroxydans_lithotrophicus_CbbQ2_|D5CTF9|
 Sulfuricella_denitrificans_CbbQ2_|S6ABC2|
 Gallionella_capsiferriiformans_CbbQ2_|D9SHP7|
 Mariprofundus_ferrooxydans_CbbQ2_|Q0F2Z9|
 Acidithiobacillus_caldus_CbbQ2_|F9ZR06|
 Thiopalladium_brandeum_CbbQ2_|W0TLR2|
 Hydrogenovibrio_marinus_CbbQ2_|Q9WXI1|
 Hydrogenovibrio_kuenenii_CbbQ2_|WP_024851802.1|
 Thiomicrobacter_sp_CbbQ2_|WP_029406689.1|
 Hydrogenovibrio_halophilus_CbbQ2_|WP_019895619.1|
 Thiomicrobacter_chilensis_CbbQ2_|WP_028486930.1|
 Hydrogenovibrio_crunogenus_CbbQ2_|Q31IK4|
 Thiothrix_nivea_CbbQ2_|I3BN91|
 consensus>70

197 APSIAAQLVTVAAATARRLKGHGLDEGISTRLLVYAAAMLMDDGVAPRAACRMALVQPIT
 193 NKEVADKLVSIAERARNLKGHGLDEGISTRLLVYAGSLINKGVEAKAACRVTLVRPIT
 193 SVEIAGKLVSIAERARNLKGHGLDEGISTRLIYAGSLITKGVAPLAACHMTLVRPIT
 193 GVEIAGKLVSIAERARNLKGHGLDEGISTRLIYAGSLITQGVAPLAACRMTLVRPIT
 192 SLNATNKPISFTKHACKLKGHGLEKGISIRMLIYADKLIASDINLLFLDVTLVQPID
 193 NEEMAAGKLVSIAERARNLKGHGLDEGISTRLLVYAAASLIAGKVEPLGACRMALVVPIT
 193 SLDVASKLVSIAERARNLKGHGLDEGISTRLLVYAGSLIHKGVAPLAACRMALVVPIT
 194 GHDDASRLVSIAERARNLKGHGLDEGISTRLIYAAASLIAGKVDRIAACEMALVSPIT
 193 STDVAGKLVSIAERSRNLKGHGLDEGLSTRMLIYAGALIKQGIAPITACSVLVVPIT
 193 DKDTAEKLVIQAQRARNLKGHGLDEGLSTRLLVYAGKLIAGVDARAACMTLVNVPIT
 204 ALDAAKKLVIHAERARNLKGHGLDEGLSTRMLIYAGSLISRGVAPLAACRVALVVPIT
 204 SMDVAKNLVHIAERARNLKGHGLDEGLSTRMLIYAGRLIAQGVAPLAACRVALVVPIT
 231 GLEVAKNLVHIAERSRNLKGHGLDEGLSTRMLIYAGSLIAQGVAPLACRMALVVPIT
 193 GKDVAGKLVSIAERARNLKGHGLDEGLSTRMLVYAGSLISKGVEAKAACRVTLVVPIT
 193 DKAISTKLVSIAERARNLKGHGLDEGLSTRMLIYAGNLIAGKVEAHAAACRVALVVPIT
 193 DSAIADKLVSIAERARNLKGHGLDEGLSTRMLIYAGNLIAGKVDVHAACRVALVVPIT
 193 STEIAEKLVSIAERARNLKGHGLDEGLSTRMLIYAGSLIVKKVDAHAACRVALVVPIT
 193 DADTAGKLVSIAERSRNLKGHGLDEGLSTRMLVYAGSLIAQGVAPLAACRVALVVPIT
 193 NADVAGKLVSIAERSRNLKGHGLDEGLSTRMLIYAGSLIAKGVNPLAACRVALVVPIT
 193 PKDVAEKLVSIAERSRNLKGHGLDEGLSTRMLIYAGSLIAKGVDAKAAACRVALVVPIT
 197 EMSIAEKLVIQAQRARNLKGHGLDEGLSTRLLVYAGQLIKGVSPNKAACRMALITPLT
 193 KADVAGKLVSIAERARNLKGHGLDEGLSTRMLVYAGSLIATGVDPVKAACRVALVVPIT
 193 NAEVAGKLVSIAERARNLKGHGLDEGLSTRMLIYAGSLIATGVDPVKAACRVALVVPIT
 197 DMSIAEKLVIQAQRARNLKGHGLDEGLSTRLLVYAGQLIKGVSPNKAACRMALITPLT
 197 DKATAEKLVIQAQRARNLKGHGLDEGLSTRLLVYAAQLINKGVPAKQACTVALITPLT
 197 DLATAEKLVIQAQRARNLKGHGLDEGLSTRLLVYAAQLINKGVDPKKAACRMALITPLT
 197 DKETAEKLVIQAQRARNLKGHGLDEGLSTRLLVYAAQLINKGVDPKKAACRMALITPLT
 197 DKAIAEKLVIQAQRARNLKGHGLDEGLSTRLLVYAAQLINKGVDPKKAACRMALITPLT
 194 ESEIAVKLVIQAQRARNLKGHGLDEGLSTRLLVYAGNMIAGVEPMAACMMTLVTPLT
 194 DQKTAGKLVIQAHRARNLKGHGLDEGLSTRLLVYAGQLIKGISPQAACMTMTVTPLT
 197 ASTIAAQLVTVAAATARRLKGHGLDEGISTRLLVYAAALLIDDGVAPRSACRMALVQPIT
 193 DETTAGRLVKIGGVARNLKGHGLDEGISTRLLVYAAATLMKDGVDAGDACRMALVVPIT
 192 DAAVAAQLVKIGSVARNLKGHGLDEGISTRLLVYAAATLIKDGVEPRDACRMALVVPIT
 192 DADTVAKLVKIGNVARNLKGHGLDEGISTRLLVYAAATLIKDVSPQDACRMALVVPIT
 193 DAETAALKLVKIGHAARNLKGHGLDEGISTRLMVYAAQLIRDGVDAGDACRMALVVPIT
 193 DAETAALKLVKIGHAARNLKGHGLDEGISTRLMVYAAQLIRDGVDAGDACRMALVVPIT
 192 DPGLAAGLVKIGQAARNLKGHGLDEGISTRLLVYAAATLIKGVVEPRDACRMALVVPIT
 192 DEALAALKLVKIGTTARNLKGHGLDEGISTRLLVYAAATLIKGVVEPRDACRMALVVPIT
 192 DIEVCGKLVIQAQRARNLKGHGLDEGISTRLLVYAAAILIKNGVDAKAAACRMALVVPIT
 193 DTELAALKLVITIGGAARNLKGHGLDEGISTRLIYAAALIKGVVEAKAACRMALVVPIT
 192 PAHVAAGLVKIGGVARNLKGHGLDEGISTRLLVYAGVLIQDGVTPQDACRMALVVPIT
 194 DAGTAARLVQVGEAARNLKGHGLDEGISTRLLVYAAASLISKGISPRDAARMAMVVPIT
 191 DADVAKKLVIQAQRARNLKGHGLDEGISTRLMVYAAAMLIQGITPVEACKMALVVPIT
 191 DEATAKKLVQIGEAARNLKGHGLDEGISTRLMVYAAATLINKGITPVEACKMALVVPIT
 191 DAETAALKLVQIGETARNLKGHGLDEGISTRLMVYAAATLINKGITPVEACKMALVVPIT
 191 DADTAALKLVQIGETARNLKGHGLDEGISTRLMVYAAAVLINQGITPVEACKMALVVPIT
 191 DADTAALKLVQIGETARNLKGHGLDEGISTRLMVYAAATLINQGITPVEACKMALVVPIT
 191 DSDTAQKLVIQAQRARNLKGHGLDEGISTRLMVYAAATLINQGITPVEACKMALVVPIT
 192 GADIAADKLVIQAHTARNLKGHGLDEGISTRLLVYAAANLIKDGIEAGDACRMALVSPIT
a.klv.i.e.arnLKGHGL#eGiStr\$\$vYA..\$.g!...ac.ma\$!.Pit


```

AfQ2_|B7J5E4|
AfQ1_|B7JA26|
AfQ3|B7JB17|
Acidithiobacillus_ferrivorans_CbbQ3_|G0JL86|
Ferrovum_sp_CbbQ3_|A0A0R0LVR7|
Thioalkalivibrio_sp_CbbQ3_|D3SG06|
Thioalkalivibrio_sulfidophilus_CbbQ3_|B8GQN5|
Acidimicrobium_ferrooxidans_CbbQ3_|C7M202|
Acidithiobacillus_caldus_CbbQ3_|F9ZLQ7|
Halothiobacillus_neapolitanus_CbbQ3_|D0KZ75|
Nitrobacter_hamburgensis_CbbQ3_|Q1QFQ9|
Nitrosomonas_eutropha_CbbQ3_|Q0AHV0|
Thiomonas_intermedia_CbbQ3_|D5X343|
Acidithiobacillus_ferrivorans_CbbQ1_|G0JNK4|
Nitrosomonas_communis_CbbQ1_|A0A0F7KI56|
Nitrosomonas_europaea_CbbQ1_|Q82TG8|
Nitrosomonas_sp_CbbQ1_|F9ZF97|
Thiobacillus_denitrificans_CbbQ1_|Q3SFN1|
Thiobacillus_thioparus_CbbQ1_|WP_018507362.1|
Thiomonas_sp_CbbQ1_|A0A078BG06|
Hydrogenovibrio_marinus_CbbQ1_|Q75W43|
Sideroxydans_lithotrophicus_CbbQ1_|D5CQ61|
Sulfuricella_denitrificans_CbbQ1_|S6AM60|
Hydrogenovibrio_kuenenii_CbbQ1_|WP_024851894.1|
Hydrogenovibrio_halophilus_CbbQ1_|WP_019895627.1|
Thiomicrothabodus_chilensis_CbbQ1_|WP_028486808.1|
Hydrogenovibrio_crunogenus_CbbQ1_|Q31IJ8|
Thiomicrothabodus_sp_CbbQ1_|WP_029406683.1|
Mariprofundus_ferrooxydans_CbbQ1_|Q0EX20|
Thiothrix_nivea_CbbQ1_|I3BR11|
Acidithiobacillus_ferrivorans_CbbQ2_|G0JPS2|
Thiobacillus_denitrificans_CbbQ2_|Q3SFL6|
Thiobacillus_thioparus_CbbQ2_|WP_018507397.1|
Halothiobacillus_neapolitanus_CbbQ2_|D0KZQ7|
Thiomonas_arsenitoxydans_CbbQ2_|D6CTS9|
Thiomonas_intermedia_CbbQ2_|O84918|
Sideroxydans_lithotrophicus_CbbQ2_|D5CTF9|
Sulfuricella_denitrificans_CbbQ2_|S6ABC2|
Gallionella_capsiferriformans_CbbQ2_|D9SHP7|
Mariprofundus_ferrooxydans_CbbQ2_|Q0F2Z9|
Acidithiobacillus_caldus_CbbQ2_|F9ZR06|
Thiolapillus_brandeum_CbbQ2_|W0TLR2|
Hydrogenovibrio_marinus_CbbQ2_|Q9WXI1|
Hydrogenovibrio_kuenenii_CbbQ2_|WP_024851802.1|
Thiomicrothabodus_sp_CbbQ2_|WP_029406689.1|
Hydrogenovibrio_halophilus_CbbQ2_|WP_019895619.1|
Thiomicrothabodus_chilensis_CbbQ2_|WP_028486930.1|
Hydrogenovibrio_crunogenus_CbbQ2_|Q31IK4|
Thiothrix_nivea_CbbQ2_|I3BN91|
consensus>70

```

```

255 DDADIRATLEHAI DMTFA...
251 DDPPDMRDALDAAVTTYF....
251 DDPPDMRDALDAAVSTFF....
251 DDPPDMRDALDAAVSTFF....
250 GELDRRDALDVVINTF....
251 DDPPDMRDALDGAVTTF....
251 DDPPDMRDALDAAVTTF....
252 DDPPDMRDALDAAVTTYF....
251 DDPPDMRDALDAAVTTF....
251 DDVDMRDALDTVVKTF....
262 DDLDMRDALDGAVTTF....
262 DDPPDMRDALDAAVTTYF....
289 DDPPDMRDALDAAVTTYF....
251 DDPPDMRDALDAAVTTYF....
251 DDPPDMRDALDAAVATFF....
251 DDPPDMRDALDAAVTTF....
251 DDPPDMRDALDAAVSTFF...
251 DDPPDMRDALDAAVTTF....
251 DDPPDMRDALDAAVSTFF....
251 DDPPDMRDALDAAVSTFF....
255 DDLIDLTLMITVEITFE...
251 DDPPDMRDALDAAVTTF....
251 DDPPDMRDALDAAVTTF....
255 DDLIDLTLMITVEITFE...
255 DDDIIDLTLTTVDITFE...
255 DDDIIDLTLTTVDITFE...
255 DDDIIDLTLTTVDITFE...
255 DDDIIDLTLTTVDITFE...
252 DDPPDMYETLEAAVQTFEVE...
252 DDPPDMRDTLNAAVQTFEG...
255 DDADIRATLEHAI DMTFA...
251 DDADIRETLDHAIDATFA...
250 DDADIRETLDHAIDATFA...
250 DDIDIRETLDHAIDATFA...
251 DDADIRQTLNHAIDVTF...
251 DDADIRQTLNHAIDVTF...
250 DDADIRDITLDHAIDITFA...
250 DDADIRDITLDHAIDAITFA...
250 DDADIRDITLDHAIDASFA...
251 DDVIDIRDITLDHAIDATFS...
250 DDADIRATLEHAI RATFV...
252 DDFDIRDITLDNAIDAVFTEGS
249 DDADIRQTLNNTIEMI FG...
249 DDADIRQTLNNTIEMI FG...
249 DDADIRETLDNAIEMIIG...
249 DDLDIRKTLNNAIEMVFG...
249 DDADIRQTLNNAIEMI FN...
249 DDADIRQTLNNAIEMI FG...
250 DDADIRSTLNHSLDAIEFG...
d#.D.rd.Ld.av...f....

```

8.3 Sequence alignment of 50 CbbO orthologs

```

AfO2_|B7J5E5|
AfO1_|B7JA27|
AfO3_|B7JB16|
Acidithiobacillus_ferriivorans_O3_|G0JL85|
Ferrovum_sp_O3_|A0A0R0LL45|
Thioalkalivibrio_sp_O3_|D3SG07|
Thioalkalivibrio_sulfidiphilus_O3_|B8GQN4|
Acidimicrobium_ferrooxidans_O3_|C7M203|
Acidithiobacillus_caldus_O3_|F9ZLQ2|
Thiothrix_nivea_O3_|I3BSU7|
Halothiobacillus_neapolitanus_O3_|D0KZ80|
Nitrobacter_hamburgensis_O3_|Q1QFQ8|
Nitrosomonas_eutropha_O3_|Q0AHU9|
Thiomonas_arsenitoxydans_O3_|D6CQL4|
Thiomonas_intermedia_O3_|D5X344|
Acidithiobacillus_ferriivorans_O1_|G0JNK3|
Nitrosomonas_communis_O1_|A0A0F7KHE0|
Nitrosomonas_europaea_O1_|Q82TG9|
Nitrosomonas_sp_O1_|F9ZF96|
Thiobacillus_denitrificans_O1_|Q3SFN2|
Thiobacillus_thioparus_O1_|UPI0003621E4A|
Thiomonas_sp_O1_|A0A078BJA6|
Hydrogenovibrio_marinus_O1_|Q75W42|
Sideroxydans_lithotrophicus_O1_|D5CQ62|
Sulfuricella_denitrificans_O1_|S6AAH0|
Thiomicrospira_kuenenii_O1_|UPI000466D9A2|
Thiomicrospira_halophila_O1_|UPI000378C5A3|
Thiomicrospira_chilensis_O1_|UPI000429B796|
Thiomicrospira_crunogena_O1_|Q31IJ7|
Thiomicrospira_sp_O1_|UPI000493B602|
Mariprofundus_ferrooxydans_O1_|Q0EX19|
Thiothrix_nivea_O1_|I3BR12|
Acidithiobacillus_ferriivorans_O2_|G0JPS1|
Thiobacillus_denitrificans_O2_|Q3SFL7|
Thiobacillus_thioparus_O2_|UPI00035FE92D|
Halothiobacillus_neapolitanus_O2_|D0KZQ5|
Thiomonas_arsenitoxydans_O2_|D6CTS8|
Thiomonas_intermedia_O2_|D5X1M7|
Sideroxydans_lithotrophicus_O2_|D5CTG0|
Sulfuricella_denitrificans_O2_|S6AJV8|
Gallionella_capsiferriiformans_O2_|D9SHP8|
Mariprofundus_ferrooxydans_O2_|Q0F300|
Acidithiobacillus_caldus_O2_|F9ZR07|
Thiolapillus_brandeum_O2_|W0TP28|
Hydrogenovibrio_marinus_O2_|Q75W24|
Thiomicrospira_kuenenii_O2_|UPI000467D206|
Thiomicrospira_sp_O2_|UPI000493F397|
Thiomicrospira_halophila_O2_|UPI000399C451|
Thiomicrospira_chilensis_O2_|UPI00040A3D56|
Thiomicrospira_crunogena_O2_|Q31IK5|
consensus>50
1 .....MTDHPRDASLPSRLAA.YRKQLDCRFPRVGEVFSDCIAKASARLGPAGVSA.YV
1 .....MTDHPD.....MKHLEDYAEELQAFGEQSAIQLQSKWQEAATRVFSPSGLDQ.YL
1 .....MAVELEKYQDILAELGEHAEVLRASWDEAARVFTPRGLENY.YL
1 .....MAVELEKYQDILAELGEHAEVLRASWDEAARVFTPRGLENY.YL
1 .....MAVELENYQDILEELGEHAEVLRASWGEAARVFSSRGIEKY.FL
1 .....MSVHLEDYQDLLEDLGNHAAEVLSSWQEAARAFSAGGLKRY.YL
1 .....MSIRLEDYQEILDDLGPAAEVMESAWQEAATRAFSSHGLKAY.YL
1 .....MAIELERYQGLLDSLGSQAEEVLRASWDEATRSFSPRGLEEV.YL
1 .....MAIQLSDFDILEALGDHAAELLASQWHEAARVFSSHSALEA.YL
1 .....MAIQLSDYQDVLDELGETANAVLESNWHEAARVFTPRGLDT.YL
1 .....MNPATEEYQALFEDLGEPIAEVLETHWIELSRLFSMPGFFKA.FL
1 MRSAKAASPGSPNLHMAIHLEDEFQELLDLDELPGGAQELLRASWNEAARALSSRGLDN.YL
1 .....MAIHLEEFQSLLDLDELPGGAQELLRASWNEAARALTPRGLDN.YL
1 .....MAIHLEEFQELVDELPLSGAQDVLRASWNEAARAFSSRGLDN.FL
1 .....MAIHLEEFQELVDELPLSGAQDVLRASWNEAARAFSSRGLDN.FL
1 .....MKRLLEDYAEELLEVFGESQAVLQSNWQEAARIFSPAGLDQ.YL
1 .....MSVNLTDYKELLESLPSESMSLLENTWHDATKVFSPRGLDN.YL
1 .....MSLNLTDYSELLESLOAESRSLLENTWSDAIRTLSPRGIDN.YL
1 .....MSINLDDYKELLEDLAPESVELLSHAWPEVVKTFARGLDN.YL
1 .....MSIKLEDYADLLEDLSEHSRAALDANWHEATKVFSPRGLDN.YL
1 .....MSIKLEDYAEELLEDLSPHTRDTLNATWHEATKVFSPRGLDN.YL
1 .....MAITLDDYADYLEDLSPHSLLEVLHAWPDAAKVLSPPRGLDN.YL
1 .....MSSEIEEIKKALIEAVPELSKNLDSMISEASHHMNEATRKI.WL
1 .....MAINLDDYQELLAALSPELQERVRSWPEAAKVFSARGLDN.FL
1 .....MAINLEDYQELLDLSLSPALQESLHSAWHEATKVFSPRGLDN.YL
1 .....MSSEIEEIKQALIEAVPELSDTLDSMIQEGSRHMNEATRKI.WL
1 .....MSSEIEEIKASLIESVPQLEEMLDLSLMEAAHSMNEATRKT.WL
1 .....MSFDFIVEVKESLIESVPQLEEMLDLSLIQEASHYMNEASRET.WL
1 .....MSFDFIEEVKASLIESVPQLEEMLDLSLIQEASHYMNEASRET.WL
1 .....MSFDFLEEVKQSLIKSVQLEETLDALIQEASHYMNEHSREE.WL
1 .....MNLDDYQEQFAEASPOLKEVLESSIQDAARIMSPAGLQE.YL
1 .....MPIHREYQEFLLGKFSPEETREVFNSTYQEAATRVMSPPAGLTD.YL
1 .....MTDHLHNTALPPQLAAAYRDQLDCRFPRVGEVFPDCIAKAAASLSPAGMSA.YI
1 .....MSTDTAYQHPLMAAYWKALDTRFAQVEEVFDDCMAEALTVLSAEGVAA.YL
1 .....MAAYWKQLDTRFPQVEEVFEDVMAEALAVLTREGIDA.YL
1 .....MDNTLSQDPQLQVYWTCKLDRFPLAEAVFEDCMTEALASLSPGIDA.YL
1 .....MSDALTDSDIRASWAQLDTRFPLVEAIFPACMAADAQGALTPDGLSA.FL
1 .....MVS DALMDDSIASWARLDTRFPLVEAIFPACMADAQSVLTPDGLSA.FL
1 .....MSVQATDIPPEELQPYWKELACGFPVVAEVFADHMQAALATLSDAGVEA.YI
1 .....MPKVNAEYEHPLMKGYREQLNCGFPQVKDVFDDCMEALSILSKEGITA.YL
1 .....MTEIVTQRTTELMSYWEKLSRCFARVDVDFDECMKEAQALLSKAGLEA.YI
1 .....MNVEMYWQKLGCNFAQVDRIFDECMKEAMALLSKAGVNA.YI
1 .....MNGFSEAAQEVVRAQAWERLDTSFAEVRAVFDLDFNAALRILSPADIDA.YI
1 .....MTEVHCNYREVQAVLADCLEEVELVLVSGKGLKD.FI
1 .....MNEELFAEYQEKFTCKFPKAVELFPDLIENAAKLSPPQGVQA.YL
1 .....MNEELFAEYQEKFTCKFPKAVELFPDLIENAVSKLSPQGVDA.YL
1 .....MNEELFAEYQEKFTCNFPKAVELFPDLIEHAVNTLTPKGVGT.YL
1 .....MNDALLAEYQAKFTCNFPQAREVFPHCLADAVNKVSPQGVNA.YL
1 .....MDAEFLAEYQTFQTCGFEKAKDVFPACTIEAHQILSPQGVQA.YL
1 .....MNADLITEVSEYQDKFTCNFPKALEVFPVCTIEAAQQLSDEGVTAYI
1 .....m...leey.e.ld...p...evld...qea...lsp.gld...yl

