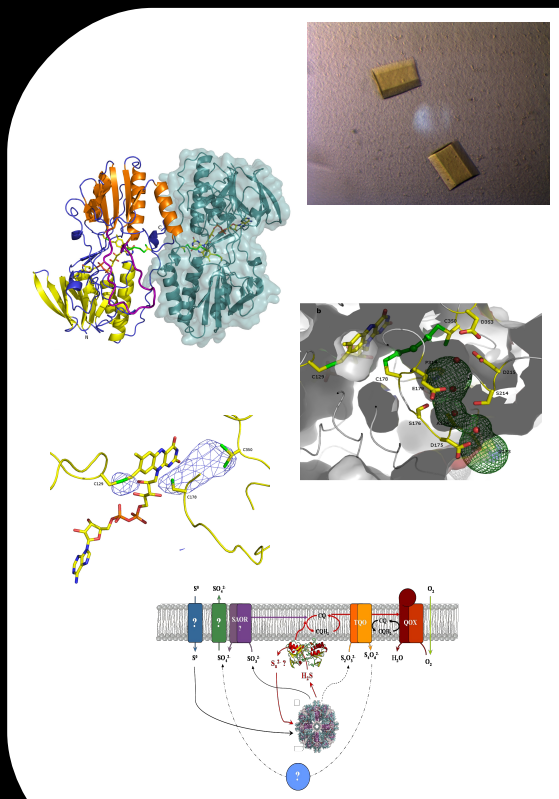
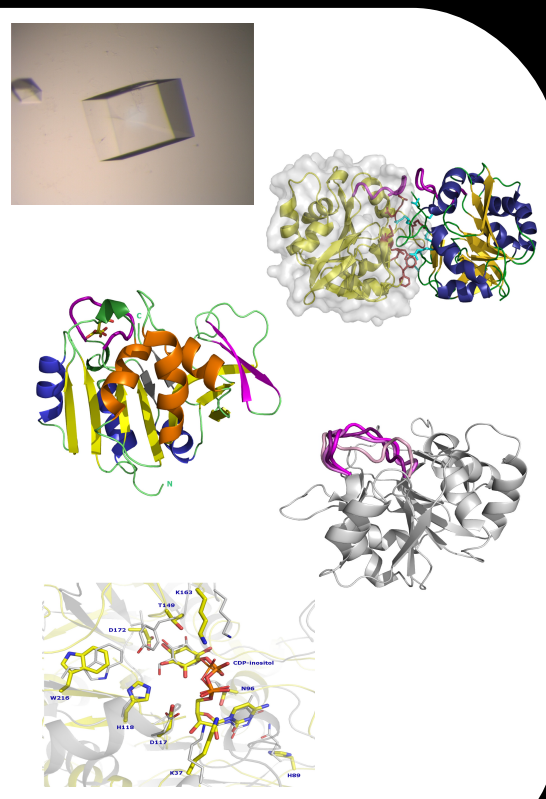
Crystallographic studies on two hyperthermophilic enzymes

Sulfide:quinone oxidoreductase from *Acidianus ambivalens* and CTP:inositol-1-phosphate cytidylyltransferase from *Archaeoglobus fulgidus*

José Artur Alves de Brito



Sulfide:quinone oxidoreductase from *Acidianus ambivalens*



CTP:inositol-1-phosphate cytidylyltransferase from *Archaeoglobus fulgidus*

Dissertation presented to obtain the Ph.D degree in Biochemistry

Instituto de Tecnologia Química e Biológica | Universidade Nova de Lisboa

Oeiras,

September, 2011

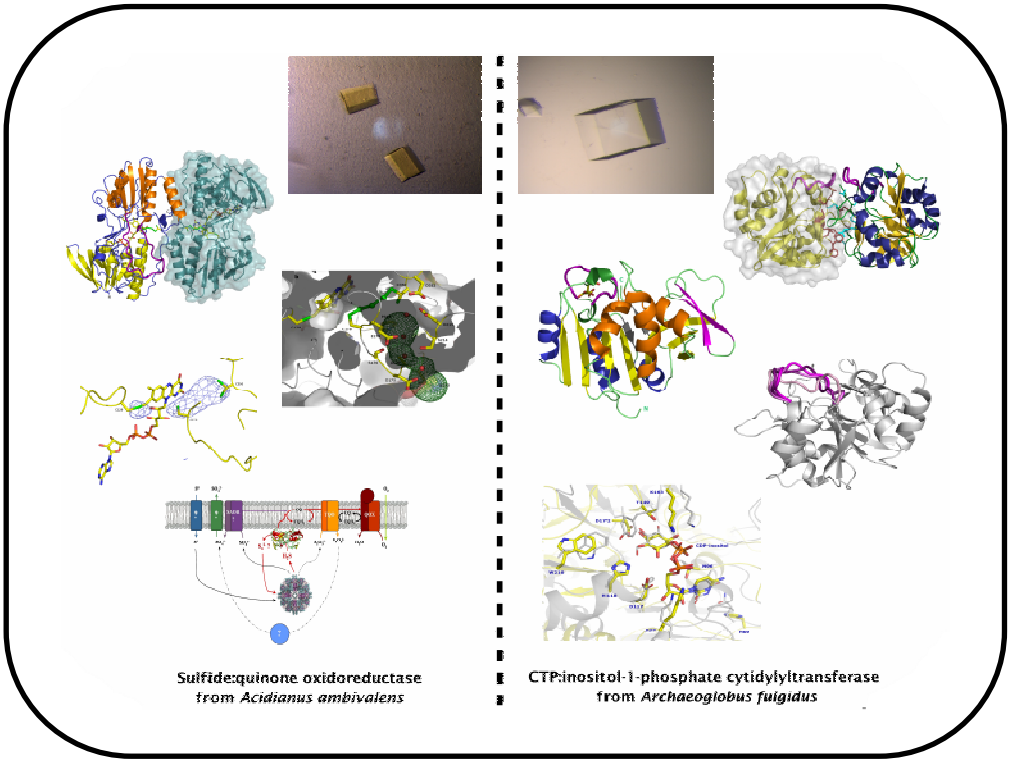
INSTITUTO DE TECNOLOGIA QUÍMICA E BIOLÓGICA /UNL

Knowledge Creation

Crystallographic studies on two hyperthermophilic enzymes

Sulfide:quinone oxidoreductase from *Acidianus ambivalens* and CTP:inositol-1-phosphate cytidylyltransferase from *Archaeoglobus fulgidus*

José Artur Alves de Brito



Dissertation presented to obtain the Ph.D degree in Biochemistry

Instituto de Tecnologia Química e Biológica | Universidade Nova de Lisboa

Oeiras,

September, 2011

INSTITUTO DE TECNOLOGIA QUÍMICA E BIOLÓGICA /UNL

Knowledge Creation

From left to right: **Doctor José Trincão** (Faculdade de Ciências e Tecnologia – Universidade Nova de Lisboa), **Professor Paulo Martel** (Faculdade de Ciências e Tecnologia – Universidade do Algarve), **Professor Maria Arménia Carrondo** (Instituto de Tecnologia Química e Biológica – Universidade Nova de Lisboa), **Doctor José Artur Brito**, **Doctor Margarida Archer** (Instituto de Tecnologia Química e Biológica – Universidade Nova de Lisboa, Supervisor), **Doctor Isabel de Moraes** (Diamond Light Source and Imperial College, United Kigdom).

2nd Edition, October 2011

ISBN: 978-989-20-2657-2

Membrane Protein Crystallography Laboratory

Macromolecular Crystallography Unit

Instituto de Tecnologia Química e Biológica

Universidade Nova de Lisboa

Av. da República, EAN

2781-901 Oeiras

Portugal

Phone: +351.214.469.761, Fax: +351.214.463.644

URL: http://mx.itqb.unl.pt/

Foreword

The present dissertation is the result of the research done at the Membrane Protein Crystallography Laboratory of the Macromolecular Crystallography Unit at Instituto de Tecnologia Química e Biológica (ITQB), from Universidade Nova de Lisboa (UNL), under the supervision of Doctor Margarida Archer.

The work was performed during the term of a 4 year Ph.D grant from Fundação para a Ciência e a Tecnologia (FCT), and concerns the X-ray structure determination of two extremozymes (enzymes from extremophilic microorganisms): the sulfide:quinone oxidoreductase from *Acidianus ambivalens* and the CTP:inositol-1-phosphate cytidylyltransferase from *Archaeoglobus fulgidus*.

The thesis comprises a first chapter that introduces the reader to the major features of extremophiles, types of extremophily, domains of Life and importance of studying this type of enzymes.

A second chapter is devoted to the crystallographic studies performed on the sulfide:quinone oxidoreductase from *Acidianus ambivalens*. An introduction to the sulfur cycle and sulfur metabolism, as well as to *Acidianus ambivalens*, a model organism to perform sulfur-related metabolic studies, topics is presented.

Chapter 3 refers to the CTP:inositol-1-phosphate cytidylyltransferase from *Archaeoglobus fulgidus* and the topic of osmolytes (also called compatible solutes). A brief description of this type of molecules and its function in cells submitted to osmotic shock is performed.

The thesis ends with a short chapter presenting some concluding remarks as well as future perspectives on the study of these two enzymes.

Acknowledgments

During this long walk on Macromolecular Crystallography (not only during my Ph.D but also before that), I am grateful to a great deal of people which I would like to express my profound gratitude:

To my Supervisor, Doctor Margarida Archer: with you I did not learn only Crystallography; I learned how to write papers, projects and grants, to organize congresses and workshops, and to manage the day-to-day tasks of a Membrane Protein Crystallography Laboratory (purchasing lab reagents and equipments, organize lab space and bench work, plan synchrotron trips, manage computer resources, …). Thank you for the opportunity you gave me to become a Linux Systems Administrator (two intensive courses on the subject at the Instituto Nacional da Administração), to present my work in several scientific meetings all around Europe and the chance to work in two World-class Science Centres, the University of Madison – Wisconsin (USA), and the University of Leeds (UK). Thank you "Chefe".

To the Macromolecular Crystallography Unit Coordinator, Professor Maria Arménia Carrondo. It has been a privilege to work next to you and to witness your continuous effort to always give the Unit the best equipments (we were the first Crystallography Group in Portugal to have a crystallization robot), the best resources (we were members of SPINE and SPINE2complexes which gave us access to world-leading experts and laboratories), to be around renowned experts in Macromolecular Crystallography (the Erice Crystallography School and the BioCrys meetings), and to access top-class research facilities (namely synchrotrons). I am also honoured to have you several times as my partner dancing the "Vira Minhoto".

To Doctor Pedro Matias, for everything you teach my regarding Linux (namely Fedora core stuff), data collection (devising a strategy and then collect the data!!), reduction (with XDS, the horrible "black box"), and phasing (your tricks in *Phaser* became legendary!!). For your patience during those long nights at the synchrotron and for always sorting out a solution to a problem that seemed unsolvable. Also for showing me where to eat an exquisite "Fish Soup" and a delicious "Shark Steak".

To Doctor Clemens Vonrhein, for your assistance solving the structures of SQR and IPCT, and for the valuable lessons regarding experimental phasing and refinement. Together with Pedro Matias, you are my Guru.

To Doctor Carlos Frazão, for all the "Crystallography 101" crash-courses you gave me whenever I had a question, and for the all assistance when running *SHELX* or interpreting the log files.

I would also like to acknowledge:

Professor Miguel Teixeira, Doctor Manuela Pereira and Doctor Filipa Sousa, for all the hard work in the SQR project. At some point, it seemed like a "Never Ending Story" but together, with your help and assistance, we managed to get a "Happy Ending".

Professor Helena Santos and Doctor Nuno Borges, for your valuable contribution, inputs and suggestions in the IPCT project. We proved that Crystallographers and "NMRers", often regarded as "opposite sides of the same coin", can work together and get very interesting and complementary research results.

Professor Peter Henderson and all the colleagues in the Astbury Centre for Structural Molecular Biology in the University of Leeds (UK), namely Doctor Nick Rutherford, Doctor Pikyee Ma and Doctor Kim Bettaney. Thank you for receiving me there and teach me all the molecular biology and protein expression and purification techniques that are now part of the day-to-day work in the Membrane Protein Crystallography Laboratory at ITQB. A special thanks to Doctor Nick Rutherford for his friendship and support.

Master Mário Correia, for his true friendship and honesty at all times. You are a good friend and I miss our sushi lunches and "Casal Garcia" afternoons.

Doctor Tiago Bandeiras and Doctor Célia Romão, for their friendship and guidance when the future (or a project), seemed rather "fainted" or "cloudy". Thank you for being my "Pool of Sorrows" when I was feeling "blue" and for all the comforting words that kept me going.

All my colleagues, past and present, in the Macromolecular Crystallography Unit for their support and friendship.

Aos meus Pais, José e Lurdes, por sempre terem conseguido, muitas vezes com prejuízo pessoal e dificuldades financeiras, proporcionar-me todas as condições para estudar e perseguir a carreira em Ciência que sempre desejei. É graças a vocês que cheguei onde cheguei e espero que vos tenha conseguido fazer ter orgulho em mim.

À Liliana, minha mulher, pelo Amor, carinho e compreensão demonstrado em todos estes anos, especialmente nesta recta final da escrita da tese. Não foi fácil mas com o teu ânimo, força e palavras de encorajamento fui arranjando força e vontade para continuar a lutar por este sonho. Nem sempre to digo ou expresso da melhor forma mas espero que saibas o quanto te adoro. És o Amor da Minha Vida e sem ti não sei nem quero viver.

Aos meus filhos, José Miguel e Maria Leonor, por permitirem e compreenderem que o Pai nem sempre estivesse com eles, especialmente quando passava dias em sincrotrão ou num qualquer país "esquisito" num congresso ou conferência. Obrigado pelas recepções calorosas e, especialmente, pelos beijos e abraços que me "sabiam a mel" quando chegava dessas viagens. Vocês são a razão do meu viver e espero que saibam o quanto o Pai vos adora.

Por fim, a três pessoas que já não estão connosco. Ao Engenheiro João Santos e ao Doutor José Sant'Ana Pereira, foi um privilégio privar convosco e ganhar de vocês o gosto pelo desconhecido, pela descoberta e pela Ciência. Com o Engenheiro João Santos aprendi o Método Científico e com o Doutor José Sant'Ana Pereira aprendi a levá-lo ao extremo e a indagar mais e mais. À minha tia Prazeres por ter sido uma segunda mãe para mim quando estava a fazer investigação na Universidade do Minho (Braga), e ter sempre uma palavra amiga e de carinho quando eu desanimava por não ter resultados. O meu muito obrigado aos três e que Deus vos tenha em descanso.

À Fundação para a Ciência e a Tecnologia (FCT), e Fundo Social Europeu (FSE), o apoio financeiro no âmbito do Quadro Comunitário de apoio (Bolsa de Doutoramento SFRH/BD/30512/2006).

Para José e Lurdes

Para Liliana, José Miguel
e Maria Leonor

Thesis publications

The work presented in this dissertation is based on the following publications:

José A. Brito, Tiago M. Bandeiras, Miguel Teixeira, Clemens Vonrhein, Margarida Archer (2006). "Crystallization and preliminary structure determination of NADH: quinone oxidoreductase from the extremophile *Acidianus ambivalens*", *Biochimica et Biophysica Acta - Proteins and Proteomics* **1764**, 842-845.

José A. Brito, Filipa L. Sousa, Meike Stelter, Tiago M. Bandeiras, Clemens Vonrhein, Miguel Teixeira, Manuela M. Pereira and Margarida Archer (2009). Structural and functional insights into sulfide:quinone oxidoreductase, *Biochemistry* **48**, 5613-5622.

José A. Brito, Nuno Borges, Helena Santos and Margarida Archer (2010). "Production, crystallization and preliminary X-ray analysis of CTP:inositol-1-phosphate cytidylyltransferase from *Archaeoglobus fulgidus*", *Acta Cryst* **F66**, 1463-1465.

José A. Brito, Nuno Borges, Clemens Vonrhein, Helena Santos and Margarida Archer (2011). "Crystal structure of *Archaeoglobus fulgidus* CTP:inositol-1-phosphate cytidylyltransferase, a key enzyme for di-*myo*-inositol-phosphate synthesis in (hyper)thermophiles", *J Bacteriol* **193**, 2177-2185.

Dissertation Abstract

While Aristotle cautioned "*everything in moderation*", the Romans, known for their eccentricities, coined the word "extremus", the superlative of exter, "being on the outside". By the fifteenth century "extreme" had arrived to English, via Middle French. At the beginning of the 21st century, we know that Earth contains environmental extremes unimaginable to our ancestors of the 19th century. Even more unimaginable to them would be the fact that there are organisms that live, and grow, in these environmental extremes. R. D. MacElroy named these organisms lovers (from the Greek "philos"), "extremophiles" as in "lovers of extreme environments".

The discovery of extremophiles has put vitality in the biotechnology industry as this discipline has exploded in the past 20 years. Several reviews have been published on extremophiles and an increasing number of meetings and conferences are organised around the theme. Genomes of extremophiles have been sequenced, patents have been filed and several funding programmes have been launched namely the US National Science Foundation and NASA's programmes in "Life in Extreme Environments, Exobiology and Astrobiology", and the European Union's "Biotechnology of Extremophiles" and "Extremophiles as Cell Factories".

Environmental extremes where we can find extremophiles include physical extremes (e.g., temperature, radiation or pressure), and geochemical extremes (e.g., desiccation, salinity or pH). Common designations of extremophiles are thermophiles, psychrophiles, acidophiles, alkaliphiles, piezophiles, xerophiles and halophiles (which thrive at high temperature, low temperature, low pH, high pH, and under extremes of pressure, dessication and salinity, respectively) (see Chapter 1 for details).

Proteins from hyperthermophiles, organisms that thrive at temperatures of 80 ^{o}C and above, and that do not propagate at temperatures of 50 ^{o}C and below, are in general intrinsically far more stable than their counterparts in

mesophilic organisms. There is enormous interest in revealing the structural basis of these stability differences, not only to understand mechanisms of high-temperature adaptation but also to be able to design proteins with a desired combination of stability and function.

Most cultivated hyperthermophiles are chemolithoautotrophes, these organisms can obtain their cell carbon from carbon dioxide and display respiratory metabolism, gaining energy from the oxidation of sulfur, sulfides, hydrogen, or ferrous iron. The most important electron acceptors are oxygen (aerobic/microaerophilic respiration), nitrate, ferric iron, sulphate, sulfur, or carbon dioxide (anaerobis respiration). Interestingly, novel metabolic pathways have been detected in some hyperthermophilic archaea. For example, glucose fermentation in *Pyrococus furiosus* proceeds via an Embden-Meyerhof-Parnas pathway involving a glucokinase and phosphofructokinase, which are ADP-dependent as opposed to the classic ATP-dependent enzymes found in other microorganisms. Thus, it appears that hyperthermophiles are excellent sources for novel enzymes and metabolic pathways.

In this thesis, we describe the work done on two hyperthermophilic enzymes: sulfide:quinone oxidoreductase (SQR) from *Acidianus* (*A.*) *ambivalens* and CTP:inositol-1-phosphate cytidylyltransferase (IPCT) from *Archaeoglobus* (*Arc.*) *fulgidus*. These are two *de novo* structures, solved in collaboration with the Metalloproteins and Bioenergetics Laboratory, headed by Professor Miguel Teixeira, and the Cell Physiology and NMR Laboratory, headed by Professor Helena Santos, both from ITQB, that contribute a little bit to the effort of understanding life and the mechanisms that underlie thermostability.

The SQR enzyme was isolated from the membranes of the hyperthermoacidophilic archaeon *A. ambivalens* and was functionally and structurally characterized. The enzyme was previously characterized as a type-2 NADH:quinone oxidoreductase but the X-ray structure prompted for a new characterization. The structure was solved to 2.6 Å resolution by the single isomorphous replacement with anomalous scattering (SIRAS) method using two iodide and platinum derivatives and

showed a covalently bound FAD cofactor and three sulfur atoms bridging two active site cysteine residues. A catalytic mechanism for sulfide oxidation is presented and the sulfur dependent bioenergetics of *A. ambivalens* is discussed.

IPCT from *Arc. fulgidus* is the cytoplasmic N-terminal domain of a bifunctional enzyme (IPCT/DIPPS), that catalyzes the conversion of L-*myo*-inositol-1-phosphate and CTP to di-*myo*-inositol-1,3'-phosphate-1'-phosphate (DIPP). The IPCT domain catalyzes the initial reaction of L-*myo*-inositol-1-phosphate with CTP to produce CDP-inositol, and the DIPPS C-terminal membrane domain catalyzes the reaction of CDP-inositol with L-*myo*-inositol-1-phosphate to produce the final product DIPP. This molecule is further dephosphorylated leading to the production of di-*myo*-inositol-1,3'-phosphate (DIP), which is the most widespread organic solute in (hyper)thermophiles. The enzyme was solved at 1.9 Å resolution by the SIRAS method using a mercury derivative. IPCT is composed of a central seven-stranded mixed β-sheet surrounded by six α-helices, a fold reminiscent of the dinucleotide-binding Rossmann fold and typical of nucleotide-diphospho-sugar transferases (SCOP Superfamily 53448). The enzyme shares structural homology with other pyrophosphorylases and based on this homology, CTP, inositol-1P and CDP-inositol were docked into the catalytic site.

Resumo da Dissertação

Enquanto que Aristóteles advertia "*tudo com moderação*", os Romanos, conhecidos pelas suas excentricidades, cunhavam a palavra "Extremus", o superlativo de Exter, "estar do lado de fora". No século XV "extreme" já tinha chega ao Inglês, via Francês Médio. No início deste século XXI, conhecemos na Terra condições ambientais extremas inimagináveis para nossos antepassados do século XIX. Ainda mais inimaginável para eles seria o facto de haverem organismos que vivem, e crescem, nestas condições ambientais extremas. R. D. MacElroy chamou a estes amantes (do grego, "philos"), "extremófilos", como em "amantes de ambientes extremos".

A descoberta dos extremófilos causou grande vitalidade na indústria biotecnológica à medida que esta disciplina emergia nos últimos 20 anos. Várias revisões têm sido publicadas sobre extremófilos e um número crescente de reuniões e conferências têm sido feitas em torno do tema. Genomas de extremófilos foram sequenciados, patentes foram depositadas e vários programas de financiamento têm sido lançados nomeadamente os programas da National Science Foundation e da NASA (EUA), "Vida em ambientes extremos, Exobiologia e Astrobiologia", e os programas da União Europeia "Biotecnologia de Extremófilos" e "Extremófilos como Fábricas Celulares".

As condições ambientais extremas onde podemos encontrar organismos extremófilos incluem extremos físicos (por exemplo, temperatura, radiação ou pressão), e extremos geoquímicos (por exemplo, dissecação, salinidade ou pH). Entre as designações comuns de extremófilos encontramos termófilos, psicrófilos, acidófilos, alcalófilos, piezófilos, xerófilos ou halófilos (que prosperam em alta temperatura, baixa temperatura, pH baixo, pH alto, e sob condições extremas de dissecação, pressão e salinidade, respectivamente) (ver Capítulo 1 para detalhes).

Proteínas de hipertermófilos, organismos que se desenvolvem a temperaturas acima de 80 ^{o}C, e que não proliferam a temperaturas abaixo

dos 50 ºC, são, em geral, intrinsecamente muito mais estáveis do que suas homólogas de organismos mesófilos. Existe um enorme interesse em revelar a base estrutural destas diferenças de estabilidade, não só para compreender os mecanismos de adaptação a altas temperaturas, mas também para permitir a engenharia de proteínas com uma combinação desejada de estabilidade e função.