```


AfO2_|B7J5E5|
 AfO1_|B7JA27|
 AfO3_|B7JB16|
 Acidithiobacillus_ferrivorans_O3_|G0JL85|
 Ferrovum_sp_O3_|A0A0R0LL45|
 Thioalkalivibrio_sp_O3_|D3SG07|
 Thioalkalivibrio_sulfidiphilus_O3_|B8GQN4|
 Acidimicrobium_ferroxidans_O3_|C7M203|
 Acidithiobacillus_caldus_O3_|F9ZLQ2|
 Thiobacillus_nivea_O3_|I3BSU7|
 Halothiobacillus_neapolitanus_O3_|D0KZ80|
 Nitrobacter_hamburgensis_O3_|Q1QFQ8|
 Nitrosomonas_eutropha_O3_|Q0AHU9|
 Thiomonas_arsenitoxydans_O3_|D6CQL4|
 Thiomonas_intermedia_O3_|D5X344|
 Acidithiobacillus_ferrivorans_O1_|G0JNK3|
 Nitrosomonas_communis_O1_|A0A0F7KHE0|
 Nitrosomonas_europaea_O1_|Q82TG9|
 Nitrosomonas_sp_O1_|F9ZF96|
 Thiobacillus_denitrificans_O1_|Q3SFN2|
 Thiobacillus_thioparus_O1_|UPI0003621E4A|
 Thiomonas_sp_O1_|A0A078BJA6|
 Hydrogenovibrio_marinus_O1_|Q75W42|
 Sideroxydans_lithotrophicus_O1_|D5CQ62|
 Sulfuricella_denitrificans_O1_|S6AAH0|
 Thiomicrospira_kuenenii_O1_|UPI000466D9A2|
 Thiomicrospira_halophila_O1_|UPI000378C5A3|
 Thiomicrospira_chilensis_O1_|UPI000429B796|
 Thiomicrospira_crunigena_O1_|Q31IJ7|
 Thiomicrospira_sp_O1_|UPI000493B602|
 Mariprofundus_ferroxidans_O1_|Q0EX19|
 Thiobacillus_nivea_O1_|I3BR12|
 Acidithiobacillus_ferrivorans_O2_|G0JPS1|
 Thiobacillus_denitrificans_O2_|Q3SFL7|
 Thiobacillus_thioparus_O2_|UPI00035FE92D|
 Halothiobacillus_neapolitanus_O2_|D0KZQ5|
 Thiomonas_arsenitoxydans_O2_|D6CTS8|
 Thiomonas_intermedia_O2_|D5X1M7|
 Sideroxydans_lithotrophicus_O2_|D5CTG0|
 Sulfuricella_denitrificans_O2_|S6AJV8|
 Gallionella_capsiferriiformans_O2_|D9SHP8|
 Mariprofundus_ferroxidans_O2_|Q0F300|
 Acidithiobacillus_caldus_O2_|F9ZR07|
 Thiobacillus_brandenburgensis_O2_|W0TP28|
 Hydrogenovibrio_marinus_O2_|Q75W24|
 Thiomicrospira_kuenenii_O2_|UPI000467D206|
 Thiomicrospira_sp_O2_|UPI000493F397|
 Thiomicrospira_halophila_O2_|UPI000399C451|
 Thiomicrospira_chilensis_O2_|UPI00040A3D56|
 Thiomicrospira_crunigena_O2_|Q31IK5|
 consensus>50

113 LAPVSRALLSR EQLGH YLNL TLRDMMERTT GSIHGHHTTSPS PGLPELLRQAPT LLQSLTV
 103 SPTAANRLADP ELFKS YLGM LDRLLAQVP RALRPMLENIDQLFGY LTL
 105 SPVAAA RLGDPELFRG YLHL LDTLLAQAP RGVVRPMLDHLST LLGQLTL
 105 SPVAAVRLGDP ELFRAYLH LDTLLAQAP RGVVRPMLDHLST LLGQLTL
 105 SPVAAA RLGDPELFRS YLHL LDTLLAQAP RGLKPMLDHLV LLSQLTL
 105 SPVAAQRLGDYSLFEG YLRL LDTLLAQAP RGVVRPMLDHLV LLSQLTL
 105 VNVAAARLGDQSLFS GYLR LLDHLLVQAP RGVVRPMLDHLV LLSQLTL
 105 SPRAAERLGDLELFRS YLGL IDQLLLGLAP RGVVRPMLDHLV LLSQLTL
 104 APFVAR RLGEI ELFRG YLRL LDTLLAQAP RGVVRPMLGKIEALLEQMTL
 104 APTAATRMGEL ELFKG YLGL LNTLLAQAP RSLRPMLSKLDV LLLH LTL
 104 APVAVSRLGEA ELFVA YLRL LLDYIAAKAP RALRPLLGQIEQLLGVL LTL
 120 AP LVA YRVGDA DLFRR YLKL VLEI VAAQVP RGLRPMLENLDRLNQLTL
 104 AP LAAQRLSDAGLFVS YLKL VLEI MLLAQAP RGLRPMLENLDRLNQLTL
 104 AP LAGQRLGDE ELFRQ FLVSLMLAQAP RGLRPMLENLDRLNQLTL
 104 AP LAGQRLGDE ELFRQ FLVSLMLAQAP RGLRPMLENLDRLNQLTL
 103 SPTAANRLADP ELFKS YLGM LDRLLAQVP RAVRPMLENIDQLFGY LTL
 104 AVTAAKRLGDE DLFRR YLQL FVSTLVQAP RGVVRPMLDKLDI LLSQLTL
 104 SVTAA SRLGDPELFRN YLQL FNTFTAQAP RGVVRPMLDNLEVLFSQLTL
 104 SPTAAKRLGDAQLFQNYL QL FNTLLAQAP RGMVRPMLNKLID LFGQLTL
 104 AATAANRLGDGQLFLN YLQL FINTLLAQAP RGVVRPMLDKLEVLFQHLTL
 104 APTAASRLGDA DLFLL KYL QL FINTLLAQAP RGVVRPMLDKLEVLFQHLTL
 104 APTAANRLGDA ELFSN YLQL FINTLLAQAP RGVVRPMLDKLEVLFQHLTL
 104 LPLAAERTGDI DILLR QYL KLIYQIGSKSP RGLRPMMLNIDEILSKLTV
 104 APTAANRLGDE ALFRA YLQL FLANLLAQAP RGVVRPMLENLEK LFSQLTL
 104 APTAANRLGDE ELFRAYL QL FLANLLAQAP RGVVRPMLENLEK LFSQLTL
 104 LPLAAERTGDI DILLR QYL KLIYQIGSKSP RGLRPMMLNIDEILSKLTV
 104 LPLAAERTGDI DILLR QYL KLIYQIGSKSP RGLRPMMLNIDEILSKLTV
 104 LPLVAERTGDI DILLR QYL KLIYQIGSKTP RGLRPMMLNIDEILSKLTV
 104 LPLVSERTGDI DILLR QYL KLIYQIGSKTP RGMVRPMLNIDEILSKLTV
 104 LPTIAEKTNDI DILLR QYL KLIYQIGSKTP RGMVRPMLNIDEILSKLTV
 102 LPAARRLGDP ELLR GYL TLVHQISA KAS RGLRPMMLNIDEILSKLTV
 104 LPAARRLGDA ELVRG YLTL FTHQFSSTAA RGLRPMMLNIDEILSKLTV
 114 LAPVTRALQSRALLRD YLNL TLRDMMERTT GSIHGHHTTSPS PGLPDLLRQAPT LLQCLTV
 111 LAPVARRLQSA EQQLQHYV DVTLD FMTRTT GSIHGHHTTSPS PGLPEFFAQAPN LLNQLTL
 100 LAPVARRLQSQGLLQHYL DITLD LKARTT GSIHGHHTTSPS PGLPEFFAQAPN LLGQLTL
 110 LAPVARRLQSV ELLQEH YLDTIQELMGRTT GSIHGHHTTSPS PGLPDFFLAQAPN LLSQLTL
 109 LAPVAQRLQAG EQQLQHYL DITLD LKARTT RSIHGFHTTSPS PGLPDFFNQAPT LLAQLDI
 110 LAPVAQRLQAG EQQLQHYL DITLD LKARTT RSIHGFHTTSPS PGLQDFFNQAPT LLAQLDI
 111 LPATARRLQTR ELLAEYLA IALND LMERTT GSVHGHHTTSPS PGLPEMFQVPA LLSQLSL
 112 LAAIARRLHSQERMQR YL D LALDVMNRTT GSIHGHHTTSPS PGLPELYAQAPY LLSQISL
 111 SGAVARRLNND KMYGR YIE LIRDM LDRTS GSIHGHHTTSPS PGLPELLKQAPK LLGVLSL
 102 SGVARRLHSY DVFKQ YVEVIRDMMDRTS GSVHGHHTTSPS PGLPDLLKQAPH LLSALS L
 113 LIPVARRL E LSTELVRD YLEVVLDLMVQTS GTIHGRQATIPSPVMAEFFVQAP LLLDQVSL
 96 LAEIAARLE EPKLLQGYL D IIVLEMVRTS ISIHGHSHATMASP GLIPFLEKAPV LLGLTI
 105 LSTVCRIDTV EDLQHYL D IIDEYVDKTQ TVIHGHHSLEYSP GMVALLLESMPQLISKLCL
 105 LSTVCRIDTV EDLQHYL D IIDEYVDKTQ TVIHGHHSLEYSP GMVALLLESMPQLISKLCL
 105 LSTVCRIDTV EDLQHYL D IIDEYVDKTQ TVIHGHHSLEYSP GMVALLLESMPQLISKLCL
 105 LSSVCRIDTV EDLQHYL D IIDEYVDKTQ TVIHGHHSLEYSP GMVALLLESMPQLISKLCL
 105 LSSVCRIDTV EDLQHYL D IIDEYVDKTQ TVIHGHHSLEYSP GMVALLLESMPQLISKLCL
 105 LSSVCRIDTV EDLQHYL D IIDEYVDKTQ TVIHGHHSLEYSP GMVALLLESMPQLISKLCL
 105 LSSVCRIDTV EDLQHYL D IIDEYVDKTQ TVIHGHHSLEYSP GMVALLLESMPQLISKLCL
 109 LSSVCRIDTV EDLQHYL D IIDEYVDKTQ TVIHGHHSLEYSP GMVALLLESMPQLISKLCL

lp.aa.rlgd.elf..%ld..ddll...p.....rglrxpmlenld.11..ltl

AfO2_|B7J5E5|
 AfO1_|B7JA27|
 AfO3_|B7JB16|
 Acidithiobacillus_ferriivorans_O3_|G0JL85|
 Ferrovum_sp_O3_|A0A0R0LL45|
 Thioalkalivibrio_sp_O3_|D3SG07|
 Thioalkalivibrio_sulfidophilus_O3_|B8GQN4|
 Acidimicrobium_ferroxidans_O3_|C7M203|
 Acidithiobacillus_calidus_O3_|F9ZLQ2|
 Thiobacillus_nivea_O3_|I3BSU7|
 Halothiobacillus_neapolitanus_O3_|D0KZ80|
 Nitrobacter_hamburgensis_O3_|Q1QFQ8|
 Nitrosomonas_eutropha_O3_|Q0AHU9|
 Thiomonas_arsenitoxydans_O3_|D6CQL4|
 Thiomonas_intermedia_O3_|D5X344|
 Acidithiobacillus_ferriivorans_O1_|G0JNK3|
 Nitrosomonas_communis_O1_|A0A0F7KHE0|
 Nitrosomonas_europaea_O1_|Q82TG9|
 Nitrosomonas_sp_O1_|F9ZF96|
 Thiobacillus_denitrificans_O1_|Q3SFN2|
 Thiobacillus_thioparus_O1_|UPI0003621E4A|
 Thiomonas_sp_O1_|A0A078BJA6|
 Hydrogenovibrio_marinus_O1_|Q75W42|
 Sideroxydants_lithotrophicus_O1_|D5CQ62|
 Sulfuricella_denitrificans_O1_|S6AAH0|
 Thiomicrospira_kuenenii_O1_|UPI000466D9A2|
 Thiomicrospira_halophila_O1_|UPI000378C5A3|
 Thiomicrospira_chilensis_O1_|UPI000429B796|
 Thiomicrospira_crunogena_O1_|Q31IJ7|
 Thiomicrospira_sp_O1_|UPI000493B602|
 Mariprofundus_ferroxidans_O1_|Q0EX19|
 Thiobacillus_nivea_O1_|I3BR12|
 Acidithiobacillus_ferriivorans_O2_|G0JPS1|
 Thiobacillus_denitrificans_O2_|Q3SFL7|
 Thiobacillus_thioparus_O2_|UPI00035FE92D|
 Halothiobacillus_neapolitanus_O2_|D0KZQ5|
 Thiomonas_arsenitoxydans_O2_|D6CTS8|
 Thiomonas_intermedia_O2_|D5X1M7|
 Sideroxydants_lithotrophicus_O2_|D5CTG0|
 Sulfuricella_denitrificans_O2_|S6AJV8|
 Gallionella_capsiferriiformans_O2_|D9SHP8|
 Mariprofundus_ferroxidans_O2_|Q0F300|
 Acidithiobacillus_calidus_O2_|F9ZR07|
 Thiobacillus_brandenburg_O2_|W0TP28|
 Hydrogenovibrio_marinus_O2_|Q75W24|
 Thiomicrospira_kuenenii_O2_|UPI000467D206|
 Thiomicrospira_sp_O2_|UPI000493F397|
 Thiomicrospira_halophila_O2_|UPI000399C451|
 Thiomicrospira_chilensis_O2_|UPI00040A3D56|
 Thiomicrospira_crunogena_O2_|Q31IK5|
 consensus>50

173 DGLRNWVDYGVNRYLHHPERQKDFFSLSQSA DSRVAVLQRRERHGTLLADVE RKLDLYLRGLW
 151 GGLRRWALWGAHAHRTDFAEQIAYFSLKSP EALAMLQQRKGTLLFVDVQRRLMYLRALW
 153 GGLRRWALWGAQAHTD FDAQYFSLKSP ESIGVLQKERRGTLLFVDVQRRLSMYLRALW
 153 GGLRRWALWGAQAHTD FDAQYFSLKSP ESMGVVLQKERRGTLLFVDVQRRLGMYLRALW
 153 GGLRRWIMWGAQAHTN FDEQLKYFSLDSESMGVVLQKERRGTLLFVDVQRRLGMYLRALW
 153 GGLRRWATWGAQAHTDLEAQIAYFGLKSP ESRVSVFQKERRGTLLFVDVQRRLNLYLRSLW
 153 GGLRRWAMWGAQAHTDLEAQVAYFSLKSP ESKSVFQKERRGTLLFVDVQRRLNMYLRALW
 153 SGLTRWALWGAQVHKVD FEAQQRIFYFSLKSP DALAMLQQRKGTLLFVDVQRRLQAYLRALW
 152 GGLRRWAAWGISAYRNDFAAQAEYFGLQSEQSRVAVLQQRERRGTLLFVDVQRRLIMYLRALW
 152 GGLRRWAMWGVAAAYRNK FDEQTAIFYGLDSDSRTIMKKEQRGTLLFVDVQRRLIMYLRALW
 152 GGLRRWITWGIQAHTDFAAQEAIFYSLATPESRAVLQRRERRGTLLFVDVQRRLVIMYLRALW
 168 GGLRRWVQWGAQAYKTDFEQQAVYFGLQSDADAQAVLQQRKGTLLFVDVQRRLNIMYLRALW
 152 GGLRRWVQWGAQAYRSDLDGQVAYFSLQSDADAVLQQRKGTLLFVDVQRRLNIMYLRALW
 152 GGLRRWAQWGAQAYKADFQGVAYFSLQSPDALSMQQRKGTLLFVDVQRRLNIMYLRALW
 152 GGLRRWAQWGAQAYKADFQGVAYFSLQSPDALSMQQRKGTLLFVDVQRRLNIMYLRALW
 151 GGLRRWALWGAHAHRTDFAEQIAYFSLKSP EALAMLQQRKGTLLFVDVQRRLMYLRALW
 152 GGLRRWALWGAHAHRTNYGEQIAYFSLKSP ESQAVLQKERRGTLLFVDVQRRLNIMYLRALW
 152 GGLRRWALWGAHAHRTDYEAQLQYFSLKSP ESHAVLRKERRGTLLFVDVQRRLNIMYLRALW
 152 GGLRRWAMWGAHAHRTNYEEQMKYFGLKSP ESKSLAVLQKERRGTLLFVDVQRRLNIMYLRALW
 152 GGLRRWALWGAHAHRTNYEEQIKYFSLKSP ESMAMLQKERRGTLLFVDVQRRLNIMYLRALW
 152 GGLRRWALWGAHAHRTNYEEQIRYFSLKSP ESKSLAVLQKERRGTLLFVDVQRRLNIMYLRALW
 152 GGLRRWALWGAHAHRTNYEEQIRYFSLKSP ESKSLAVLQKERRGTLLFVDVQRRLNIMYLRALW
 152 SGLRRWAMWGAQAHAHNFQAQMEYFGLKSP ADSKAVFQQQRKGTLLFIDFHR SINFYLRALW
 152 GGLRRWALWGAHAHRTNYPEQASIFYSLKSP ESKSLAVLQKERRGTLLFIDVQRRLNIMYLRALW
 152 GGLRRWALWGAHAHRTNYPEQGNIFYSLKSP ESKSLAVLQKERRGTLLFIDVQRRLNIMYLRALW
 152 SGLRRWAMWGAQAHAHNFQAQMEYFGLKSP ADSKAVFQQQRKGTLLFIDFHR SINFYLRALW
 152 SGLRRWAQWGAQAHAHNFKAQIAYFGLKSP ADSKAVFQQQRKGTLLFIDHHR SINFYLRALW
 152 SGLRRWAQWGAQAHAHNFKAQMDYFGLKSP ADSKAVFQQQRKGTLLFIDHHR SINFYLRALW
 152 SGLRRWAQWGAQAHAHNFQAQIDYFGLKSP ADSKAVFQQQRKGTLLFIDYHR SINFYLRALW
 152 SGLRRWAQWGAQAHAHNFQAQIDYFGLKSP ADSKAVFQQQRKGTLLFIDYHR SINFYLRALW
 150 SGLRRWANFGAEAYRRDLQNLVKKYFALETSDSKMLQQRERRGTLLFIDTQRRLNIMYLRALW
 152 SGLRRWANFGAQAYRRDFNNLTAYFNLES ADSKATLQKERRGTLLFVKTQRRLNIMYLRALW
 174 DGLRNWVNYGVNRYLHHPERQKDFFSLSQSA DSRVAVLQRRERHGTLLVDVE RKLDLYLRGLW
 171 AGLRNWVEYGI RNYGTHPERQDIFYSLQSDADAVLQRRERHGTLLVDVE RKLDLYLRGLW
 160 AGLKNWVEYGI RNYRTHPERQKDYFSLQSDADAVLQRRERHGTLLVDVE RKLDLYLRGLW
 170 SGMKNWVDYGVNRYNNHHPERQKDYFSLKSP ADSRAVLQRRERHGTLLVDNE RNLDLYLRALW
 169 GGLRRWVDYGVNRYATHPERQKDYFSLQSDADAVLQRRERHGTLLVDVE RHLDLYLRALW
 170 GGLRRWVDYGVNRYATHPERQKDYFSLQSDADAVLQRRERHGTLLVDVE RHLDLYLRALW
 171 AGLRNWADYGI RYDHPERQKDYFSLQSDADRAVLQRRERHGTLLVDNE RKLDLYLRGLW
 172 QGLKNWVEYGI RNYGDHPERQIDIFYSLQSDADKAVMQRERHGTLLVDVE RKLDLYLRGLW
 171 QGLQNWIYGI KYNNHHPERQEEYFRLKSP DAVAVMQRERHGTLLVNNE RKLDYIRALW
 162 EGFRRNWVEYGI TYNTHPERQVDYFTLQSDADAHAVMQRERHGTLLIDNE RKLDSYMRAW
 173 TGLARWVDHGI RHYANNPDRQKDYFGLHSHADSHAVLRERHGTLLADVE RQLGLYLLGLW
 156 GGLRRWMDYGI AQSHHHPDLAEFFSLKSP DSRVAVLQRRERHGTLLVDMER RLDFELALW
 165 NGIRNFIDYGI RNYLNAPDEQIAYFKLES HDAKSIITRERKGTTFKDVRRHMDMLKECMW
 165 NGIRNFIDYGI RNYLNAPDEQIAYFKLES HDAKSIITRERKGTTFKDVRRHMDMLKECMW
 165 NGIRNFIDYGI RNYLNAPDEQIAYFKLES HDAKSIITRERHGTTFKDVRRHMDMLKSCW
 165 AGIRNFVDYGI RNYKSDPQIIEYFSLSSHADAKSIIRERERGTTFKDVRRHMDMLKDCMW
 169 AGIRNFIDYGI RNYNDAPDQIAYFSLSSHADAKSIIRERERGTTFKDVRRHMDMLKDCMW
 gGlrwwadwGa.ay..dfe.q..%FsLes.#s.avlq.#rkGtlfvdvqR.lnmylralw

Afo2_|B7J5E5|
 Afo1_|B7JA27|
 Afo3_|B7JB16|
 Acidithiobacillus_ferrivorans_O3_|G0JL85|
 Ferrofum_sp_O3_|A0A0R0LL45|
 Thioalkalivibrio_sp_O3_|D3SG07|
 Thioalkalivibrio_sulfidiphilus_O3_|B8GQN4|
 Acidimicrobium_ferrooxidans_O3_|C7M203|
 Acidithiobacillus_caldus_O3_|F9ZLQ2|
 Thiobacillus_nivea_O3_|I3BSU7|
 Halothiobacillus_neapolitanus_O3_|D0KZ80|
 Nitrobacter_hamburgensis_O3_|Q1QFQ8|
 Nitrosomonas_eutropha_O3_|Q0AHU9|
 Thiomonas_arsenitoxydans_O3_|D6CQL4|
 Thiomonas_intermedia_O3_|D5X344|
 Acidithiobacillus_ferrivorans_O1_|G0JNK3|
 Nitrosomonas_communis_O1_|A0A0F7KHE0|
 Nitrosomonas_europaea_O1_|Q82TG9|
 Nitrosomonas_sp_O1_|F9ZF96|
 Thiobacillus_denitrificans_O1_|Q3SFN2|
 Thiobacillus_thioparus_O1_|UPI0003621E4A|
 Thiomonas_sp_O1_|A0A078BJA6|
 Hydrogenovibrio_marinus_O1_|Q75W42|
 Sideroxydans_lithotrophicus_O1_|D5CQ62|
 Sulfuricella_denitrificans_O1_|S6AAH0|
 Thiomicrospira_kuenenii_O1_|UPI000466D9A2|
 Thiomicrospira_halophila_O1_|UPI000378C5A3|
 Thiomicrospira_chilensis_O1_|UPI000429B796|
 Thiomicrospira_crunogena_O1_|Q31IJ7|
 Thiomicrospira_sp_O1_|UPI000493B602|
 Mariprofundus_ferrooxydans_O1_|Q0EX19|
 Thiobacillus_nivea_O1_|I3BR12|
 Acidithiobacillus_ferrivorans_O2_|G0JPS1|
 Thiobacillus_denitrificans_O2_|Q3SFL7|
 Thiobacillus_thioparus_O2_|UPI00035FE92D|
 Halothiobacillus_neapolitanus_O2_|D0KZQ5|
 Thiomonas_arsenitoxydans_O2_|D6CTS8|
 Thiomonas_intermedia_O2_|D5X1M7|
 Sideroxydans_lithotrophicus_O2_|D5CTG0|
 Sulfuricella_denitrificans_O2_|S6AJV8|
 Gallionella_capsiferriiformans_O2_|D9SHP8|
 Mariprofundus_ferrooxydans_O2_|Q0F300|
 Acidithiobacillus_caldus_O2_|F9ZR07|
 Thiobacillus_brandenburgensis_O2_|W0TP28|
 Hydrogenovibrio_marinus_O2_|Q75W24|
 Thiomicrospira_kuenenii_O2_|UPI000467D206|
 Thiomicrospira_sp_O2_|UPI000493F397|
 Thiomicrospira_halophila_O2_|UPI000399C451|
 Thiomicrospira_chilensis_O2_|UPI00040A3D56|
 Thiomicrospira_crunogena_O2_|Q31IK5|
 consensus>50

233 QDSEVLVPYSTAFATLRTPQPPYDALGMRLPDVLLD.....LP
 211 GRDFFLRPTSDFEDREGYRPIYEGIFIHLPDAYDD.....WED
 213 GRDFFMRPTSDFEQREGYRPSIEGYIIHLPDAYDDLWVQ.....APS
 213 GRDFFMRPTSDFERREGYRPSIEGYIIHLPDAYDDLWVQ.....APS
 213 ARDFFMRPTSDFEKREGYQPIYIDGYIIHLPDAYDDYVYT.....TVE
 213 ARDFFMRPTSDFESREGYRPIYEGFTIFHLPDAYDD.....VEG
 213 ARDFFMRPTSDFESREGYRPIYEGFTIFHLPDAYDD.....EAG
 213 GRDFFLRPTSADPQTAAVVRPFIDGFVHLPDAYDDT.....D
 212 GRDFFLRPTSDFETRQGYRPIYEQYFIHVPDAYDDWEN.....AHGQR
 212 QDFFLRPTSDFETRQGYRPIYEHYFIHLPDAFDDFSEQ.....ENA
 212 GRDFFLRPTSDFESRAGYRPFIDELFINLPDAYDDYAVE.....GEQT
 228 GRDFFLRPTAGDYESRDGLRPIYEHYVIHLPDAYDDYK.....PQGDAT
 212 GRDFFLRPTSDFESREELKPIYEHYVIHLPDAYDNYK.....PLGPD
 212 GRDFFMRPTAGDYENREGLKPIYERYVIHLPDAYDDFR.....PLGPD
 212 GRDFFMRPTAGDYENREGLKPIYERYVIHLPDAYDDFR.....PLGPD
 211 GRDFFLRPTSDFEDREGYRPIYEGIFIHLPDAYDD.....WEG
 212 ARDFFMRPTSDFETRQGYRPIYENYLIHLPDAYDA.....HGE
 212 ARDFFMRPTSDFENREGYRPIYEDYIIHLPDAYDN.....YGD
 212 GRDFFMRPTSDFESREGYKPIYENYFIHLPDAYDS.....YGD
 212 ARDFFMRPTSDFESREGYKPIYENYFIHLPDAYDD.....FAPE
 212 ARDFFMRPTSDFESREGYKPIYADYLLHVPDAYDD.....FAIE
 212 ARDFFMRPTSDFESREGYKPIYEDYLLHLPDAYDD.....HEG
 212 GRDFFLRPTSADDFEG...FKPFIYENHAMHLPDALNDLG.....D
 212 ARDFFMRPTSDFESREGYRPFVEGYFIHLPDAYDDFILPAGDRSAPLPQAGEGS
 212 ARDFFMRPTSDFETRQGYRPIYEGYFIHLPDAYDD.....QSV
 212 GRDFFLRPTSADDFEG...FKPFIYENHAMHLPDALNDLG.....D
 212 GRDFFLRPTSADDFED...FKPFIYENHAMHLPDALNDLG.....D
 212 ARDFFLRPTSADDFED...FKPFIYENHAMHLPDALNDLG.....D
 212 ARDFFLRPTSADDFED...FKPFIYENHAMHLPDALNDLG.....D
 212 ARDFFLRPTSADDFED...FKPFIYENHAMHLPDALNDLG.....D
 210 GRDFFLRPTSADDFEG...FKPFIYENHAMHLPDALNDLG.....D
 212 ARDFFLRPTSADDFED...FKPFIYENHAMHLPDALNDLG.....D
 234 QDSEVLVPYSTAFATLRTPQPPYDSLGMRLPDVLLD.....LP
 231 QDHDHLVPYSTAFDEIRKPVPPYDKLGMRLPDVYDDLVLPCP.....AGRGAG
 220 QDSDQLVPYSTAFDELRKPVPPYDKLGIHLPDVYDD.....AGD
 230 QDSEQLVPYSTAFDEIRKPVPPYDKLGIHLPDVCD.....SN
 229 RDGDLIPYSTAFDELRKPVPPYDKLGIHLPDVQDD.....AS
 230 RDGDLIPYSTAFDELRKPVPPYDKLGIHLPDVQDD.....AS
 231 QDDMLIPYSTSMYDELRKPVPPYDKLGIHLPDVYDD.....AQ
 232 EHDHVLVPYSTLGFDELRKPVPPYDKLGIHLPDVYDD.....FN
 231 NDSDYLIPYSTVAFDEQRKPMPPYTDGIRLPDVYDD.....FN
 222 SDHDLIPYSTLAFDELRKPMPPYDEIGIRLPDVYDD.....QH
 233 EDVAPLVYSTSGFHLRQPIPPYDQGFHLPDVYDD.....AR
 216 KIDASIPYSTRLMDESRDPPYDDQGGIFLPDVYDD.....LDN
 225 DCDLPFQTFATAFDQLRKPVPPYLDLTLIAIPDVYDE.....EN
 225 ECDLPFQTFATAFDQLRKPVPPYLDLTLIAIPDVYDA.....EN
 225 DCDLPFQTFATAFDQLRKPVPPYIDELTIAVDPVYES.....EN
 225 GCDLPFSVFATAFDQLRKPIPPYLDLTLIAVDPVYDQ.....EH
 225 DCDLPFSVFATAFDQLRKPIPPYLDLTLIAVDPVYDT.....ENT
 229 HCDRPFVSFSTAFDQLRKPVPPYLDLTLIAVDPVYED.....EN
 .rdfflrp.s.dfee...y.Pyie...ihlpDay#d

Afo2_|B7J5E5|
 Afo1_|B7JA27|
 Afo3_|B7JB16|
 Acidithiobacillus_ferrovorans_O3_|G0JL85|
 Ferrovum_sp_O3_|A0A0R0LL45|
 Thioalkalivibrio_sp_O3_|D3SG07|
 Thioalkalivibrio_sulfidophilus_O3_|B8GQ4|
 Acidimicrobium_ferrooxidans_O3_|C7M203|
 Acidithiobacillus_caldus_O3_|F9ZLQ2|
 Thiobacillus_nivea_O3_|I3BSU7|
 Halothiobacillus_neapolitanus_O3_|D0KZ80|
 Nitrobacter_hamburgensis_O3_|Q1QFQ8|
 Nitrosomonas_eutropha_O3_|Q0AHU9|
 Thiomonas_arsenitoxydans_O3_|D6CQ4|
 Thiomonas_intermedia_O3_|D5X344|
 Acidithiobacillus_ferrovorans_O1_|G0JNK3|
 Nitrosomonas_communis_O1_|A0A0F7KHE0|
 Nitrosomonas_europaea_O1_|Q82TG9|
 Nitrosomonas_sp_O1_|F9ZF96|
 Thiobacillus_denitrificans_O1_|Q3SFN2|
 Thiobacillus_thioparus_O1_|UPI0003621E4A|
 Thiomonas_sp_O1_|A0A078BJA6|
 Hydrogenovibrio_marinus_O1_|Q75W42|
 Sideroxydans_lithotrophicus_O1_|D5CQ62|
 Sulfuricella_denitrificans_O1_|S6AAH0|
 Thiomicrospira_kuenenii_O1_|UPI000466D9A2|
 Thiomicrospira_halophila_O1_|UPI000378C5A3|
 Thiomicrospira_chilensis_O1_|UPI000429B796|
 Thiomicrospira_crunogena_O1_|Q31IJ7|
 Thiomicrospira_sp_O1_|UPI000493B602|
 Mariprofundus_ferrooxydans_O1_|Q0EX19|
 Thiobacillus_nivea_O1_|I3BR12|
 Acidithiobacillus_ferrovorans_O2_|G0JPS1|
 Thiobacillus_denitrificans_O2_|Q3SFL7|
 Thiobacillus_thioparus_O2_|UPI00035FE92D|
 Halothiobacillus_neapolitanus_O2_|D0KZQ5|
 Thiomonas_arsenitoxydans_O2_|D6CTS8|
 Thiomonas_intermedia_O2_|D5X1M7|
 Sideroxydans_lithotrophicus_O2_|D5CTG0|
 Sulfuricella_denitrificans_O2_|S6AJV8|
 Gallionella_capsiferriiformans_O2_|D9SHP8|
 Mariprofundus_ferrooxydans_O2_|Q0F300|
 Acidithiobacillus_caldus_O2_|F9ZR07|
 Thiobacillus_brandenburgensis_O2_|W0TP28|
 Hydrogenovibrio_marinus_O2_|Q75W24|
 Thiomicrospira_kuenenii_O2_|UPI000467D206|
 Thiomicrospira_sp_O2_|UPI000493F397|
 Thiomicrospira_halophila_O2_|UPI000399C451|
 Thiomicrospira_chilensis_O2_|UPI00040A3D56|
 Thiomicrospira_crunogena_O2_|Q31IK5|
 consensus>50

272 I G A L D H Y R A I L A H M V G H R R W S T P Q I A D N . W S P F Q R L A V E F F E D A R I D T L L I R T Y P G L R T L
 250 I D G M T R Y R A V A V H A A A H V A Y M R K A I S A E A L N P E Q M A V I G L F E D A R V E A L A M R A F P G I R A I
 261 I H G I E L Y R A S A A H A A C H Q V Y T T D Q F D S T G L S T L Q S A L I G L I E D A R V E T I A L K Q F P G L R H I
 261 I S G I E L Y R A S A A H A A C H Q V Y T T H Q F D S T G L T A M Q V T L V G L V E D A R V E A I A L K Q F P G L G Q I
 261 V T G I E L Y R A S A A H A A C H Q V Y T S K Q F D D T G L S A L Q K A L C G L I E D A R I E K L A M E K F P G L R Q T
 252 L T G L D L Y R A A A A H A A S H V V H T R E M Q S E E G P R G V V G I L T E L F E D A R I E A L A I A A F P G L R Q I
 252 L T G L D L Y R A A A A H A A A H M V Y T R E M L P E A K P K G V P G A L V G L I E D A R V E A L A M A A F P G L R Q I
 256 I S A L E Q Y R A A A A H A A A H Q V A T T H A V A R A E R S R L E E L V V G I L E D A R V E R A A I A A F P G L A A L
 256 V T A L E L F R A A V A H A A A H V V L S Q Y D L R L E L A P A E Q K L W I G L L E D A W V E R E A I A R F P G L L P L
 260 I A G M E L Y R A A A A H A S A H I M H S R Y D P L A E D L S S L Q K A M I G L V E D A R V E A L A I K A F P G L K Q I
 256 V D G L T V Y R A A A A H A A A H Q I Y S R L D S D A K N A S P L D R A T I G L F E D A R V E K L A M A D F P G L V P L
 277 V S A L E I Y R A A A N H C A A H L V F T Q T P L S M Q Q L L P V Q M A V I E L F E D A R V E E L A I R R F P G M R Q A
 261 V P A L E I Y R A A V N H C A A H L V F T Q Q P L S M Q M L H P A Q M A V I E V F E D A R V E E L S I R Q F P N M R E A
 261 V S A L D V Y R A A A N H C A A H L V F T T E P L S M Q A L S P L Q I G L I E V I E D A R V E A L A I A Q F P N M R E A
 261 V S A L D V Y R A A A N H C A A H L V F T T E P L S M Q A L S P L Q I G L I E V I E D A R V E A L A I A Q F P N M R E A
 250 I D G M T R Y R A V A A H A A A H V A Y M R K A I S A E A L N P E Q M A V I G L F E D A R V E A L A M R A F P G I R A I
 251 I A A S E V Y R A A A A H A A A H L V E T K P I S A E A L N P L Q M A V I S V I E D A R V E A L S I R R F P G L R Q I
 251 I A A S E V Y R A A A A H A A A H L V E T K P I S A E A L N P L Q M A V I S V I E D A R V E A L S I R R F P G L Q H T
 251 V T A T E V Y R A A A A H A A A H L V E T K P I S A E A L N P M Q M A V I S V I E D A R I E E L A I R R F P G L R K T
 256 V P G L E L Y R A T A A H C A A H V V E T R V P I S A E A L N P L Q M A V I S V I E D A R V E T L S I R R F P G L K Q L
 256 V S G L E L Y R A T A A H C A A H V V E T K E P I S A E A L N P M Q M A V I S V I E D A R V E A L S I R R F P G L K Q L
 251 V S G L E L Y R A A A A H C A A H V V E T Q T P I S A E A L T P Y Q M A V I S V I E D A R V E A L S I R R F P G L K Q L
 248 I K G I E L Y R A M S A H I A A H L M Y T T Q A I S M E Q L N P Q M F F I E L M E D A R V E H K A I K A F P G L K K L
 272 L P G L E L Y R A A A A H G A A H L V Y T R E P I S A E A L N P F Q M A V I S V I E D A R V E T L A A N R F P G L R K L
 255 L P G L E M Y R A A A A H S A A H I V Y T R D P I S A E S L N P F Q M A V I S V I E D A R V E T L A V R K F P G L K H L
 248 I K G I E L Y R A M S A H I A A H L M Y T N E P I S M E Q L N P Q M F F I E L M E D A R V E Y K A I K A F P G L K K L
 248 I K G L E I Y R A M A A H Q A A H L M Y S S Q A I S M E Q L N P Q M F F I E L I E D A R V E Y N A I K K F P G L K K L
 248 I K G V D L Y R A M A A H L A S H I A Y T K E A I S M E Q L N P Q M F F I E L I E D A R V E Y N A I K H F P G L K K L
 248 I K G G E L Y R A M A A H M A S H L A Y T K E A I S M E Q L N P Q M F F I E L I E D A R V E Y N A I K N F P G L K G L
 248 V K G G E L Y R A M A A H L A S H L C Y S K A A I S M E Q L T P Q M F F I E I I E D A R V E Y N A I K D F P G L K N L
 246 V S G V E L Y R A Q A A H L A A H M H Y T R S A I S A E E L S Q A Q M F F I A L I E D A R V E Y C A M Q D F P G L K Q L
 248 I P G L E L Y R A T A A H M A S H L M Y A R A A I S A E Q L S P A Q M F F I G F L E D A R V E Y K A V Q A F P G L K N L
 273 A R A L D R Y R A T L A H M V G H R R W S T P Q I A D N . W S P F Q R L A V E F F E D A R I D T L L I Q T Y P G L R A L
 285 V S G L D R Y R A T L A H M V G H R R W S E A Q I A D N . W S P F Q R M A V E F F E D C R V E T L L M R E Y P G L A R I
 259 V S G L D R Y R A T L A H I V G H R R W S E A Q V A D N . W S P F Q R M A V E F F E D C R I E T L L I R E Y P G L R R L
 269 V R G I D R Y R A M L A H I V G H R R W S E A Q I A D N . W S P F Q R M A V E F F E D C R V D T L L I R E Y P G L R R L
 268 V S G L D R Y R A M L A H M V G H R R W S T P Q I A D N . W S P F Q R M A V E F F E D C R V E T L L I R E Y A G L R R L
 269 V S G L D R Y R A M L A H M V G H R R W S T P Q I A D N . W S P F Q R M A V E F F E D C R V E T L L I R E Y A G L R R L
 270 V S G L D R Y R A T L A H I A G H R R W T T A I F A D N . L S P F Q R M A A E F F E D S R V E C L A M R E Y P G L R R L
 271 I A G I D R Y R A T L A H I I G H K R W T T A I F A D N . F S P F Q R M A I E F F E D A R I E H L V M Q Q Y P G L R R I
 270 V S G I D R Y R A A L A H M A A H R R W S Q K M I V D N . W S P F Q R A S V E I F E D S R V E Y L A C Q K Y P G L R K L
 261 V S G I D R Y R A S L A H M A A H R R W S Q K M V V D N . W S P F Q R I A V E V F E D S R V E Y L A M Q R Y P G L R K L
 272 V A A I D R Y R L A L A H M A G H R R W S Q P I A D N . W S P F Q R L T V E C F E D C R I D T L L L R Y P G L R P I
 256 I S G M D Q Y R A A L A H M A G H R R W T A P V F A D N . R S P M Q R I A I E T F E D S R I E Y L L M Q E Y P G L R R I
 264 I S G I D R Y R A T L A H M A A H H K W S T K L M A D N . Y A P H M Q L F I S I F E D S R V E Y L A I Q E Y P G L K K L
 264 V S G A D R F R A T L A H M M A H Y R W S T Q L M A D N . Y A P H M Q L F I S I F E D S R V E Y L A I Q K Y P G L K K L
 264 V S G L D R Y R A V A H M L A H K T W S Q L M A D N . F A P H M Q L F I S M F E D S R V E Y L A M Q K Y P G L K K L
 264 V S G L D R Y R A M L A H M A H K R W S T K L M A D N . F A P H M Q L F V S V F E D A R V E Y L A M Q K Y P G L K R L
 264 V P G I D R Y R A M L A H M M A H R K W S T K L M A D N . F A P H M Q L F I S V F E D S R V E Y L A M Q K Y P G L K K L
 268 V S G L D R Y R A M L A H M M A H K E W S T K L M A D N . F A P H M Q L F I S T F E D C R V E R L A I Q R F P G L K P L
 v.gld.%Ra.aaH.aaH...t...i...e.l.p.q...i...EDar!#.lai...%pglr.l

AfO2_|B7J5E5|
 AfO1_|B7JA27|
 AfO3_|B7JB16|
 Acidithiobacillus_ferrivorans_O3_|G0JL85|
 Ferrovum_sp_O3_|A0A0R0LL45|
 Thioalkalivibrio_sp_O3_|D3SG07|
 Thioalkalivibrio_sulfidiphilus_O3_|B8GQN4|
 Acidimicrobium_ferrooxidans_O3_|C7M203|
 Acidithiobacillus_caldus_O3_|F9ZLQ2|
 Thiiothrix_nivea_O3_|I3BSU7|
 Halothiobacillus_neapolitanus_O3_|D0KZ80|
 Nitrobacter_hamburgensis_O3_|Q1QFQ8|
 Nitrosomonas_eutropha_O3_|Q0AHU9|
 Thiomonas_arsenitoxydans_O3_|D6CQL4|
 Thiomonas_intermedia_O3_|D5X344|
 Acidithiobacillus_ferrivorans_O1_|G0JNK3|
 Nitrosomonas_communis_O1_|A0A0F7KHE0|
 Nitrosomonas_europaea_O1_|Q82TG9|
 Nitrosomonas_sp_O1_|F9ZF96|
 Thiobacillus_denitrificans_O1_|Q3SFN2|
 Thiobacillus_thioparus_O1_|UPI0003621E4A|
 Thiomonas_sp_O1_|A0A078BJA6|
 Hydrogenovibrio_marinus_O1_|Q75W42|
 Sideroxydans_lithotrophicus_O1_|D5CQ62|
 Sulfuricella_denitrificans_O1_|S6AAH0|
 Thiomicrospira_kuenenii_O1_|UPI000466D9A2|
 Thiomicrospira_halophila_O1_|UPI000378C5A3|
 Thiomicrospira_chilensis_O1_|UPI000429B796|
 Thiomicrospira_crunogena_O1_|Q31IJ7|
 Thiomicrospira_sp_O1_|UPI000493B602|
 Mariprofundus_ferrooxydans_O1_|Q0EX19|
 Thiiothrix_nivea_O1_|I3BR12|
 Acidithiobacillus_ferrivorans_O2_|G0JPS1|
 Thiobacillus_denitrificans_O2_|Q3SFL7|
 Thiobacillus_thioparus_O2_|UPI00035FE92D|
 Halothiobacillus_neapolitanus_O2_|D0KZQ5|
 Thiomonas_arsenitoxydans_O2_|D6CTS8|
 Thiomonas_intermedia_O2_|D5X1M7|
 Sideroxydans_lithotrophicus_O2_|D5CTG0|
 Sulfuricella_denitrificans_O2_|S6AJV8|
 Gallionella_capsiferriiformans_O2_|D9SHP8|
 Mariprofundus_ferrooxydans_O2_|Q0F300|
 Acidithiobacillus_caldus_O2_|F9ZR07|
 Thiola pillus_brandeum_O2_|W0TP28|
 Hydrogenovibrio_marinus_O2_|Q75W24|
 Thiomicrospira_kuenenii_O2_|UPI000467D206|
 Thiomicrospira_sp_O2_|UPI000493F397|
 Thiomicrospira_halophila_O2_|UPI000399C451|
 Thiomicrospira_chilensis_O2_|UPI00040A3D56|
 Thiomicrospira_crunogena_O2_|Q31IK5|
 consensus>50