A maioria dos hipertermófilos crescidos em laboratório são quimolitoautotróficos. Estes organismos podem obter o carbono celular a partir de dióxido de carbono e apresentam metabolismo respiratório obtendo energia a partir da oxidação do enxofre, sulfuretos, hidrogénio, ou ferro ferroso. Os receptores de electrões mais importantes são o oxigénio (respiração aeróbia / microaerófilos), nitrato, ferro férrico, sulfato, enxofre, ou dióxido de carbono (respiração anaeróbia). Curiosamente, novas vias metabólicas foram detectadas em algumas hipertermófilos do domíno Archaea. Por exemplo, a fermentação da glucose em *Pyrococus furiosus* prossegue através da via de Embden-Meyerhof-Parnas envolvendo uma glucoquinase e uma fosfofrutoquinase que são ADP-dependentes, em comparação com as clássicas enzimas ATP-dependentes encontradas em outros microorganismos. Assim, os hipertermófilos aparecem como excelentes fontes de novas enzimas e vias metabólicas.

Nesta tese, descrevemos o trabalho realizado em duas enzimas de hipertermófilos: a sulfureto:quinona oxidorredutase (SQR) de *Acidianus* (*A.*) *ambivalens* e a CTP: inositol-1-fosfato citidililtransferase (IPCT) de *Archaeoglobus* (*Arc.*) *fulgidus*. Estas são duas estruturas *de novo*, determinadas em colaboração com o Laboratório de Metaloproteínas e Bioenergética, liderado pelo Professor Miguel Teixeira, e o Laboratório de Fisiologia Celular e RMN, liderado pela Professora Helena Santos, ambos do ITQB, que contribuem para este esforço de compreensão da vida e dos mecanismos subjacentes à termoestabilidade. A enzima SQR foi isolada das membranas do archaeon hipertermoacidófilo *A. ambivalens* e foi funcional e estruturalmente caracterizada. Esta enzima tinha sido anteriormente caracterizada como uma NADH:quinona oxidorredutase do tipo-2 mas a estrutura de raio-X apontou para uma nova caracterização A estrutura foi resolvida a 2.6 Å de resolução

pelo método SIRAS (*do ingl.*, Single Isomorphous Replacement with Anomalous Scattering) usando dois derivados de iodeto e platina, e mostrou uma molécula de FAD como cofactor covalentemente ligada, e três átomos de enxofre ligando duas cisteínas no centro activos. É sugerido um mecanismo catalítico para a oxidação do sulfureto e é discutida a bioenergética dependente do enxofre em *A. ambivalens*.

A IPCT de *Arc. fulgidus* é o domínio N-terminal citoplasmático de uma enzima bifuncional (IPCT / DIPPS), que catalisa a conversão de L-*myo*-inositol-1-fosfato e CTP a di-*myo*-inositol-1,3'-fosfato-1'-fosfato (DIPP). O domínio IPCT catalisa a reação inicial de L-*myo*-inositol-1-fosfato com CTP para produzir CDP-inositol, e o domínio membranar C-terminal DIPPS catalisa a reação do CDP-inositol com L-*myo*-inositol-1-fosfato para produzir o produto final DIPP. Esta molécula é posteriormente desfosforilada levando à produção de di-*myo*-inositol-1,3'-fosfato (DIP), sendo que este é o soluto orgânico mais difundido em (hiper)termófilos. A esterutura da enzima foi determinada a 1.9 Å de resolução por SIRAS usando um derivado de mercúrio. A IPCT é composta de uma folha-β central rodeada por seis hélices-α, uma reminiscência do arranjo de Rossmann de ligação a nucleótidos, um arranjo típico das difosfotransferases de nucleótido-açúcar (Superfamília SCOP 53448). A enzima partilha homologia estrutural com outras pirofosforilases e com base nesta homologia, as moléculas de CTP, inositol-1P e CDP-inositol foram modeladas no centro activo.

Abbreviations

A	Adenine
A.	*Acidianus*
Å	Angström
a.a.	amino acid
Ac.	*Acidithiobacillus*
All.	*Allochromatium*
Aq.	*Aquifex*
Arc.	*Archaeoglobus*
C	Cytosine
CDP	Cytidine diphosphate
CDP-inositol	Cytidine diphosphate-L-*myo*-inositol
CTP	Cytidine triphosphate
Da	Dalton
DDM	n-dodecyl-β-D-maltoside
DIP	Di-*myo*-inositol-1,3'-phosphate
DIPP	Di-*myo*-inositol-1,3'-phosphate-1-phosphate
DIPPS	Di-*myo*-inositol-1,3'-phosphate-1-phosphate synthase
DNA	Deoxyribonucleic Acid
dTDP	Deoxythymidine diphosphate
dTTP	Deoxythymidine triphosphate
dUQ	Decylubiquinone
EDTA	2,2',2",2'"-(ethane-1,2-diyldinitrilo)tetraacetic acid
e.g.	example given (for example, for instance)
ESRF	European Synchrotron Radiation Facility
et al.	And other people
FAD	Flavin adenine dinucleotide
FCC	Flavocytochrome *c*
FDRs	flavoprotein disulfide reductases
FMN	Flavin mononucleotide
G	Guanine

GlmU	N-acetylglucosamine-1-phosphate uridylyltransferases
Gy	Gray
Gyr	Billion year
HEPES	n-(2-hydroxyethyl)-piperazine-n'-2-ethanesulfonic acid
HQNO	2-heptyl-4-hydroxy-quinolone-n-oxide
Inositol-1P	L-*myo*-inositol-1-phosphate
IPCT	CTP:inositol-1-phosphate cytidylyltransferase
J	Joule
LUCA	Last Universal Common Ancestor
M	Molar
MALDI	Matrix-assisted laser desorption ionization
MW	Molecular Weight
Myr	Million year
NADH	Reduced nicotinamide adenine nucleotide
NCS	Non-crystallographic symmetry
NDH	NADH dehydrogenase
NMR	Nuclear Magnetic Resonance
OD	Optical density
Pa	Pascal
PAGE	Polyacrilamide gel electrophoresis
PDB	Protein Data Bank
RmlA	Glucose-1-phosphate thymidylyltransferase
RNA	Ribonucleic Acid
rRNA	Ribossonal ribonucleic cid
SDS	Sodium dodecylsulfate
SQR	Sulfide:quinone oxidoreductase
SSU	Small subunit
T	Thymine
Tris	Tris(hydroxymethyl)-aminomethane
U	Uracil
UV	Ultraviolet

Aminoacids

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

Table of Contents

Chapter 1

INTRODUCTION

1.1. WHAT IS LIFE?

Schrödinger (Figure 1.1), was the first one to ask it formally but many others before him, as most likely some of us today, have made the question: "What is Life?" (Schrödinger, 1944). Not only this is a scientific and intellectual challenging question but also one that closes an esoteric and holistic meaning. However, if many have asked the question, only a few have tried to answer it; and even fewer have tried to give "Life" a definition. The failures are considerable: physiological, metabolic, biochemical, genetic and thermodynamic definitions for Life were attempted but all of them faced serious problems (Chyba & Hand, 2005). As examples, we can refer to the metabolic, which finds it hard to exclude fire (which grows and reproduces via chemical reactions), and the biochemical definitions of Life, which does not exclude enzymes (which are biologically functional but not living systems), as well as the thermodynamic definition which does not exclude mineral crystals (which create and sustain local order and may reproduce) (Jortner, 2006).

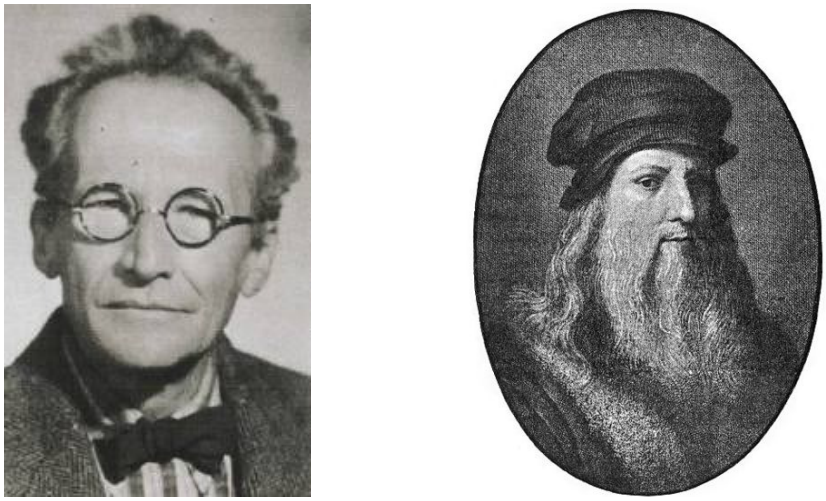


Figure 1.1 – Edwin Schrödinger (1887 – 1961, left), and Leonardo da Vinci (1452 – 1519, right), two of History's greatest thinkers.
Schrödinger received the Nobel Prize in Physics in 1933 for "*the discovery of new productive forms of atomic theory*", namely his contribution to the quantum mechanism field with the so-called «Schrödinger's equation»; Leonardo da Vinci was one of the most versatile and productive men ever being a painter, sculptor, architect, musician, scientist, mathematician, engineer, inventor, anatomist, geologist, cartographer, botanist and writer.

More than four centuries before Schrödinger, Leonardo da Vinci (Figure 1.1), also tried to answer a simple, yet similarly complex at that time, question: "What is water?" (da Vinci, 1513). It was not only until the twentieth century, with the establishment of the proper molecular composition and the structure of the H_2O molecule, together with the globally condensed phase properties of the liquid (*e.g.*, H-bonding, local order, radial and angular distribution, solvation and nuclear dynamics, among others), that a conceptual framework of an appropriate scientific theory was provided (Jortner, 2006). The same happens with "Life": we do not need a definition to the term but rather a scientific theory to it (Cleland & Chyba, 2002).

A notable attempt to provide an unified description of Life was provided by Lars Onsager and Harold Morowitz (Figure 1.2): "Life is that property of matter that results in the cycling of bioelements in aqueous solution, ultimately driven by radiant energy to attain maximum complexity" (Folsome, 1979). However, the Onsager-Morowitz definition overrides the characterization of complex biological matter and how it differs from complex chemical matter. Eigen (Eigen, 1971, Eigen, 1996), addressed this issue and the basic differences between a chemically coupled system and a living system. The three essential characteristics of all living systems, yet known, are self-reproduction (without which information would be lost), mutations (which allow evolution), and metabolism (which allows an optimal choice of a system for a certain function). It was the same author, together with others [for review see (Jortner, 2006)], that advanced and developed the concept of self-organization (self-assembly), and proposed that it resulted in the evolution of biological complex matter (Eigen, 1971, Eigen, 1996).

1.2. LIFE ON EARTH

Earth has approximately 4.5 billion years (Gyr), and it is assumed that Life arose within 1 Gyr of its formation, i.e, 3.5 Gyr ago. In its early stages, Earth was a fairly inhospitable place when we compare it to present day: it was hot, lacked

oxygen in the atmosphere and, as a consequence, it could only form little ozone to protect the emerging biota from harmful ultraviolet (UV) light. However, studies in the geological layers of the Issua formation in Greenland (Mojzsis *et al.*, 1996, Schidlowski, 1988), showed traces of metabolic activity (carbon metabolism), indicating that Life existed on Earth as early as 3.8 Gyr ago. These observations suggested that Life began rather rapidly, roughly within 200 million years (Myr), of Earth formation, when the planet cooled and thereby became habitable for carbon-based Life (Nealson & Conrad, 1999).

Figure 1.2 – Lars Onsanger (1903 – 1976, left), and Harold Morowitz (1927 – , right). Onsanger was a norwegian (naturalized American), physical-chemist and theoretical physicist, winner of the 1968 Nobel Prize in Chemistry for his work "*for the discovery of the reciprocal relations bearing his name, which are fundamental for the thermodynamics of irreversible processes*". Morowitz is an american biophysicist noted for a number of publications, beginning with his 1963 "*Life and the Physical Sciences*", on the topic of the use and application of thermodynamics in the study of biology, with focus on the origin of life topic. Morowitz' believes that life arose on earth by processes that may be understood by considering the laws of chemistry and physics as applied to complex adaptive systems.

These discoveries triggered speculation about Life in general but it is widely accepted that a set of conditions had to be present at a certain moment for Life to begin. These are (i), the existence of liquid water as an universal solvent (Blum, 1962, Darwin, 1871, Oparin, 1924); geological evidence indicates that liquid water might

have been present on Earth 4.3 Gyr ago (Watson & Harrison, 2005); (ii) the availability of raw materials like biotic elements (*e.g.*, C, H, N, O, P, S, Fe, Mg), simple biotic molecules (*e.g.*, H_2, H_2O, NH_3, CO, CO_2, H_2S, CH_4, S and SO_2), and minerals containing Si, Fe, Ni, P and S (Crick, 1981, Miller & Orgel, 1974, Folsome, 1979); terrestrial sources for these elements can be minerals, hydrothermal sources, marine aerosols, volcanic exhalations and the atmosphere of the primitive Earth itself (Jortner, 2006); (iii) the availability of sources of energy, like solar light, lightning, cosmic rays and heat; and, (iv) the prevalence of appropriate chemical and thermal conditions, like the existence of proper sites for biosynthesis and temperatures high enough to drive activated chemical reactions (Chyba & Hand, 2005). For the sake of simplicity, we will not refer to Panspermia, as proposed by Richter, Lord Kelvin and Arrhenius, which claims that reproductive bodies of living organisms can exist throughout the Universe and develop wherever the environment is favourable (Richter, 1865, Arrhenius, 1903, Thomson, 1894). In this case, interstellar dust particles might have been the source of prebiotic C, H and N (Chyba & Sagan, 1992, Chyba & Hand, 2005), pointing towards the connection between interstellar chemistry and the terrestrial origin of Life.

The bulk of evidence of the early Life on Earth is in the form of chemical records, since there are very few well preserved ancient fossils (Knoll, 1992, Schopf & Klein, 1992). Chemistry tells us that until about 2 Gyr ago there was very little oxygen on the planet and the development of eukaryotic cells (which thrive through oxygenic respiration), was not possible. Based on the study of ancient soils, it was concluded that oxygen first appeared in the atmosphere approximately 2 Gyr ago (Rye & Holland, 1998), and reached current levels approximately 500 Myr ago (Knoll, 1992, Schopf & Klein, 1992). The appearance of atmospheric oxygen made possible the development of eukaryotic organism but even before the rise of oxygen, Earth was already crowded with Life.

Since the mid-1700s, from the work of Linnaeus (Figure 1.3), Life was seen as the "traditional five kingdoms" tree (Monera, Protista, Fungi, Animalia and Plantae, Figure 1.3). That diagram, called phylogenetic tree but also known as the "Tree of Life", was used to illustrate the evolutionary relationships that may exist, or

have existed (fossil records), among organisms. Of course, that Tree was a logical consequence of the taxonomic system used, a system that relied on the observation of the visible features of organisms to give each one a name (*e.g.*, *Homo sapiens*, for humans; *Canis familiaris*, for dogs), and to group together organisms of similar appearance (Nealson & Conrad, 1999). It is clear, that this classification was strongly influenced by the available tools at the time: human eyes, hand lens and, after the work of Leeuwenhoek (Figure 1.3), the optical microscope. Hence, it is not surprising that this tree was dominated by the macroscopic, multicellular organisms such as fungi, animals and plants.

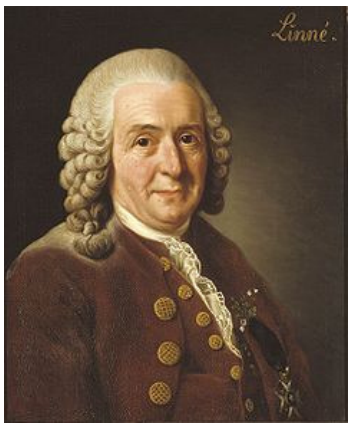
The groups Fungi, Plantae and Animalia, being eukaryotic, were considered to be the most developed organisms and topped the tree. The tiny Protists (amoeba, paramecia, giardia), although also eukaryotes, being visible but not understood, were relegated to a second level in the tree, while the microscopic prokaryotes (the Monera group also called Bacteria), were relegated to the very bottom. This view of the biosphere has dramatically changed in the past three decades with the advent of molecular taxonomy and phylogeny. From an eukaryote-centric view (five kingdoms, three of them eukaryotic), we moved towards a prokaryote-centric view (with three domains/kingdoms, two of them prokaryotic).

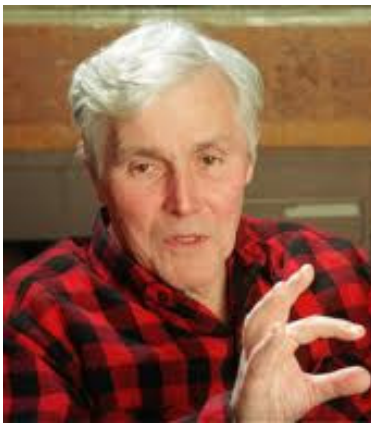
NEXT PAGE: Figure 1.3 - Carl Linnaeus (1707 – 1778, top left), and Carl Woese (1928 - , top right). Classical five kingdoms Tree of Life as biodiversity was seen since the 1700s (middle left), and phylogenetic tree based on the SSU rRNA as proposed by Woese (middle right). Antoine von Leeuwenhoek (1632-1723, bottom left), and examples of 18th century microscopes (Musée des Arts et Métiers, Paris).

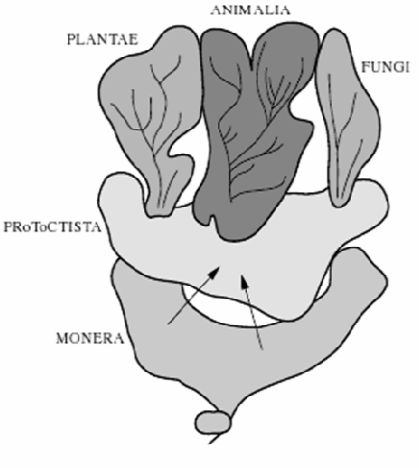
Linnaeus (or Lineus or Linné), was a swedish botanist, physician, and zoologist, who laid the foundations for the modern scheme of binomial nomenclature; he is known as the father of modern taxonomy, and is also considered one of the fathers of modern ecology.

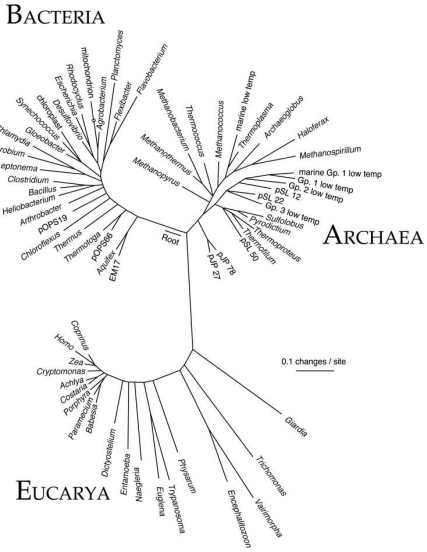
Woese is an American microbiologist and physicist known to conceive a new concept for the Tree of Life and Kingdoms, or Domains, of Life; he was the first to introduce the Archaea domain in 1977 by phylogenetic taxonomy of SSU ribosomal RNA, a technique he pioneered.

Leeuwenhoek was a dutch scientist considered to be the first microbiologist and the great improver of the optical microscope.


Linné




ANIMALIA
PLANTAE
FUNGI
PROTOCTISTA
MONERA


BACTERIA
ARCHAEA
EUCARYA
Root
0.1 changes / site
Homo
Zea
Cryptomonas
Giardia
Trichomonas
Physarum
Euglena
Entamoeba
Dictyostelium
Paramecium
Vairimorpha
Encephalitozoon

Molecular phylogeny, as proposed by Carl Woese (Figure 1.3), in 1977 (Woese & Fox, 1977), used comparative analysis of the small subunit (SSU) ribosomal RNA (rRNA), a slowly evolving molecule common to nearly all Life on Earth, to create this new view of prokaryotic diversity. Woese's analysis showed that the four eukaryotic kingdoms were a rather homogeneous single group, while the prokaryotes were found to be sufficiently diverse that they should be expanded (Nealson & Conrad, 1999). Moreover, Woese proposed the creation of a separate group of prokaryotes, besides Bacteria, which he called, originally, Archaebacteria (Woese & Fox, 1977).

1.3. THE ARCHAEA DOMAIN

The proposal of Woese (Woese & Fox, 1977), for the existence of a third domain in Life, the Archaea, took the discussion to a whole new ground and reignited the controversy on the basic dichotomy "prokaryote *versus* eukaryote". Prokaryotic cells are simple and usually small (1-2 μm), with few or no intracellular structures, no nuclear membrane surrounding the genetic material (DNA), and (often) rigid cell walls. On the other hand, eukaryotic cells are larger (10-25 μm), more complex and with intracellular structures, a nucleus (surrounded by a nuclear membrane), and (for the animals) usually non-rigid cell walls. Moreover, prokaryotes are unicellular while eukaryotes can be either unicellular (algae and protists), or multicellular. For most of the eukaryotes, their energy source is organic carbon and the respiratory metabolism is oxygen-based. In clear contrast, prokaryotes can have a great diversity of both organic and inorganic energy sources and respiratory oxidants, surviving perfectly in the absence of oxygen and with a wide variety of oxygen "substitutes" for respiration (Figure 1.4) (Gottschalk, 1994, Nealson, 1997, Schlegel, 1993).

Although, the assignment of Archaea as a third domain of Life has been validated by comparative genomics, for several years this domain was puzzling since numerous components of archaeal informational processes are more similar to their

eukaryotic than bacterial homologues. Moreover, some of them are uniquely shared between Archaea and Eukarya to the exclusion of Bacteria. Striking cases are the DNA replication apparatus, with archaeal/eukaryal primases, helicases and replicative polymerases being totally unrelated to their bacterial counterparts (Olsen & Woese, 1996), or a number of ribosomal proteins that are uniquely shared between Archaea and Eukarya while none is uniquely shared with Bacteria or between Bacteria and Eukarya (Lecompte *et al.*, 2002). However, Archaea are also quite similar to Bacteria in many aspects, like the size and organization of their chromosomes, the existence of polycistronic transcription units and the utilization of Shine-Dalgarno sequences for the initiation of translation (although nor exclusively) (Londei, 2005).

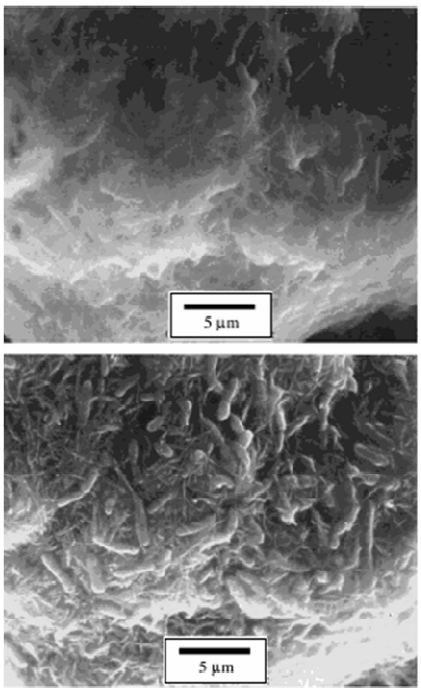


Figure 1.4 - Dissimilatory metal reducing bacteria *Shewanella oneidensis* (also known as *Shewanella putrefaciens*) growing under anaerobic conditions.
In the absence of molecular oxygen, this bacteria can use either iron oxides or manganese oxides as `oxygen substitutes' for respiration. It attaches to such oxides, uses them as its respiratory oxidant, eventually dissolving the oxides and producing reduced metals (ferrous or manganous ions) in solution: taken with an environmental scanning electron microscope shows what the organism looks like in nature – a particle of metal oxide coated with bacteria that are invisible because of the polysaccharide film produced by the bacteria (top), and taken in the same orientation and with the same magnification, but with a high vacuum scanning electron microscope, the bacteria and the metal oxide on which it sits are clearly visible (down) [from (Nealson & Conrad, 1999)].