331 FLALHPKPGEDACDPETT...SCLRHRLAMLSRALLDPEHG...YRNPVLDHDFVDR.
 310 WLPFHTAVAGP...VASMGHLSRLARALIDPDYV...DDNAWVQEGR...
 321 WLPHTATPA...SGNEAASLMARLAHALLDPDYR...DDHPWVAMG...
 321 WLPHTATPA...SGNDAPLMARLARALLDPEYS...DEHPWVAMG...
 321 WSELHTATPQ...SGESTVELMKRLARRLQDKHY...DNHAWVTILG...
 312 WVPHTATAE...DGETPQALLARTARALLDPDY...DPHAFVQKA...
 312 WQPLHTVTP...AGDAPAVLLARAARALIEPEYQ...DSNAFVARV...
 316 WARFHVVDLA...ADD.PSSLVRLVRSLLVGC...EAHPWVREACE...
 316 WQRLL...LQVPATLERERE...AAFPVLDRCALALLDPDY...DDHPLVARARS...
 320 WLN...LPT...DQSKT...DSTGALLDRMARAMADENYCN...DDHPLVEAAR...
 316 WRNLNQLTVAALPADART.GENPTLGFALDRIALALCDENSS...DTHPLVEWSR...
 337 WLALHPVFDYSA...QPQTVGELDLRLARALLDLDST...DPHPFVRQGV...
 321 WLMHLHPVFDYSR...EPQSIGELDLRLGRALLEEQSH...DPHPWVREAV...
 321 WLALQPPADYS...EPKTVGDLNLRLARALLEEHS...DPHPWVQQGV...
 321 WLALQPPADYS...EPKTVGDLNLRLARALLEEHS...DPHPWVQQGV...
 310 WLPFHTAVAGP...VASMGHLSRLARALIDPAYE...DGNWVQEGR...
 311 WSSLHDATPDM...KTIGDFLNRLARALLDPTYE...DNDDWVMEGR...
 311 WSAHLIATADM...NTTIGDYLNRLARALIDPDYS...DPDEWVIQGR...
 311 WATLHTATPDM...AATMGDYLNRLARALLDPDYH...DTDHWIAEGR...
 316 WAKLHVATPEM...NATVGDYDLRLARALLDETYV...DDDPWIAEGR...
 316 WSRHLHVATPEM...NASVGDYLNRLARALLDETYQ...DNDPWVVAEGR...
 311 WAKLHLSSTPDM...THSVGDYLNRLARALLDSSH...DADPWIAEGR...
 308 WKKVSAQAME...HPESTAGRLQLALALMDPKH...PFEDVLSEIRDR...
 332 WVQLHTIEAGQ...GDSAGDFLNRLARALLDHDYQ...DNDPWIAQGR...
 315 WAQLHTIESGQ...GKSVGDYLNRLARALLDNDYQ...DNDPWIAQGR...
 308 WKKVSSQAME...SPESTAGRLQLALSMDPKH...AFEDVLQIHER...
 308 WLKVFANDREAVEL...PEGSTAWRLQLALSMDPDYDLSALNDNALNLVHER...
 308 WMKVIRASME...ESEL...DPRSTAA...RLQLALALMDKDY...TLEDTQM...
 308 WKKVIKASME...ESEL...PEKSTAY...RLQLALALMDVKH...DLQDEQMMV...
 308 WMKVIRASQDQIVL...DPKSTAY...RLQLALALMDPKF...DLQDEQMLIVADK...
 306 WGRLLFAAEYSAA...VEHESIVTLEHFAHMLLDSSIS...CDEEELFLASK...
 308 WSRLLTVDYEGN...VEHPTMKILEHVALMLLDASVR...SGDDEINHFVDS...
 332 FLALHPKPAEAEACDPETT...SCLRHRLTMSRALLDPEHG...YRNPVLDKDFVER...
 344 FRALHPKPVAEACDGETT...SCLRHRLAMLSRAFLIDPDHG...YAAPVLDNDFVAR...
 318 FLALHPKPLETACDPQTT...SCLRHRLAMLSRALLDPDHG...YLDADLNDYVQR...
 328 FLAMHPNPQENACDPETT...SCLRHRLVMSRALLDPNHG...YKDANLNEFAAR...
 327 FLALHPRPVEDACDPATT...SCLRHRLAMLSRALLDPEHG...YRDADLNTFVAR...
 328 FLALHPRPVEDACNPETT...SCLRHRLAMLSRALLDPEHG...YRDADLNAFVTH...
 329 FKAHLHPVPQEQACDAEKE...TVRHLRLAMLSRAILDPQHE...YRNAATIEFVLR...
 330 FTSLHPKPVENACNPETT...SCLRHRLAMLSRALLDTPHG...YQDADLNTFVGR...
 329 FLALHPIPKEEDLQLEEQG...ISVFRRLAMLSRAILDPAYI...CQNEHVREFTKR...
 320 FLALHPIPNEADLQLQDR...SLINLRLAMLSRAILDPDYS...YENADVREFVGR...
 331 LLAHLHPQPEDDCNDATT...SCLRHRLVMSRALLDPAACS...YRHPVIREFTVR...
 315 FQALHPVPVEDECNDAEE...ACVRHLRLTMSRAILDPAHG...YQNPIVILDFVQQ...
 323 FLALHPYPEKGACDDTKQ...SCLRYRATRMSRALLDENFD...PENPLLEEFQR...
 323 FLALHPYPEKGACDDKKQ...SCLRYRATRMSRALLDENFD...PENPLLEEFQR...
 323 FLALHPYPEKGACDDKKF...SCLRYRATRMSRALLDENFD...PQNPLLEDFRQR...
 323 FLALHPYPEKGACDDAKQ...SCLRYRATRLSRALLDQDFD...PENPLLENFRER...
 323 FLALHPYPEKGACDDQKQ...SCLRYRAARLSRALLDDET...SQNPLEDFFRQQ...
 327 FLALHPVPEKGACDPIQQ...ACLYRSTRLSRALMDDET...PENPLIEDFRITQ...
 wl.lh...e...r1.rlaralldpdy...denp.v.e...

AfO2_|B7J5E5|
 AfO1_|B7JA27|
 AfO3_|B7JB16|
 Acidithiobacillus_ferrivorans_O3_|G0JL85|
 Ferrofum_sp_O3_|A0A0R0LL45|
 Thioalkalivibrio_sp_O3_|D3SG07|
 Thioalkalivibrio_sulfidiphilus_O3_|B8GQN4|
 Acidimicrobium_ferrooxidans_O3_|C7M203|
 Acidithiobacillus_caldus_O3_|F9ZLQ2|
 Thiothrix_nivea_O3_|I3BSU7|
 Halothiobacillus_neapolitanus_O3_|D0KZ80|
 Nitrobacter_hamburgensis_O3_|Q1QFQ8|
 Nitrosomonas_eutropha_O3_|Q0AHU9|
 Thiomonas_arsenitoxydans_O3_|D6CQL4|
 Thiomonas_intermedia_O3_|D5X344|
 Acidithiobacillus_ferrivorans_O1_|G0JNK3|
 Nitrosomonas_communis_O1_|A0A0F7KHE0|
 Nitrosomonas_europaea_O1_|Q82TG9|
 Nitrosomonas_sp_O1_|F9ZF96|
 Thiobacillus_denitrificans_O1_|Q3SFN2|
 Thiobacillus_thioparus_O1_|UPI0003621E4A|
 Thiomonas_sp_O1_|A0A078BJA6|
 Hydrogenovibrio_marinus_O1_|Q75W42|
 Sideroxydans_lithotrophicus_O1_|D5CQ62|
 Sulfuricella_denitrificans_O1_|S6AAH0|
 Thiomicrospira_kuenenii_O1_|UPI000466D9A2|
 Thiomicrospira_halophila_O1_|UPI000378C5A3|
 Thiomicrospira_chilensis_O1_|UPI000429B796|
 Thiomicrospira_crunogena_O1_|Q31IJ7|
 Thiomicrospira_sp_O1_|UPI000493B602|
 Mariprofundus_ferrooxydans_O1_|Q0EX19|
 Thiothrix_nivea_O1_|I3BR12|
 Acidithiobacillus_ferrivorans_O2_|G0JPS1|
 Thiobacillus_denitrificans_O2_|Q3SFL7|
 Thiobacillus_thioparus_O2_|UPI00035FE92D|
 Halothiobacillus_neapolitanus_O2_|D0KZQ5|
 Thiomonas_arsenitoxydans_O2_|D6CTS8|
 Thiomonas_intermedia_O2_|D5X1M7|
 Sideroxydans_lithotrophicus_O2_|D5CTG0|
 Sulfuricella_denitrificans_O2_|S6AJV8|
 Gallionella_capsiferriiformans_O2_|D9SPH8|
 Mariprofundus_ferrooxydans_O2_|Q0F300|
 Acidithiobacillus_caldus_O2_|F9ZR07|
 Thiolaipillus_brandeum_O2_|W0TP28|
 Hydrogenovibrio_marinus_O2_|Q75W24|
 Thiomicrospira_kuenenii_O2_|UPI000467D206|
 Thiomicrospira_sp_O2_|UPI000493F397|
 Thiomicrospira_halophila_O2_|UPI000399C451|
 Thiomicrospira_chilensis_O2_|UPI00040A3D56|
 Thiomicrospira_crunogena_O2_|Q31IK5|
 consensus>50

381 .FHGELAGGN..ADTATMARIALDYVTTTRRQ.SDQFAKVHFADTEISYRDDNRGLWRFI
 352 .RVFMEAEARWDDPQLSLWDLGVHLAARL.RNQRVSYHARTDI.PDITYRDDNRYIWEFE
 362 .RQIFSEQQERLVSSTEWSEVGLRLAEM.LQLGVAYSATTDVDPIDYRDDNRYMWFEFE
 362 .RQMFSQQQDRLGSTEWAREVGLHFAEM.QQLGVVYSGSSGLPDIDYRDDNRYMWFEFE
 362 .LRLFNEKNKQDNVTEWVIEIGKELAEL.QLLGVTYSQSKDQIDIPYRDDNRYMWFEFE
 353 .VSAFAERVDDLEAAGLARLEAELAPEF.PEQGYMN.PAKVAPDVL.YRDDNRYLWNFG
 353 .RAFAAERMNDLNEQPQLAWELGQELAKHF.PLLGGFN.ASKDQPDIT.YRDDNRYIWAFG
 358 .HARVLDDELGGAEAAAALELARTLIDAA.RRFVGLDQVDSRAVLPSYRDDNRYVFERD
 365 AALLAAQPQDASARYTEILQRAQELAAAA.AEIGLELPGYGDDGAILYRDDNRYLWTAP
 363 .ALFQQAASDRLSEQRLAWEIGRELAQQA.ESLGIRFNPNKDLPANPYRDDNRYLWVDQ
 367 .EAFKAALDRLEDQEMARELGIALAEKI.QQNLVFSRRTDIPSSPYRDDNRYLWQTP
 381 .AFAFRQTSD.LDSNQVSWNLGVDSLHRLFELDLPYHKLTDALSAIYRDDNRAVWHLE
 365 .VAFRAQTDLSTNQTSWNLGIDLSHKLFLFELPPEFNNRTDHRLSAIYRDDNRAVWQLE
 364 .AHFRAQPD.LLDAKFSWNVGIDLSHPLAALDLPEYNNRTDG.LRALYRDDNRAIWELE
 364 .AHFRAQPD.LLDAKFSWNVGIDLSHPLAALDLPEYNNRTDG.LRALYRDDNRAIWELE
 352 .QAFMAAEARWDDPQLSLWDLGVRLADSL.RKEKVTYHARTDI.PDVIYRDDNRYIWEFE
 353 .SLFTQAQGRLENNQTSWDIGVRLAAKI.MDKRIAFNPRTDI.LTAPYRDDNRYFWKFE
 353 .LFSQVQDRLTDNQICWDIGVQLAHEI.MAKKIPFNPRTDV.LTALYRDDNRYFWFEFE
 353 .RLFAAAQDKLDSHQISWDIGVQLAHSF.REKNIPFAMLTDL.LTAPYRDDNRYFWFEFE
 358 .TLFAAATDRLGNNQTSWDIGVQLAHSI.ADKKIPFNPRTDV.LSAPYRDDNRYFWFEFE
 358 .ALFAAAQDRLGDNQVSWDIGVQLAHNL.TEKRIPFNPRTDI.LAAPYRDDNRYFWFEFE
 353 .ALFAAAQNRLHNDQISWDIGVQLAHLQADKRLPFSPRTDL.LTAPYRDDNRYFWFEFE
 356 .FYAEVEEQDDNEKWSWDMGITLYNTLNKSLTRWESLSTEISKIRLAYRDDNRLVWASD
 374 .VLFKQAAEDRLDSNQISWDIGVQLAHEF.LAKRIPFNPRSDL.LTTPYRDDNRYFWFEFE
 357 .VLFKQAAESLDTNQVSWDIGVQLAHEF.MAKRIAFNPRTDI.LTAPYRDDNRYFWFEFE
 356 .FYAEIEENQDNEKWSWDMGITLYNLLNKSLTRWESLSTEISKIRLAYRDDNRLVWASD
 359 .FYNEIDHLEDEKWSWDMGLTLYNALNQAVTKWESLTDISRTREFGYRDDNRLVWASE
 356 .FHEQIEDNLDNEKWSWDMGITLYNVLNQATTKWESLTDISRLRFAYRDDNRLVWASE
 356 .FHNIEIENLDNEKWSWDLGITLYNVLNKATSKWESLTDISQREFGYRDDNRLVWASE
 356 .FYAEIDNLDNEKWSWDLGITLYNVLNISTNKWVSLTELNKQREFGYRDDNRFVWASE
 352 .FHDGIELKKLDSFSWHLGMDLYHLFVQRK.QVPSLRILQSIKIPYRDDNRFVWFEFE
 354 .FHANIDRQDDQHFSWLMGMELFNIFSARR.EVPSLRILERIRIPYRDDNRIVWFEFE
 382 .FQAEIVGGN..ADTASMARIALDYVTTTRRQ.SDQFAKMHFADTEIN.YRDDNRGLWRFI
 394 .FHARLADGT..STSEMDLALSYVAKTRRP.SDQFAKVHFDDTVVDYRDDNRQLWKFI
 368 .FHAVMEAGA..STKEIAALALSYVARTTRRQ.SDQFAKVHFDDTVVDYRDDNRQLWKFI
 378 .FHALMEQPT..TTTNEVALLALSYVARTTRRQ.SDQFSKVHFDDTVVDYRDDNRQLWKFI
 377 .FHALMAQGT..STRDIAALALAYVAKTRRQ.SDQFAKVHFDDTVVDYRDDNRHLWRFI
 378 .FHALMAQEA..STQDIAALALAYVAKTRRQ.SDQFAKVHFDDTVVDYRDDNRHLWRFI
 379 .FHALLENGA..ASTQEMADLGVAFIAKTRLN.SDNFANTYFKDITVDYRDDNRQMWKFI
 380 .FHEIMAKGK..SGTEEIAADLAISYVAKTRRQ.SDQLAKMYFDNTVVDYRDDNRHMRFH
 381 .FHELMAGGQ..STSAAMASVALAFIARTRQ.RDLSANMYFDNTVVDYRDDNRHLWFI
 370 .FFAEMKEGE..SSSSAIAIALSFIAARTRQ.TDASPNTYFENTEVDYRDDNRHLWFI
 381 .FQALLRTGP..SNSSEIAALALDFVTRTRLP.SDQFARVHFADTIVDYRDDNRHFRFI
 365 .FFAVMEQGE..TTTMDMETLGIQFMARTRLQ.SDALPNVWFDDTEISYRDDNRNLWQFY
 373 .FHRLMAEKGENSTTADMAEIGIDFYVKSRRKSSDSMPSVFFDDTEISYRDDNRYIWFY
 373 .FHKLMAEKGENSTTADMAEIGIDFYVKSRRKSSDSMPSVFFDDTEISYRDDNRYIWFY
 373 .FHKLMAEKGEDSTTQDMSLGIQFMARTRLQ.SDALPNVWFDDTEISYRDDNRNLWQFY
 373 .FQAVMSEKGEHSTTQDMAKLGIDFYVKSRRKSSDSLPDIFDDTEISYRDDNRYIWFHH
 373 .FHTILQEKGEKSTTQDMAKLGIDFYVKSRRKSSDSLPDIFDDTEISYRDDNRYIWFHH
 377 .FNAIMAEKGDASTTQDMAKLGIDFYVKSRRKSSDSLPNVFFDDTEISYRDDNRFIWFHY
f.e.e.d.....q...dlgidl.....q....f.....YrDDNry.w.fe

AfO2_|B7J5E5|
 AfO1_|B7JA27|
 AfO3_|B7JB16|
 Acidithiobacillus_ferrovorans_O3_|G0JL85|
 Ferrovum_sp_O3_|A0A0R0LL45|
 Thioalkalivibrio_sp_O3_|D3SG07|
 Thioalkalivibrio_sulfidiphilus_O3_|B8GQN4|
 Acidimicrobium_ferrooxidans_O3_|C7M203|
 Acidithiobacillus_caldus_O3_|F9ZLQ2|
 Thiiothrix_nivea_O3_|I3BSU7|
 Halothiobacillus_neapolitanus_O3_|D0KZ80|
 Nitrobacter_hamburgensis_O3_|Q1QFQ8|
 Nitrosomonas_eutropha_O3_|Q0AHU9|
 Thiomonas_arsenitoxydans_O3_|D6CQL4|
 Thiomonas_intermedia_O3_|D5X344|
 Acidithiobacillus_ferrovorans_O1_|G0JNK3|
 Nitrosomonas_communis_O1_|A0A0F7KHE0|
 Nitrosomonas_europaea_O1_|Q82TG9|
 Nitrosomonas_sp_O1_|F9ZF96|
 Thiobacillus_denitrificans_O1_|Q3SFN2|
 Thiobacillus_thioparus_O1_|UPI0003621E4A|
 Thiomonas_sp_O1_|A0A078BJA6|
 Hydrogenovibrio_marinus_O1_|Q75W42|
 Sideroxydans_lithotrophicus_O1_|D5CQ62|
 Sulfuricella_denitrificans_O1_|S6AAH0|
 Thiomicrospira_kuenenii_O1_|UPI000466D9A2|
 Thiomicrospira_halophila_O1_|UPI000378C5A3|
 Thiomicrospira_chilensis_O1_|UPI000429B796|
 Thiomicrospira_crunogena_O1_|Q31IJ7|
 Thiomicrospira_sp_O1_|UPI000493B602|
 Mariprofundus_ferrooxydans_O1_|Q0EX19|
 Thiiothrix_nivea_O1_|I3BR12|
 Acidithiobacillus_ferrovorans_O2_|G0JPS1|
 Thiobacillus_denitrificans_O2_|Q3SFL7|
 Thiobacillus_thioparus_O2_|UPI00035FE92D|
 Halothiobacillus_neapolitanus_O2_|D0KZQ5|
 Thiomonas_arsenitoxydans_O2_|D6CTS8|
 Thiomonas_intermedia_O2_|D5X1M7|
 Sideroxydans_lithotrophicus_O2_|D5CTG0|
 Sulfuricella_denitrificans_O2_|S6AJV8|
 Gallionella_capsiferriiformans_O2_|D9SHP8|
 Mariprofundus_ferrooxydans_O2_|Q0F300|
 Acidithiobacillus_caldus_O2_|F9ZR07|
 Thiola pillus_brandeum_O2_|W0TP28|
 Hydrogenovibrio_marinus_O2_|Q75W24|
 Thiomicrospira_kuenenii_O2_|UPI000467D206|
 Thiomicrospira_sp_O2_|UPI000493F397|
 Thiomicrospira_halophila_O2_|UPI000399C451|
 Thiomicrospira_chilensis_O2_|UPI00040A3D56|
 Thiomicrospira_crunogena_O2_|Q31IK5|
 consensus>50

437 E S G D E . E E A F D A E Q R K A P D L
 408 E I D W E H E T T I L P G . . . A Q V R K Y V N V M E M V N E V D T E L A G . D D A Q E V W V L Q S E L F P Y E D M G .
 419 D I R A . D G Q V I A G V T . T Q Q I R K Y V S V M E M V N A L D V P G A G . D D A N E I W V L S S E F F R D E E P T .
 419 D V R E . V G Q V I A G V T . T Q Q I R K Y V S V M E M V N A I D V P G A G . D D A N E I W V L S S E F F R D E E P T .
 419 D I R E . T T Q V I A G M N . S H Q I R K T V S V M E M I N A L D V P G A G . D D A N E I W V L K S E F F R D E E T T .
 409 R E E E . A A E V L A A H G . S G Q E R F Y V T P M Q M I N T L D V P N A G . D D A Q Q V M V L A T E W F R D G E E E .
 409 R E E E . E P E H V R L G D . A R Q V R K Y V S L M E M L N A V D V P G A G . D D A Q E I W V L A T E F F R D E E E T .
 416 E A I R . A Q A A N L D L A . P S Q V R K Y V S V M E M I N E V D V P S A G . D D A E I W V L A T E F F R D G E A E .
 423 A S Q . . L G D I L A G G P . P . Q V R K Y V S L M E M I N A V D V E T A G . D D A Q E I W V L A T E F F D D . D G T .
 419 T E E . . A D G I L L P G Q . T K Q V R K Y V S V M E M W N N L D V E F A G . D D A Q E V W V L A T E W F H D . D G K .
 423 E T G . . K S N A H T A A S . I Q E K R K Y V S V M Q M L H A L D V E Y A T A E S A D E I W V L A T E L F D D . D G T .
 437 E Y D E A Q V L A A T W E T . R Q Q V R K K V N V M E M V N E V D N E F A G . D D A Q E V W E L P T E F Y L D Q E D . .
 422 E Y D E A Q A L E A T W Q T . R Q Q V R K K V N L M E M V N E V D N E F A G . D D A Q E V W E L P T E Y L F D D G . . .
 420 E Y D E A Q A L Q A T W E T . Q Q Q V R R K V S V M E M V N E V D N E F A G . D D A Q E V W I L P T E F Y L D Q E G . .
 420 E Y D E A Q A L Q A T W E T . Q Q Q V R R K V S V M E M V N E V D N E F A G . D D A Q E V W I L P T E F Y L D Q E G . .
 408 E V D W E H E A T I L P G . . . A Q V R K Y V N V M E M V N E V D T E L A G . D D A Q E V W V L Q S E L F P Y E D M G .
 409 E F D F E Q S L T A G Y E E . P K Q I R K Y V S V M E M I N E V D V E T A G . D D A E I W V M G T E F F P Y E D S G .
 409 E F D F E R S I A A G Y E P . L K Q V R K H V S V M E F I N E L D V E T A G . D D A Q E I W V L G T E L F P Y E D Y G .
 409 E F D L N K S L S A G Y E A . P K Q V R K Y V N V M E F A N E V E V E S A G . D D A Q E I W V M G T E F F P Y E D M G .
 414 E F D F E Q S L T A G Y E E . V K Q V R K N V S V M E M I N E I D V E T A G . D D A E I W V L S T E L F P Y E N I G D
 414 E F D F D K A A G A G Y E S . V K Q V R K N V S V M E M V N A I D V E T A G . D D A G E I W V L S T E L F P Y E N I G D
 410 E F D F D K A A N A G Y E S . I K Q V R K N V S V M E M A N E I D V E N A G . D D A E I W V L S T E L F P Y E D M G .
 413 E W G E V E G . . . G G A P . T E T V R K H V S L M E M I N E I D S E L V D . V D H D E V W V L G T E L Y P Y E D . . .
 430 E F D F D Q A L A A G Y E A . P K Q V R K H V S L M E F A N E L D V E T A G . D D A E I W V L G T E L F P Y E D M G .
 413 E F D F E K A M S A G Y E A . P K Q V R K H V S L M E F A N E I D V E N A G . D D A E I W V L G T E L F P Y E D M G .
 413 E W G E A E G . . . S G A P . T E T V R K H V S L M E M I N E I D S E L V D . V D H D E V W V L G T E L Y P Y E D . . .
 416 E W G E M A G . . . M G A A Q Q E T V R K H V S L M E M V N E I D S E L E D . V D H D E V W V L G T E L F P Y E D . . .
 413 E W S E M E G . . . G S G Y K E T V R K H V S V M E M I N E I D S E L L D . V D H D E V W V L G S E F F P Y E D . . .
 413 E W A E M E G . . . G G A P H Q E T V R K N V S L M E M I N E I D S E L V D . V D H E V W V L G S E L Y P Y E D . . .
 413 E W A E M E G . . . S G S A Y Q E T V R K H V S V M E M I N E I D S E L V D . V D H D E V W V L G S E L M P Y E D . . .
 408 A F N F A E H G L E Y I P A S Q R Q V R R Y V S A V E M A N E V D C E L A G . D D A Q E V W I C A T E F F P Y E D A G E
 410 E L T W D V G . F E Y V P A S Q R Q V R K T V S V M E I A N E V D C E L A G . D D A Q E I W T C S T E M Y P Y E D D L E
 438 E A G D E . E E S F D E E R R T A P D P
 450 E E G D E . E E A F D A K R K I E P G E
 424 E A G D E . E E V F D E E R K I E P G E
 434 E A G D E . E E A F D E Q R K I K P . Q
 433 E Q G D E . E E R F D E T R R N D A Q D
 434 E Q G D E . E E R F D E T R R N D A Q D
 435 E E G D E . E E A F D E P R K V E P G E
 436 E Q S D D . E E M F D D H R K I E L E E K
 437 E E S D E . D E M F E H L D Q R K N S E V
 426 E N S D E . E E S Y E K . E P K V Q E E E
 437 E P E G G . E T S T S D T F R P S S S P
 421 E L D D D . E D F H V E P T E L E E A E
 432 E E Y D E P D E I L P N E Y E N E E K V V D
 432 E E Y D E P D E I L P N E Y E N E E R V V E
 432 E E Y D E P D E I L P N E Y E N E E R I V D
 432 E E N D E S E D F H Q N E Y E A E E K I V D
 432 E E N D E A E D F H N N E Y E S D E K V V D
 435 E E N D E A E D F H N N E Y H S D E K I V D
 e . . d e . . e e q v r k . v s . m e m . n . . . d . e . a . . . d d a q e i w v l . . e . f . . e d . . .

AfO2_|B7J5E5|
 AfO1_|B7JA27|
 AfO3_|B7JB16|
 Acidithiobacillus_ferrivorans_O3_|G0JL85|
 Ferrofum_sp_O3_|A0A0R0LL45|
 Thioalkalivibrio_sp_O3_|D3SG07|
 Thioalkalivibrio_sulfidiphilus_O3_|B8GQN4|
 Acidimicrobium_ferrooxidans_O3_|C7M203|
 Acidithiobacillus_caldus_O3_|F9ZLQ2|
 Thiobacillus_nivea_O3_|I3BSU7|
 Halothiobacillus_neapolitanus_O3_|D0KZ80|
 Nitrobacter_hamburgensis_O3_|Q1QFQ8|
 Nitrosomonas_eutropha_O3_|Q0AHU9|
 Thiomonas_arsenitoxydans_O3_|D6CQL4|
 Thiomonas_intermedia_O3_|D5X344|
 Acidithiobacillus_ferrivorans_O1_|G0JNK3|
 Nitrosomonas_communis_O1_|A0A0F7KHE0|
 Nitrosomonas_europaea_O1_|Q82TG9|
 Nitrosomonas_sp_O1_|F9ZF96|
 Thiobacillus_denitrificans_O1_|Q3SFN2|
 Thiobacillus_thioparus_O1_|UPI0003621E4A|
 Thiomonas_sp_O1_|A0A078BJA6|
 Hydrogenovibrio_marinus_O1_|Q75W42|
 Sideroxydans_lithotrophicus_O1_|D5CQ62|
 Sulfuricella_denitrificans_O1_|S6AAH0|
 Thiomicrospira_kuenenii_O1_|UPI000466D9A2|
 Thiomicrospira_halophila_O1_|UPI000378C5A3|
 Thiomicrospira_chilensis_O1_|UPI000429B796|
 Thiomicrospira_crunigena_O1_|Q31IJ7|
 Thiomicrospira_sp_O1_|UPI000493B602|
 Mariprofundus_ferrooxydans_O1_|Q0EX19|
 Thiobacillus_nivea_O1_|I3BR12|
 Acidithiobacillus_ferrivorans_O2_|G0JPS1|
 Thiobacillus_denitrificans_O2_|Q3SFL7|
 Thiobacillus_thioparus_O2_|UPI00035FE92D|
 Halothiobacillus_neapolitanus_O2_|D0KZQ5|
 Thiomonas_arsenitoxydans_O2_|D6CTS8|
 Thiomonas_intermedia_O2_|D5X1M7|
 Sideroxydans_lithotrophicus_O2_|D5CTG0|
 Sulfuricella_denitrificans_O2_|S6AJV8|
 Gallionella_capsiferriiformans_O2_|D9SHP8|
 Mariprofundus_ferrooxydans_O2_|Q0F300|
 Acidithiobacillus_caldus_O2_|F9ZR07|
 Thiobacillus_brandenburg_O2_|W0TP28|
 Hydrogenovibrio_marinus_O2_|Q75W24|
 Thiomicrospira_kuenenii_O2_|UPI000467D206|
 Thiomicrospira_sp_O2_|UPI000493F397|
 Thiomicrospira_halophila_O2_|UPI000399C451|
 Thiomicrospira_chilensis_O2_|UPI00040A3D56|
 Thiomicrospira_crunigena_O2_|Q31IK5|
 consensus>50

456ETQGLPFRHYPEWDYQSQSYRPDWVSLYEGLHASAPAAATIDQLLLKHA
 463L.SYNQMEGKEPVSDPFHYPEWDYQIQASRPDWVSLYEGLHASAPAAATIDQLLLKHA
 475SLNQQEGREPPDPFHYPEWDYQMQLEPPEWCTVLEKRAKSGTLETIHEIVAKHK
 475SLNQQEGREPPDPFHYPEWDYQMQLEPPEWCTVLEKRAKSGALEIDEIVAKHK
 475SINQQEGREPPDPFHYPEWDYQMQLEPPEWCTVLEKRAKSGDVKIINDIVVKKH
 465SINQQEGKEPISLPPHYPEWDYQMQLEPPEWCTVLEKRAKSGEITEAIVYDKHK
 465SINEQEGKAPISLPPHYPEWDYQMQLEPPEWCTVLEKRAKSGDLSVLEGLAEKHH
 472SVNAQEGREPPDPFHYPEWDYQVQLEPPEWCTVLEKRAKSGPLGSLEEAEAILAKHR
 476TFNEREGKAPVSQPVSYPEWDYQVQLEPPEWCTVLEKRAKSGDPEVTHAFLEAQK
 473SLNEQEGKPPMASPVHYPEWDYQVQLEPPEWCTVLEKRAKSGEVAEIDAATERNK
 478TFNELYGTPTASPRHYPEWDYQVQLEPPEWCTVLEKRAKSGDPAVVDNIFLEHA
 493VINSLEGRPPVSEPFHYPEWDYQIQLEPPEWCTVLEKRAKSGELVIEGIIIEHK
 477VINSLEGRPPVSEPFHYPEWDYQIQLEPPEWCTVLEKRAKSGELVIEGIIIEHK
 476VINSLEGRPPVSEPFHYPEWDYQIQLEPPEWCTVLEKRAKSGELVIEGIIIEHK
 476VINSLEGRPPVSEPFHYPEWDYQIQLEPPEWCTVLEKRAKSGELVIEGIIIEHK
 463L.SYNQMEGKEPISLPPHYPEWDYQIQASRPDWVSLYEGLHASAPAAATIDQLLLKHA
 466V.SYNQMEGKEPISLPPHYPEWDYQIQLEPPEWCTVLEKRAKSGDLSVLEGLAEKHH
 466V.SYNQMEGKEPISLPPHYPEWDYQIQLEPPEWCTVLEKRAKSGDLSVLEGLAEKHH
 466V.SYNQMEGKEPISLPPHYPEWDYQIQLEPPEWCTVLEKRAKSGDLSVLEGLAEKHH
 472V.SYNQMEGKEPISLPPHYPEWDYQIQLEPPEWCTVLEKRAKSGDLSVLEGLAEKHH
 472V.SYNQMEGKEPISLPPHYPEWDYQIQLEPPEWCTVLEKRAKSGDLSVLEGLAEKHH
 467K.SYNQMEGKEPISLPPHYPEWDYQIQLEPPEWCTVLEKRAKSGDLSVLEGLAEKHH
 465HGV.SYNQMEGKEPISLPPHYPEWDYQIQLEPPEWCTVLEKRAKSGDLSVLEGLAEKHH
 487K.SYNQMEGKEPISLPPHYPEWDYQIQLEPPEWCTVLEKRAKSGDLSVLEGLAEKHH
 470K.SYNQMEGKEPISLPPHYPEWDYQIQLEPPEWCTVLEKRAKSGDLSVLEGLAEKHH
 465NGV.SYNQMEGKEPISLPPHYPEWDYQIQLEPPEWCTVLEKRAKSGDLSVLEGLAEKHH
 469HGQ.SYNQMEGKEPISLPPHYPEWDYQIQLEPPEWCTVLEKRAKSGDLSVLEGLAEKHH
 466NGI.SYNQMEGKEPISLPPHYPEWDYQIQLEPPEWCTVLEKRAKSGDLSVLEGLAEKHH
 466NGI.SYNQMEGKEPISLPPHYPEWDYQIQLEPPEWCTVLEKRAKSGDLSVLEGLAEKHH
 466NGI.SYNQMEGKEPISLPPHYPEWDYQIQLEPPEWCTVLEKRAKSGDLSVLEGLAEKHH
 467ATT.SYNQMEGKEPISLPPHYPEWDYQIQLEPPEWCTVLEKRAKSGDLSVLEGLAEKHH
 468NTK.SYNQMEGKEPISLPPHYPEWDYQIQLEPPEWCTVLEKRAKSGDLSVLEGLAEKHH
 457E.AQGLPFRHYPEWDYQSQSYRPDWVSLYEGLHASAPAAATIDQLLLKHA
 469E.IQGLPFRHYPEWDYQSQSYRPDWVSLYEGLHASAPAAATIDQLLLKHA
 443E.IRGLPFRHYPEWDYQSQSYRPDWVSLYEGLHASAPAAATIDQLLLKHA
 452E.VQGLPFRHYPEWDYQSQSYRPDWVSLYEGLHASAPAAATIDQLLLKHA
 452E.DLRLPFRHYPEWDYQSQSYRPDWVSLYEGLHASAPAAATIDQLLLKHA
 453E.DLRLPFRHYPEWDYQSQSYRPDWVSLYEGLHASAPAAATIDQLLLKHA
 454E.LKGLPFRHYPEWDYQSQSYRPDWVSLYEGLHASAPAAATIDQLLLKHA
 456E.EKGLPFRHYPEWDYQSQSYRPDWVSLYEGLHASAPAAATIDQLLLKHA
 457E.LQALPFRHYPEWDYQSQSYRPDWVSLYEGLHASAPAAATIDQLLLKHA
 445E.LKGLPFRHYPEWDYQSQSYRPDWVSLYEGLHASAPAAATIDQLLLKHA
 456E.EKGLPFRHYPEWDYQSQSYRPDWVSLYEGLHASAPAAATIDQLLLKHA
 440P.DDTLPFRHYPEWDYQSQSYRPDWVSLYEGLHASAPAAATIDQLLLKHA
 454N.EGGLPFRHYPEWDYQSQSYRPDWVSLYEGLHASAPAAATIDQLLLKHA
 454N.EGGLPFRHYPEWDYQSQSYRPDWVSLYEGLHASAPAAATIDQLLLKHA
 454N.EGGLPFRHYPEWDYQSQSYRPDWVSLYEGLHASAPAAATIDQLLLKHA
 454DANS.LPFRHYPEWDYQSQSYRPDWVSLYEGLHASAPAAATIDQLLLKHA
 454TADSLPFRHYPEWDYQSQSYRPDWVSLYEGLHASAPAAATIDQLLLKHA
 457DANS.LPFRHYPEWDYQSQSYRPDWVSLYEGLHASAPAAATIDQLLLKHA
s.ne.eg.ep...pfhy.EWDYq.ql.RPDw.tvy#.r...gd...id.ilekhh

AfO2_|B7J5E5|
 AfO1_|B7JA27|
 AfO3_|B7JB16|
 Acidithiobacillus_ferrivorans_O3_|G0JL85|
 Ferrovum_sp_O3_|A0A0R0LL45|
 Thioalkalivibrio_sp_O3_|D3SG07|
 Thioalkalivibrio_sulfidiphilus_O3_|B8GQ4|
 Acidimicrobium_ferrooxidans_O3_|C7M203|
 Acidithiobacillus_caldus_O3_|F9ZLQ2|
 Thiiothrix_nivea_O3_|I3BSU7|
 Halothiobacillus_neapolitanus_O3_|D0KZ80|
 Nitrobacter_hamburgensis_O3_|Q1QFQ8|
 Nitrosomonas_eutropha_O3_|Q0AHU9|
 Thiomonas_arsenitoxydans_O3_|D6CQL4|
 Thiomonas_intermedia_O3_|D5X344|
 Acidithiobacillus_ferrivorans_O1_|G0JNK3|
 Nitrosomonas_communis_O1_|A0A0F7KHE0|
 Nitrosomonas_europaea_O1_|Q82TG9|
 Nitrosomonas_sp_O1_|F9ZF96|
 Thiobacillus_denitrificans_O1_|Q3SFN2|
 Thiobacillus_thioparus_O1_|UPI0003621E4A|
 Thiomonas_sp_O1_|A0A078BJA6|
 Hydrogenovibrio_marinus_O1_|Q75W42|
 Sideroxydants_lithotrophicus_O1_|D5CQ62|
 Sulfuricella_denitrificans_O1_|S6AAH0|
 Thiomicrospira_kuenenii_O1_|UPI000466D9A2|
 Thiomicrospira_halophila_O1_|UPI000378C5A3|
 Thiomicrospira_chilensis_O1_|UPI000429B796|
 Thiomicrospira_crunogena_O1_|Q31IJ7|
 Thiomicrospira_sp_O1_|UPI000493B602|
 Mariprofundus_ferrooxydants_O1_|Q0EX19|
 Thiiothrix_nivea_O1_|I3BR12|
 Acidithiobacillus_ferrivorans_O2_|G0JPS1|
 Thiobacillus_denitrificans_O2_|Q3SFL7|
 Thiobacillus_thioparus_O2_|UPI00035FE92D|
 Halothiobacillus_neapolitanus_O2_|D0KZQ5|
 Thiomonas_arsenitoxydants_O2_|D6CTS8|
 Thiomonas_intermedia_O2_|D5X1M7|
 Sideroxydants_lithotrophicus_O2_|D5CTG0|
 Sulfuricella_denitrificans_O2_|S6AJV8|
 Gallionella_capsiferriiformans_O2_|D9SHP8|
 Mariprofundus_ferrooxydants_O2_|Q0F300|
 Acidithiobacillus_caldus_O2_|F9ZR07|
 Thioliapillus_brandeuum_O2_|W0TP28|
 Hydrogenovibrio_marinus_O2_|Q75W24|
 Thiomicrospira_kuenenii_O2_|UPI000467D206|
 Thiomicrospira_sp_O2_|UPI000493F397|
 Thiomicrospira_halophila_O2_|UPI000399C451|
 Thiomicrospira_chilensis_O2_|UPI00040A3D56|
 Thiomicrospira_crunogena_O2_|Q31IK5|
 consensus>50

504 ALAKHLKRLLDLLKPKQDKVRIRYQEEGSELDLDVALRSWIDFKSGSTPDPRIINMSHRTAG
 519 PVSSRLRYLIEALQPQGVQRLRKQEDGDELDVEAAVRALVDIRMGLAPDTRIHMRS LRKV
 530 PIIGRLKF LIEALQPQGVQRLRKQEDGDELDLNAAVRAMIEMRMGEQPDPRIMMRNVRKV
 530 PVIGRLKY LIEALQPQGVQRLRKQEDGDELDLNAAVRAMIEMRMGEQPDPRIMMRNVRKV
 530 PIIGRLKH LIEAMQPQGVQRLRKVEDGDELDLNAAVLAMTEMRRGEQPNPRIMIRNVRKV
 520 PLVSRLKF LIEAMQPQGVQRLRKQFEGDELDVNALVEAQIDVRMRRQPDRIYVRN LRHV
 520 PLIARIKY LIEAMQPQGVQRLRKQFEGDELDVNALVHAMVDQRMGQQPDPRVFIRN IRKV
 527 PLLGRMRIMEALAPEGVQRLRKQEEGEELDINAADVFAVDLRRAEEDL RVMVRT RRV
 531 PLVQRLKK LIEAVQPQGVIRQRKVEDGDELDINAADVLAFTDIRMGQQPDPRIGIRT RLHT
 528 PIISRLKH LIEAVQPQGVVRQRKQEDGDTIDINAADVAMVSIRMGSQPDPRINIRT HLQI
 533 PMVRLKY LIEGLQPQGVVRQKKIEDGDEVDLNAATGAMIDIRMNRQPDPRIGIRRIHV
 549 PVISRLKF LIESMVPQGMQRI RHMEDGDELDINAAVRAMIDIRIGQQHDP RIMIRY KRAL
 533 PVISRLKF LIESMVPQGMQRI RHVEDGDDLDINAAVRAMIDIRMGPDP RIMMRY KRTL
 532 PVVSRLKF LIESMIPQGLQRI RQEDGDDLDINAAVRAMIDVRMGVQPDPRIMVRL KRAQ
 532 PVVSRLKF LIESMIPQGLQRI RQEDGDDLDINAAVRAMIDVRMGVQPDPRIMVRL KRAQ
 519 PVSSRLKY LIEALQPQGVQRLRKQEDGDEIDMEAAALRALVDIRMGLAPDTRIHMRT LRKL
 522 REISRMKF LLDAMQPQGVQRI RKLDEGDEIDINAATIHSTIDIRMGLQPDPRIMMRS VRKI
 522 REIHRMKF LLDAMQPQGVQRI RKLDEGDEIDINAATISSLVDIRMGHDP RVMMAVRKV
 522 REIARMKF LLDAMQPQGTRI RKLDEGDEIDINAATIRSMIDIRMGMQPDPRIMMRS VRKV
 532 REIHRMKF LLDAMQPQGVQRI RKLDEGDEIDINAATIQSFDIRLGNQPDPRIMMRS VRKT
 532 REIHRMKF LLDAMQPQGVQRI RKLDEGDEIDINAATISFDIRLGNQPDPRIMMRS VRKT
 523 REIHRMKF LLDAMQPQGVQRI RKLDEGDEIDINAATISFIALRMGQQPDPRIMMRS VRTV
 523 QIAHRIKQ IVDKLSVGLQRI RRIEDGDEIDLNACVEAITALRMGQEPDP RITMKN VIRT
 543 RTIGRLKF LLDALQPQGVTRIRKLEDEGDELDVNAAIRAMIDMRMGQPDPRIMMRS VRKV
 526 RTIGRLKF LLDALQPQGVTRIRKLEDEGDELDVNAAIRAMIDIRMGQQPDPRIMMRS VRKI
 523 QIAHRIKQ IVDKLSVGLQRI RRIEDGDEIDLNACVEAITALRMGQEPDP RITMKN VINT
 527 GIAHRIKQ IVDKLSVGLQRI RRIEDGDEIDLNACVEAITSLRMGQEPDP RITMKN VIRS
 524 GIAHRIKQ IVDKLQAVGLQRI RRIEDGDEIDLNACVEAITSLRMGQEPDP RITMKN VIRS
 524 GIAHRIKQ IVDKLQAVGLQRI RRIEDGDEIDLNACVEAITSLRMGHEPDP RITMKN VIRS
 524 GIAHRIKQ IVDKLQAVGLQRI RRIEDGDEIDLNACVEAITAIRMGEPPDP RITMKN VIRS
 525 PVAHRIKQ IIDLLSPQGVQRI RNMEDGDEIDLNAAVDAMI AIRMGEQPNPRITMRN VIKN
 526 PVAHRIKQ IIDLLTPEGVQRI RQGLDEGDEVDINAADVAMISIRMGEQPNPRITMRN VLKS
 505 TLARRLKRLLDLLKPKQDKVRIRYQEEGSELDLDVALRSWIDFKSSPPDPRINMSHRTAG
 517 ALAKRLKKMLDLLKPKQDKVVRVRYQEEGSELDLDVAIRSLIDFKGGATPDPRIINMSHRTAG
 491 ALAKRLKKMLDLLKPKQDKVRIRYQEEGSELDLDVAIRSLIDFKGGATPDPRIINMSHRTAG
 500 ALAKQLKKMLDALPKQDKVRIRYQEEGSELDLDVAIRSLIDYKSGANPDPRIINMSHRTAG
 500 PLAKRLQRMLELLKPKQDKVRIRYQEEGSELDLDVAIRSLIDFKAGSNPDPRIINLSHRTAG
 501 PLAKRLQRMLELLKPKQDKVRIRYQEEGSELDLDVAIRSLIDFKAGSNPDPRIINLSHRTAG
 502 GLAKRLKKMLDLLKPKQDKVRIRYQEEGSELDLDVAIRSLIDYKSGAQPDPRINMSHRTAG
 504 ALAKRLKKMLDLLKPKQDKVRIRYQEEGSELDLDVAIRSLIDYKSGANPDPRIINMSHRTAG
 505 ALAKRLKKMLDLLKPKQDKVRIRYQEEGSELDLDVAIRSLIDFKGGSPDPRINMSHRTAG
 493 GLAKRLKKMLEMLKPKQDKVRIRYQEEGSELDLDVAIRSLIDFKGGCAPDPRINMSHRTAG
 504 VLSNRLRLFDLLKPKQEREIRRQEDGSELDLMALRALIDYQMGMPDPRISLSHSHSNG
 488 RLAKQIKRMLDLLKPKQNYQRIRYQEEGTELDLDIAIRSLIDYKTGHDPDPRINISHKHND
 502 DLANRLKKMIEMLKPKQNKRI RFQEEGEELDLIALRSVIDFKSGQAPDPRINYSHTTDS
 502 TLAKRLKKMIEMLKPKQNKRI RFQEEGEELDLVALRSVIDFKSGQAPDPRINYSHTTDS
 502 ALAKRLKKMIEMLKPKQNKRI RFQEEGEELDLVALRSVIDFKSGQAPDPRINYSHTTDS
 502 DLAKRLKKMIEMLKPKQNKRI RFQEEGEELDLVAIRSLIDFRSGHAPDPRINYSHTTDS
 502 GLAKRLKKMIEMLKPKQNKRI RFQEEGEELDLVALRSVIDFKSGQAPDPRINYSHTTDS
 502 GLAKRLKKMIEALKPKQNKRI RFQEEGEELDLIAIRSVDFKSGQAPDPRINYSHTTDS
 505 ..a.rlk.li#.l.pqg.qRir.qe#GdelDl#aavralid.rmgqqp#pR!.mr....