In summary, the Archaea could be regarded as organisms that use eukaryotic-like proteins in a bacterial context (Bell *et al.*, 2002, Grabowski & Kelman,

2003, Myllykallio *et al.*, 2000). This is easily perceptible if we analyze some of the universal rooted phylogenetic trees that were produced between the end of the 1980s and the end of the 1990s [check (Gribaldo & Brochier-Armanet, 2006) for an excellent review on the origin and evolution of the Archaea]. All of these studies agree that Bacteria derived directly from the Last Universal Common Ancestor (LUCA), while Archaea and Eukarya share a common ancestor more recent than LUCA and, hence, are sister lineages (Woese *et al.*, 1990).

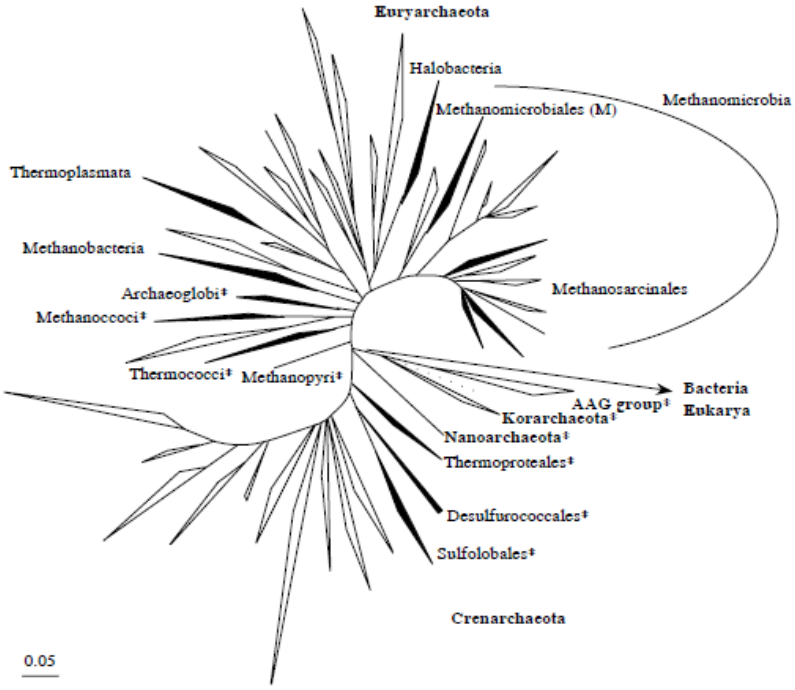


Figure 1.5 – SSU rRNA archaeal phylogeny [adapted from (Schleper *et al.*, 2005)].
Triangles are proportional to the diversity of groups; filled triangles represent groups for which cultivated species are available and empty triangles indicate groups represented only by environmental sequences; asterisks indicate groups with hyperthermophilic species (taken from (Gribaldo & Brochier-Armanet, 2006).

As shown in Figure 1.5, SSU rRNA sequences from Archaea are currently assigned to four phyla: the Crenarchaeota, the Euryarchaeota, the Nanoarchaeota

and the Korarchaeota. However, only the Crenarchaeota and the Euryarchaeota phyla are recognized by the *Bergey's Manual of Systematic Bacteriology* (Boone & Castenholz, 2001).

The Nanoarchaeota phylum was proposed by Huber *et al.* (Huber *et al.*, 2002) and currently has only one representative, the hyperthermophilic coccoid-shaped *Nanoarchaeum* (*N.*) *equitans* (Figure 1.6). This organism was isolated from a submarine hydrothermal system at the Kolbeinsey Ridge (Iceland), is the smallest living organism known with a cell diameter of *ca.* 400 nm and grows attached to the surface of a crenarchaeotal host, a member of the genus *Ignicoccus*. Cells of *N. equitans* were never found isolated, i.e., without being attached to its host, and the fact that it has a limited biosynthetic and catabolic capacity suggests that this symbiotic relationship might be parasitic (Stetter, 2006). Two recently found SSU rRNA sequences from the Uzon Caldeira (Kamchatka, Russia), and Yellowstone National Park (USA), exhibit 83 % sequence similarity to *N. equitans* and, therefore, should represent a distinct family within this phylum. These two novel Nanoarchaeota are again tiny cocci, about the size of *N. equitans*, but grow attached to a different host, in this case a rod-shaped *Pyrobaculum*-like host (Stetter, 2006). According to Kandler (Kandler, 1994), symbiotic systems like these may represent remains from pre-Archaea communities.

The phylum Korarchaeota was proposed more than ten years ago and for a long time it did not had any representative including only sequences from uncultivated (environmental) species. Analysis of their SSU rRNA gene sequences suggests that they are a deeply-branching lineage that does not belong to the main archaeal groups Crenarchaeota and Euryarchaeota (Barns *et al.*, 1996). Analysis of the genome of one korarchaeote (*Candidatus Korarchaeum cryptofilum* OPF8), that was enriched from a mixed culture, revealed a number of both Crenarchaeota- and Euryarchaeota-like features that supports the hypothesis of deep-branching ancestry (Elkins *et al.*, 2008).

Recently, a new phylum was proposed based on phylogenetic data and the presence of a form of type I topoisomerase that was previously thought to be unique to eukaryotes (Brochier-Armanet *et al.*, 2008a, Brochier-Armanet *et al.*, 2008b). The

newly-proposed phylum is called Thaumarchaeota and contains four species: *Nitrosopumilus maritimus*, *Cenarchaeum symbiosum*, *Nitrososphaera viennensis* and *Nitrososphaera gargensis* (Brochier-Armanet *et al.*, 2008a, Tourna *et al.*, 2011) All organisms from this lineage are chemolithoautotrophic ammonia-oxidizers and may play important roles in biogeochemical cycles such as the nitrogen and carbon cycles.

Figure 1.6 – Four cells of *N. equitans* (small spheres), attached to *Ignicoccus* strain Kin 4M (large).
Freeze-etched image from transmission electron micrographs; scale bar at bottom represents 1 μm [taken from (Huber *et al.*, 2002)].

1.4. EXTREMOPHILES

As mentioned above, organisms from the domain Eukarya are almost entirely limited to growth on organic carbon with oxygen as the oxidant and any set of conditions in which organic carbon or oxygen is absent is potentially life threatening. For Archaea and Bacteria, on the other hand, these environments simple address a challenge to continue living with a different metabolic system. Having a simple structure confers prokaryotes a degree of environmental toughness not seen in the more complex eukaryotes. The word "extremophile" (from the latin, "extremus" meaning extreme, out of range; and "philos", meaning love or lover), was coined to refer to these organisms that endure tough/extreme conditions; these "extremes" can either be physical (*e.g.*, temperature, radiation or pressure), or geochemical extremes (*e.g.*, desiccation, salinity, pH, oxygen species or redox potential) (Table 1.1) (Nealson & Conrad, 1999, Rothschild & Mancinelli, 2001).

Table 1.1 – Classification and examples of extremophiles [adapted from (Rothschild & Mancinelli, 2001)]

Environmental parameter	Type	Definition	Examples
Temperature	Hyperthermophile	Growth > 80 °C	*Pyrolobus fumarii*, 113 °C
	Thermophile	Growth 60 – 80 °C	*Synechococcus lividis*
	Mesophile	16 – 60 °C	Homo sapiens
	Psychrophile	< 15 °C	*Psychrobacter*, some insects
Radiation			*Deinococcus radiodurans*
Pressure	Barophile	Weight-loving	Unknown
	Piezophile	Pressure-loving	For microbe, 130 MPa
Vacuum		Tolerates vacuum	Tardigrades, insects, microbes and seeds
Desiccation	Xerophiles	Anhydrobiotic	*Artemia salina*; nematodes, microbes, fungi, lichens
Salinity	Halophile	Salt-loving (2-5 M)	Halobacteriaceae, *Dunaliella salina*
pH	Alkaliphile	pH > 9	*Natronobacterium*, *Bacillus firmus* OF4, *Spirulina* spp. (all pH 10.5)
	Acidophile	Low pH-loving	*Cyanidium caldarium*, *Ferroplasma* sp. (both pH 0)
Oxygen tension	Anaerobe	Cannot tolerate O_2	*Methanococcus jannaschii*
	Microaerophile	Tolerates some O_2	*Clostridium*
	Aerobe	Requires O_2	*Homo sapiens*
	Gases		*Cyanidium caldarium* (pure CO_2)
Chemical extremes	Metals	Can tolerate high concentrations of metals (metalotolerant)	*Ferroplasma acidarmanus* (Cu, As, Cd, Zn) *Ralstonia* sp. CH34 (Zn, Co, Cd, Hg, Pb)

An organism that thrives in an extreme environment is an extremophile; in more than one extreme it is a polyextremophile (Rothschild & Mancinelli, 2001). It should be kept in mind that it is seldom in nature that an organism encounters just one extreme (Nealson & Conrad, 1999). The words "extremophilic organisms",

build-up in the past 30 years to the synonym of "prokaryotic organism", yet the taxonomic range of extremophiles spans all three domains of Life (see below).

TEMPERATURE

Temperature creates a series of problems, from the structural destruction through ice crystals to the denaturation of biomolecules. Temperatures approaching 100 °C normally denature proteins and nucleic acids, and increase membranes fluidity to lethal levels. Moreover, chlorophyll degrades above 75 °C hampering photosynthesis. However, in nature we find thermal preferences of microorganisms ranging from hyperthermophilic (maximum growth above 80 °C, see Table 1.1), to psychrophilic (maximum growth below 15 °C) (Rothschild & Mancinelli, 2001).

All hyperthermophilic organisms belong to the Archaea domain with *Pyrolobus fumarii*, a crenarchaeote capable of growing up to 113 °C, being the current "record-holder" (Blochl *et al.*, 1997). However, hyperthermophilic enzymes can have an even higher optimum temperature like amylopullulanase which has activity up to 142 °C (Schuliger *et al.*, 1993).

Thermophilic species (optimum growth between 60 – 80 °C), are found among phototrophic bacteria (*e.g.*, cyanobacteria, purple and green bacteria), eubacteria (*e.g.*, *Bacillus*, *Clostridium*, *Thiobacillus* and *Thermus* among numerous other genera), and archaea (*e.g.*, *Acidianus*, *Archaeoglobus*, *Pyrocuccus* and *Sulfolubus* genera, just to mention a few). In contrast, for eukaryotes the upper limit is *ca.* 60 °C (Figure 1.7), a temperature suitable for some protozoa, algae and fungi. The maximum temperature for mosses is about 50 °C, for vascular plants about 48 °C and for fish about 40 °C (possibly owing to the low solubility of oxygen at high temperatures (Rothschild & Mancinelli, 2001).

Representatives of all major taxa inhabit temperatures just below 0 °C. Many microorganisms and cell lines can be successfully preserved in liquid nitrogen (-196 °C), but the lowest recorded temperature for active microbial communities is -18 °C (Clarke, 2003). For animals, the Himalayan midge is active also at -18 °C (Kohshima, 1984).

At low temperatures and with nucleation, water freezes. The freezing of intracellular water and the resulting ice crystals can destroy cell membranes which is invariably lethal. The only exception to this rule, besides cryopreservation, is the nematode *Panagrolaimus davidii* which can withstand freezing of all body water (Figure 1.9) (Wharton & Ferns, 1995).

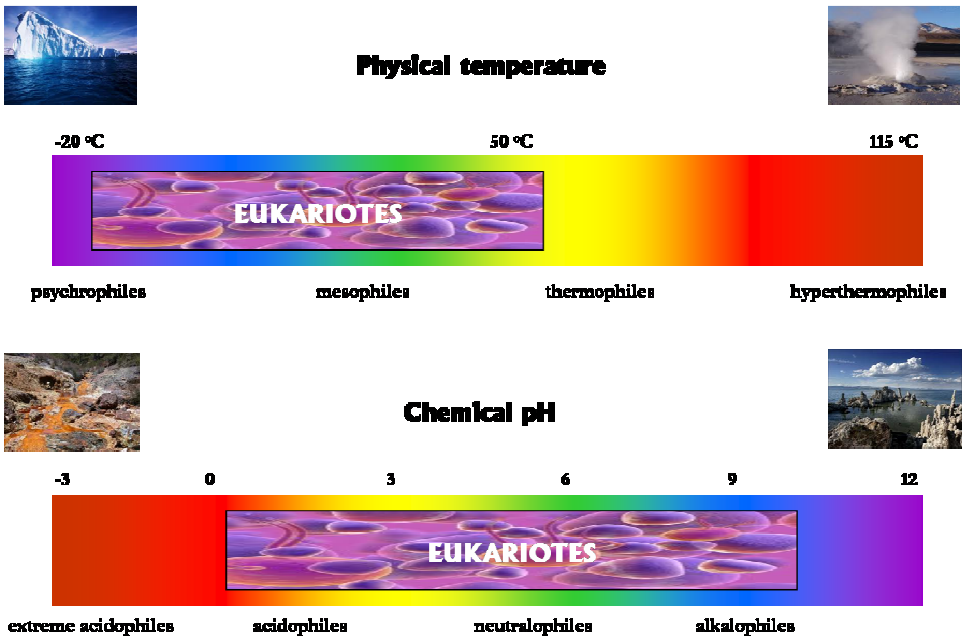


Figure 1.7 – Physical and chemical extremophiles.
This diagram shows one physical variable (temperature), and one chemical variable (pH), the ranges that life is known to tolerate for both eukaryotes and prokaryotes, and the names that have been erected to describe the organisms that live over these ranges. Similar profiles can be shown for many other variables, including radiation, dryness and salinity (Nealson, 1997). The pictures depict environments on earth: the Antarctic (upper left), a thermal pool in Yellowstone Park (upper right), an acidic pond resulting from acid mine drainage (lower left); and, Mono Lake (California, USA, lower right), an alkaline lake of pH approximately 10. Adapted from (Nealson & Conrad, 1999).

pH

Biological processes tend to occur in the middle of the pH spectrum and intracellular and environmental pH often fall within this range. However, pH can be higher than 14 (like in soda lakes or drying ponds), or as low as 0 and below.

Acidophiles thrive at low pH and alkaliphiles prefer high pH. Fish and cyanobacteria have never been found at pH's lower than 4 while plants and insects below pH 2-3 (Figure 1.8). Several unicellular eukaryotes live below pH 1 with one example being the red alga *Cyanidium caldarium* which as been described to live in nature at pH as low as 0.5 (optimum pH for growth is 2-3) (Doemel & Brock, 1971). Several archaea also thrive in these harsh environments like the aerobic heterotrophs *Picrophilus oshimae* and *Picrophilus torridus* isolated from Japanese soils permeated with solfataric gases that grow at pH 0.7 and 60 ºC (Schleper *et al.*, 1995).

Ferroplasma acidarmanus has been described growing at pH 0 in acid mine drainage in Iron Mountain (California, USA), thriving in a brew of sulphuric acid and high levels of Cu, Ar, Cd and Zn with only a cell membrane and no cell wall (Krulwich *et al.*, 1998). Several representatives of all domains are also capable of tolerating pH's as high as 11 (Jones *et al.*, 1998).

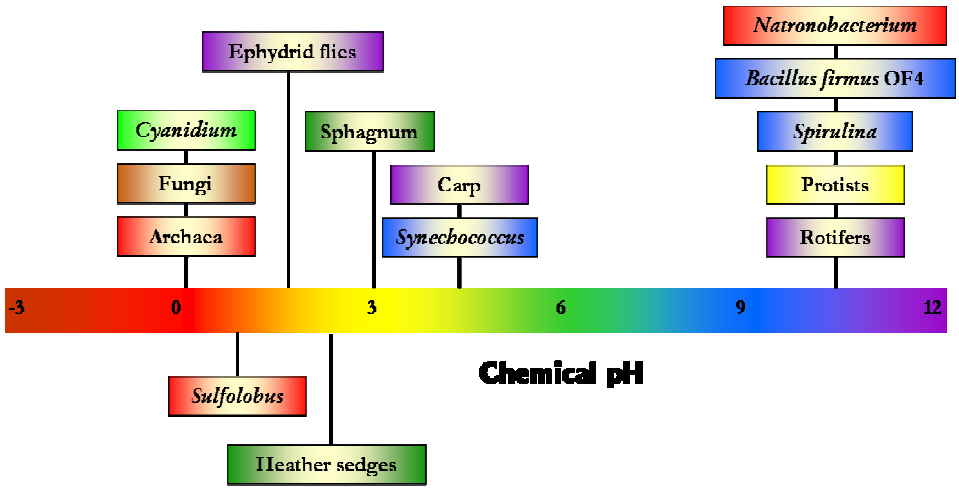


Figure 1.8 – pH limits for Life.
Examples of known pH limits for Life are shown. Archaea are in red, bacteria in blue, algae in light green, assorted protists in yellow, fungi in brown, plants in dark green and animals in purple [adapted from (Rothschild & Mancinelli, 2001)].

RADIATION

Exceptional levels of radiation, sufficient to qualify for "extremophile" status, rarely occur naturally on Earth; rather, intense levels of UV and ionizing radiation are well-studied because of their importance to medicine, energy production, warfare and space travel (Rothschild & Mancinelli, 2001). The dangers of radiation relate to decreased motility in microorganism as well to inhibition of photosynthesis and damage to nucleic acids (either being direct damages to DNA or indirect through the production of reactive oxygen species that create modified bases and strand breaks).

The bacterium *Deinococcus radiodurans* (Figure 1.9), has the ability to withstand ionizing radiation (up to 20 kGy), and UV radiation (doses up to 1.000 $J.m^2$), but this ability is thought to be a by-product of resistance to extreme desiccation (Battista, 1997). Other organisms that tolerate high levels of radiation are two *Rubrobacter* species (Ferreira *et al.*, 1999), and the green alga *Dunaliella bardawil* (Ben-Amotz & Avron, 1990).

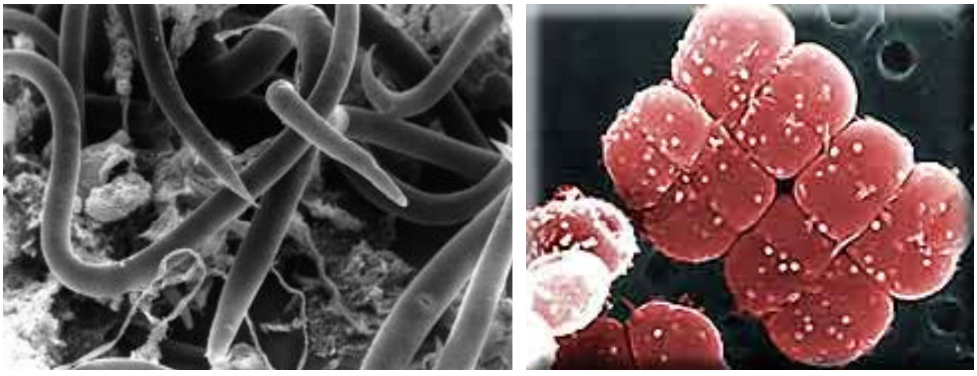


Figure 1.9 - *Panagrolaimus davidii* (left), and *Deinococcus radiodurans* (right).
P. davidii is around 1 mm in length, lives in Antarctica's soils and can freeze completely and get back to life; *D. radiodurans* can not only withstand devastating levels of radiation (both ionising and ultraviolet), but can also resist very harsh genotoxic chemicals and oxidative damage.

PRESSURE

Pressure is known to alter gene expression (Nakasone *et al.*, 1998), and when pressure increases, or temperature decreases, the molecules in lipid membranes pack tighter, resulting in decreased membrane fluidity (Pledger *et al.*, 1994). Pressure can also stabilize enzymes (Pledger *et al.*, 1994), but high pressure can also damage DNA and proteins (Abe *et al.*, 1999).

The Mariana Trench is the world's deepest sea floor with approximately 11 km of depth. However, it harbours organisms that can thrive at standard pressure and temperature. It also harbours obligatory piezophilic species that can grow at 70 – 80 MPa, but not below 50 MPa (Kato *et al.*, 1998).

DESICCATION AND SALINITY

Water limitation is an extreme environment. Organisms that manage to thrive under extreme desiccation enter anhydrobiosis, a state characterized by little intracellular water and no metabolic activity (Rothschild & Mancinelli, 2001). A variety of organisms can become anhydrobiotic, including bacteria, yeast, fungi, plants, insects, tardigrades, mycophagous nematodes and the shrimp *Artemia salina* (Glasheen & Hand, 1988, Potts, 1994, Wright, 1989, Crowe, 1971). Mechanisms of death due to anhydrobiosis include irreversible phase changes to lipids, proteins and nucleic acids, such as denaturation and structural breakage through Mayllard reactions, and accumulation of reactive oxygen species during drying, especially under solar radiation (Cox, 1993, Dose *et al.*, 1995, Dose & Gill, 1995).

Organisms live in a wide range of salinities, from (essentially) distilled water to saturated salt solutions. Osmophily refers to the aspects of life at high salt concentrations (turgor pressure, cellular dehydration and desiccation), and halophily refers to the ionic requirements for life at high salt concentrations. Although these phenomena are physiologically different, they are environmentally linked. Thus, an halophile must cope with osmotic stress. Halophiles include a range of microorganism, namely some archaea, cyanobacteria and the green algae *Dunaliella salina* which tolerates periods in saturated NaCl (Rothschild & Mancinelli, 2001).

OXYGEN

Extant organisms inhabit environments ranging from strictly anaerobic to aerobic. Although Earth has been anaerobic throughout most of Life's history, aerobic metabolism is much more efficient than anaerobic. However, the use of aerobic metabolism has its costs namely the oxidative stress resulting from the reduced forms of molecular oxygen, especially the hydroxyl radical ($^{\cdot}OH$) (Rothschild & Mancinelli, 2001). Oxidative damage has been already implicated in a wide range of health problems, from ageing to cancer (Beckman & Ames, 1998, Pourzand & Tyrrell, 1999).

Reactive oxygen species are a recursive and invasive threat. There is photochemical production of these species as H_2O_2 by UVA radiation (320-400 nm), within cells (Tyrell, 1991), and metabolic production during aerobic metabolism and photosynthesis. In eukaryotes, there are also other endogenous sources like mitochondrial respiration (a valuable source of O_2^-), cytochrome P450 metabolism of hydroperoxides (an important source of $^1O_2$, singlet oxygen), production of uric acid and oxidative bursts used in fighting pathogens in animals and plants (Rothschild & Mancinelli, 2001). Exogenous sources include the photochemical production of H_2O_2 in aquatic systems (Cooper & Lean, 1992), and the production of hydroxyl radical by ionizing radiation.

The presence of oxygen can enhance radiation-induced DNA damage (Tyrell, 1991).

1.5. EXTREMOPHILIC STRATEGIES

The term "extreme", or "extremophilic" is always in the optics of the beholder. Our cozy body temperature (37 ^oC), would be impossible for hyperthermophiles to thrive in; so, in *strictu sensu*, human body could be considered an "extreme environment". A little creative thinking could enlarge the list of "extreme conditions" presented above, like redox potential, toxic or xenobiotic (synthetic) compounds, heavy metal concentration and vacuum [link to Panspermia

again where "living bodies" would stand traveling throughout of Space colonizing meteors, asteroids and other planets (Rampelotto, 2010)]. Moreover, there are organisms that thrive immersed in high levels of organic solvents (Isken & de Bont, 1998), and the electric eel (*Electrophorus electricus*), can produce, and hence withstand, strong electric currents.

Understanding these processes and the strategies that extremophiles use to cope with the environment can help us to better understand evolutionary processes on Earth, contribute to the present Knowledge (basic science), and be useful in the commercial exploitation of these organisms.

TEMPERATURE ADAPTATION

High temperature increases the fluidity of membranes and, hence, microorganism thriving in high temperature environments have to maintain optimal membrane fluidity through the adjustment of the composition of the membrane, namely the amount and type of lipids (*e.g.*, saturated *versus* unsaturated). Temperature also affects the structure and function of proteins (Jaenicke, 1996).

Throughout evolution, proteins have evolved in a way that leads to the development of strategies to cope with high temperature. The increase of ion-pair content, formation of high-order oligomers and decrease of flexibility at room temperature are examples of such strategies. The reduction in the length of surface loops is also a known strategy, specially loops connecting secondary structure elements, as well as the optimization of electrostatic and hydrophobic interactions, and the exchange of amino acids to increase internal hydrophobicity and helix propensity of residues in α-helices. This feature also leads to a smaller surface-to-volume ratio, which is a function of folding, improving thermostability by leading to a more compact form of the protein (Ladenstein & Antranikian, 1998).

Protein stability can also be achieved by extrinsic factors to the proteins, namely chaperonins. One of these chaperonins, called thermosome (Figure 1.10), has been extensively studied in *Methanopyrus kandleri* and *Thermoplasma acidophilum*, where

it was shown that it binds to heat denatured proteins, preventing their agregation and refolding them into their active form (Andra *et al.*, 1998, Ditzel *et al.*, 1998).

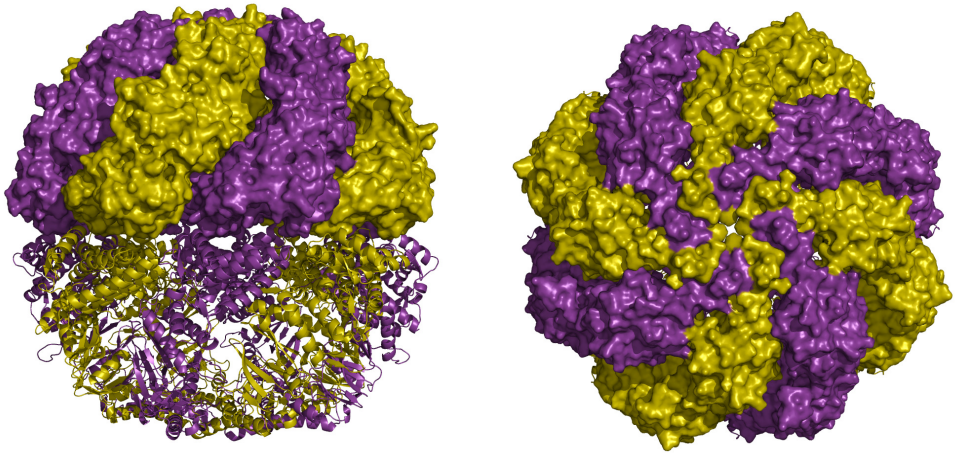


Figure 1.10 – X-ray crystal structure of the thermossome of *Thermoplasma acidophilum* [PDB code 1A6D (Andra *et al.*, 1998, Ditzel *et al.*, 1998)].

The thermossome is a chaperonin which helps in the thermostabilization of proteins *in vivo*. The thermossome of *Thermoplasma acidophilum* was determined to 2.6 Å resolution and has two molecules in the asymmetric unit: the alpha-subunit (coloured in yellow), the beta-subunit (coloured in purple); the biological unit, shown above, is an hexadecameric assembly of eight alpha- and eight beta-subunits.
(Left) Sideview of the hexadecameric thermossome structure with the molecules above the equatorial plane shown in surface representation, and below in cartoon representation.
(Right) Top view [corresponds to a 90º rotation from the view in the (left) pane], of the thermossome.

DNA at temperatures above 70 ºC is normally subjected to thermal denaturation and chemical modifications. However, the DNA of (hyper)thermophiles, such as *Pyrococcus furiosus*, is known to be more stable *in vivo* than the DNA of the mesophilic *Escherichia coli* (Peak *et al.*, 1995). Monovalent and divalent salts (*e.g.*, KCl and $MgCl_2$), enhance the stability of nucleic acids stabilizing the negative charges of the phosphate groups as well as protecting DNA from depurination and hydrolysis (Marguet & Forterre, 1998). Positive supercoiling of the DNA (Forterre *et al.*, 1996, van der Oost *et al.*, 1998), and the existence of

histone-like proteins (Sandman *et al.*, 1998), are also strategies adopted by (hyper)thermophilic archaea to stabilize DNA.

The G-C pair of nucleic acids is more thermostable than the A-T pair (or A-U pair in RNA), because of the extra hydrogen bond when those bases bind (Figure 1.11) (Galtier *et al.*, 1999). However, high G+C contents are not found among thermophilic prokaryotes because of the stability of their chromosomal DNA, although thermostability is correlated with the G+C content of their ribosomal and transfer RNAs (Galtier & Lobry, 1997).

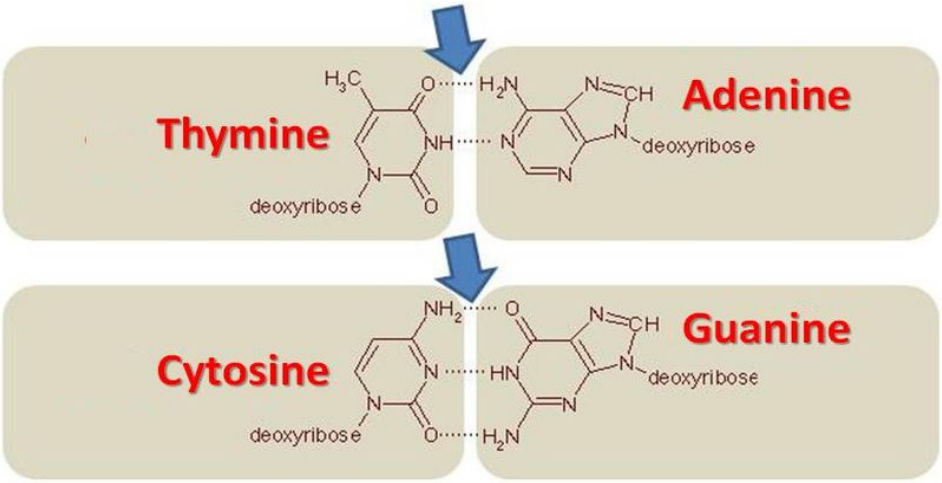


Figure 1.11 - Nucleotide bonds showing adenine-thymine and guanine-cytosine pairs.
Arrows point to the hydrogen bonds that are established when pairs bond.

Membrane fluidity decreases with the decreasing of the temperature. Similarly to the high temperature aggression, organisms respond to the lowering of the temperature by increasing the ratio of unsaturated *versus* saturated lipids. Moreover, the ability of coping with temperatures below the freezing point of water settles upon two strategies: protecting the cells from ice formation by avoiding freezing, and, if ice does form, protect them from damage during thawing (Clarke, 2003).

Cold-temperature adaptation of proteins occurs in slight different ways that are observed for thermophilic enzymes. At low temperatures, there are low levels of free energy so in order to decrease the activation of an enzyme, it must have a high degree of conformational complementarity with its substrate (Cummings & Black, 1999). At cold temperatures, a protein becomes more rigid, implying that enhancing flexibility can restore function. Studies on α-amylase from the psychrophile *Alteromonas haloplanctis*, an enzyme with a less rigid protein core and fewer interdomain interactions than those from its mesophilic counterparts, have supported this idea (Aghajari *et al.*, 1998), as well as studies on tubulin structure (Willem *et al.*, 1999).

SALINITY AND DESICCATION ADAPTATION

It is critical for an organism to maintain intracellular function, and the easiest way of achieving this is to keep the external environment out. As an example, *Cyanidium caldarium* or *Dunaliella acidophila* grow at pH 0.5 but show a neutral pH in the cytoplasm (Beardall & Entwisle, 1984, Pick, 1999). However, if it is impossible to keep the environment out, evolutionary responses promoted protective mechanisms, altering physiologies and enhancing repair capabilities.

The two main strategies used by halophilic microorganisms to cope with osmotic stress are the accumulation of inorganic salts in the cytoplasm and the maintenance of low levels of salt in cytoplasm but higher levels of other solutes. Regarding the accumulation of inorganic salts, K^+ rather than Na^+ is the dominant intracellular cation, and Cl^- the dominant anion (Madigan & Oren, 1999).

The presence up to molar levels of these inorganic ions in the cytoplasm requires long-term adaptations of the intracellular enzymatic machinery (Dennis & Shimmin, 1997). Most proteins from halophilic microorganisms contain a large excess of acidic a.a. when comparing to basic a.a.. The high cation concentrations in the cell can act as a counter-measure needed to shield the negative charges on the protein surfaces. Moreover, high contents of glutamate were shown to be very favourable in this strategy since glutamate has the greatest water binding capacity of

all a.a. thus enabling the maintenance of a protective hydration shell (Eisenberg *et al.*, 1992). However, the requirement for high salt concentrations for structural stability can also be attributed to the low content of hydrophobic residues since high salt is needed to maintain weak hydrophobic interactions (Lanyi, 1974). Accordingly, most proteins from halobacteria denature when suspended in solutions containing less than 1 – 2 M salt (Eisenberg *et al.*, 1992).

The second strategy for adaptation to high salt concentrations is present in most halophilic Bacteria and Archaea and eukaryotic algae and fungi, and involves the maintenance of a cytoplasm low in salt but high in other solutes, namely compatible solutes. Compatible solutes, or osmolytes, are "small organic compounds used for osmotic adjustment that do not interfere with cell function" (Brown, 1976), and their intracellular concentrations are regulated according to the external salt concentration (see Chapter 3 for details on this subject). The use of organic osmotic solutes confers a high degree of flexibility and adaptability to the cells preventing unfolding caused by heating, freezing and drying (Madigan & Oren, 1999).

A few years back, the genes involved in the biosynthesis of ectoine, a compatible solute, in *Halomonas elongata* have been cloned and characterized (Canovas *et al.*, 1997). This microorganism can be exploited to produce large quantities of ectoine and hydroxyectoine, very useful as moisturizers in skin care products, in a process called "bacterial milking". In this process, dense cultures of the microorganism are exposed to an osmotic down-shock, causing the excretion of excess osmotic solutes. Addition of salt to the culture induces (re)synthesis of ectoine and hydroxyectoine where-after the process can be repeated (Sauer & Galinski, 1998).

1.6. ECONOMY AND BIOTECNHOLOGY OF EXTREMOPHILES

The industrial application of "extremozymes" – enzymes from extremophilic organism – has greatly increased in the past two decades. Extremophiles have endeared themselves to multibillion-dollar industries including

agriculture, chemical synthesis, laundry detergents, beauty products and pharmaceuticals, and human health may benefit from these organisms indirectly through biotechnology and bioremediation (Table 1.2).

Direct uses of extremozymes include marketing of dried *Dunaliella* as a nutritional supplement, primarily as an antioxidant, and antifreeze proteins as cryoprotectants of frozen organs (Rothschild & Mancinelli, 2001).

The discovery of new extremophilic organisms and their enzymes had a great impact on the field of biocatalysis. The use of (hyper)thermophilic enzymes became common practice when mesophilic enzymes, not suited for the harsh conditions often used in industrial processes due to the lack of stability, were impossible to use. Psychrophilic enzymes, produced by cold-adapted microorganisms, display a high catalytic efficiency that offers considerable potential to the biotechnology industry (*e.g.*, detergent and food industries as well as for the production of fine chemicals) (Gerday *et al.*, 2000).

The industrial potential for halophilic enzymes increased in the past years as the approaches to study the genetic processes of halophiles and our understanding of haloadaptation becomes more sophisticated (Ventosa *et al.*, 1998). Acidophiles, which often concomitantly present other properties such as thermophilicity, halophilicity or heavy-metal resistance, are getting ahead in the field of bioprocessing of minerals (Norris *et al.*, 2000). Alkaliphiles have also made a great impact in industrial applications, namely alkaline proteases and cellulases in biological detergents.

Understanding the mechanism that underlie adaptation to high concentrations of heavy-metals is in its infancy following studies on the homeostasis for metal ions (Nies, 1999). Obvious applications in the field of bioremediation and biomining are now being complemented by the biosynthesis of novel composites in the field of materials science (Klaus-Joerger *et al.*, 2001). Studies are also being pursued with *Deinococcus radiodurans* for the engineering of special strains for the bioremediation of radioactive waste (Brim *et al.*, 2000).

Table 1.2 – Examples of extremophiles in industry and biotechnology [adapted from (Rothschild & Mancinelli, 2001)].

Industrial process	Biomolecule	Advantages	Source organism
Hydrolysis of starch to produce soluble (malto) dextrins and corn syrups	α-amylase	High stability, aciduric, bacterial amylase	*Bacillus stearothermophilus* G-ZYME G995
Paper bleaching	Xylanases	Decrease amount of bleach needed	Thermophiles
Prevent stalling in range of baked products	α-amylase	Gives boost to yeast fermentation	Highest-stability bacterial amylase available, G-ZYME G995
Food processing, baking, brewing , detergents	Proteases	Stable at high temperatures	Thermophiles
PCR reaction	DNA polymerase	No need to add additional enzyme during each cycle	Thermophiles
Cheese maturation, dairy production	Neutral proteases	Stable at low temperatures	Psychrophiles
Degradation of polymers in detergents	Proteases, amylases, lipases	Improved performance of detergent	Psychrophiles
Degradation of polymers in detergents	Cellulases, proteases, amylases, lipases	Stable at high pH	Alkaliphiles
Mariculture	Polyunsaturated fatty acids	Produced in cold temperatures	Psychrophiles
Bioremediation	Reduction of oil spills	Works efficiently in cold waters	Psychrophiles
Pharmaceuticals	Polyunsaturated fatty acids		Psychrophiles
Biosensors	Dehydrogenases		Psychrophiles
Desulphurication of coal	Sulphur oxidation		Acidophiles
Antibiotic production	Antibiotics		Alkaliphiles
Food colouring	Carotene	Inexpensive to produce	Halophiles/*Dunaliella*
Pharmaceuticals	Glycerol, compatible solutes	Inexpensive to produce	Halophiles
Surfactants for pharmaceuticals	Membranes		Halophiles

On the field of basic science, a novel subgroup of alkaliphiles, the anaerobic alkalithermophiles, is having much attention recently due to its very interesting feature of short doubling time (Wiegel, 1998). Piezophilic organisms are also an interesting field of research focusing on the identification of pressure-regulated operons showing the relationship between pressure and microbial growth (Horikoshi, 1998).

1.7. REFERENCES

Abe, F., C. Kato & K. Horikoshi, (1999) Pressure-regulated metabolism in microorganisms. *Trends in microbiology* **7**: 447-453.

Aghajari, N., G. Feller, C. Gerday & R. Haser, (1998) Structures of the psychrophilic *Alteromonas haloplanctis* alpha-amylase give insights into cold adaptation at a molecular level. *Structure* **6**: 1503-1516.

Andra, S., G. Frey, R. Jaenicke & K. O. Stetter, (1998) The thermosome from *Methanopyrus kandleri* possesses an NH4+-dependent ATPase activity. *Eur J Biochem* **255**: 93-99.

Arrhenius, S., (1903) Die Verbreitung des Lebens im Weltenraum. *Umschau* **7**: 481-485.

Barns, S. M., C. F. Delwiche, J. D. Palmer & N. R. Pace, (1996) Perspectives on archaeal diversity, thermophily and monophyly from environmental rRNA sequences. *Proc Natl Acad Sci U S A* **93**: 9188-9193.

Battista, J. R., (1997) Against all odds: the survival strategies of *Deinococcus radiodurans*. *Annu Rev Microbiol* **51**: 203-224.

Beardall, J. & L. Entwisle, (1984) Internal pH of the obligate acidophile *Cyanidium caldarium* Geitler (Rhodophyta?). *Phycologia* **23**: 397-399.

Beckman, K. B. & B. N. Ames, (1998) The free radical theory of aging matures. *Physiol Rev* **78**: 547-581.

Bell, S. D., C. H. Botting, B. N. Wardleworth, S. P. Jackson & M. F. White, (2002) The interaction of Alba, a conserved archaeal chromatin protein, with Sir2 and its regulation by acetylation. *Science* **296**: 148-151.

Ben-Amotz, A. & M. Avron, (1990) The biotechnology of cultivating the halotolerant alga *Dunaliella*. *Trends in Biotechnology* **8**: 121-126.

Blochl, E., R. Rachel, S. Burggraf, D. Hafenbradl, H. W. Jannasch & K. O. Stetter, (1997) *Pyrolobus fumarii*, gen. and sp. nov., represents a novel group of archaea, extending the upper temperature limit for life to 113 degrees C. *Extremophiles* **1**: 14-21.

Blum, H. F., (1962) *Time's arrow and evolution*. Harper & Brothers, New York, USA.

Boone, D. R. & R. W. Castenholz, (2001) *Bergey's Manual of Systematic Bacteriology*. Springer, New York, USA.

Brim, H., S. C. McFarlan, J. K. Fredrickson, K. W. Minton, M. Zhai, L. P. Wackett & M. J. Daly, (2000) Engineering *Deinococcus radiodurans* for metal remediation in radioactive mixed waste environments. *Nat Biotechnol* **18**: 85-90.

Brochier-Armanet, C., B. Boussau, S. Gribaldo & P. Forterre, (2008a) Mesophilic Crenarchaeota: proposal for a third archaeal phylum, the Thaumarchaeota. *Nat Rev Microbiol* **6**: 245-252.

Brochier-Armanet, C., S. Gribaldo & P. Forterre, (2008b) A DNA topoisomerase IB in Thaumarchaeota testifies for the presence of this enzyme in the last common ancestor of Archaea and Eucarya. *Biol Direct* **3**: 54.

Brown, A. D., (1976) Microbial water stress. *Bacteriol Rev* **40**: 803-846.

Canovas, D., C. Vargas, F. Iglesias-Guerra, L. N. Csonka, D. Rhodes, A. Ventosa & J. J. Nieto, (1997) Isolation and characterization of salt-sensitive mutants of the moderate halophile *Halomonas elongata* and cloning of the ectoine synthesis genes. *J Biol Chem* **272**: 25794-25801.

Chyba, C. & C. Sagan, (1992) Endogenous production, exogenous delivery and impact-shock synthesis of organic molecules: an inventory for the origins of life. *Nature* **355**: 125-132.

Chyba, C. F. & K. P. Hand, (2005) Astrobiology: The study of the living universe. *Annu. Rev. Astron. Astrophys.* **43**: 31-74.

Clarke, A., (2003) *Evolution on Planet Earth: the impact of the physical environment*. Elsevier Science, London, UK.

Cleland, C. E. & C. F. Chyba, (2002) Defining 'life'. *Orig Life Evol Biosph* **32**: 387-393.

Cooper, W. & D. Lean, (1992) *Encyclopedia of Earth System Science*. Academic, San Diegi, USA.

Cox, C. S., (1993) Roles of water molecules in bacteria and viruses. *Orig Life Evol Biosph* **23**: 29-36.

Crick, F. H. C., (1981) *Life itself: its origin and nature*. Simon & Schuster, New York, USA.

Crowe, J. H., (1971) Abhydrobiosis: an unsolved problem. *Am Nat* **105**: 563-574.

Cummings, S. P. & G. W. Black, (1999) Polymer hydrolysis in a cold climate. *Extremophiles* **3**: 81-87.

da Vinci, L., (1513) Il Codice Arundel. In: The notebooks of Leonardo da Vinci. K. Konecky (ed). Old Saybrook, Connecticut, USA, pp. 1180.

Darwin, C., (1871) *Letter to Hooker*. NASA.

Dennis, P. P. & L. C. Shimmin, (1997) Evolutionary divergence and salinity-mediated selection in halophilic archaea. *Microbiol Mol Biol Rev* **61**: 90-104.

Ditzel, L., J. Lowe, D. Stock, K. O. Stetter, H. Huber, R. Huber & S. Steinbacher, (1998) Crystal structure of the thermosome, the archaeal chaperonin and homolog of CCT. *Cell* **93**: 125-138.

Doemel, W. N. & T. D. Brock, (1971) The Physiological Ecology of *Cyanidium caldarium*. *Journal of General Microbiology* **67**: 17-32.

Dose, K., A. Bieger-Dose, R. Dillmann, M. Gill, O. Kerz, A. Klein, H. Meinert, T. Nawroth, S. Risi & C. Stridde, (1995) ERA-experiment "Space Biochemistry". *Adv Space Res* **16**: 119-129.

Dose, K. & M. Gill, (1995) DNA stability and survival of *Bacillus subtilis* spores in extreme dryness. *Orig Life Evol Biosph* **25**: 277-293.

Eigen, M., (1971) Selforganization of matter and the evolution of biological macromolecules. *Naturwissenschaften* **58**: 465-523.

Eigen, M., (1996) *Steps towards Life*. Oxford University Press, Oxford (UK).

Eisenberg, H., M. Mevarech & G. Zaccai, (1992) Biochemical, structural, and molecular genetic aspects of halophilism. *Adv Protein Chem* **43**: 1-62.

Elkins, J. G., M. Podar, D. E. Graham, K. S. Makarova, Y. Wolf, L. Randau, B. P. Hedlund, C. Brochier-Armanet, V. Kunin, I. Anderson, A. Lapidus, E. Goltsman, K. Barry, E. V. Koonin, P. Hugenholtz, N. Kyrpides, G. Wanner, P. Richardson, M. Keller & K. O. Stetter, (2008) A

korarchaeal genome reveals insights into the evolution of the Archaea. *Proceedings of the National Academy of Sciences* **105**: 8102-8107.

Ferreira, A. C., M. F. Nobre, E. Moore, F. A. Rainey, J. R. Battista & M. S. da Costa, (1999) Characterization and radiation resistance of new isolates of *Rubrobacter radiotolerans* and Rubrobacter xylanophilus. *Extremophiles* **3**: 235-238.

Folsome, C. E., (1979) *The origin of Life*. Freeman and Co., San Francisco, USA.

Forterre, P., A. Bergerat & P. Lopez-Garcia, (1996) The unique DNA topology and DNA topoisomerases of hyperthermophilic archaea. *FEMS Microbiol Rev* **18**: 237-248.

Galtier, N. & J. R. Lobry, (1997) Relationships between genomic G+C content, RNA secondary structures, and optimal growth temperature in prokaryotes. *J Mol Evol* **44**: 632-636.

Galtier, N., N. Tourasse & M. Gouy, (1999) A nonhyperthermophilic common ancestor to extant life forms. *Science* **283**: 220-221.

Gerday, C., M. Aittaleb, M. Bentahir, J. P. Chessa, P. Claverie, T. Collins, S. D'Amico, J. Dumont, G. Garsoux, D. Georlette, A. Hoyoux, T. Lonhienne, M. A. Meuwis & G. Feller, (2000) Cold-adapted enzymes: from fundamentals to biotechnology. *Trends Biotechnol* **18**: 103-107.

Glasheen, J. S. & S. C. Hand, (1988) Anhydrobiosis in Embryos of the Brine Shrimp Artemia: Characterization of Metabolic Arrest During Reductions in Cell-Associated Water. *Journal of Experimental Biology* **135**: 363-380.

Gottschalk, G., (1994) *Microbial metabolism*. Springer, New York, USA.

Grabowski, B. & Z. Kelman, (2003) Archeal DNA replication: eukaryal proteins in a bacterial context. *Annu Rev Microbiol* **57**: 487-516.

Gribaldo, S. & C. Brochier-Armanet, (2006) The origin and evolution of Archaea: a state of the art. *Philos Trans R Soc Lond B Biol Sci* **361**: 1007-1022.

Horikoshi, K., (1998) Barophiles: deep-sea microorganisms adapted to an extreme environment. *Curr Opin Microbiol* **1**: 291-295.

Huber, H., M. J. Hohn, R. Rachel, T. Fuchs, V. C. Wimmer & K. O. Stetter, (2002) A new phylum of Archaea represented by a nanosized hyperthermophilic symbiont. *Nature* **417**: 63-67.

Isken, S. & J. A. de Bont, (1998) Bacteria tolerant to organic solvents. *Extremophiles* **2**: 229-238.

Jaenicke, R., (1996) Stability and folding of ultrastable proteins: eye lens crystallins and enzymes from thermophiles. *FASEB J* **10**: 84-92.

Jones, B. E., W. D. Grant, A. W. Duckworth & G. G. Owenson, (1998) Microbial diversity of soda lakes. *Extremophiles* **2**: 191-200.

Jortner, J., (2006) Conditions for the emergence of life on the early Earth: summary and reflections. *Philos Trans R Soc Lond B Biol Sci* **361**: 1877-1891.

Kandler, O., (1994) *The early diversification of life*. Columbia University Press, New York, USA.