Af02_|B7J5E5|
Af01_|B7JA27|
Af03_|B7JB16|
Acidithiobacillus_ferrovirans_O3_|G0JL85|
Ferrovum_sp_O3_|A0A0R0LL45|
Thioalkalivibrio_sp_O3_|D3SG07|
Thioalkalivibrio_sulfidiphilus_O3_|B8GQN4|
Acidimicrobium_ferroxidans_O3_|C7M203|
Acidithiobacillus_caldus_O3_|F9ZLQ2|
Thiothrix_nivea_O3_|I3BSU7|
Halothiobacillus_neapolitanus_O3_|D0KZ80|
Nitrobacter_hamburgensis_O3_|Q1QFQ8|
Nitrosomonas_eutropha_O3_|Q0AHU9|
Thiomonas_arsenitoxydans_O3_|D6CQL4|
Thiomonas_intermedia_O3_|D5X344|
Acidithiobacillus_ferrovirans_O1_|G0JNK3|
Nitrosomonas_communis_O1_|A0A0F7KHE0|
Nitrosomonas_europaea_O1_|Q82TG9|
Nitrosomonas_sp_O1_|F9ZF96|
Thiobacillus_denitrificans_O1_|Q3SFN2|
Thiobacillus_thioparus_O1_|UPI0003621E4A|
Thiomonas_sp_O1_|A0A078BJA6|
Hydrogenovibrio_marinus_O1_|Q75W42|
Sideroxydans_lithotrophicus_O1_|D5CQ62|
Sulfuricella_denitrificans_O1_|S6AAH0|
Thiomicrospira_kuenenii_O1_|UPI000466D9A2|
Thiomicrospira_halophila_O1_|UPI000378C5A3|
Thiomicrospira_chilensis_O1_|UPI000429B796|
Thiomicrospira_crunogena_O1_|Q31IJ7|
Thiomicrospira_sp_O1_|UPI000493B602|
Mariprofundus_ferroxydans_O1_|Q0EX19|
Thiothrix_nivea_O1_|I3BR12|
Acidithiobacillus_ferrovirans_O2_|G0JPS1|
Thiobacillus_denitrificans_O2_|Q3SFL7|
Thiobacillus_thioparus_O2_|UPI00035FE92D|
Halothiobacillus_neapolitanus_O2_|D0KZQ5|
Thiomonas_arsenitoxydans_O2_|D6CTS8|
Thiomonas_intermedia_O2_|D5XIM7|
Sideroxydans_lithotrophicus_O2_|D5CTG0|
Sulfuricella_denitrificans_O2_|S6AJV8|
Gallionella_capsiferriiformans_O2_|D9SHP8|
Mariprofundus_ferroxydans_O2_|Q0F300|
Acidithiobacillus_caldus_O2_|F9ZR07|
Thiolapillus_brandeam_O2_|W0TP28|
Hydrogenovibrio_marinus_O2_|Q75W24|
Thiomicrospira_kuenenii_O2_|UPI000467D206|
Thiomicrospira_sp_O2_|UPI000493F397|
Thiomicrospira_halophila_O2_|UPI000399C451|
Thiomicrospira_chilensis_O2_|UPI00040A3D56|
Thiomicrospira_crunogena_O2_|Q31IK5|
consensus>50

AfO2_|B7J5E5|
 AfO1_|B7JA27|
 AfO3_|B7JB16|
 Acidithiobacillus_ferrivorans_O3_|G0JL85|
 Ferrovum_sp_O3_|A0A0R0LL45|
 Thioalkalivibrio_sp_O3_|D3SG07|
 Thioalkalivibrio_sulfidiphilus_O3_|B8GQN4|
 Acidimicrobium_ferrooxidans_O3_|C7M203|
 Acidithiobacillus_calculus_O3_|F9ZLQ2|
 Thiobacillus_nivea_O3_|I3BSU7|
 Halothiobacillus_neapolitanus_O3_|D0KZ80|
 Nitrobacter_hamburgensis_O3_|Q1QFQ8|
 Nitrosomonas_eutropha_O3_|Q0AHU9|
 Thiomonas_arsenitoxydans_O3_|D6CQL4|
 Thiomonas_intermedia_O3_|D5X344|
 Acidithiobacillus_ferrivorans_O1_|G0JNK3|
 Nitrosomonas_communis_O1_|A0A0F7KHE0|
 Nitrosomonas_europaea_O1_|Q82TG9|
 Nitrosomonas_sp_O1_|F9ZF96|
 Thiobacillus_denitrificans_O1_|Q3SFN2|
 Thiobacillus_thioparus_O1_|UPI0003621E4A|
 Thiomonas_sp_O1_|A0A078BJA6|
 Hydrogenovibrio_marinus_O1_|Q75W42|
 Sideroxydans_lithotrophicus_O1_|D5CQ62|
 Sulfuricella_denitrificans_O1_|S6AAH0|
 Thiomicrospira_kuenenii_O1_|UPI000466D9A2|
 Thiomicrospira_halophila_O1_|UPI000378C5A3|
 Thiomicrospira_chilensis_O1_|UPI000429B796|
 Thiomicrospira_crunigena_O1_|Q31IJ7|
 Thiomicrospira_sp_O1_|UPI000493B602|
 Mariprofundus_ferrooxydans_O1_|Q0EX19|
 Thiobacillus_nivea_O1_|I3BR12|
 Acidithiobacillus_ferrivorans_O2_|G0JPS1|
 Thiobacillus_denitrificans_O2_|Q3SFL7|
 Thiobacillus_thioparus_O2_|UPI00035FE92D|
 Halothiobacillus_neapolitanus_O2_|D0KZQ5|
 Thiomonas_arsenitoxydans_O2_|D6CTS8|
 Thiomonas_intermedia_O2_|D5X1M7|
 Sideroxydans_lithotrophicus_O2_|D5CTG0|
 Sulfuricella_denitrificans_O2_|S6AJV8|
 Gallionella_capsiferriiformans_O2_|D9SHP8|
 Mariprofundus_ferrooxydans_O2_|Q0F300|
 Acidithiobacillus_calculus_O2_|F9ZR07|
 Thiobacillus_brandenburg_O2_|W0TP28|
 Hydrogenovibrio_marinus_O2_|Q75W24|
 Thiomicrospira_kuenenii_O2_|UPI000467D206|
 Thiomicrospira_sp_O2_|UPI000493F397|
 Thiomicrospira_halophila_O2_|UPI000399C451|
 Thiomicrospira_chilensis_O2_|UPI00040A3D56|
 Thiomicrospira_crunigena_O2_|Q31IK5|
 consensus>50

624 RHEVR Y Q H I K G F S E P W G D V V K G R L A A L Q A G Y S T R M G A A M R H A G H Y L A T R K A D K K L M L V L T
 636 RHDVA Y Y R F K D F D E A Y S E E A M G R I A G M T G K F S T R M G A A L R H A G H H L R H Q R Q S K K L L L V I T
 647 RHDVE Y F R Y K D F G A S Y N D Q A K S R L A G M S G Q I S T R M G A A L R H A G S I L K R Q P S N K K L L L V I T
 647 RHDVE Y F R Y K D F G A S Y N D R A K A R L A G M S G Q I S T R M G A A L R H A G S I L K R Q P S N K K L L L V I T
 647 RHDIE Y F R Y K D F G A S Y N D H A K A R L A G M S G Q I S T R M G A A M R H A G S I L K R Q P S S K K L L L V I T
 637 RHDVE Y Y R F K D F D A P F D D T A K A R L A G M K G Q I S T R M G A A L R H A G Q L L H N Q G T A K K L L L I T
 637 RHDVE Y F R F K D F D A P Y D D K V K A R L A G M S G Q M S T R M G T A I R H A G E I L K R Q P S H K K L L L I L T
 644 RHAVE Y V R Y K D F D E R F G D E A R A R V A G M T G Q I S T R M G A A L R H A G A H L A R Q P Q R R R L L L V L T
 651 RHDVE Y Y T F K D F A A E Y D D K V K G R L A G M T G Q I S T R M G A A L R H A G H L L D Q Q P S Q K K L L L L L T
 647 RHDVE Y Y R I K D F D A P Y S D V T K G R L A G M K G Q I S T R M G A A L R H A G S L L A Q R A S Q K K L V I L L T
 652 RSDVE Y F R F K D F D D A F D D L A K A R L A G M T G Q I S T R M G T A L R H A G H L L D S R P S Q K K L I L L L T
 666 RHDVH Y Q R F K D F D Q P Y D D S V K T R L A G M K G R I S T R M G A A L R H A A H Y L A E Q P K S K R V V F V V T
 650 RHDVH Y H R F K D F D Q P Y G D I A K A R L A G M K G H F S T R M G A A L R H A G Y Y L A Q P K S K R V I F V I T
 649 RHDVE Y Y R Y K D F D Q P Y G D L A R A R L A G M K G S Y S T R M G A A L R H A G H Y L E Q P K R K K I C F V V T
 649 RHDVH Y Y R F K D F D Q P Y G D L A R A R L A G M K G S Y S T R M G A A L R H A G H Y L E Q P K R K K I C F V V T
 636 RHDVA Y Y R F K D F D E A Y G D E A M G R I A G M T G K F S T R M G A A L R H A G H Y L R Q R Q S K K L L L V I T
 639 RHDVE Y Y R Y K D F D Q P Y N E I P K A K L A G M A A Q I S T R M G A A L R H A S Y Y L K S Q R S K K L L L V I T
 639 RHDVE Y Y R F K D F D Q P Y N E V P K M R L A G M T G Q I S T R M G A A L R H A S H F L K L Q R A A K K L L L V I T
 639 RHDVE Y Y R F K D F D Q P Y N E I P K A R L A G M T G Q I S T R M G A A M R H A T H Y L K L Q K S A K K L L L V I T
 649 RHDVE Y Y R F K D F D Q H W N E V P K A K L A G M T G Q I S T R M G A A L R H A G H H L K L R Q S A K K L L I V I T
 649 RHDVE Y Y R F K D F D Q H W D E G P K A K L A G M T G Q I S T R M G A A L R H A G H H L K L Q R S A K K L L I V I T
 640 RHDVQ Y L R F K D F D Q H W N E E P K A K L A G M T G Q I S T R M G A A L R H A G H H L Q Q Q K S A K K L L L V I T
 640 RHDVQ Y T R F K Q F D E P F N A D V H A K M A G M T G G I S T R M G A A M R H A G A Y L D R Q T S K Q K L M L V I T
 660 RHNVE Y S R F K D F D Q P Y N E V P K S R L A G M R G Q I S T R M G A A L R H A G H Y L K L Q K S A K K L L L V I T
 643 RHNVE Y S R F K D F D Q P Y N E V P K S R L A G M S G Q I S T R M G A A L R H A G H Y L K L Q K S A K K L L L V I T
 640 RHDVQ Y T R F K Q F D E P F D A D V H A K M A G M T G G I S T R M G A A M R H A G A Y L D R Q A S K Q K L M L V I T
 644 RHDVQ Y T R F K Q F E E P F D A D V K A L A G M Q G G I S T R M G A A M R H A G H Y L D Q Q S K Q K L L L I T
 641 RHDVQ Y V R F K Q F E E P F D A D V H A K L A G M Q G G I S T R M G A A M R H A G S Y L E K Q S S K K L L L V I T
 641 RHDVQ Y T R F K Q F D E P F D Q D V H S R L A G M K G G I S T R M G G A M R H A G S Y L E K Q S S K Q K L L L V I T
 641 RHDVQ Y V R F K Q F E S F D S E V H A R L A G M Q G G I S T R M G G A M R H A G H Y L E Q Q N S K Q K L L L V I T
 642 RHDVQ Y F R F K D F N Q T F G D E A K A R L A G M K G G I S T R M G A A M R H A G E H L L R Q P E R K K L L L I V T
 643 RHDVQ Y Y R L K D F N Q H F D D E A R S R L A G M K G G I S T R M G A A L R H A A H H L H K Q P E R K K L I L L V T
 624 RHAVE Y Q H I K G F S E P W G D T V K G R L A A L Q A G Y S T R M G A A L R H A G H Y L A A R Q A D K K L M L V L T
 634 RHDVR Y F H I K G Y S E R W N D D V K A R L A A M E A G Y S T R M G A A M R H A A H Y L S A R P A D K K L M L I L T
 608 RHDVR Y L H I K G Y G E H F N D D V K A R L A A M E A G Y S T R M G A A M R H A A H Y L G A R S A D K K L M L I L T
 617 RHEVR Y Q H I K G Y S E R W N D D V K A R L A K I E A G Y S T R M G A A M R H A A H Y L S A Q K A D K K L M L I L T
 617 RHEVR Y Q H F K G F G E P W G D E V K A R L A A M Q A N Y S T R M G A A L R H A A H S L A R K S D K K L L L I L T
 618 RHEVR Y L H I K G F G E P W G D E V K A R L A A M Q A N Y S T R M G A A L R H A A H S L A R K S D K K L L L V L T
 619 RHDVR Y F L H I K G Y S E K W D D V K G R L A A M E A S Y S T R M G A A M R H A A H Y L E K Q A D K K L M L I L T
 621 RHDVR Y L H I K G Y S E H W N D E V K G R L A A M E A G Y S T R M G A A M R H A A H Y L E A Q K A D K K L M L I L T
 622 RHDVR Y F M H I K G F S E K W D E V K G R L A A M Q A S Y S T R M G A A M R H A S H Y L E Q Q A D K K L M L I L T
 610 RHDVR Y Y H I K G F S E R W D E V K S R I A E M E A G F S T R M G G A M R H A A H Y L K A Q A D K K L M L I L T
 621 RHDVLR Y L H I K G F S E A W D D T V K A R L A G M Q A R W S T R M G A A L R H A T H F L E A R K T S K K L L L I L T
 605 RHEVR Y Y H I K G F S E R W N L D V K S R L A A M A G A Y S T R M G A A M R H A A H Y L E H Q K S D K K L L L V I T
 619 RHEVR Y Q H I K G Y S E G Y T S D V K A R I A A M E A S Y S T R M G A A M R H A A H Y L S A Q A E K K L M L I L T
 619 RHEVR Y Q H I K G Y S E G Y T D E V K A R I A S M E A S Y S T R M G A A M R H A A H Y L E A Q Q A E K K L M L V L T
 619 RHEVR Y Q H I K G Y S E N Y N D E V K A R I A G M E A S Y S T R M G A A M R H A A H Y L E A Q Q A E K K L M L V L T
 619 RHEVR Y Q H I K G F S E H Y G D E V K A R I S A M E A S Y S T R M G A A M R H A A H Y L A S Q E A E K K L M L V L T
 619 RHEVR Y Q H I K G Y S E H Y G D E V K A R I A M E A S Y S T R M G A A M R H A A H Y L E A Q K A E K K L M L V L T
 622 RHEVR Y Q H I K G Y S E H Y G D E V K S R I A M E A S Y S T R M G A A M R H A A H Y L E A Q Q A E K K L M L I L T
 Rhdve%.rfKd%de.yndevkarlagm.g.lSTRMGaAmRHAgHyL..q...kklllviT

Afo2_|B7J5E5|
 Afo1_|B7JA27|
 Afo3_|B7JB16|
 Acidithiobacillus_ferrivorans_O3_|G0JL85|
 Ferrofum_sp_O3_|A0A0R0LL45|
 Thioalkalivibrio_sp_O3_|D3SG07|
 Thioalkalivibrio_sulfidiphilus_O3_|B8GQN4|
 Acidimicrobium_ferrooxidans_O3_|C7M203|
 Acidithiobacillus_calidus_O3_|F9ZLQ2|
 Thiobacillus_nivea_O3_|I3BSU7|
 Halothiobacillus_neapolitanus_O3_|D0KZ80|
 Nitrobacter_hamburgensis_O3_|Q1QFQ8|
 Nitrosomonas_eutropha_O3_|Q0AHU9|
 Thiomonas_arsenitoxydans_O3_|D6CQL4|
 Thiomonas_intermedia_O3_|D5X344|
 Acidithiobacillus_ferrivorans_O1_|G0JNK3|
 Nitrosomonas_communis_O1_|A0A0F7KHE0|
 Nitrosomonas_europaea_O1_|Q82TG9|
 Nitrosomonas_sp_O1_|F9ZF96|
 Thiobacillus_denitrificans_O1_|Q3SFN2|
 Thiobacillus_thioparus_O1_|UPI0003621E4A|
 Thiomonas_sp_O1_|A0A078BJA6|
 Hydrogenovibrio_marinus_O1_|Q75W42|
 Sideroxydans_lithotrophicus_O1_|D5CQ62|
 Sulfuricella_denitrificans_O1_|S6AAH0|
 Thiomicrospira_kuenenii_O1_|UPI000466D9A2|
 Thiomicrospira_halophila_O1_|UPI000378C5A3|
 Thiomicrospira_chilensis_O1_|UPI000429B796|
 Thiomicrospira_crunigena_O1_|Q31IJ7|
 Thiomicrospira_sp_O1_|UPI000493B602|
 Mariprofundus_ferrooxydans_O1_|Q0EX19|
 Thiobacillus_nivea_O1_|I3BR12|
 Acidithiobacillus_ferrivorans_O2_|G0JPS1|
 Thiobacillus_denitrificans_O2_|Q3SFL7|
 Thiobacillus_thioparus_O2_|UPI00035FE92D|
 Halothiobacillus_neapolitanus_O2_|D0KZQ5|
 Thiomonas_arsenitoxydans_O2_|D6CTS8|
 Thiomonas_intermedia_O2_|D5X1M7|
 Sideroxydans_lithotrophicus_O2_|D5CTG0|
 Sulfuricella_denitrificans_O2_|S6AJV8|
 Gallionella_capsiferriformans_O2_|D9SHP8|
 Mariprofundus_ferrooxydans_O2_|Q0F300|
 Acidithiobacillus_calidus_O2_|F9ZR07|
 Thiobacillus_brandenburg_O2_|W0TP28|
 Hydrogenovibrio_marinus_O2_|Q75W24|
 Thiomicrospira_kuenenii_O2_|UPI000467D206|
 Thiomicrospira_sp_O2_|UPI000493F397|
 Thiomicrospira_halophila_O2_|UPI000399C451|
 Thiomicrospira_chilensis_O2_|UPI00040A3D56|
 Thiomicrospira_crunigena_O2_|Q31IK5|
 consensus>50