Kato, C., L. Li, Y. Nogi, Y. Nakamura, J. Tamaoka & K. Horikoshi, (1998) Extremely Barophilic Bacteria Isolated from the Mariana Trench, Challenger Deep, at a Depth of 11,000 Meters. *Appl. Environ. Microbiol.* **64**: 1510-1513.

Klaus-Joerger, T., R. Joerger, E. Olsson & C. Granqvist, (2001) Bacteria as workers in the living factory: metal-accumulating bacteria and their potential for materials science. *Trends Biotechnol* **19**: 15-20.

Knoll, A. H., (1992) The early evolution of eukaryotes: a geological perspective. *Science* **256**: 622-627.

Kohshima, S., (1984) A novel cold-tolerant insect found in a Himalayan glacier. *Nature* **310**: 225-227.

Krulwich, T. A., M. Ito, D. B. Hicks, R. Gilmour & A. A. Guffanti, (1998) pH homeostasis and ATP synthesis: studies of two processes that necessitate inward proton translocation in extremely alkaliphilic Bacillus species. *Extremophiles* **2**: 217-222.

Ladenstein, R. & G. Antranikian, (1998) Proteins from hyperthermophiles: Stability and enzymatic catalysis close to the boiling point of water. In: Biotechnology of Extremophiles. G. Antranikian (ed). Springer Berlin / Heidelberg, pp. 37-85.

Lanyi, J. K., (1974) Salt-dependent properties of proteins from extremely halophilic bacteria. *Bacteriol Rev* **38**: 272-290.

Lecompte, O., R. Ripp, J. C. Thierry, D. Moras & O. Poch, (2002) Comparative analysis of ribosomal proteins in complete genomes: an example of reductive evolution at the domain scale. *Nucleic Acids Res* **30**: 5382-5390.

Londei, P., (2005) Evolution of translational initiation: new insights from the archaea. *FEMS Microbiol Rev* **29**: 185-200.

Madigan, M. T. & A. Oren, (1999) Thermophilic and halophilic extremophiles. *Curr Opin Microbiol* **2**: 265-269.

Marguet, E. & P. Forterre, (1998) Protection of DNA by salts against thermodegradation at temperatures typical for hyperthermophiles. *Extremophiles* **2**: 115-122.

Miller, S. L. & L. E. Orgel, (1974) *The origins of Life on Earth*. Prentice Hall, Englewood Cliffs, NJ, USA.

Mojzsis, S. J., G. Arrhenius, K. D. McKeegan, T. M. Harrison, A. P. Nutman & C. R. L. Friend, (1996) Evidence for life on Earth before 3,800 million years ago. *Nature* **384**: 55-59.

Myllykallio, H., P. Lopez, P. Lopez-Garcia, R. Heilig, W. Saurin, Y. Zivanovic, H. Philippe & P. Forterre, (2000) Bacterial mode of replication with eukaryotic-like machinery in a hyperthermophilic archaeon. *Science* **288**: 2212-2215.

Nakasone, K., A. Ikegami, C. Kato, R. Usami & K. Horikoshi, (1998) Mechanisms of gene expression controlled by pressure in deep-sea microorganisms. *Extremophiles* **2**: 149-154.

Nealson, K. H., (1997) Sediment bacteria: who's there, what are they doing, and what's new? *Annu Rev Earth Planet Sci* **25**: 403-434.

Nealson, K. H. & P. G. Conrad, (1999) Life: past, present and future. *Philos Trans R Soc Lond B Biol Sci* **354**: 1923-1939.

Nies, D. H., (1999) Microbial heavy-metal resistance. *Appl Microbiol Biotechnol* **51**: 730-750.

Norris, P. R., N. P. Burton & N. A. Foulis, (2000) Acidophiles in bioreactor mineral processing. *Extremophiles* **4**: 71-76.

Olsen, G. J. & C. R. Woese, (1996) Lessons from an Archaeal genome: what are we learning from *Methanococcus jannaschii*? *Trends Genet* **12**: 377-379.

Oparin, A. I., (1924) *Proiskhodenie Zhizny*. Jones & Bertlett, Boston, USA.

Peak, M. J., F. T. Robb & J. G. Peak, (1995) Extreme resistance to thermally induced DNA backbone breaks in the hyperthermophilic archaeon *Pyrococcus furiosus*. *J Bacteriol* **177**: 6316-6318.

Pick, U., (1999) *Microorganims and Life in Extreme Environments*. Kluwer, Dordrecht, Germany.

Pledger, R. J., B. C. Crump & J. A. Baross, (1994) A barophilic response by two hyperthermophilic, hydrothermal vent Archaea: An upward shift in the optimal temperature and acceleration of growth rate at supra-optimal temperatures by elevated pressure. *FEMS Microbiology Ecology* **14**: 233-241.

Potts, M., (1994) Desiccation tolerance of prokaryotes. *Microbiol Rev* **58**: 755-805.

Pourzand, C. & R. M. Tyrrell, (1999) Apoptosis, the role of oxidative stress and the example of solar UV radiation. *Photochem Photobiol* **70**: 380-390.

Rampelotto, P. H., (2010) Panspermia: A promising field of research. In: Astrobiology Science Conference. pp. Abs 5224.

Richter, H., (1865) Zur Darwinschen Lehre. *Schmidts Jahrb Ges Med* **126**: 243-249.

Rothschild, L. J. & R. L. Mancinelli, (2001) Life in extreme environments. *Nature* **409**: 1092-1101.

Rye, R. & H. D. Holland, (1998) Paleosols and the evolution of atmospheric oxygen: a critical review. *Am J Sci* **298**: 621-672.

Sandman, K., S. L. Pereira & J. N. Reeve, (1998) Diversity of prokaryotic chromosomal proteins and the origin of the nucleosome. *Cell Mol Life Sci* **54**: 1350-1364.

Sauer, T. & E. A. Galinski, (1998) Bacterial milking: A novel bioprocess for production of compatible solutes. *Biotechnol Bioeng* **57**: 306-313.

Schidlowski, M., (1988) A 3,800-million-year isotopic record of life from carbon in sedimentary rocks. *Nature* **333**: 313-318.

Schlegel, H. G., (1993) *General microbiology*. Cambridge University Press, Cambridge, UK.

Schleper, C., G. Jurgens & M. Jonuscheit, (2005) Genomic studies of uncultivated archaea. *Nat Rev Microbiol* **3**: 479-488.

Schleper, C., G. Puehler, I. Holz, A. Gambacorta, D. Janekovic, U. Santarius, H. P. Klenk & W. Zillig, (1995) *Picrophilus* gen. nov., fam. nov.: a novel aerobic, heterotrophic, thermoacidophilic genus and family comprising archaea capable of growth around pH 0. *J Bacteriol* **177**: 7050-7059.

Schopf, J. W. & C. Klein, (1992) *The proterozoic biosphere*. Cambridge University Press, Cambridge, UK.

Schrödinger, E., (1944) *What is Life?* Cambridge University Press, Cambridge, UK.

Schuliger, J. W., S. H. Brown, J. A. Baross & R. M. Kelly, (1993) Purification and characterization of a novel amylollytic enzyme from ES4, a marine hyperthermophilic archaeum. *Mol Mar Biol Biotech* **2**: 76-87.

Stetter, K. O., (2006) Hyperthermophiles in the history of life. *Philos Trans R Soc Lond B Biol Sci* **361**: 1837-1842; discussion 1842-1833.

Thomson, W., (1894) *Popular Lectures and Addresses*. Macmillan, New York, USA.

Tourna, M., M. Stieglmeier, A. Spang, M. Könneke, A. Schintlmeister, T. Urich, M. Engel, M. Schloter, M. Wagner, A. Richter & C. Schleper, (2011) *Nitrososphaera viennensis*, an ammonia oxidizing archaeon from soil. *Proceedings of the National Academy of Sciences* **108**: 8420-8425.

Tyrell, R. M., (1991) *Oxidative Stress: Oxidants and Antioxidants*. Academic, London, UK.

van der Oost, J., M. Ciaramella, M. Moracci, F. Pisani, M. Rossi & W. de Vos, (1998) Molecular biology of hyperthermophilic *Archaea*. In: Biotechnology of Extremophiles. G. Antranikian (ed). Springer Berlin / Heidelberg, pp. 87-115.

Ventosa, A., J. J. Nieto & A. Oren, (1998) Biology of moderately halophilic aerobic bacteria. *Microbiol Mol Biol Rev* **62**: 504-544.

Watson, E. B. & T. M. Harrison, (2005) Zircon thermometer reveals minimum melting conditions on earliest Earth. *Science* **308**: 841-844.

Wharton, D. & D. Ferns, (1995) Survival of intracellular freezing by the Antarctic nematode *Panagrolaimus davidi*. *J Exp Biol* **198**: 1381-1387.

Wiegel, J., (1998) Anaerobic alkalithermophiles, a novel group of extremophiles. *Extremophiles* **2**: 257-267.

Willem, S., M. Srahna, N. Devos, C. Gerday, R. Loppes & R. F. Matagne, (1999) Protein adaptation to low temperatures: a comparative study of alpha-tubulin sequences in mesophilic and psychrophilic algae. *Extremophiles* **3**: 221-226.

Woese, C. R. & G. E. Fox, (1977) Phylogenetic structure of the prokaryotic domain: the primary kingdoms. *Proc Natl Acad Sci U S A* **74**: 5088-5090.

Woese, C. R., O. Kandler & M. L. Wheelis, (1990) Towards a natural system of organisms: proposal for the domains Archaea, Bacteria, and Eucarya. *Proc Natl Acad Sci U S A* **87**: 4576-4579.

Wright, J. C., (1989) Desiccation Tolerance and Water-Retentive Mechanisms in Tardigrades. *Journal of Experimental Biology* **142**: 267-292.

Chapter 2

THE CRYSTAL STRUCTURE OF SULFIDE:QUINONE OXIDOREDUCTASE FROM *ACIDIANUS AMBIVALENS*

The results presented in this chapter were published in

José A. Brito, Tiago M. Bandeiras, Miguel Teixeira, Clemens Vonrhein, Margarida Archer (2006). "Crystallization and preliminary structure determination of NADH: quinone oxidoreductase from the extremophile *Acidianus ambivalens*", *Biochimica et Biophysica Acta - Proteins and Proteomics* **1764**, 842-845.

José A. Brito, Filipa L. Sousa, Meike Stelter, Tiago M. Bandeiras, Clemens Vonrhein, Miguel Teixeira, Manuela M. Pereira and Margarida Archer (2009). Structural and functional insights into sulfide:quinone oxidoreductase, *Biochemistry* **48**, 5613-5622.

NOTE: This work was done in collaboration with the Metalloproteins and Bioenergetics Laboratory (ITQB-UNL), headed by Prof. Doctor Miguel Teixeira. JAB, FLS and TMB purified protein and performed activity assays; JAB and MA crystallized protein and collected diffraction data; JAB, CV and MA determined and refined the structure; JAB, FLS, MS, MT, MMP and MA analyzed the structure and wrote the manuscripts.

2.1. ABSTRACT

Oxidation of hydrogen sulphide to sulphate is one of the major reactions of the global sulfur cycle playing a critical role in climate and Health. The biological involvement in the dissimilatory oxidation of elemental sulfur and inorganic sulfur compounds has been established for more a century. Inorganic sulfur compounds can serve as electron donors in many phototrophic and chemiotrophic bacteria as well as in archaea, mostly with sulfate as the final oxidation product of these energy-yielding processes (Friedrich, 1998). Although many photo and chemolithothrophic bacteria are usually used as model organisms to elucidate this process, little is known about the sulfur oxidation in Archaea.

A sulfide:quinone oxidoreductase (SQR), was isolated from the membranes of the hyperthermoacidophilic archaeon *Acidianus ambivalens* and its X-ray structure, the first reported for an SQR, was determined to 2.6 Å resolution. A truncated form of the enzyme was firstly characterized as a type-2 NADH:quinone oxidoreductase (NDH-2), but the crystal structure prompted for a new characterization.

SQR was functionally and structurally characterized and was shown to have two redox active sites: a covalently bound FAD and an adjacent pair of cysteine residues. The X-ray structure also revealed the presence of a chain of three sulfur atoms bridging those two cysteine residues. The possible implications of this observation in the catalytic mechanism for sulfide oxidation are discussed and the role of SQR in the sulfur dependent bioenergetics of *A. ambivalens*, linked to oxygen reduction, is addressed.

2.2. THE SULFUR CYCLE

Sulfur (S), is the tenth most abundant element in the Universe. As a brittle, yellow, tasteless and odourless non-metallic element (Figure 2.1), sulfur is a main constituent of vitamins, proteins and hormones and plays pivotal roles in both climate and health of various ecosystems. The majority of the Earth's sulfur is stored underground in rocks and minerals, including as sulfate salts buried deep within ocean sediments.

In Nature, sulfur can be found in oxidation states ranging from +6 (as in sulfate), to -2 (as in sulfide), and together with oxygen, nitrogen and carbon constitute some of Life's essential elements. Interconversions between various species of these elements constitute their global biogeochemical cycles which are sustained by complex biological processes where microorganisms play prominent roles (Morel & Price, 2003, Anbar & Knoll, 2002). Moreover, sulfur can form stable compounds with all elements in the Periodic Table, except with the noble gases (Greenwood & Earnshaw, 1997).

It is thought that Earth's primordial atmosphere had very likely CO_2, CO and N_2 as main components by the time of the origin of Life. NO, H_2, H_2S and sulfuric gases could also been present, as well as trace amounts of O_2. The evolution of atmospheric O_2, intimately related with the evolution of the Earth both from a biological and geological perspective, is controlled by the long-term (multi-million-year), geochemical cycles of carbon and sulfur (Kronneck, 2005). The first documented description of this control was published in 1845 by J.J. Ebelmen which deduced the following global reactions (Ebelmen, 1845):

$$CO_2 + H_2O \leftrightarrow CH_2O + O_2 \quad \text{(Eq. 2.1)}$$

$$2Fe_2O_3 + 8SO_4^{2-} \leftrightarrow 15O_2 + 4FeS_2 + 8H_2O \quad \text{(Eq. 2.2)}$$

The level of atmospheric O_2 began to raise ~2.4 – 2.6 gigayears (Gyr), ago and the present concentration of free oxygen was reached ~500 million years ago. Although still a matter of debate, the generally accepted theory for dioxygen

accumulation is that the emergence of organisms capable of performing oxygenic photosynthesis, probably ancestors of the present-day cyanobacteria, was responsible for this (Kasting, 1993).

The sulfur cycle (Figure 2.1), comprises an atmospheric and a terrestrial component. Within the terrestrial component, the cycle begins with the weathering of rocks releasing the stored sulfur. When in contact with atmospheric oxygen, sulfur is converted into sulfate (SO_4^{2-}), which is taken up by plants and microorganisms, and converted to organic forms. Moving through the food chain, some of this sulfur is again released as SO_4 as organisms die and decompose (Cunningham & Saigo, 1999).

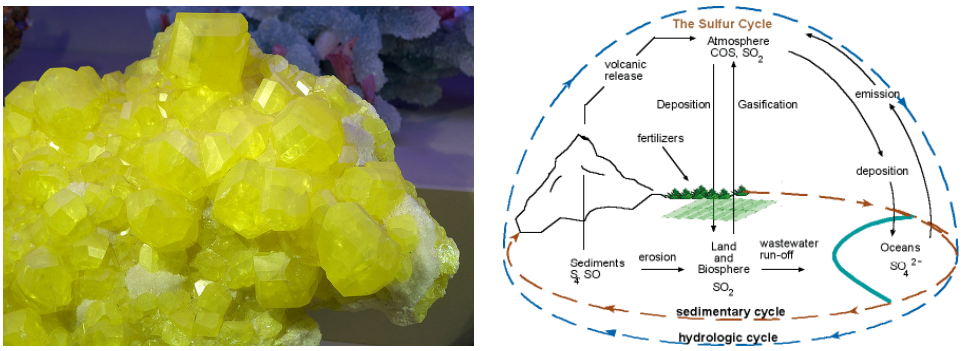


Figure 2.1 – Sulfur crystal collected at Agrigento (Sicily - Italy, left), and schematic representation of the hydrologic and sedimentary sulfur cycle (right).

There are a variety of natural sources that emit sulfur directly into the atmosphere like volcanic eruptions and breakdown of organic matter in swamps and tidal flats. Evaporation of water is also considered as an important component of emission with sulfur settling back to the surface or coming down within rainfall. A continuous loss of sulfur from terrestrial ecosystem runoff occurs through drainage into lakes, streams and, ultimately, oceans. Sulfur also enters the oceans through fallout from the atmosphere cycling through marine communities and moving through the food chain. Part of this sulfur is emitted back to the atmosphere from

sea spray, the remaining being lost to ocean depths combined with iron to form ferrous sulfide (Cunningham & Saigo, 1999).

In Nature, three types of sulfur storage products can be distinguished (Figure 2.2): gypsum ($CaSO_4$), metal sulfides and elemental sulfur (S^0), which have formed in different geological periods. Increased anthropogenic activity has resulted in significant local imbalances of the cyclic conversions of the sulfur cycle (Lens & Kuenen, 2001).

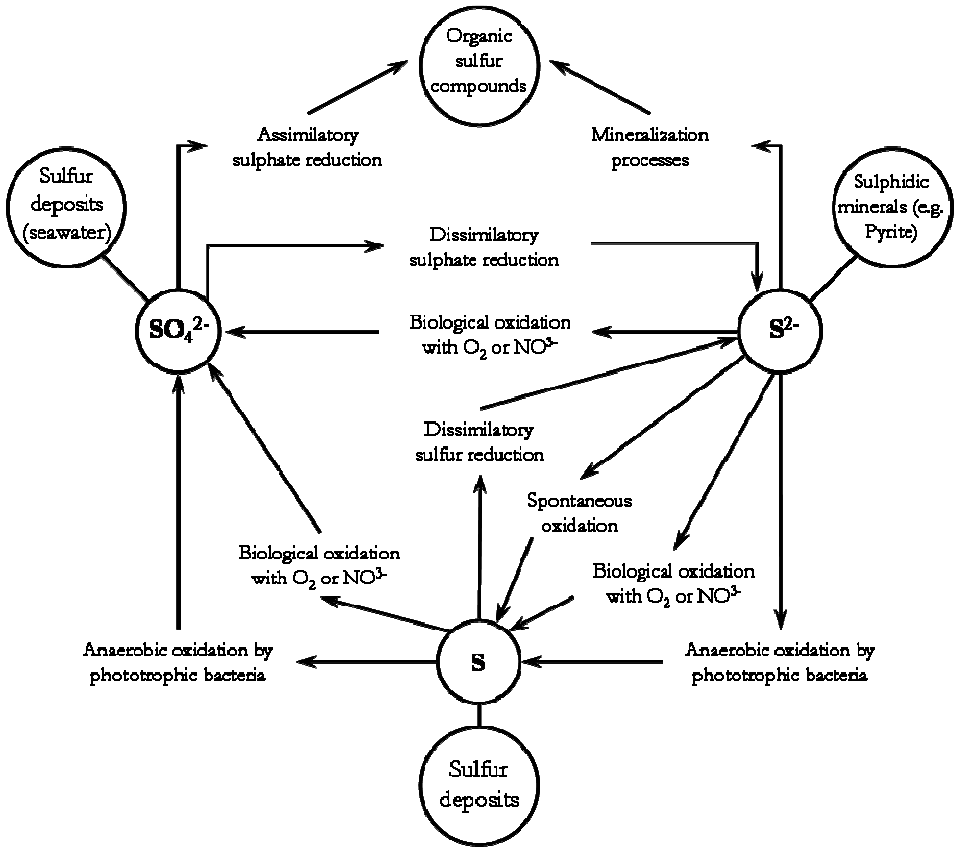


Figure 2.2 – The sulfur cycle [adapted from (Robertson & Kuenen, 1992)].

Since the Industrial Revolution, human activities have contributed to the amount of sulfur that enters the atmosphere, primarily through the burning of fossil fuels and the processing of metals. Almost one third of all sulfur that reaches the

atmosphere stems from human activities. Emissions from these activities, along with nitrogen emissions, react with other chemicals in the atmosphere to produce tiny particles of sulfate salts which fall as acid rain, causing a variety of damage to both natural as well as man-made environments. However, as particles and tiny airborne droplets, sulfur can also act as a regulator of global climate: sulfur dioxide and sulfate aerosols absorb ultraviolet radiation creating cloud covers that cool cities and may offset global warming caused by the greenhouse effect (Jackson & Jackson, 1996).

2.3. SULFUR METABOLISM

Elemental sulfur (S^o), and reduced inorganic sulfur compounds, like sulfate, sulfite, thiosulfate, organic sulfoxides, polysulfide and organic disulfides, are metabolized by a large number of microorganisms (Urich *et al.*, 2004, Hedderich *et al.*, 1998). Most of these microorganisms are hyperthermophilic, i.e. grow optimally above 80 ^oC, belong to the Bacteria and Archaea domains, and are found in water-containing volcanic areas, such as terrestrial solfataric fields and hot springs, and shallow and abyssal submarine hydrothermal systems (Hedderich *et al.*, 1998). In recent years, hyperthermophiles have also been found in oil-bearing, deep-subterranean rocks, about 4 km below the surface of the Earth (Stetter *et al.*, 1993).

Within volcanic environments, sulfur may be formed in different concentrations at the surface by oxidation of hydrogen sulfide (H_2S), escaping from the depths. In the hot biotopes they live in, hyperthermophiles form complex ecosystems comprising a variety of primary producers and decomposers of organic matter. Mesophilic and thermophilic sulfur reducers, mostly from the bacterial domain, have been isolated from such hot biotopes (Stetter, 1996). Among these sulfur-reducing microorganisms, *Acidianus*, *Thermoproteus*, *Igneococcus*, *Pyrodictium* and *Wolinella* genera, among others, are able to gain ATP by lithotrophic sulfur respiration; *Thermococcus*, *Pyrococcus*, *Hyperthermus* (Archaea), or *Thermotoga* and *Fervidobacterium* (Bacteria), genera, are strictly fermentative sulfur reducers (Widdel & Pfennig, 1991, Stetter, 1996).

The hyperthermophilic bacterium *Aquifex pyrophilus*, an aerobic chemolithoautotroph that uses sulfur in addition to hydrogen and thiosulfate as electron donor to reduce oxygen and nitrate, has the ability of producing high levels of H_2S from S^o and H_2 in the late exponential growth phase (Huber *et al.*, 1992). If in the presence of sulfur, methanogenic (hyper)thermophilic Archaea of the genera *Methanopyrus* and *Methanococcus* produce significant amounts of H_2S, while methanogenesis is significantly reduced (Stetter & Gaag, 1983).

Ecosystems where the sulfur transformations, namely sulfate reduction and sulfide oxidation, occur in a cyclic manner are called sulfureta (Jorgensen, 1982). In Nature, sulfureta are found in many different environments, like marine sediments, microbial mats and hydrothermal vents, and often show strong diurnal, or even tidal, cycles, resulting in moving O_2-H_2S interfaces (van den Ende *et al.*, 1997, Lens & Kuenen, 2001). In such systems, oxidation of HS^- in the light may be catalyzed both by aerobic and anaerobic phototrophs. In the absence of light, "colourless sulfur bacteria" can also contribute to the sulfur turnover.