684 DGRPSDVQDERLIE DARQAVNELDRNGIFTYCISLDPH... ADAVYADIFG ROYF
 696 DGEPAIDVDRDPQYLRH DARRAVEDLGRDGVITYCLSLDPH... ADEYVSRIFGARNYQ
 707 DGEPAIDVDRDPQYLRH DARRAVEDLGRDGVITYCLSLDPH... ADQYVSRIFGARNYV
 707 DGEPAIDVDRDPQYLRH DARRAVEDLGRDGVITYCLSLDPH... ADQYVSRIFGARNYV
 707 DGEPAIDVDRDPQYLRH DARRAVEDLGRDGVITYCLSLDPH... SDQYVSRIFGARNYM
 697 DGEPAIDVDRDPQYLRH DARRAVEDLGRDGVITYCLSLDPH... ADEYVSRIFGAQMFM
 697 DGEPAIDVDRDPQYLRH DARRAVEDLGRDGVITYCLSLDPH... ADEYVSRIFGARNYA
 704 DGEPAIDVDRDPQYLRH DARRAVEDLGRDGVITYCLSLDPH... ADQYVSRIFGPTGFI
 711 DGEPSDNDVDRDPQYLRH DARRAVEDLGRDGVITYCLSLDPH... ADEYVSRIFGARNYL
 707 DGEPAIDVDRDPQYLRH DARRAVEDLGRDGVITYCLSLDPH... ADQYVSRIFGARNYL
 712 DGEPSDNDVDRDPQYLRH DARRAVEDLGRDGVITYCLSLDPH... ADDYVSRIFGPKHYM
 726 DGEPAIDVDRDPQYLRH DARRAVEDLGRDGVITYCLSLDPH... ADQYVSRIFGMKNFT
 710 DGEPAIDVDRDPQYLRH DARRAVEDLGRDGVITYCLSLDPH... ADQYVSRIFGAKNFT
 709 DGEPAIDVDRDPQYLRH DARRAVEDLGRDGVITYCLSLDPH... ADLYVSRIFGVKNFT
 709 DGEPAIDVDRDPQYLRH DARRAVEDLGRDGVITYCLSLDPH... ADLYVSRIFGVKNFT
 696 DGEPAIDVDRDPQYLRH DARRAVEDLGRDGVITYCLSLDPH... ADEYVSRIFGARNYQ
 699 DGEPAIDVDRDPQYLRH DARRAVEDLGRDGVITYCLSLDPH... ADQYVSRIFGARNYM
 699 DGEPAIDVDRDPQYLRH DARRAVEDLGRDGVITYCLSLDPH... ADQYVSRIFGMARNYM
 699 DGEPAIDVDRDPQYLRH DARRAVEDLGRDGVITYCLSLDPH... ADQYVSRIFGERNYL
 709 DGEPAIDVDRDPQYLRH DARRAVEDLGRDGVITYCLSLDPH... ADNYVSRIFGQKNYM
 709 DGEPAIDVDRDPQYLRH DARRAVEDLGRDGVITYCLSLDPH... ADNYVSRIFGQKNYM
 700 DGEPAIDVDRDPQYLRH DARRAVEDLGRDGVITYCLSLDPH... ADQYVSRIFGQNNFM
 700 DGEPAIDVDRDPQYLRH DARRAVEDLGRDGVITYCLSLDPH... ADKVVQIFGANRFA
 720 DGEPAIDVDRDPQYLRH DARRAVEDLGRDGVITYCLSLDPH... ADQYVSRIFGARNYM
 703 DGEPAIDVDRDPQYLRH DARRAVEDLGRDGVITYCLSLDPH... ADQYVSRIFGAKNYM
 700 DGEPAIDVDRDPQYLRH DARRAVEDLGRDGVITYCLSLDPH... ADKVVQIFGANRFA
 704 DGEPSDIDEQDSQYLRH DARRAVEDLGRDGVITYCLSLDPH... ADKVVQIFGHNRFA
 701 DGEPAIDVDRDPQYLRH DARRAVEDLGRDGVITYCLSLDPH... ADKVVQIFGHNRFA
 701 DGEPAIDVDRDPQYLRH DARRAVEDLGRDGVITYCLSLDPH... ADKVVHNIFGQNRFA
 701 DGEPAIDVDRDPQYLRH DARRAVEDLGRDGVITYCLSLDPH... ADKVVQIFGHNRFA
 702 DGEPAIDVDRDPQYLRH DARRAVEDLGRDGVITYCLSLDPH... ADAVYKRIFGANNYT
 703 DGEPAIDVDRDPQYLRH DARRAVEDLGRDGVITYCLSLDPH... ADDYVKRIFGQNNYT
 684 DGRPSDVQDERLIE DARQAVNELDRNGIFTYCISLDPH... ADAVYADIFG ROYF
 694 DGRPSDVQDAERLIE DARQAVNELDRNGIFTYCISLDPH... ADAVYAEIFG ROYF
 668 DGRPSDVQDERLIE DARQAVNELDRNGIFTYCISLDPH... ADEYVSRIFG ROYF
 677 DGRPSDVQDERLIE DARQAVNELDRNGIFTYCISLDPH... DDYVREIFG QOYT
 677 DGEPAIDVDRDPQYLRH DARRAVEDLGRDGVITYCLSLDPH... ADAVYSDIFG RRYT
 678 DGEPAIDVDRDPQYLRH DARRAVEDLGRDGVITYCLSLDPH... ADAVYSDIFG RRYT
 679 DGRPSDVQSKDGERLIE DARQAVNELDRNGIFTYCISLDPH... ADEYVADIFG KOYT
 681 DGRPSDVQSKDGERLIE DARQAVNELDRNGIFTYCISLDPH... ADEYVTDIFG KOYT
 682 DGEPAIDVDRDPQYLRH DARRAVEDLGRDGVITYCLSLDPH... ADEYVADIFG KOYT
 670 DGEPSDIDEQDSQYLRH DARRAVEDLGRDGVITYCLSLDPH... ADEYVADIFG ROYF
 681 DGRPSDVQDAERLIE DARQAVNELDRNGIFTYCISLDPH... ADSYVAQIFG RRFs
 665 DGEPSDIDEQDSQYLRH DARRAVEDLGRDGVITYCLSLDPH... ADEYVADIFG NOYS
 679 DGEPAIDVDRDPQYLRH DARRAVEDLGRDGVITYCLSLDPH... ADEYVETIFG NOYT
 679 DGEPAIDVDRDPQYLRH DARRAVEDLGRDGVITYCLSLDPH... ADDYVESIFG NHYT
 679 DGEPAIDVDRDPQYLRH DARRAVEDLGRDGVITYCLSLDPH... ADEYVADIFG NOYT
 679 DGEPAIDVDRDPQYLRH DARRAVEDLGRDGVITYCLSLDPH... ADDYVADIFG NHYT
 679 DGEPAIDVDRDPQYLRH DARRAVEDLGRDGVITYCLSLDPH... ADEYVADIFG NOYT
 682 DGEPAIDVDRDPQYLRH DARRAVEDLGRDGVITYCLSLDPH... ADEYVETIFG NHYT
 DGePaDiDvrDpqqYlr.DakkaVe#1...giy.%clsIDp...aDeYV...!Fg..n%.

9. REFERENCES

- Afonine, P.V., Grosse-Kunstleve, R.W., Echols, N., Headd, J.J., Moriarty, N.W., Mustyakimov, M., Terwilliger, T.C., Urzhumtsev, A., Zwart, P.H., Adams, P.D., 2012. Towards automated crystallographic structure refinement with phenix.refine. *Acta crystallographica. Section D, Biological crystallography* 68, 352-367.
- Alonso, H., Blayney, M.J., Beck, J.L., Whitney, S.M., 2009. Substrate-induced Assembly of *Methanococcoides burtonii* d-Ribulose-1,5-bisphosphate Carboxylase/Oxygenase Dimers into Decamers. *Journal of Biological Chemistry* 284, 33876-33882.
- Andersson, I., Backlund, A., 2008. Structure and function of Rubisco. *Plant Physiology and Biochemistry* 46, 275-291.
- Ashida, H., Saito, Y., Nakano, T., Tandeau de Marsac, N., Sekowska, A., Danchin, A., Yokota, A., 2008. RuBisCO-like proteins as the enolase enzyme in the methionine salvage pathway: functional and evolutionary relationships between RuBisCO-like proteins and photosynthetic RuBisCO. *Journal of experimental botany* 59, 1543-1554.
- Ashkenazy, H., Erez, E., Martz, E., Pupko, T., Ben-Tal, N., 2010. ConSurf 2010: calculating evolutionary conservation in sequence and structure of proteins and nucleic acids. *Nucleic acids research* 38, W529-W533.
- Badger, M., 2003. The roles of carbonic anhydrases in photosynthetic CO₂ concentrating mechanisms. *Photosynthesis Research* 77, 83-94.
- Badger, M.R., Bek, E.J., 2008. Multiple Rubisco forms in proteobacteria: their functional significance in relation to CO₂ acquisition by the CBB cycle. *Journal of experimental botany* 59, 1525-1541.
- Bai, X.-c., Rajendra, E., Yang, G., Shi, Y., Scheres, S.H.W., 2015a. Sampling the conformational space of the catalytic subunit of human γ -secretase. *eLife* 4, e11182.
- Bai, X.C., McMullan, G., Scheres, S.H., 2015b. How cryo-EM is revolutionizing structural biology. *Trends Biochem Sci* 40, 49-57.
- Baker, N.A., Sept, D., Joseph, S., Holst, M.J., McCammon, J.A., 2001. Electrostatics of nanosystems: application to microtubules and the ribosome. *Proceedings of the National Academy of Sciences of the United States of America* 98, 10037-10041.
- Baker, R.T., Catanzariti, A.M., Karunasekara, Y., Soboleva, T.A., Sharwood, R., Whitney, S., Board, P.G., 2005. Using Deubiquitylating Enzymes as Research Tools, *Methods in Enzymology*. Academic Press, pp. 540-554.
- Banerjee, S., Bartesaghi, A., Merk, A., Rao, P., Bulfer, S.L., Yan, Y., Green, N., Mroczkowski, B., Neitz, R.J., Wipf, P., Falconieri, V., Deshaies, R.J., Milne, J.L.S., Huryn, D., Arkin, M., Subramaniam, S., 2016. 2.3 Å resolution cryo-EM structure of human p97 and mechanism of allosteric inhibition. *Science (New York, N.Y.)* 351, 871-875.

- Barrio-Garcia, C., Thoms, M., Flemming, D., Kater, L., Berninghausen, O., Bassler, J., Beckmann, R., Hurt, E., 2016. Architecture of the Rix1-Real checkpoint machinery during pre-60S-ribosome remodeling. *Nature structural & molecular biology* 23, 37-44.
- Bassham, J.A., Benson, A.A., Calvin, M., 1950. The path of carbon in photosynthesis. *The Journal of biological chemistry* 185, 781-787.
- Bauwe, H., 2001. Photorespiration, eLS. John Wiley & Sons, Ltd.
- Berg, I.A., 2011. Ecological aspects of the distribution of different autotrophic CO₂ fixation pathways. *Applied and environmental microbiology* 77, 1925-1936.
- Berry, J.A., Lorimer, G.H., Pierce, J., Seemann, J.R., Meek, J., Freas, S., 1987. Isolation, identification, and synthesis of 2-carboxyarabinitol 1-phosphate, a diurnal regulator of ribulose-bisphosphate carboxylase activity. *Proceedings of the National Academy of Sciences of the United States of America* 84, 734-738.
- Bhat, J.Y., Thieulin-Pardo, G., Hartl, F.U., Hayer-Hartl, M., 2017a. Rubisco Activases: AAA+ Chaperones Adapted to Enzyme Repair. *Frontiers in molecular biosciences* 4, 20.
- Bhat, J.Y., Milicic, G., Thieulin-Pardo, G., Bracher, A., Maxwell, A., Ciniawsky, S., Mueller-Cajar, O., Engen, J.R., Hartl, F.U., Wendler, P., Hayer-Hartl, M., 2017b. Mechanism of Enzyme Repair by the AAA(+) Chaperone Rubisco Activase. *Molecular cell* 67, 744-756.e746.
- Bienkowska, J., Cruz, M., Atiemo, A., Handin, R., Liddington, R., 1997. The von willebrand factor A3 domain does not contain a metal ion-dependent adhesion site motif. *The Journal of biological chemistry* 272, 25162-25167.
- Bracher, A., Sharma, A., Starling-Windhof, A., Hartl, F.U., Hayer-Hartl, M., 2015. Degradation of potent Rubisco inhibitor by selective sugar phosphatase. *Nature Plants* 1, 14002.
- Catanzariti, A.-M., Soboleva, T.A., Jans, D.A., Board, P.G., Baker, R.T., 2004. An efficient system for high-level expression and easy purification of authentic recombinant proteins. *Protein Science : A Publication of the Protein Society* 13, 1331-1339.
- Celikel, R., Varughese, K.I., Madhusudan, Yoshioka, A., Ware, J., Ruggeri, Z.M., 1998. Crystal structure of the von Willebrand factor A1 domain in complex with the function blocking NMC-4 Fab. *Nature structural biology* 5, 189-194.
- Chen, V.B., Arendall, W.B., 3rd, Headd, J.J., Keedy, D.A., Immormino, R.M., Kapral, G.J., Murray, L.W., Richardson, J.S., Richardson, D.C., 2010. MolProbity: all-atom structure validation for macromolecular crystallography. *Acta crystallographica. Section D, Biological crystallography* 66, 12-21.
- Chen, Z., Suzuki, H., Kobayashi, Y., Wang, A.C., DiMaio, F., Kawashima, S.A., Walz, T., Kapoor, T.M., 2018. Structural Insights into Mdn1, an Essential AAA Protein Required for Ribosome Biogenesis. *Cell* 175, 822-834.e818.

Chenna, R., Sugawara, H., Koike, T., Lopez, R., Gibson, T.J., Higgins, D.G., Thompson, J.D., 2003. Multiple sequence alignment with the Clustal series of programs. *Nucleic acids research* 31, 3497-3500.

Cleland, W.W., Andrews, T.J., Gutteridge, S., Hartman, F.C., Lorimer, G.H., 1998. Mechanism of Rubisco: The Carbamate as General Base. *Chemical reviews* 98, 549-562.

Dangel, A.W., Tabita, F.R., 2015. CbbR, the Master Regulator for Microbial Carbon Dioxide Fixation. *Journal of bacteriology* 197, 3488-3498.

de la Peña, A.H., Goodall, E.A., Gates, S.N., Lander, G.C., Martin, A., 2018. Substrate-engaged 26S proteasome structures reveal mechanisms for ATP-hydrolysis-driven translocation. *Science (New York, N.Y.)* 362, eaav0725.

de Marco, A., Deuerling, E., Mogk, A., Tomoyasu, T., Bukau, B., 2007. Chaperone-based procedure to increase yields of soluble recombinant proteins produced in *E. coli*. *BMC Biotechnology* 7, 1-9.

Delwiche, C.F., Palmer, J.D., 1996. Rampant horizontal transfer and duplication of rubisco genes in eubacteria and plastids. *Molecular biology and evolution* 13, 873-882.

Dodd, A.N., Borland, A.M., Haslam, R.P., Griffiths, H., Maxwell, K., 2002. Crassulacean acid metabolism: plastic, fantastic. *Journal of experimental botany* 53, 569-580.

Dong, Y., Zhang, S., Wu, Z., Li, X., Wang, W.L., Zhu, Y., Stoilova-McPhie, S., Lu, Y., Finley, D., Mao, Y., 2019. Cryo-EM structures and dynamics of substrate-engaged human 26S proteasome. *Nature* 565, 49-55.

Duff, A.P., Andrews, T.J., Curmi, P.M., 2000. The transition between the open and closed states of rubisco is triggered by the inter-phosphate distance of the bound bisphosphate. *Journal of molecular biology* 298, 903-916.

Edmondson, D.L., Kane, H.J., Andrews, T.J., 1990. Substrate isomerization inhibits ribulosebisphosphate carboxylase-oxygenase during catalysis. *FEBS Letters* 260, 62-66.

El Bakkouri, M., Gutsche, I., Kanjee, U., Zhao, B., Yu, M., Goret, G., Schoehn, G., Burmeister, W.P., Houry, W.A., 2010. Structure of RavA MoxR AAA+ protein reveals the design principles of a molecular cage modulating the inducible lysine decarboxylase activity. *Proceedings of the National Academy of Sciences of the United States of America* 107, 22499-22504.

Ellis, R.J., 1979. The most abundant protein in the world. *Trends in Biochemical Sciences* 4, 241-244.

Emsley, J., Knight, C.G., Farndale, R.W., Barnes, M.J., Liddington, R.C., 2000. Structural basis of collagen recognition by integrin $\alpha 2\beta 1$. *Cell* 101, 47-56.

Emsley, P., Lohkamp, B., Scott, W.G., Cowtan, K., 2010. Features and development of Coot. *Acta crystallographica. Section D, Biological crystallography* 66, 486-501.

- Engel, B.D., Schaffer, M., Kuhn Cuellar, L., Villa, E., Plitzko, J.M., Baumeister, W., 2015. Native architecture of the Chlamydomonas chloroplast revealed by in situ cryo-electron tomography. *eLife* 4, e04889.
- Erb, T.J., Evans, B.S., Cho, K., Warlick, B.P., Sriram, J., Wood, B.M., Imker, H.J., Sweedler, J.V., Tabita, F.R., Gerlt, J.A., 2012. A RubisCO-like protein links SAM metabolism with isoprenoid biosynthesis. *Nature chemical biology* 8, 926-932.
- Faini, M., Stengel, F., Aebersold, R., 2016. The Evolving Contribution of Mass Spectrometry to Integrative Structural Biology. *Journal of the American Society for Mass Spectrometry* 27, 966-974.
- Field, C.B., Behrenfeld, M.J., Randerson, J.T., Falkowski, P., 1998. Primary Production of the Biosphere: Integrating Terrestrial and Oceanic Components. *Science (New York, N.Y.)* 281, 237.
- Freeman Rosenzweig, E.S., Xu, B., Kuhn Cuellar, L., Martinez-Sanchez, A., Schaffer, M., Strauss, M., Cartwright, H.N., Ronceray, P., Plitzko, J.M., Förster, F., Wingreen, N.S., Engel, B.D., Mackinder, L.C.M., Jonikas, M.C., 2017. The Eukaryotic CO₂-Concentrating Organelle Is Liquid-like and Exhibits Dynamic Reorganization. *Cell* 171, 148-162.e119.
- Fuchs, G., 2011. Alternative pathways of carbon dioxide fixation: insights into the early evolution of life? *Annual review of microbiology* 65, 631-658.
- Gates, S.N., Yokom, A.L., Lin, J., Jackrel, M.E., Rizo, A.N., Kendsersky, N.M., Buell, C.E., Sweeny, E.A., Mack, K.L., Chuang, E., Torrente, M.P., Su, M., Shorter, J., Southworth, D.R., 2017. Ratchet-like polypeptide translocation mechanism of the AAA+ disaggregase Hsp104. *Science (New York, N.Y.)* 357, 273-279.
- Gibson, J.L., Tabita, F.R., 1997. Analysis of the cbbXYZ operon in *Rhodobacter sphaeroides*. *Journal of bacteriology* 179, 663-669.
- Gottesman, S., Roche, E., Zhou, Y., Sauer, R.T., 1998. The ClpXP and ClpAP proteases degrade proteins with carboxy-terminal peptide tails added by the SsrA-tagging system. *Genes Dev* 12, 1338-1347.
- Graham, J.W., Lei, M.G., Lee, C.Y., 2013. Trapping and Identification of Cellular Substrates of the *Staphylococcus aureus* ClpC Chaperone. *Journal of bacteriology* 195, 4506-4516.
- Gunn, L.H., Valegard, K., Andersson, I., 2017. A unique structural domain in *Methanococcoides burtonii* ribulose-1,5-bisphosphate carboxylase/oxygenase (Rubisco) acts as a small subunit mimic. *The Journal of biological chemistry* 292, 6838-6850.
- Guzman, L.M., Belin, D., Carson, M.J., Beckwith, J., 1995. Tight regulation, modulation, and high-level expression by vectors containing the arabinose PBAD promoter. *Journal of bacteriology* 177, 4121-4130.
- Haas, T.A., Plow, E.F., 1994. Integrin-ligand interactions: a year in review. *Current opinion in cell biology* 6, 656-662.

- Hansen, S., Vollan, V.B., Hough, E., Andersen, K., 1999. The crystal structure of rubisco from *Alcaligenes eutrophus* reveals a novel central eight-stranded beta-barrel formed by beta-strands from four subunits. *Journal of molecular biology* 288, 609-621.
- Hanson, P.I., Whiteheart, S.W., 2005a. AAA+ proteins: have engine, will work. *Nature Reviews Molecular Cell Biology* 6, 519.
- Hanson, P.I., Whiteheart, S.W., 2005b. AAA+ proteins: have engine, will work. *Nature reviews. Molecular cell biology* 6, 519-529.
- Hartl, F.U., Hayer-Hartl, M., 2002. Molecular chaperones in the cytosol: from nascent chain to folded protein. *Science (New York, N.Y.)* 295, 1852-1858.
- Hartman, F.C., Harpel, M.R., 1994. Structure, function, regulation, and assembly of D-ribulose-1,5-bisphosphate carboxylase/oxygenase. *Annu Rev Biochem* 63, 197-234.
- Hasse, D., Larsson, A.M., Andersson, I., 2015. Structure of *Arabidopsis thaliana* Rubisco activase. *Acta crystallographica. Section D, Biological crystallography* 71, 800-808.
- Hauser, T., Popilka, L., Hartl, F.U., Hayer-Hartl, M., 2015. Role of auxiliary proteins in Rubisco biogenesis and function. *Nature Plants* 1, 15065.
- Hayashi, N.R., Arai, H., Kodama, T., Igarashi, Y., 1997. The novel genes, *cbbQ* and *cbbO*, located downstream from the RubisCO genes of *Pseudomonas hydrogenothermophila*, affect the conformational states and activity of RubisCO. *Biochemical and biophysical research communications* 241, 565-569.
- Hayashi, N.R., Arai, H., Kodama, T., Igarashi, Y., 1999. The *cbbQ* genes, located downstream of the form I and form II RubisCO genes, affect the activity of both RubisCOs. *Biochemical and biophysical research communications* 265, 177-183.
- Hughes, A.L., 2001. Evolution of the integrin alpha and beta protein families. *Journal of molecular evolution* 52, 63-72.
- Iyer, L.M., Leipe, D.D., Koonin, E.V., Aravind, L., 2004. Evolutionary history and higher order classification of AAA+ ATPases. *Journal of structural biology* 146, 11-31.
- Joshi, J., Mueller-Cajar, O., Tsai, Y.C., Hartl, F.U., Hayer-Hartl, M., 2015. Role of small subunit in mediating assembly of red-type form I Rubisco. *The Journal of biological chemistry* 290, 1066-1074.
- Jurgen, B., Breitenstein, A., Urlacher, V., Buttner, K., Lin, H., Hecker, M., Schweder, T., Neubauer, P., 2010. Quality control of inclusion bodies in *Escherichia coli*. *Microbial cell factories* 9, 41.
- Kahle, M., Ter Beek, J., Hosler, J.P., Adelroth, P., 2018. The insertion of the non-heme FeB cofactor into nitric oxide reductase from *P. denitrificans* depends on NorQ and NorD accessory proteins. *Biochimica et biophysica acta. Bioenergetics*.
- Kearse, M., Moir, R., Wilson, A., Stones-Havas, S., Cheung, M., Sturrock, S., Buxton, S., Cooper, A., Markowitz, S., Duran, C., Thierer, T., Ashton, B., Meintjes, P.,

- Drummond, A., 2012. Geneious Basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. *Bioinformatics* (Oxford, England) 28, 1647-1649.
- Kon, T., Oyama, T., Shimo-Kon, R., Imamula, K., Shima, T., Sutoh, K., Kurisu, G., 2012. The 2.8 Å crystal structure of the dynein motor domain. *Nature* 484, 345-350.
- Kreuzer, K.N., Jongeneel, C.V., 1983. *Escherichia coli* phage T4 topoisomerase. *Methods in enzymology* 100, 144-160.
- Kubien, D.S., Brown, C.M., Kane, H.J., 2011. Quantifying the amount and activity of Rubisco in leaves. *Methods in molecular biology* (Clifton, N.J.) 684, 349-362.
- Kucukelbir, A., Sigworth, F.J., Tagare, H.D., 2013. Quantifying the local resolution of cryo-EM density maps. *Nature methods* 11, 63.
- Kühlbrandt, W., 2014. The Resolution Revolution. *Science* (New York, N.Y.) 343, 1443-1444.
- Kumar, A., Li, C., Portis, A.R., Jr., 2009. *Arabidopsis thaliana* expressing a thermostable chimeric Rubisco activase exhibits enhanced growth and higher rates of photosynthesis at moderately high temperatures. *Photosynth Res* 100, 143-153.
- Kurek, I., Chang, T.K., Bertain, S.M., Madrigal, A., Liu, L., Lassner, M.W., Zhu, G., 2007. Enhanced Thermostability of *Arabidopsis* Rubisco Activase Improves Photosynthesis and Growth Rates under Moderate Heat Stress. *The Plant Cell* 19, 3230-3241.
- Larson, E.M., O'Brien, C.M., Zhu, G., Spreitzer, R.J., Portis, A.R., Jr., 1997. Specificity for activase is changed by a Pro-89 to Arg substitution in the large subunit of ribulose-1,5-bisphosphate carboxylase/oxygenase. *The Journal of biological chemistry* 272, 17033-17037.
- Lee, J.O., Rieu, P., Arnaout, M.A., Liddington, R., 1995. Crystal structure of the A domain from the alpha subunit of integrin CR3 (CD11b/CD18). *Cell* 80, 631-638.
- Leitner, A., Faini, M., Stengel, F., Aebersold, R., 2016. Crosslinking and Mass Spectrometry: An Integrated Technology to Understand the Structure and Function of Molecular Machines. *Trends Biochem Sci* 41, 20-32.
- Li, C., Salvucci, M.E., Portis, A.R., Jr., 2005a. Two residues of rubisco activase involved in recognition of the Rubisco substrate. *The Journal of biological chemistry* 280, 24864-24869.
- Li, H., Sawaya, M.R., Tabita, F.R., Eisenberg, D., 2005b. Crystal structure of a RuBisCO-like protein from the green sulfur bacterium *Chlorobium tepidum*. *Structure* (London, England : 1993) 13, 779-789.
- Li, X., Mooney, P., Zheng, S., Booth, C.R., Braunfeld, M.B., Gubbens, S., Agard, D.A., Cheng, Y., 2013. Electron counting and beam-induced motion correction enable near-atomic-resolution single-particle cryo-EM. *Nature methods* 10, 584.