Over the years, the term "colourless sulfur bacteria" has been used to classify lithotrophic sulfur compounds oxidizers taxonomically. Nowadays, this term has no taxonomic meaning anymore; nevertheless, the term is still used to describe the physiological group of bacteria that are commonly encountered in sulfide rich environments. "Traditional" colourless sulfur bacteria belong to the Bacteria genera *Beggiatoa*, *Thiobacillus*, *Thiotrix*, *Thiomicrospira* and *Thioploca*. However, many organoheterotrophic bacteria from other genera, as well as from Archaea (*e.g.*, *Sulfolobus* and *Acidianus* genera), can obtain energy from the oxidation of sulfur compounds (Lens & Kuenen, 2001)

2.4. *ACIDIANUS AMBIVALENS*

Acidianus (A.), ambivalens is a hyperthermoacidophilic archaeon of the *Sulfolobales* order which was isolated by Zillig and coworkers in the Leirhnjúkur fissure near the Krafla volcano in Iceland (Figure 2.3) (Zillig *et al.*, 1985). It is a

facultative aerobic or anaerobic chemolithoautotroph capable of utilizing CO_2 as sole carbon source and S^o as energy source, oxidizing or reducing it to H_2SO_4 (aerobic conditions), or H_2S (anaerobic conditions) (Zillig *et al.*, 1986, Gomes *et al.*, 2001a). *A. ambivalens* grows between pH 1-4 and at 84-87 ºC but displays its growth optimum at around 81 ºC and pH 2.5 (Zillig *et al.*, 1986)

Figure 2.3 – Krafla volcano (left), and Leirhnjúkur fissure (right), in Iceland where Zillig *et al.* isolated *Acidianus ambivalens* (Zillig *et al.*, 1986).

A. ambivalens was firstly classified in 1985 by Zillig and coworkers (Zillig *et al.*, 1985) as *Sulfolobus ambivalens* but the same authors reclassified the organism the next year as *Desulfurolobus ambivalens* (Zillig *et al.*, 1986). Due to the coccoid morphology of their members, all genera within the *Crenarchaeota* class (*Sulfolobus*, *Desulfurolobus*, *Metallosphaera* and *Stygiolobus*), were distinguished mainly by its metabolic properties and GC-content; but it was already clear at that time that the genus *Sulfolobus*, with three recognized species of sulfur-dependent Archaea, *S. acidocaldarius*, *S. solfataricus* and *S. brierleyi*, displayed significant differences in physiology and phylogenetic distance among its members that they should be reconsidered for reclassification. Then, in 1996, Fuchs and coworkers published a 16S rDNA-based phylogenetic analysis of all species within the order *Sulfolobales* (not only the three species mentioned above but also several other strains and isolates obtained meanwhile), where they reclassified *Desulfurolobus ambivalens* as *Acidianus ambivalens* and establish the physiological and morphological features that members of this genus should display (Fuchs *et al.*, 1996). The genus *Acidianus* gets it name

from the juxtaposition of *Acidus* (from the Latin, acid), and *Ianus* (the mythological Roman God of gates, doors and transitions, beginning and ending of time). Ianus is normally portrayed as having two heads facing opposite directions (Figure 2.4), and the "acidic bifacial" definition of *Acidianus* reflects the dual ability of growth conditions and metabolism of the organism. The specific epithet *ambivalens*, meaning "ambivalent", just corroborates this dual ability of the metabolism and growth of the microorganism.

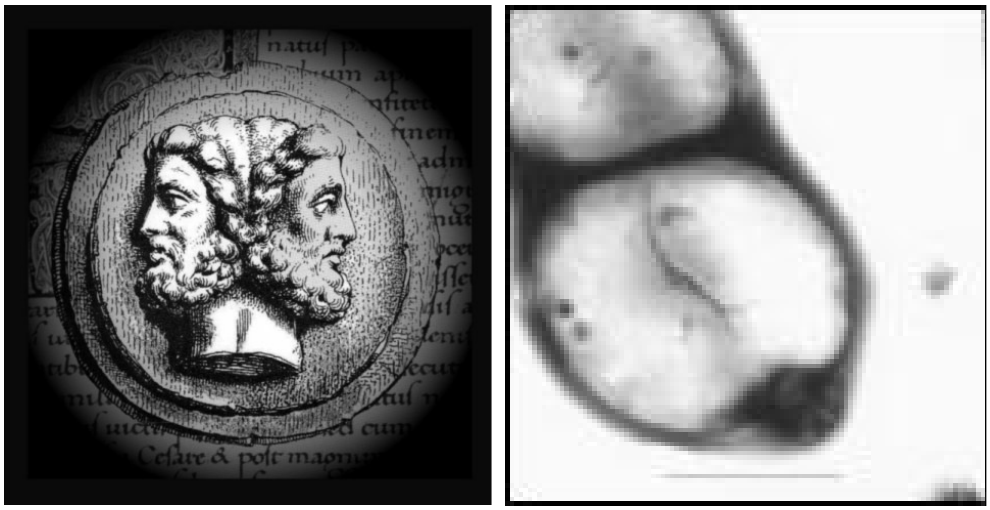


Figure 2.4 – Illustration of the mythological Roman God Ianus (left), and electron micrograph of *A. ambivalens* (right, courtesy of T. Urich, Darmstadt, Germany).

A. ambivalens has a GC-content of 32.7 % (lower than for other *Sulfolobus* species which is around 38 %), and contains three plasmids one of which, pSL10, suggested to be a pro-virus, is amplified during anaerobic growth (Zillig *et al.*, 1986). *Acidianus* species are resistant to vancomycin, ampicillin and kanamycin (150 $\mu g.mL^{-1}$ each), and, when grown in aerobic conditions, contain caldariella quinone, sulfolobus quinone and tricyclic quinone as the main electron carriers in the cell membrane.

Up until recently, *A. ambivalens* was thought to have the simplest membrane-bound aerobic respiratory chain known so far (Kletzin *et al.*, 2004). This minimal respiratory chain was composed by a noncanonical type-II NADH dehydrogenase (Bandeiras *et al.*, 2002, Gomes *et al.*, 2001a), and by an atypical succinate

dehydrogenase (Lemos *et al.*, 2001), both having the ability to reduce caldariella quinone. This pool of caldariella quinol was used to reduce a terminal oxygen reductase, a proton-pumping aa_3-type quinol oxidase. Besides these, several enzymes involved in the sulfur metabolism and aerobic respiratory chain of *A. ambivalens* have been characterized, namely a heme:copper oxygen reductase (Gomes *et al.*, 2001a), a thiosulfate:quinone oxidoreductase (Muller *et al.*, 2004) and a sulfur oxygenase reductase (SOR), of which sulfide is one of the products (Urich *et al.*, 2004). These features, all together, make *A. ambivalens* an interesting model microorganism for the study of dissimilatory oxidation and reduction of elemental sulfur in (hyper)thermophilic Archaea.

2.5. SULFIDE:QUINONE OXIDOREDUCTASE (SQR)

Hydrogen sulfide (H_2S), was discovered by Carl Wilhelm Scheele in 1777 (Dobbin, 1931). It is considered a very toxic substance for aerobic organisms hampering oxygen transport and inhibiting oxygen reduction by heme:copper oxygen reductases, thus preventing energy production by oxidative phosphorylation. Furthermore, sulfide is a strong nucleophile and may react with disulfide bridges and bind to metal centres. Despite its toxicity, which is five-fold higher than that for carbon monoxide, hydrogen sulfide is a fundamental molecule in both aerobic and anaerobic organisms. H_2S has now been proposed to be the third signalling "gas" in eukaryotes (Wang, 2002), being of vital importance in the brain, heart and smooth muscle (Lloyd, 2006). In mammalian cells the presence of hydrogen sulfide results mainly from the activity of two enzymes, the cystathionine γ-lyase and cystathionine β-synthetase (Qu *et al.*, 2008). Also, hydrogen sulfide is a metabolite produced by prokaryotes present in the lumen of the large intestine. Oxidation of sulfide by animal mitochondria has been demonstrated more than 20 years ago (Powell & Somero, 1986), which was shown to be associated with the respiratory chain and coupled to ATP production (Ouml *et al.*, 1997); recently sulfide has been reported to be the first inorganic substrate of human cells (Goubern *et al.*, 2007).

In Archaea and Bacteria, H_2S may be an electron donor to the respiratory chain. Two enzymatic systems are known to be involved in sulfide oxidation: flavocytochrome *c* (FCC), and sulfide:quinone oxidoreductase (SQR), (Griesbeck *et al.*, 2000). Genes coding for the latter are present in mitochondrial genomes, including the human one, and a bacterial origin of eukaryotic SQR has been proposed (Theissen *et al.*, 2003). The first evidence of the participation of quinones in sulfide oxidation came from the work of Knaff and Buchanan (Knaff & Buchanan, 1975). However, it was not only until 10 years later that Brune and Trüper (Brune & Trüper, 1986), following their work with the bacterium *Rhodobacter sulfidophilus*, first demonstrated that electrons from sulfide enter the membrane-bound photosynthetic electron transport chain at the level of quinone.

SQR is single polypeptide membrane-bound enzyme with an apparent molecular weight of 55 kDa and possibly active as dimer (Arieli *et al.*, 1994, Schutz *et al.*, 1997). SQR, like FCC, is a member of the flavin disulfide reductases (FDR), family, which includes glutathione and thioredoxin reductases, and dihydrolipoamide dehydrogenases. Members of this family are characterized by having two redox centres, one of these being FAD. The other redox centre, located close to the flavin, can be a pair of cysteine residues, a cysteine sulfenic acid or a mixed C-S-S-CoA disulfide. In the case of a redox pair of cysteine residues, the sequence position of these residues may vary but is conserved within each subfamily (Argyrou & Blanchard, 2004). Protein sequence comparison leads to the conclusion that SQR is a phylogenetically very old enzyme that was acquired early in evolution (Shahak *et al.*, 1999): the sulfide-dependent quinone reduction was essential for the early organisms to thrive in the sulfidic anoxic conditions of oceanic environments; genes coding for SQR were then acquired by eukaryotic cells from the endosymbiontic ancestor and were conserved in these organisms due to the importance of sulfide (Theissen *et al.*, 2003).

To date, several SQRs have been preliminarily characterized but only one, from *R. capsulatus*, has been extensively characterized (Theissen & Martin, 2008, Arieli *et al.*, 1994, Griesbeck *et al.*, 2002, Shibata & Kobayashi, 2006, Wakai *et al.*, 2007, Shibata *et al.*, 2007, Vande Weghe & Ow, 1999); however, its subcellular

localization is still a matter of debate. Most SQR enzymes are tightly bound to the membrane and are only extracted by detergent treatment (Reinartz *et al.*, 1998, Arieli *et al.*, 1994, Nübel *et al.*, 2000). In contrast, SQR from *R. capsulatus* (Schutz *et al.*, 1997), as well as a recombinant form of SQR from *A. ambivalens* expressed in *E. coli* (our own observations, not published), are detached from the membrane by treatment with 2 M sodium bromide. Since *R. capsulatus* deposits sulfur outside the cells, it seems reasonable to assume that SQR is attached to the periplasmic surface of the cytoplasmic surface. However, proteins that become translocated across the cytoplasmic membrane, exhibit a characteristic N-terminal signal peptide for translocation, and no signal peptide is found in the amino acid (a.a.) sequence of SQR. Moreover, the sulfide oxidation site of SQR was suggested to be located at the cytoplasmic surface of thylakoid membranes of *Oscillatoria limnetica* (Arieli *et al.*, 1991).

Up to recently, there was no three-dimensional structure of SQRs available. In fact, only the X-ray structure of FCC from *Allochromatium* (*All.*) *vinosum* was known to date (Chen *et al.*, 1994, Van Driessche *et al.*, 1996). Herein we report a detailed functional and structural characterization of SQR isolated from the membranes of *A. ambivalens*. We have initially crystallized a truncated form of SQR (proteolytic cleaved at the C-terminal), (Brito *et al.*, 2006), first assigned as a type II NADH dehydrogenase due to its NADH oxidase activity (Gomes *et al.*, 2001b). More recently, the complete form of SQR was isolated, which showed exclusively sulfide:quinone oxidoreductase and no NADH dehydrogenase activity. The three-dimensional structures of the truncated (SQR_T), and complete (SQR), forms of sulfide:quinone oxidoreductase were determined to 2.7 and 2.6 Å resolution, respectively. Both structures revealed a covalently bound FAD and a pair of adjacent cysteine residues bridged by a chain of three sulfur atoms. A possible sulfide oxidizing mechanism is proposed and the role of SQR in the global sulfur-linked bioenergetics of *A. ambivalens* is discussed, showing how this enzyme allows *A. ambivalens* to obtain maximal energy from sulfur. A comparison with FCC from *All. vinosum* (Chen *et al.*, 1994, Van Driessche *et al.*, 1996), and two new SQR crystal structures from *Aquifex aeolicus* (Marcia *et al.*, 2009), and *Acidithiobacillus ferrooxidans* (Cherney *et al.*, 2010), as well as the mechanisms therein proposed, is also performed.

2.6. MATERIALS AND METHODS

CELL GROWTH AND PROTEIN PURIFICATION

Cells of *Acidianus ambivalens* (DSM 3772), were grown aerobically in 10 L glass flasks by batch culture, at 80 ºC and pH 2.5 as described in (Zillig *et al.*, 1986). The growth media was completed with the addition of 0.02 % (w/V) of yeast extract. Growth was followed by monitoring the absorbance band at 546 nm, after removal of the suspended sulfur by filtration. The pH was also measured in order to follow the sulfuric acid production. After two or three days the culture medium reaches an absorbance of 0.3-0.4 at 546 nm. To harvest the cells, the cell suspension was cooled to room temperature followed by filtration through a 10-16 μm pore sintered glass filter and then centrifuged at 5400g for 7 min, at 4 ºC. The olive-green cell pellet was washed several times in 100 mM potassium phosphate buffer pH 6.5, to keep the cells at a neutral pH. If not for immediate use, cells were stored at -80 ºC.

The cells were then suspended in 40 mM potassium phosphate buffer pH 6.5 and broken at 6000 Psi using a French press. Unbroken cells were removed by centrifugation at 18000 rpm during 15 min. The cell extract was ultracentrifuged for 6 hours at 138000g. The resulting green pellet, the membrane fraction, was homogenized and solubilized with n-dodecyl-β-D-maltoside (DDM), in the proportion of 1 g detergent per g of protein, and the resulting suspension was centrifuged at 138000 g for 6 hours at 4 °C. All the chromatographic steps carried out during protein purification were performed on a Pharmacia HiLoad™ system, at 4 ºC and NADH oxidase activity was monitored using the NADH driven ferricyanide reduction. The green membrane solubilized extract was applied to a 330 mL DEAE-Sepharose Fast flow column. After elution, the NADH oxidizing fraction was applied to a 70 mL home-packed HTP Micro-beads column, followed by a second ionic exchanger, Q-Sepharose and a 330 mL gel filtration Superdex S-200 column. In the end a second HTP column step was performed in order to increase purity. The purified enzyme was divided into aliquots and stored at -80°C. This resulted in a truncated form of SQR (SQR_T), lacking ~50 a.a. residues in the

C-terminal region. This enzyme was shown to have NADH:quinone oxidoreductase activity and was thus first assigned as a type II NADH dehydrogenase (Gomes *et al.*, 2001b). However, the three-dimensional structure of this form prompted a new functional characterization, which led to the conclusion that it is in fact a sulfide:quinone oxidoreductase. The complete form of SQR (SQR), was then purified by similar procedures, but in the presence of a cocktail of protease inhibitors (Complete Protease inhibitor cocktail tablets from Roche), and by monitoring the sulfide:quinone oxidoreductase activity. To ensure homogeneity of the purified enzyme, mass spectrometry and N-terminal sequencing (performed by ITQB services), and SDS-PAGE (Laemmli, 1970), analysis were carried out. Protein concentration was determined using the BCA method (Smith *et al.*, 1985). Flavin extraction was attempted by incubating the protein with 10 % trichloroacetic acid as in (Susin *et al.*, 1993). Caldariella quinone, the microorganism native quinone, was extracted from lyophilized *A. ambivalens* membranes using a (1:1), mixture of chloroform/methanol (Hu *et al.*, 1999).

SPECTROSCOPIC CHARACTERIZATION

Electronic spectra were obtained on a Shimadzu UV1603 spectrophotometer at room temperature. Unless otherwise stated, the sample was buffered in 50 mM potassium phosphate buffer pH 6.5, 0.1 % (w/V) DDM.

MOLECULAR MASS DETERMINATION

The protein molecular mass determination by gel permeation chromatography was performed in a 24 mL bed volume Superdex S200 column (GE healthcare), using both high and low molecular mass protein standards (67-669 kDa and 14.4-97 kDa, GE healthcare). Elution was made with 40 mM potassium phosphate buffer at pH 6.5, 150 mM NaCl and 0.1 % (w/V) DDM. For mass spectrometry analysis, *A. ambivalens* SQR with a concentration between 460 and 830 μM was diluted 1:10 in a matrix (sinapinic acid in 70 % acetonitrile, 0.1 % TFA), and subjected to MALDI-TOF analysis (PO 07MS spectrometer), at the ITQB mass spectrometry facility.

SEQUENCE ANALYSIS

Amino acid sequences of enzymes from other organisms were compared using BLAST at NCBI databases. Multiple sequence alignments were produced as previously described (Thompson *et al.*, 1997), and manually adjusted. Dendrograms were built with the Geneious software (Drummond A.J., 2008).

CATALYTIC ACTIVITY ASSAYS

NADH dehydrogenase activity was assayed at 50 ^{o}C following $K_3Fe(CN)_6$ reduction at 420 nm (ε_{420nm}=1.0 $mM^{-1}.cm^{-1}$), in a medium containing 1 mM $K_3Fe(CN)_6$ and 200 μM NDH-2 in 20 mM potassium phosphate buffer pH 6.5 and 0.1 % (w/V) DDM. After approximately 5 min incubation at the working temperature, the reaction was initiated by the addition of NADH.

Sulfide oxidase activity in *A. ambivalens* membranes was monitored at 50 ºC by measuring O_2 consumption polarographically with a Clark-type oxygen electrode, YSI model 5300 (Yellow Springs). The reaction mixture contained 40 mM potassium phosphate buffer at pH 6.5 and 300 μM Na_2S (freshly prepared). For NADH consumption assays, 0.3 to 4.6 mM NADH was used instead of Na_2S. The reactions were started by the addition of membranes (approximately 210 μg protein.mL^{-1}). For inhibition experiments, 30 mM of iodoacetoamide (in the same buffer), and an ethanolic solution of HQNO (50 mM), were used (final concentration in the assay of 300 μM and 500 μM, respectively).

SQR activity was measured under anaerobic conditions at 50 ºC in an Olis DW2 UV/Vis spectrophotometer by sulfide-dependent quinone reduction, using two beams at the following wavelengths: decylubiquinone (dUQ), 275-300 nm ($\Delta\varepsilon$=12500 $M^{-1}\cdot cm^{-1}$), 2,3-dimethyl-1,4-naphthoquinone, 270-290 nm ($\Delta\varepsilon$=15200 $M^{-1}\cdot cm^{-1}$), 2-methyl-1,4-naphthoquinone (menadione), 280-260 nm ($\Delta\varepsilon$=7800 $M^{-1}\cdot cm^{-1}$), and caldariella quinone, 351-341 nm ($\Delta\varepsilon$=1180 $M^{-1}\cdot cm^{-1}$). The reaction mixture contained 50 mM potassium phosphate pH 6.5, 20 mM glucose, 1 U glucose oxidase mL^{-1}, 10 U catalase mL^{-1}, 0.025 % (w/V) DDM, 50 μM of the

respective quinone and 5 to 14 μg protein.mL^{-1}. The enzymatic reactions were started by addition of Na_2S.

Inhibition experiments were performed by adding HQNO or KCN (final concentration was 100 μM prepared in 200 mM potassium phosphate buffer pH 6.5), after the addition of Na_2S. In the case of iodoacetoamide, the substrate was added after prior incubation with the inhibitor. Controls were performed in the absence of the enzyme.

The pH activity profile was performed between pH 3 and 8 using the following buffers (all at 50 mM concentration with 0.025 % (w/V) DDM): potassium phosphate pH 3, formic acid between pH 3.5 and 4.5, potassium acetate between pH 5 and 5.5, and MES/Bis-tris propane between pH 6 and 8. In the temperature activity profile, the solution temperature was monitored using a HIBOK 14 thermometer inside the reference cell. NADH:quinone oxidoreductase activity was monitored in the same conditions by following the decrease in absorbance of NADH (initial concentration of 260 μM), at 339 nm.

CRYSTALLIZATION AND X-RAY DATA

Initial crystallisation trials were undertaken with a protein concentration of 10 mg.mL^{-1} in 10 mM potassium phosphate buffer pH 6.5 and 0.025 % (w/V) DDM using the vapour diffusion technique. Sitting drops containing 1.5 μL of protein and 1.5 μL of precipitant solutions were equilibrated against 500 μL of precipitant solution at 4 ºC and 22 ºC. Specially designed screens for membrane proteins, MemSys from Molecular Dimensions and MembFac from Hampton Research, were used initially. Since most of the drops showed weak/no precipitate, Crystal Screen I and II (Hampton Research), containing higher precipitant concentration were also used. Three very small crystals appeared after several weeks in condition 48 of the Crystal Screen I (2.0 M $NH_4H_2PO_4$ and 100 mM Tris-HCl pH 8.5). It is noteworthy to mention that although the pH of the buffer was 8.5, the pH of the crystallisation solution was *ca.* 4.5. Optimised crystallisation conditions included 2.2 M $NH_4H_2PO_4$ and 100 mM Tris-HCl pH 8.5 or 2.0-2.4 M $NH_4H_2PO_4/K_2HPO_4$ at pH 4.0-5.0.

Yellow crystals with a hexagonal shape appeared in 2 weeks and grew to the largest dimensions ~0.2×0.15×0.1 mm^3 (Figure 2.5). Due to the high concentration of precipitant, salt crystals rapidly started to form after the sitting drop cover lid was opened. The addition of glycerol to the crystallisation drop was necessary to prevent the formation of salt crystals during necessary manipulation of the protein crystals. Suitable cryoprotectant conditions were obtained by lowering the precipitant concentration to 1.5 M and adding 25 % (V/V), glycerol. Under these conditions SQR_T crystals are stable and can be kept for several weeks without degrading.

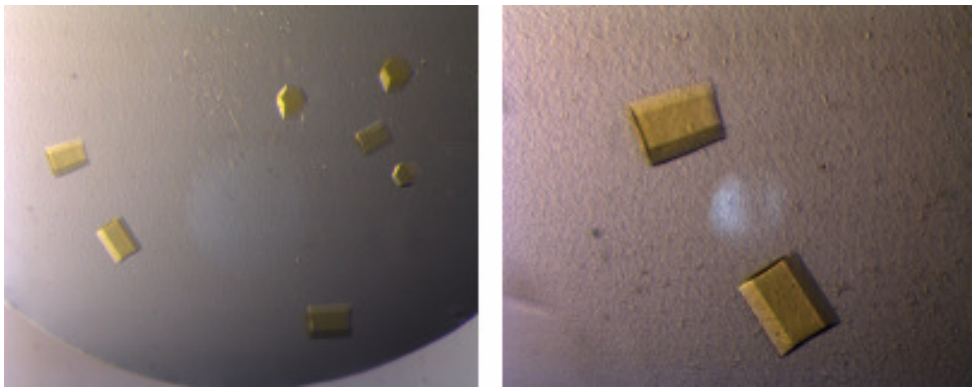


Figure 2.5 – SQR_T crystals grown in 2.2 M $NH_4H_2PO_4$ and 100 mM Tris-HCl pH 8.0.
Crystal dimensions ~ 0.2×0.15×0.1 mm^3.

Crystals were flash frozen in liquid nitrogen and X-ray diffraction data were collected using a synchrotron radiation source. The intensity data were measured on beamline ID14-1 at ESRF (European Synchrotron Radiation Facility, Grenoble), using an ADSC Q210 CCD detector. Native crystals belong to the hexagonal space group $P6_{1,5}22$ with unit cell dimensions a=178.58 Å and c=162.31 Å and diffract to 2.6 Å. Data were processed with DENZO and merged with SCALEPACK [both programs are from the HKL suite (Otwinowski *et al.*, 1997)], or with XDS (Kabsch, 1993) and SCALA (Evans, 1997) from the CCP4 suite of programs (1994). The estimated solvent content is ~65 or ~47 %, assuming 2 or 3 molecules in the asymmetric unit, respectively.