- Liddington, R.C., 2002. Will the real integrin please stand up? *Structure* (London, England : 1993) 10, 605-607.
- Liu, D., Ramya, R.C.S., Mueller-Cajar, O., 2017. Surveying the expanding prokaryotic Rubisco multiverse. *FEMS Microbiology Letters* 364, fnx156-fnx156.
- Loganathan, N., Tsai, Y.C., Mueller-Cajar, O., 2016. Characterization of the heterooligomeric red-type rubisco activase from red algae. *Proceedings of the National Academy of Sciences of the United States of America* 113, 14019-14024.
- Lorimer, G.H., Andrews, T.J., Tolbert, N.E., 1973. Ribulose diphosphate oxygenase. II. Further proof of reaction products and mechanism of action. *Biochemistry* 12, 18-23.
- Luo, B.H., Carman, C.V., Springer, T.A., 2007. Structural basis of integrin regulation and signaling. *Annual review of immunology* 25, 619-647.
- Maeda, N., Kitano, K., Fukui, T., Ezaki, S., Atomi, H., Miki, K., Imanaka, T., 1999. Ribulose biphosphate carboxylase/oxygenase from the hyperthermophilic archaeon *Pyrococcus kodakaraensis* KOD1 is composed solely of large subunits and forms a pentagonal structure. *Journal of molecular biology* 293, 57-66.
- Maier, U.G., Fraunholz, M., Zauner, S., Penny, S., Douglas, S., 2000. A nucleomorph-encoded CbbX and the phylogeny of RuBisCo regulators. *Molecular biology and evolution* 17, 576-583.
- Martin, A., Baker, T.A., Sauer, R.T., 2008. Pore loops of the AAA+ ClpX machine grip substrates to drive translocation and unfolding. *Nature structural & molecular biology* 15, 1147-1151.
- McKinlay, J.B., Harwood, C.S., 2010. Carbon dioxide fixation as a central redox cofactor recycling mechanism in bacteria. *Proceedings of the National Academy of Sciences* 107, 11669-11675.
- Meyer, M.T., Whittaker, C., Griffiths, H., 2017. The algal pyrenoid: key unanswered questions. *Journal of experimental botany* 68, 3739-3749.
- Meyer, M.T., Genkov, T., Skepper, J.N., Jouhet, J., Mitchell, M.C., Spreitzer, R.J., Griffiths, H., 2012. Rubisco small-subunit α -helices control pyrenoid formation in *Chlamydomonas*. *Proceedings of the National Academy of Sciences* 109, 19474-19479.
- Michishita, M., Videm, V., Arnaout, M.A., 1993. A novel divalent cation-binding site in the A domain of the beta 2 integrin CR3 (CD11b/CD18) is essential for ligand binding. *Cell* 72, 857-867.
- Moroney, J.V., Somanchi, A., 1999. How Do Algae Concentrate CO₂ to Increase the Efficiency of Photosynthetic Carbon Fixation? *Plant Physiology* 119, 9-16.
- Mueller-Cajar, O., 2017. The Diverse AAA+ Machines that Repair Inhibited Rubisco Active Sites. *Frontiers in molecular biosciences* 4, 31.

- Mueller-Cajar, O., Stotz, M., Bracher, A., 2014. Maintaining photosynthetic CO₂ fixation via protein remodelling: the Rubisco activases. *Photosynth Res* 119, 191-201.
- Mueller-Cajar, O., Stotz, M., Wendler, P., Hartl, F.U., Bracher, A., Hayer-Hartl, M., 2011. Structure and function of the AAA+ protein CbbX, a red-type Rubisco activase. *Nature* 479, 194-199.
- Nelson, D.L., 2005. *Lehninger principles of biochemistry*. Fourth edition. New York : W.H. Freeman, 2005.
- Ogura, T., Wilkinson, A.J., 2001. AAA+ superfamily ATPases: common structure--diverse function. *Genes to cells : devoted to molecular & cellular mechanisms* 6, 575-597.
- Olivares, A.O., Baker, T.A., Sauer, R.T., 2015. Mechanistic insights into bacterial AAA+ proteases and protein-remodelling machines. *Nature Reviews Microbiology* 14, 33.
- Parry, M.A., Keys, A.J., Madgwick, P.J., Carmo-Silva, A.E., Andralojc, P.J., 2008. Rubisco regulation: a role for inhibitors. *Journal of experimental botany* 59, 1569-1580.
- Pearce, F.G., 2006. Catalytic by-product formation and ligand binding by ribulose biphosphate carboxylases from different phylogenies. *The Biochemical journal* 399, 525-534.
- Perez-Iratxeta, C., Andrade-Navarro, M.A., 2008. K2D2: Estimation of protein secondary structure from circular dichroism spectra. *BMC Structural Biology* 8, 25.
- Pettersen, E.F., Goddard, T.D., Huang, C.C., Couch, G.S., Greenblatt, D.M., Meng, E.C., Ferrin, T.E., 2004. UCSF Chimera--a visualization system for exploratory research and analysis. *Journal of computational chemistry* 25, 1605-1612.
- Portis, A.R., Jr., 2003. Rubisco activase - Rubisco's catalytic chaperone. *Photosynth Res* 75, 11-27.
- Portis, A.R., Jr., Li, C., Wang, D., Salvucci, M.E., 2008. Regulation of Rubisco activase and its interaction with Rubisco. *Journal of experimental botany* 59, 1597-1604.
- Puchades, C., Rampello, A.J., Shin, M., Giuliano, C.J., Wiseman, R.L., Glynn, S.E., Lander, G.C., 2017. Structure of the mitochondrial inner membrane AAA+ protease YME1 gives insight into substrate processing. *Science (New York, N.Y.)* 358, eaao0464.
- Rae, B.D., Long, B.M., Whitehead, L.F., Forster, B., Badger, M.R., Price, G.D., 2013. Cyanobacterial carboxysomes: microcompartments that facilitate CO₂ fixation. *Journal of molecular microbiology and biotechnology* 23, 300-307.
- Rappsilber, J., 2011. The beginning of a beautiful friendship: cross-linking/mass spectrometry and modelling of proteins and multi-protein complexes. *Journal of structural biology* 173, 530-540.

- Reinfelder, J.R., 2011. Carbon concentrating mechanisms in eukaryotic marine phytoplankton. *Annual review of marine science* 3, 291-315.
- Rinner, O., Seebacher, J., Walzthoeni, T., Mueller, L.N., Beck, M., Schmidt, A., Mueller, M., Aebersold, R., 2008. Identification of cross-linked peptides from large sequence databases. *Nature methods* 5, 315-318.
- Robert, X., Gouet, P., 2014. Deciphering key features in protein structures with the new ENDscript server. *Nucleic acids research* 42, W320-324.
- Roberts, A.J., Kon, T., Knight, P.J., Sutoh, K., Burgess, S.A., 2013. Functions and mechanics of dynein motor proteins. *Nature reviews. Molecular cell biology* 14, 713-726.
- Rodriguez-Carmona, E., Cano-Garrido, O., Seras-Franzoso, J., Villaverde, A., Garcia-Fruitos, E., 2010. Isolation of cell-free bacterial inclusion bodies. *Microbial cell factories* 9, 71.
- Rohou, A., Grigorieff, N., 2015. CTFFIND4: Fast and accurate defocus estimation from electron micrographs. *Journal of structural biology* 192, 216-221.
- Rosenthal, P.B., Henderson, R., 2003. Optimal Determination of Particle Orientation, Absolute Hand, and Contrast Loss in Single-particle Electron Cryomicroscopy. *Journal of molecular biology* 333, 721-745.
- Sage, R.F., 2004. The evolution of C4 photosynthesis. *New Phytologist* 161, 341-370.
- Salvucci, M.E., Crafts-Brandner, S.J., 2004. Relationship between the Heat Tolerance of Photosynthesis and the Thermal Stability of Rubisco Activase in Plants from Contrasting Thermal Environments. *Plant Physiology* 134, 1460-1470.
- Santelli, E., Bankston, L.A., Leppla, S.H., Liddington, R.C., 2004. Crystal structure of a complex between anthrax toxin and its host cell receptor. *Nature* 430, 905-908.
- Satagopan, S., Chan, S., Perry, L.J., Tabita, F.R., 2014. Structure-function studies with the unique hexameric form II ribulose-1,5-bisphosphate carboxylase/oxygenase (Rubisco) from *Rhodospseudomonas palustris*. *The Journal of biological chemistry* 289, 21433-21450.
- Sato, T., Atomi, H., Imanaka, T., 2007. Archaeal Type III RuBisCOs Function in a Pathway for AMP Metabolism. *Science (New York, N.Y.)* 315, 1003-1006.
- Sauer, R.T., Baker, T.A., 2011. AAA+ Proteases: ATP-Fueled Machines of Protein Destruction. *Annual Review of Biochemistry* 80, 587-612.
- Scheres, S.H., 2012a. RELION: implementation of a Bayesian approach to cryo-EM structure determination. *Journal of structural biology* 180, 519-530.
- Scheres, S.H.W., 2012b. A Bayesian view on cryo-EM structure determination. *Journal of molecular biology* 415, 406-418.

- Schmidt, H., Zalyte, R., Urnavicius, L., Carter, A.P., 2015. Structure of human cytoplasmic dynein-2 primed for its power stroke. *Nature* 518, 435-438.
- Shimaoka, M., Takagi, J., Springer, T.A., 2002. Conformational regulation of integrin structure and function. *Annu Rev Biophys Biomol Struct* 31, 485-516.
- Shively, J.M., van Keulen, G., Meijer, W.G., 1998. Something from almost nothing: carbon dioxide fixation in chemoautotrophs. *Annual review of microbiology* 52, 191-230.
- Shivhare, D., Mueller-Cajar, O., 2017. In Vitro Characterization of Thermostable CAM Rubisco Activase Reveals a Rubisco Interacting Surface Loop. *Plant Physiology* 174, 1505-1516.
- Siddiqui, S.M., Sauer, R.T., Baker, T.A., 2004. Role of the processing pore of the ClpX AAA+ ATPase in the recognition and engagement of specific protein substrates. *Genes & Development* 18, 369-374.
- Snider, J., Houry, W.A., 2006. MoxR AAA+ ATPases: a novel family of molecular chaperones? *Journal of structural biology* 156, 200-209.
- Snider, J., Gutsche, I., Lin, M., Baby, S., Cox, B., Butland, G., Greenblatt, J., Emili, A., Houry, W.A., 2006. Formation of a distinctive complex between the inducible bacterial lysine decarboxylase and a novel AAA+ ATPase. *The Journal of biological chemistry* 281, 1532-1546.
- Song, G., Yang, Y., Liu, J.-h., Casanovas, J.M., Shimaoka, M., Springer, T.A., Wang, J.-h., 2005. An atomic resolution view of ICAM recognition in a complex between the binding domains of ICAM-3 and integrin $\alpha_L\beta_2$. *Proceedings of the National Academy of Sciences of the United States of America* 102, 3366-3371.
- Sosnowski, P., Urnavicius, L., Boland, A., Fagiewicz, R., Busselez, J., Papai, G., Schmidt, H., 2018. The CryoEM structure of the *Saccharomyces cerevisiae* ribosome maturation factor Real. *eLife* 7, e39163.
- Spreitzer, R.J., 2003. Role of the small subunit in ribulose-1,5-bisphosphate carboxylase/oxygenase. *Archives of Biochemistry and Biophysics* 414, 141-149.
- Spreitzer, R.J., Salvucci, M.E., 2002. Rubisco: structure, regulatory interactions, and possibilities for a better enzyme. *Annu Rev Plant Biol* 53, 449-475.
- Springer, T.A., 2006. Complement and the multifaceted functions of VWA and integrin I domains. *Structure (London, England : 1993)* 14, 1611-1616.
- Sprott, G.D., Ekiel, I., Patel, G.B., 1993. Metabolic Pathways in *Methanococcus jannaschii* and Other Methanogenic Bacteria. *Applied and environmental microbiology* 59, 1092-1098.
- Stein, N., 2008. CHAINSAW: a program for mutating pdb files used as templates in molecular replacement. *Journal of Applied Crystallography* 41, 641-643.

Stotz, M., Mueller-Cajar, O., Ciniawsky, S., Wendler, P., Hartl, F.U., Bracher, A., Hayer-Hartl, M., 2011. Structure of green-type Rubisco activase from tobacco. *Nature structural & molecular biology* 18, 1366-1370.

Streusand, V.J., Portis, A.R., 1987. Rubisco Activase Mediates ATP-Dependent Activation of Ribulose Bisphosphate Carboxylase. *Plant Physiol* 85, 152-154.

Sutter, M., Roberts, E.W., Gonzalez, R.C., Bates, C., Dawoud, S., Landry, K., Cannon, G.C., Heinhorst, S., Kerfeld, C.A., 2015. Structural Characterization of a Newly Identified Component of alpha-Carboxysomes: The AAA+ Domain Protein CsoCbbQ. *Scientific reports* 5, 16243.

Tabita, F.R., Satagopan, S., Hanson, T.E., Kreel, N.E., Scott, S.S., 2008. Distinct form I, II, III, and IV Rubisco proteins from the three kingdoms of life provide clues about Rubisco evolution and structure/function relationships. *Journal of experimental botany* 59, 1515-1524.

Tabita, F.R., Hanson, T.E., Li, H., Satagopan, S., Singh, J., Chan, S., 2007a. Function, structure, and evolution of the RubisCO-like proteins and their RubisCO homologs. *Microbiol Mol Biol Rev* 71, 576-599.

Tabita, F.R., Hanson, T.E., Li, H., Satagopan, S., Singh, J., Chan, S., 2007b. Function, Structure, and Evolution of the RubisCO-Like Proteins and Their RubisCO Homologs. *Microbiology and Molecular Biology Reviews* : MMBR 71, 576-599.

Toyoda, K., Ishii, M., Arai, H., 2018. Function of three RuBisCO enzymes under different CO₂ conditions in *Hydrogenovibrio marinus*. *Journal of bioscience and bioengineering* 126, 730-735.

Tsai, Y.C., Lapina, M.C., Bhushan, S., Mueller-Cajar, O., 2015. Identification and characterization of multiple rubisco activases in chemoautotrophic bacteria. *Nat Commun* 6, 8883.

van de Loo, F.J., Salvucci, M.E., 1996. Activation of Ribulose-1,5-bisphosphate Carboxylase/Oxygenase (Rubisco) Involves Rubisco Activase Trp16. *Biochemistry* 35, 8143-8148.

Varaljay, V.A., Satagopan, S., North, J.A., Witte, B., Dourado, M.N., Anantharaman, K., Arbing, M.A., Hoefft McCann, S., Oremland, R.S., Banfield, J.F., Wrighton, K.C., Tabita, F.R., 2016. Functional metagenomic selection of ribulose 1, 5-bisphosphate carboxylase/oxygenase from uncultivated bacteria. *Environmental microbiology* 18, 1187-1199.

Wachter, R.M., Salvucci, M.E., Carmo-Silva, A.E., Barta, C., Genkov, T., Spreitzer, R.J., 2013. Activation of interspecies-hybrid Rubisco enzymes to assess different models for the Rubisco-Rubisco activase interaction. *Photosynth Res* 117, 557-566.

Walzthoeni, T., Claassen, M., Leitner, A., Herzog, F., Bohn, S., Forster, F., Beck, M., Aebersold, R., 2012. False discovery rate estimation for cross-linked peptides identified by mass spectrometry. *Nature methods* 9, 901-903.

- Weibezahn, J., Schlieker, C., Bukau, B., Mogk, A., 2003. Characterization of a trap mutant of the AAA+ chaperone ClpB. *The Journal of biological chemistry* 278, 32608-32617.
- Wendler, P., Ciniawsky, S., Kock, M., Kube, S., 2012. Structure and function of the AAA+ nucleotide binding pocket. *Biochimica et Biophysica Acta (BBA) - Molecular Cell Research* 1823, 2-14.
- Whitney, S.M., Houtz, R.L., Alonso, H., 2011. Advancing our understanding and capacity to engineer nature's CO₂-sequestering enzyme, Rubisco. *Plant Physiol* 155, 27-35.
- Whitney, S.M., Baldet, P., Hudson, G.S., Andrews, T.J., 2001. Form I Rubiscos from non-green algae are expressed abundantly but not assembled in tobacco chloroplasts. *The Plant journal : for cell and molecular biology* 26, 535-547.
- Wong, K.S., Houry, W.A., 2012. Novel structural and functional insights into the MoxR family of AAA+ ATPases. *Journal of structural biology* 179, 211-221.
- Wrighton, K.C., Castelle, C.J., Varaljay, V.A., Satagopan, S., Brown, C.T., Wilkins, M.J., Thomas, B.C., Sharon, I., Williams, K.H., Tabita, F.R., Banfield, J.F., 2016. RubisCO of a nucleoside pathway known from Archaea is found in diverse uncultivated phyla in bacteria. *The ISME journal* 10, 2702-2714.
- Wrighton, K.C., Thomas, B.C., Sharon, I., Miller, C.S., Castelle, C.J., VerBerkmoes, N.C., Wilkins, M.J., Hettich, R.L., Lipton, M.S., Williams, K.H., Long, P.E., Banfield, J.F., 2012. Fermentation, hydrogen, and sulfur metabolism in multiple uncultivated bacterial phyla. *Science (New York, N.Y.)* 337, 1661-1665.
- Wunder, T., Cheng, S.L.H., Lai, S.-K., Li, H.-Y., Mueller-Cajar, O., 2018. The phase separation underlying the pyrenoid-based microalgal Rubisco supercharger. *Nature Communications* 9, 5076.
- Xiong, J.P., Stehle, T., Zhang, R., Joachimiak, A., Frech, M., Goodman, S.L., Arnaout, M.A., 2002. Crystal structure of the extracellular segment of integrin alpha Vbeta3 in complex with an Arg-Gly-Asp ligand. *Science (New York, N.Y.)* 296, 151-155.
- Yeates, T.O., Kerfeld, C.A., Heinhorst, S., Cannon, G.C., Shively, J.M., 2008. Protein-based organelles in bacteria: carboxysomes and related microcompartments. *Nature reviews. Microbiology* 6, 681-691.
- Yoshizawa, Y., Toyoda, K., Arai, H., Ishii, M., Igarashi, Y., 2004. CO₂-responsive expression and gene organization of three ribulose-1,5-bisphosphate carboxylase/oxygenase enzymes and carboxysomes in *Hydrogenovibrio marinus* strain MH-110. *Journal of bacteriology* 186, 5685-5691.
- Young, M.M., Tang, N., Hempel, J.C., Oshiro, C.M., Taylor, E.W., Kuntz, I.D., Gibson, B.W., Dollinger, G., 2000. High throughput protein fold identification by using experimental constraints derived from intramolecular cross-links and mass spectrometry. *Proceedings of the National Academy of Sciences of the United States of America* 97, 5802-5806.

Zhang, N., Portis, A.R., Jr., 1999. Mechanism of light regulation of Rubisco: a specific role for the larger Rubisco activase isoform involving reductive activation by thioredoxin-f. *Proceedings of the National Academy of Sciences of the United States of America* 96, 9438-9443.

Zhang, N., Kallis, R.P., Ewy, R.G., Portis, A.R., Jr., 2002. Light modulation of Rubisco in *Arabidopsis* requires a capacity for redox regulation of the larger Rubisco activase isoform. *Proceedings of the National Academy of Sciences of the United States of America* 99, 3330-3334.

Zhu, G., Jensen, R.G., 1991. Fallover of Ribulose 1,5-Bisphosphate Carboxylase/Oxygenase Activity : Decarbamylation of Catalytic Sites Depends on pH. *Plant Physiology* 97, 1354-1358.

Zhu, X.G., Long, S.P., Ort, D.R., 2010. Improving photosynthetic efficiency for greater yield. *Annu Rev Plant Biol* 61, 235-261.

Zivanov, J., Nakane, T., Forsberg, B.O., Kimanius, D., Hagen, W.J.H., Lindahl, E., Scheres, S.H.W., 2018. New tools for automated high-resolution cryo-EM structure determination in RELION-3. *eLife* 7, e42166.