STRUCTURE DETERMINATION AND REFINEMENT

SQR_T crystal soakings or co-crystallisation with different heavy-atom compounds were performed to search for derivatives. Most of the heavy atom compounds tested (Hg, Pt, Au, U, Sm, Ag, W, Xe, Yb, Pb, I, Br, Tl, Pd), showed no anomalous signal but a few (Pt, I, Br, Mo and Xe), were considered to be "weak" derivatives, by comparison of the χ^2 and R-factor between the native dataset and a few frames from the putative derivative, using the SCALEPACK program (Otwinowski *et al.*, 1997). Soaking times ranging from 1 hour to 16 days (except for the halides and the Xe soakings which were only a few minutes and higher concentrations), and soaking conditions ranging from 100 μM to 100 mM were tested.

SHELXC (Sheldrick, 2008), within HKL2MAP (Pape & Schneider, 2004), was used to prepare the experimental data (option "no merge original index" used in SCALEPACK to generate the .sca file), and SHELXD (Sheldrick, 2008), was used to locate heavy atom sites from the putative derivatives. Two positive hits were found out of 1000 trials, the best solution showing Correlation Coefficients - CC_{all} of 17.31 % and CC_{weak} of 10.34 %, and PATFOM 0.84. A total of 7 Pt sites were found with occupancies higher than 0.5, using space group $P6_122$ but SHELXE (Sheldrick, 2008), did not produce an interpretable electron density map. SHELXD was also used to search for heavy atom sites of other putative derivatives (I, Br, Mo and Xe), but no solutions were found out of 2500 trials.

The initial set of 7 sites found with SHELXD for the Pt derivative were refined in SHARP and subsequent analysis of log-likelihood gradient maps revealed 4 additional sites, giving a total of 11 heavy atom positions. Statistics were good at low resolution (< 5.5 Å). The overall values of Phasing power to 2.6 Å (isomorphous differences), for acentric and centric reflections were 0.795 and 0.852, respectively. The Phasing power (anomalous differences), for acentric reflections were 0.387. The figure-of-merit, for acentric and centric reflections was 0.080 and 0.125, respectively. Although these statistics were rather poor, the calculated phases allowed detection of the correct hand through a single cycle of solvent flipping with SOLOMON (Abrahams & Leslie, 1996), showing significantly better statistics for the

enantiomeric space group $P6_522$.

Difference Fourier maps were calculated with the other putative derivatives. Additional derivatives were added: I, Mo, Br and Xe. However, refinement of the heavy atom parameters in SHARP showed rather large non-isomorphism parameters. Furthermore, some derivatives gave little contribution to the overall phasing power. Since the highest resolution of the data (2.6 Å), seemed rather low for attempting automatic model building with ARP/wARP, it was decided to reduce the noise introduced by non-isomorphism to a minimum, by using only the I derivative in the final stages of experimental phasing.

To further improve the heavy atom refinement and to detect additional minor sites, the phases after solvent flattening (using 63.6 % solvent content), were used as phase restraints in SHARP. This resulted in a more stable refinement of the I derivative parameters and detection of additional 10 sites, resulting in a total of 51 I sites (with occupancies of 0.14-0.68 and temperature factors of 42-81 $Å^2$). The overall values of Phasing power to 2.6 Å (isomorphous differences), for acentric and centric reflections were 1.184 and 0.994, respectively. The Phasing power (anomalous differences), for acentric reflections was 0.308. The figure-of-merit, for acentric and centric reflections was 0.180 and 0.232, respectively.

Automatic model building with ARP/wARP and phase improvement with SOLOMON were performed in a cyclic manner; the protein part of an ARP/wARP model was used during the early cycles of density modification to improve the solvent mask determination. This resulted in better density modified phases and maps, which in turn were given to ARP/wARP to build more of the model. This resulted in 18 chains with 621 residues (out of 818 residues in two molecules), with 15 of these chains (or 589 residues), docked into the sequence. Iterative model building and crystallographic refinement were performed with the programs COOT (Emsley & Cowtan, 2004) and BUSTER-TNT (Bricogne *et al.*, 2008).

Later on, the full-length enzyme (SQR), was purified and crystals grown under similar conditions of SQR_T. Since crystals of SQR were isomorphous with SQR_T, an initial rigid body refinement was done with REFMAC5 (Murshudov *et al.*, 1997). Iterative model building and crystallographic refinement were performed with

the programs COOT (Emsley & Cowtan, 2004) and BUSTER-TNT (Bricogne *et al.*, 2008). All structural figures were drawn using PyMOL (DeLano, 2002), and the topology diagram was generated with the program TOPS (Westhead *et al.*, 1998).

2.7. RESULTS AND DISCUSSION

EVIDENCE FOR THE PRESENCE OF A SULFIDE:QUINONE OXIDOREDUCTASE IN *A. AMBIVALENS* MEMBRANES

A. ambivalens membranes were shown to consume O_2 upon addition of Na_2S at 150 nmol $O_2 .mg^{-1}.min^{-1}$, which implies the presence of a sulfide oxidation system linked to the aerobic respiratory chain. Upon addition of HQNO, a quinone competitive inhibitor, the rate of O_2 consumption decreased by ~80 %, while prior incubation with iodoacetoamide (alkylating sulfydryl reagent known to bind to cysteine residues), completely abolished O_2 consumption.

PRIMARY STRUCTURE AND SEQUENCE COMPARISON

Genes encoding SQRs are present in the three domains of Life. On the basis of a.a. sequence comparisons, Theissen and coworkers proposed that SQRs are divided into three groups: Group I comprises enzymes from Bacteria; Group II includes enzymes from Bacteria and Eukarya; Group III contains bacterial and archaeal enzymes (Theissen *et al.*, 2003). With the larger number of sequences now available, a new sequence alignment was performed, whose derived dendrogram (Figure 2.6), reveals that SQR from *A. ambivalens* and other related archaea cluster in a branch included in Group I, and not in Group III as would be expected from that classification. Moreover, some archaea have genes encoding several SQRs, belonging either to Groups I or III, meaning that the distribution of SQRs is not related to the microbial Phyla and evidencing the occurrence of lateral gene transfers throughout evolution of this enzyme family.

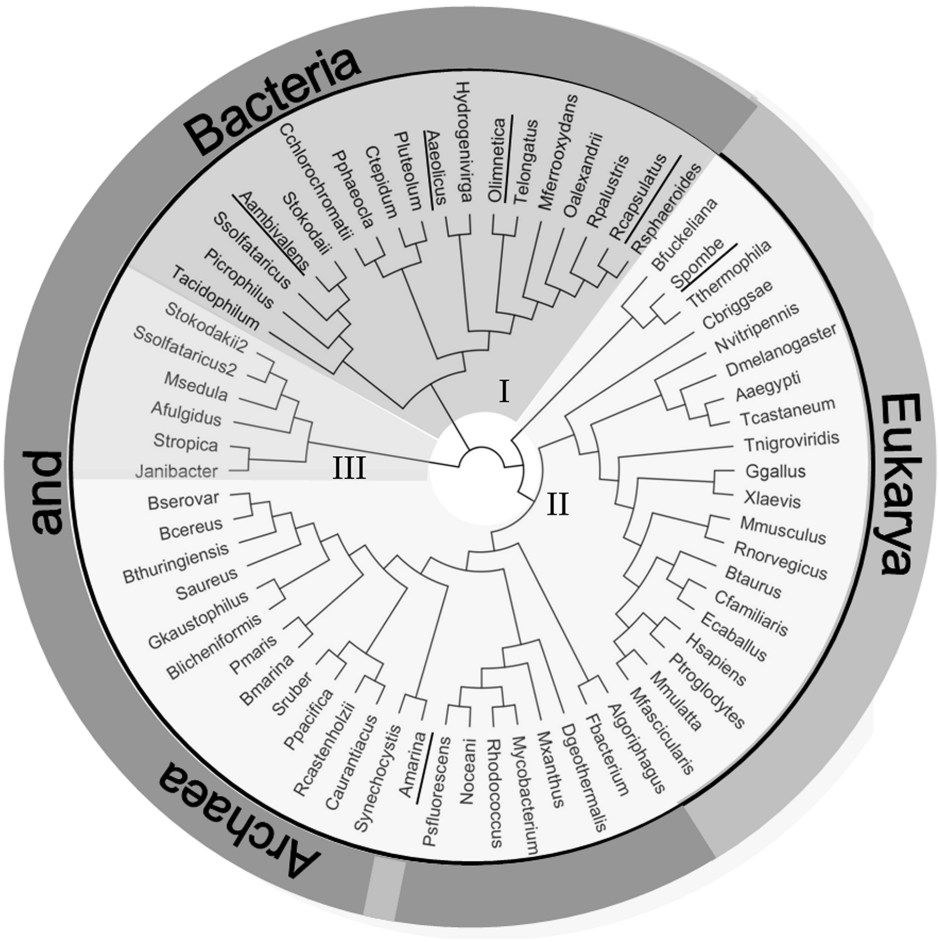


Figure 2.6 – Dendrogram of amino acid sequences from SQRs built with the Geneious software.

Sequences belonging to groups I, II, and III are indicated by different shadowing. Eukaryote and prokaryote SQRs are indicated at the periphery of the dendrogram. Underlined species names indicate those from which SQRs have been characterized.

The a.a. sequence of the *A. ambivalens* SQR is 13 to 89 % identical to those of other SQRs, which show among themselves a quite low sequence identity. The most similar sequences to the *A. ambivalens* SQR are those from archaea of the same phylogenetic order, the *Sulfolobales* (Figure 2.7).

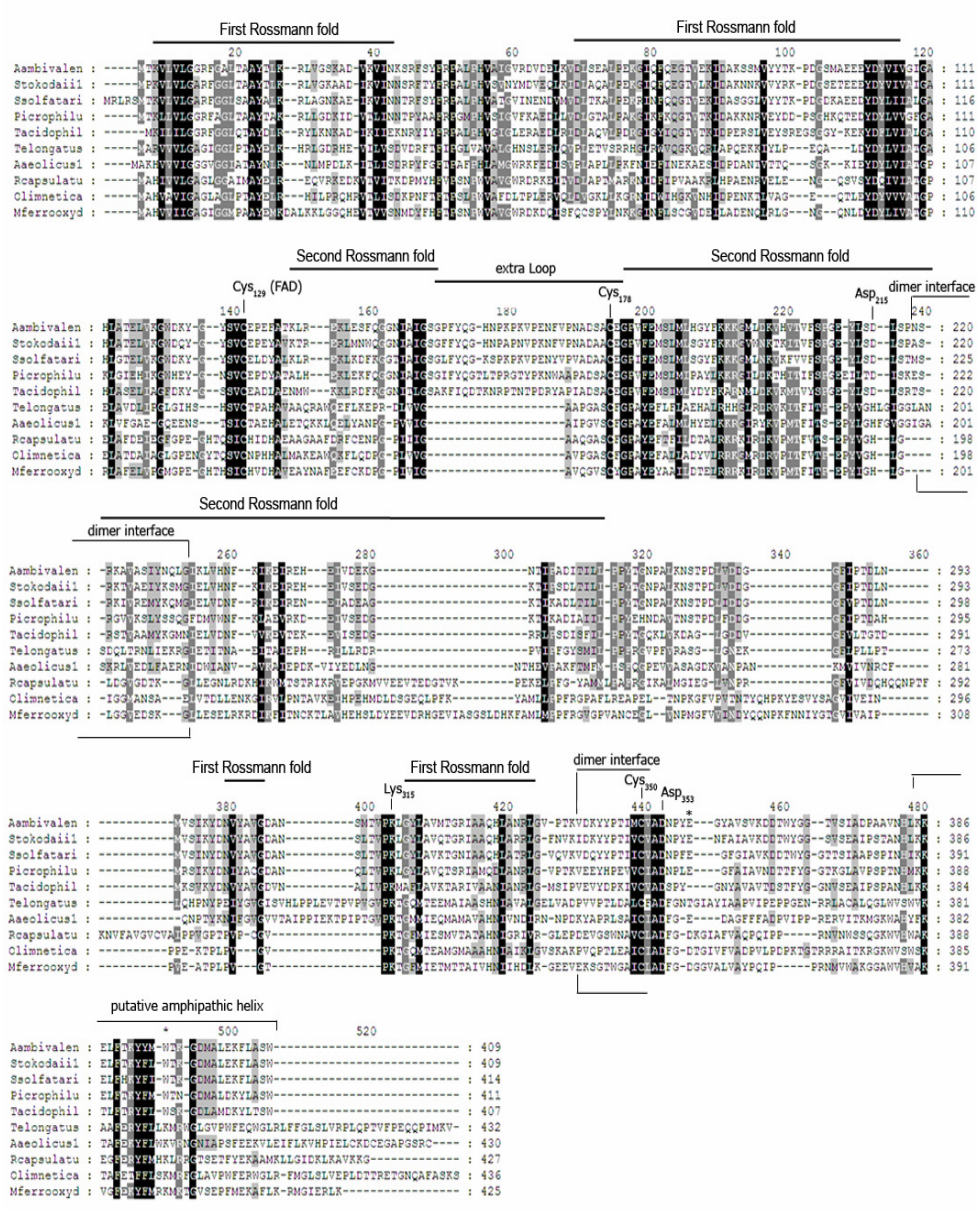


Figure 2.7 - Amino acid sequence alignment of *Acidianus ambivalens* SQR (acession numbers from NCBI database in parentheses), with sulfide:quinone oxidoreductases from Group I.

Aambivalen, *Acidianus ambivalens* (CAD33806.1); Stokodaii, *Sulfolobus tokodaii* str. 7 (NP_378484.1); Ssolfatari, *Sulfolobus solfataricus* P2 (NP_343636.1); Picrophilu, *Picrophilus torridus* DSM 9790 (YP_023317.1), Tacidophilu, *Thermoplasma acidophilum* DSM 1728 (NP_394588.1), Telongatus, *Thermosynechococcus elongatus* BP-1 (NP_681079.1), Aaeolicus, *Aquifex aeolicus* VF5 (NP_214500.1), Rcapsulatu, *Rhodobacter capsulatus* (CAA66112.1), Olimnetica, *Oscillatoria limnetica* 'Solar Lake' (AAF72962.1), Mferrooxyd, *Mariprofundus ferrooxydans* PV-1 (ZP_01453072.1); relevant secondary structure elements and residues are indicated.

The primary structures of the FDR enzymes have two glycine residue patterns that are part of two Rossmann folds, generally described as the FAD and the NAD(P)H binding domains. Depending on the subfamily, the position of the active cysteine residues can vary, but these positions are conserved within each subfamily (Argyrou & Blanchard, 2004). As for SQRs, the *A. ambivalens* enzyme has the conserved glycine residue pattern typical of the Rossmann fold near the N-terminus and has two conserved cysteine residues which form a redox-active disulfide bridge (Griesbeck *et al.*, 2000). The presence of the Rossmann fold at the N-terminus and the conserved cysteine residues compose the fingerprint of the SQR family. In contrast to other FDRs, whose second glycyl pattern is part of the second Rossmann fold involved in NAD(P)H binding, SQRs contain an insertion which blocks the NAD(P)H access to the flavin cofactor as shown in the structure described here (Figure 2.7, detailed below). Apart from the conserved a.a. residues from these motifs, few other residues are conserved. Among those are (numbering refers to *A. ambivalens* SQR): a highly conserved serine residue (S_{127}), a lysine residue (K_{315}), and a motif [K_{386}–$(X)_{4-7}$–(Y/F)-$(X)_{0-1}$-(Y/W/F)], near the C-terminus. The third cysteine residue (C_{129}), that was proposed to participate in the sulfide oxidation mechanism (Griesbeck *et al.*, 2002), although present in many SQRs from Group II, is only strictly conserved in Group I enzymes. A glutamic residue, E_{184} (E_{165} in *Rhodobacter capsulatus* SQR), suggested to act as an active base during catalysis (Griesbeck *et al.*, 2002) is not strictly conserved and may be substituted by a lysine. There is also an aspartic residue, D_{353}, strictly conserved among Group I SQRs, lying 3 to 9 residues downstream of the last conserved cysteine (C_{350}). In *R. capsulatus* SQR (Griesbeck *et al.*, 2000), two histidines (H_{131} and H_{196}, *R. capsulatus* numbering), were proposed to be involved in quinone binding but none of these are fully conserved among SQRs.

BIOCHEMICAL CHARACTERIZATION

An SQR was isolated from the membranes of *A. ambivalens*. The *A. ambivalens* SQR was initially purified in a truncated form and later a complete form was isolated. The C-terminally truncated form of the enzyme (named SQR_T), of

~40 kDa, has NADH:quinone oxidoreductase activity and was initially considered to be a type II NADH dehydrogenase (Gomes *et al.*, 2001b). The complete form of SQR consists of 409 a.a. residues (calculated mass of 45151.85 Da), and has exclusively sulfide:quinone oxidoreductase activity.

SQR was purified to homogeneity as judged both by SDS-PAGE and MALDI-TOF mass spectrometry. The oligomerization state of the protein was investigated by gel permeation chromatography and a single peak corresponding to a molecular mass of ~48 kDa was observed which indicates that the protein is a monomer in solution under the conditions tested (20 mM phosphate buffer pH 6.5 and 150 mM NaCl). The electronic absorption spectrum of the as purified SQR has the typical fingerprints of a flavoprotein (λ_{max} at 454 nm and 350 nm), (Figure 2.8a).

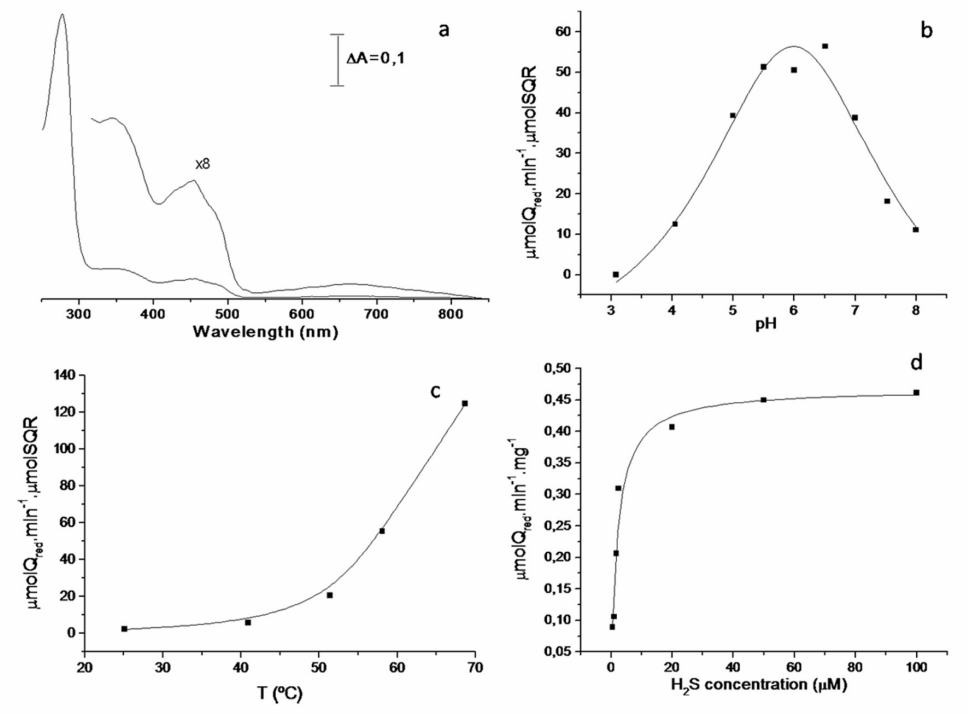


Figure 2.8 – a), UV-visible spectrum of the *A. ambivalens* SQR at pH 6.5; b), pH (at 50 ºC), and c), temperature (at pH 6.5), profiles of SQR activity using dUQ as substrate (b- and c- lines depicted in the figure are for better visualization); d), Michaelis-Menten hyperbola using sulfide as substrate (the solid curve was calculated with the parameters presented in the text).

All attempts to extract the flavin cofactor were unsuccessful as described previously (Bandeiras *et al.*, 2002), indicating the presence of a covalently bound flavin, now confirmed by the X-ray structure. The structure described here also showed that the flavin cofactor is a flavin adenine dinucleotide (FAD).

FUNCTIONAL CHARACTERIZATION

The intact form of the enzyme has exclusively sulfide:quinone oxidoreductase activity with turnovers and specific activities presented in Table 2.2. The highest sulfide:quinone oxidoreductase activity was observed with the native caldariella quinone and the reactivity of SQR with quinones seems to be correlated with their reduction potentials: activity increases with the increase of the quinones reduction potential (Table 2.2).

Table 2.1 – Sulfide:quinone oxidoreductase activity of SQR at 50 ºC and pH 6.5 with different quinones

Quinones	Specific activity (μmol quinone red.mg^{-1}.min^{-1})	Turnover (min^{-1})	Redox Potential (mV)
2,3-dimethyl-1,4-naphthoquinone	0.127	6	- 60
2-methyl-1,4-naphthoquinone	0.194	9	0
Decylubiquinone	0.470	23	+ 100
Caldariella quinone	0.531	26	+ 100

With dUQ a maximum turnover of 125 min^{-1} was obtained at 70 ºC, pH 6.5 (Figure 2.8b and 2.8c). Assays at higher temperatures were not possible to perform due to quinone instability. At 25 ºC, only 3 % of the activity is detected indicating that at room temperature, the enzyme is almost inactive, as frequently observed for enzymes from hyperthermophiles. The turnover for this enzyme is similar to the one from *Oscillatoria limnetica* (94 min^{-1}), (Arieli *et al.*, 1994).

At 50 ºC, the K_m and V_{max} for sulfide were determined to be 2 μM and 0.470 μmoles $min^{-1} \cdot mg^{-1}$, respectively, by monitoring quinol formation (Figure 2.8d). Reduction of dUQ is 55 % inhibited by the quinone analogue HQNO, indicating that quinone reduction is specific.

To investigate the possible involvement of the cysteine residues in the oxidation of sulfide, activity assays were performed in the presence of iodoacetoamide. Prior incubation of SQR with iodoacetoamide led to a complete inhibition of the sulfide-dependent quinone reduction, which suggests that the cysteine residues have indeed an active role in catalysis. The reaction is also inhibited by KCN to 42 %, as found for some other SQRs (Griesbeck *et al.*, 2000).

In the case of SQR_T, the truncated form of SQR, both NADH and sulfide:quinone oxidoreductase activities are observed, whereas the intact enzyme does not react with NADH. Both SQR and SQR_T structures have an extra loop blocking the NADH access to the flavin cofactor (see below). This means that the NADH oxidase activity of SQR_T results from a non-physiological exposure of the isoalloxazine ring of the FAD to the solvent in the truncated enzyme due to the lack of its last ~50 residues leading to unspecific binding of NADH. This conclusion is reinforced by the observations that *A. ambivalens* membranes consume O_2 upon addition of H_2S (see above), and that O_2 consumption using NADH as electron donor is negligible. These data indicate that in the membranes the truncated form is not present. NADH and sulfide:quinone oxidoreductase activities were also assayed in the presence of iodoacetoamide. As expected, iodoacetoamide completely abolished the sulfide dependent quinone reduction, but had no effect on NADH:quinone oxidoreduction.

CRYSTALLOGRAPHIC REFINEMENT AND MODEL QUALITY

The complete form of SQR crystallized under similar conditions as SQR_T and both forms are isomorphous. Crystals appeared within 2 weeks and grew up to maximum dimensions of ~0.2×0.15×0.1 mm^3. Crystals belong to the hexagonal space group $P6_522$ with two molecules in the asymmetric unit, approximate unit cell

dimensions of a=b=179.7 Å and c=163.4 Å (for SQR), and contain a high solvent content (~60 %).

The SQR_T X-ray structure was refined to 2.7 Å resolution with an R_{factor} of 19.6 % (R_{free} of 22.5 %), while the SQR final model shows an R_{factor} of 19.4 % (R_{free} of 22.2 %), to a resolution of 2.6 Å. X-ray data collection and refinement statistics as well as overall model quality parameters are depicted in Table 2.3.

The electron density maps were of good quality except for two short regions: a bulged β-strand fragment (P_{43} to A_{49}), and a α-helical stretch (E_{130} to A_{134}), which show higher thermal motion parameters (B-factors). No interpretable density was observed for the last 53 a.a. residues at the C-terminal region of the full-length SQR. Hence, similarly to the SQR_T structure, the crystallographic model of SQR consists of 356 residues per monomer (out of 409). Mass-spectrometry analysis of SQR in solution and from dissolved crystals showed that the polypeptide is complete in both samples (data not shown), suggesting high flexibility around the C-terminus region.

Next page: **TABLE 2.2 - Data collection and refinement statistics**

$$*\,R_{merge} = \sum_{hkl} \sum_{i} \left| I_i(hkl) - \overline{I(hkl)} \right| \Big/ \sum_{hkl} \sum_{i} I_i(hkl)$$

$$\#\,R_{pim} = \sum_{hkl} \left[1/(N-1) \right]^{1/2} \sum_{i} \left| I_i(hkl) - \overline{I(hkl)} \right| \Big/ \sum_{hkl} \sum_{i} I_i(hkl)$$

Calculated with the program SCA, R_{merge} and R_{pim} are indicators of the precision of the final merged and averaged data-set, where I_i *(h k l),* is the observed intensity of the i[th] measurement, $\overline{I(hkl)}$ is the average intensity of multiple observations of symmetry-related reflections and *N* is redundancy.

	SQR:K_2Cl_4Pt	**SQR:KI**	**SQR_T***	**SQR**
PDB ID code	-	-	3H8I	3H8L
Space group		$P6_522$		
Unit cell				
a, b , c (*Å*)	179.2, 163.2	178.1, 162.0	178.6, 162.3	179.7, 163.4
Soaking conditions	750 μM for 16 days	1 M for 1.5 min	-	-
Data collection				
Beamline		ESRF ID14-1		
Wavelength (*Å*)		0.934		
Resolution range (*Å*)	56.20-3.50 (3.63-3.50)†	58.30-3.00 (3.11-3.00)	47.43-2.65 (2.80-2.70)	51.44-2.57 (2.71-2.60)
Total no. of observations	58220 (5738)	90835 (9009)	157117 (22674)	383076 (29746)
No. of unique observations	19747 (1919)	30654 (3003)	44293 (6392)	49511 (6870)
R_{merge}* (%)	10.0 (62.8)	8.6 (56.0)	8.2 (50.4)	7.5 (51.0)
R_{pim}# (%)	6.8 (43.9)	58.0 (38.3)	11.3 (35.9)	8.5 (30.3)
Completeness (%)	98.6 (99.7)	99.0 (96.8)	99.6 (99.9)	99.5 (96.8)
Mean < I / σ(I) >	11.5 (1.6)	13.2 (1.7)	10.2 (2.0)	20.0 (2.6)
Refinement statistics				
R_{work} / R_{free} (%)	-	-	19.6 / 22.5	19.4 / 22.2
Geometry				
rmsd bonds (*Å*)	-	-	0.006	0.006
rmsd angles (°)	-	-	0.907	0.892
Ramachandran statistics				
Residues in (%)				
most favoured regions	-	-	94.9	96.3
allowed regions	-	-	4.4	3.0
disallowed regions	-	-	0.7	0.7
Mean B factors				
Protein (*Å²*)	-	-	56.8	55.4
FAD (*Å²*)	-	-	49.4	46.6
PS3 (*Å²*)	-	-	61.0	57.8
Solvent (*Å²*)	-	-	61.0	61.3

* Used as "native" for SIRAS phasing

† Values between parentheses refer to the highest resolution shell

OVERALL FOLD

The crystal structure of SQR contains two monomers in the asymmetric unit (Figure 2.9a). The dimer interface is approximately 1500 Å^2 and involves 20 H-bonds and a few hydrophobic contacts suggesting a possible dimeric arrangement of the enzyme in the membranes.

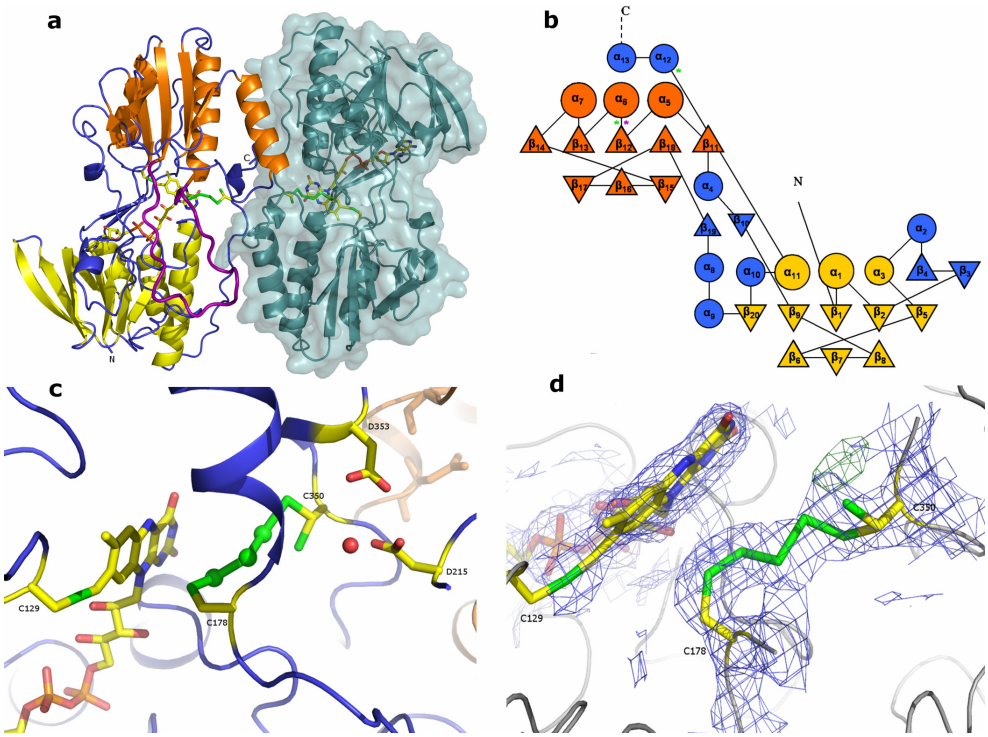


Figure 2.9 – The X-ray structure of sulfide:quinone oxidoreductase from *A. ambivalens*.
a) Cartoon and surface representation of the crystallographic dimer of SQR: in chain A, FAD domain is coloured in yellow, the second Rossman-like fold domain in orange and the loop hindering the access to the second Rossman fold in purple; chain B is in dark cyan with the molecular surface displayed; in both chains, the FAD, the redox active cysteines, the cysteine covalently binding the FAD and the trisulfide molecule are shown in sticks;
b) TOPS-generated topology diagram of SQR with numbered secondary structure elements (α-helices in circles and β-sheets in triangles), and numbered; FAD domain is coloured in yellow and the second Rossman-like fold in orange; the location of the loop blocking the access to the flavin moiety is pointed as a purple asterisk and the position of the two redox active cysteine residues pointed as green asterisks;
c) Cartoon representation of SQR redox active sites; FAD and active site residues are shown in stick representation;
d) $2F_o$-F_c contoured at 1.5σ (blue), and F_o-F_C at 5σ (green), electron density maps around the active site's residues (shown in sticks).

Both SQR_T and SQR structures are very similar showing an rmsd of 0.14 Å for 356 aligned C_α atoms [chains A, superposition performed with "Secondary Structure Matching" tool (Krissinel & Henrick, 2004), within COOT (Emsley & Cowtan, 2004)]. Hereafter we will refer to the complete SQR structure unless otherwise stated. Each monomer has two domains of similar architecture, which consist of a twisted five-stranded parallel β-sheet flanked by a three-stranded anti-parallel β-sheet on one side and by three α-helices on the other side, a Rossmann-like fold (see topology diagram, Figure 2.9b). The presence of these two domains is shared among the enzymes of the FDR family. The first (N-terminal), domain is involved in flavin binding, whereas, in most FDRs, the second domain is involved in NAD(P)H binding, except for SQR and flavocytocrome *c* sulfide dehydrogenase (FCC). FCC from *Allochromatium vinosum* is a heterodimer containing a 46 kDa flavoprotein subunit, homologous to FDRs, and a 21 kDa diheme cytochrome subunit (PDB code: 1FCD), (Chen *et al.*, 1994). Structural superposition of SQR and FCC flavoprotein subunit yields an rmsd of 2.6 Å for 287 aligned C_α atoms, where the highest deviations occur for a few loops and at the C-terminus (after N_{333}, SQR numbering), (Figures 2.10a and 2.10b).

In contrast to other FDRs, SQR contains an additional 25 amino acid-long loop (G_{154} to C_{179}, loop coloured in purple in Figure 2.9a, Figure 2.7), inserted in the second Rossmann-like fold, between the first β-strand and α-helix (β_{12} and α_6, location of the loop indicated by a purple asterisk in Figure 2.9b). This loop, with variable length among SQRs, extends towards the FAD binding domain and blocks the NAD(P)H binding cavity present in other FDRs (Figure 2.7). In FCC, a shorter loop is also observed (heptapeptide segment P_{155} to C_{161}, FCC numbering). Moreover, in both SQR and FCC the redox pair of cysteine residues is located on the *re*-side of FAD, the same side at which NADH interacts with FAD in most members of the FDR family.

In FCC, C_{161} and C_{337} form a disulfide bond, whereas C_{42} establishes a covalent linkage with the flavin. The a.a. sequence position of the cysteine residue covalently bound to the flavin and its environment differ significantly in both structures: in SQR, C_{129} is located on a turn between strand β_{11} and helix α_5 in the

second Rossmann-like fold (corresponding to the NAD(P)H binding domain for most FDRs), whereas C_{42} of FCC is situated on a loop between β_2 and α_2 in the first Rossmann fold (FAD binding domain), (Figures 2.9b and 2.10b). Moreover, in FCC the *si*-face of the flavin ring lies against the polypeptide backbone of residues C_{42} to L_{44}, in contrast to SQR, where the polypeptide chain is further away leaving enough free space to accommodate a quinone molecule. Noteworthy in FCC, the electron acceptor is a heme from the cytochrome *c* subunit.

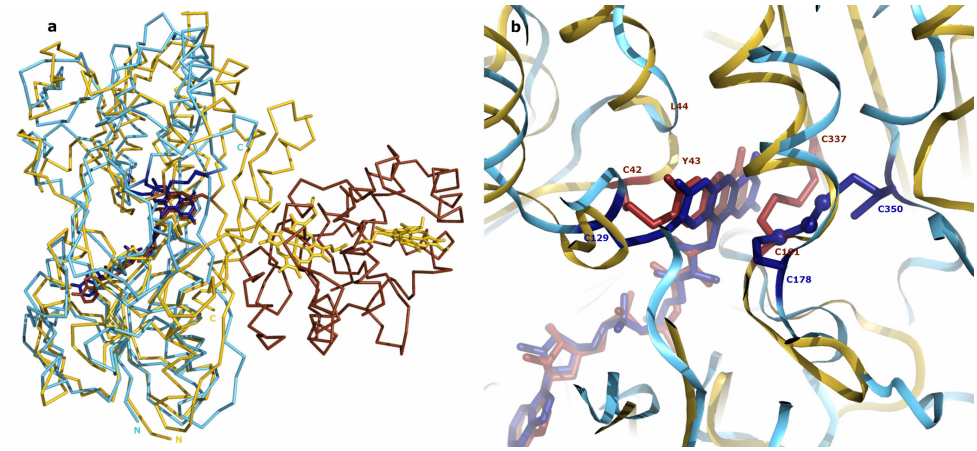


Figure 2.10 – Overall structure superposition between SQR and FCC.
a) Superposition of SQR (light blue), and FCC (yellow), monomeric structures in ribbon representation; FADs are represented in sticks (brown for FCC and dark blue for SQR);
b) Close up view of the active sites; cysteines and residues impairing quinone binding in FCC (see text), are labelled [colour scheme as in a)].

COMPARISON WITH OTHER SQRS

After the release and publication of *A. ambivalens* SQR structure, two other SQR structures were reported: the SQRs from the hyperthermophilic bacterium *Aquifex* (*Aq.*), *aeolicus* [PDB codes 3HYX, 3HYV and 3HYW (Marcia *et al.*, 2009)], and from the acidophilic bacterium *Acidithiobacillus* (*Ac.*), *ferrooxidans* [PDB codes 3KPG, 3KPI and 3KPK (Cherney *et al.*, 2010)].

Aq. aeolicus SQR was determined in the "as-purified" state to 2.30 Å resolution, and in complex with the quinone analogues dUQ and aurachin C, to

2.00 Å and 2.90 Å, respectively. The protein crystallized as an hexamer (a dimer of trimers, Figure 2.11a), and, surprisingly, the biological unit seems to be an homotrimer (Figure 2.11b), in contrast to an homodimer as found for other enzymes of the FDR family.

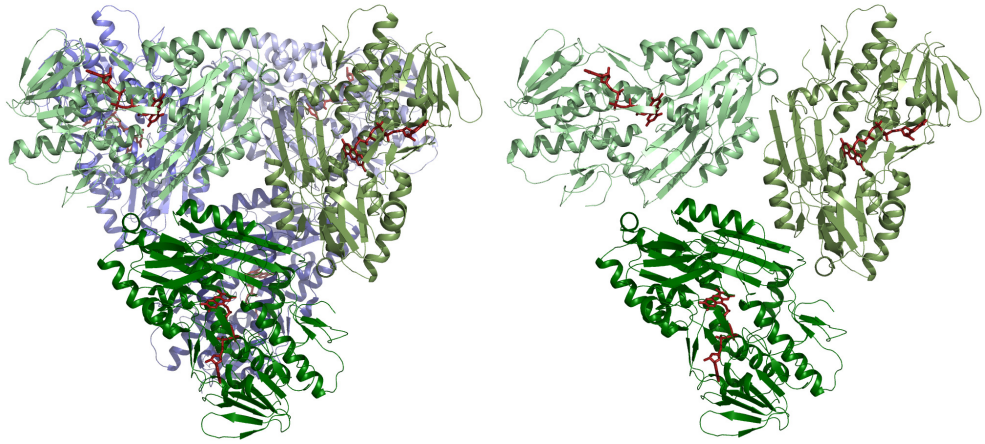


Figure 2.11 – *Aq. aeolicus* SQR asymmetric unit (left), and proposed biological trimer (right). Each chain is coloured in different shades of green and blue and the FAD cofactor is shown in sticks coloured in dark red.

Its overall arrangement is the same as for *A. ambivalens* SQR with the two Rossman-like folds constituting the core of the enzyme. *Aq. aeolicus* SQR is somewhat longer than the SQR from *A. ambivalens* (430 and 409 residues, respectively), but both structures superimpose well (rmsd of ~2.5 Å for *ca.* 300 superimposed C_α atoms), with the main differences being in the C-terminal region: the last 53 a.a. residues of *A. ambivalens* SQR are not visible in the electron density and they are clearly visible in *Aq. aeolicus* SQR. In fact, in *Aq. aeolicus* SQR the C-terminal domain shows a distinctive arrangement compared to other FDR members. It consists of 2 amphipathic helices (K_{376}-N_{395} and P_{400}-K_{412}), mediating the membrane attachment, and an 18 a.a.-long loop that forms oligomerization contacts (Figure 2.12a). In addition, there are two disulfide bridges (C_{280}-C_{422} and C_{419}-C_{430}), that are thought to corroborate the idea that the enzyme is located in the

periplasmic oxidizing environment. These disulfide bridges also help stabilizing the C-terminal portion of the *Aq. aeolicus* enzyme and are not present in *A. ambivalens* SQR, which could explain its disordered C-terminus.

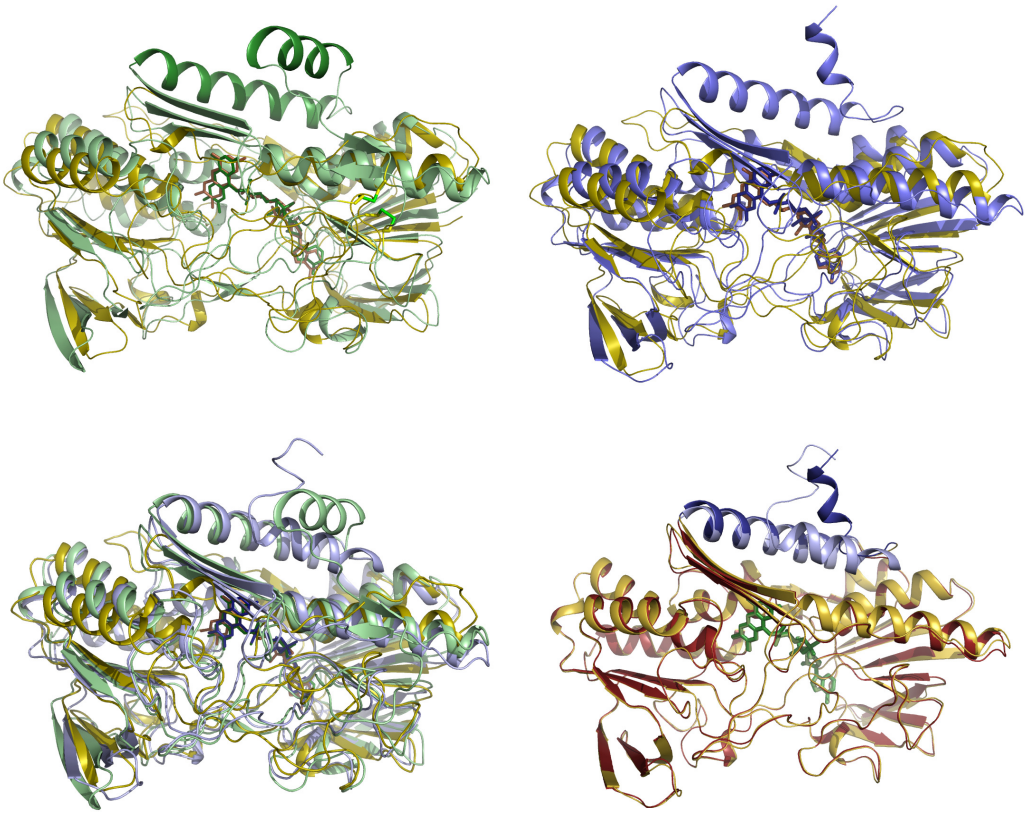


Figure 2.12 – Overall structure superposition between the three known SQRs.
(a) *Aq. aeolicus* (green), and *A. ambivalens* (dark yellow), SQR monomers superposed. FAD cofactors are shown in sticks (dark green and orange, respectively); in *Aq. aeolicus* SQR, the two disulfide bridges are show in yellow sticks with the sulfur atoms coloured in green; the two amphipathic helices are coloured in dark green;
(b) *Ac. ferrooxidans* (blue), and *A. ambivalens* (dark yellow), SQR monomers superposed. FAD cofactors are shown in sticks (dark blue and orange, respectively);
(c) *Aq. aeolicus* (green), *Ac. ferrooxidans* (blue), and *A. ambivalens* (dark yellow), SQR monomers superposed; FAD cofactors are shown in sticks (dark green, dark blue and orange, respectively);
(d) *Ac. ferrooxidans* monomer determined in the presence (gold and dark blue), and absence (red and light blue), of DDM. FAD cofactors are shown in sticks (green and light green, respectively); the two amphipathic helices are coloured in shades of blue.

Ac. ferrooxidans SQR was determined in the native state and in complex with dUQ, both to 2.30 Å resolution. A mutant form of the enzyme, $C_{160}A$, was determined to 2.05 Å and helped clarifying the involvement of the active site cysteines in the reaction mechanism (see below), (Cherney *et al.*, 2010). *Ac. ferrooxidans* SQR shares 27 and 40 % sequence identity with *A. ambivalens* and *Aq. aeolicus* SQRs, respectively, but the overall fold is the same for the three enzymes as expected (the *Ac. ferrooxidans* structure was determined by molecular replacement using *Aq. aeolicus* SQR monomer as search template). The enzyme crystallized in two crystal forms (tetragonal and hexagonal), with the major differences lying in the C-terminal region: the hexagonal crystals were grown in the presence of detergents and the first residues of the second amphipathic helix partially unwind and the helix changes its orientation (Figure 2.12d).

REDOX ACTIVE SITES

The structure of SQR shows two redox active centres: an FAD and an adjacent pair of cysteine residues, C_{178} and C_{350}, located on the *re*-side of the FAD, of which C_{178} is closer to the isoalloxazine ring of FAD (S_γ of C_{178} is only 4.1 and 4.8 Å apart from the N_{10} and N_5 atoms of FAD, respectively), and the S_γ of C_{350} has two alternate conformations. The FAD is covalently bound to the protein through a thioether bond between the S_γ of C_{129} and the 8-methylene group (C8M), of the isoalloxazine ring (Figures 2.9d, 2.10b and 2.14). There are no more contacts between the isoalloxazine ring of FAD and protein a.a. side chains, as also noted for FCC (Chen *et al.*, 1994), and the other two SQRs.

In *Aq. aeolicus* SQR, C_{124} is spatially similar positioned as C_{129} in *A. ambivalens* but the distance between its thiol group and FAD C8M atom is too long to form a thioeter bond (between 3.1-3.3 Å, depending on the chain). There was an extra electron density blob between these two atoms and an S atom was modelled following an anomalous difference Fourier map calculated at 1.907 Å. Thus, *Aq. aeolicus* SQR binds FAD through a labile persulfide and not a stable thioeter bond. In *Ac. ferrooxidans*, the FAD cofactor is not covalently bound to C_{128}, which corresponds to C_{129} and C_{124} in *A. ambivalens* and *Aq. aeolicus*, respectively, even though its S_γ atom

is *ca.* 3.9 Å away from FAD C8M group. This observation was also confirmed by the fact that the FAD cofactor could be detached from the enzyme by extraction with trichloroacetic acid. Moreover, a C_{128}A mutant have been found to be fully active being this the first observation that C_{128} is not involved in the FAD reduction step (see below), (Cherney *et al.*, 2010).

As in all other SQR structures, the most noticeable electrostatic interaction in *A. ambivalens* SQR that may affect the flavin redox potential and the type of semiquinone formed is the dipole of helix α_{11} (residues G_{317}-L_{335}), for which the NH_2 terminus is near N_1 and O_2 of FAD.

The distance between the S_γ atoms of C_{178} and C_{350} in SQR is *ca.* 6.5 Å, larger than the distance of 2.1 Å between same atoms of the FCC. This distance allows a disulfide bridge formation which is not possible in *A. ambivalens* SQR. The distances between the S_γ atoms of the redox active cysteines (C_{178} and C_{350} in *A. ambivalens*, C_{156} and C_{347} in *Aq. aeolicus,* and C_{160} and C_{356} in *Ac. ferrooxidans*), are *ca.* 6.5, 8.2 and 6.4 Å, respectively, too long for covalent attachment. The S_γ atom of C_{350} in *A. ambivalens* SQR was modelled with two conformations with approximately half occupancy each in both monomers. Quite remarkably, initial experimental and further difference Fourier maps consistently revealed continuous electron density between both cysteine residues, which was interpreted as a three sulfur atoms chain (denominated S_1, S_2 and S_3), based on geometrical and chemical considerations and further corroborated by anomalous data collected at a longer wavelength (Figures 2.9d, 2.10b and 2.14).

The refined occupancies for the atoms in this trisulfide bridge show a trend of decreasing occupancy from the C_{178}-bound S1 to the C_{350}-bound S3: 83, 88 and 47 % in monomer A, and 84, 81 and 42 % in monomer B. This decrease in occupancy relates well to the refined occupancies of the alternate-conformations (B), of C_{350} to which S_3 is bound. As in *A. ambivalens* SQR, the electron density connecting the redox active cysteines in *Aq. aeolicus* and *Ac. ferrooxidans* also extends and form adducts with additional sulfur atoms (see Table 2.3 for details). For example, in the tetragonal form of *Ac. ferrooxidans* SQR, C_{356} was modelled as a

disulfide that is attached to a trisulfide at C_{160}; the occupancy of the sulfur atoms refined to an average of 58 % (Cherney *et al.*, 2010).

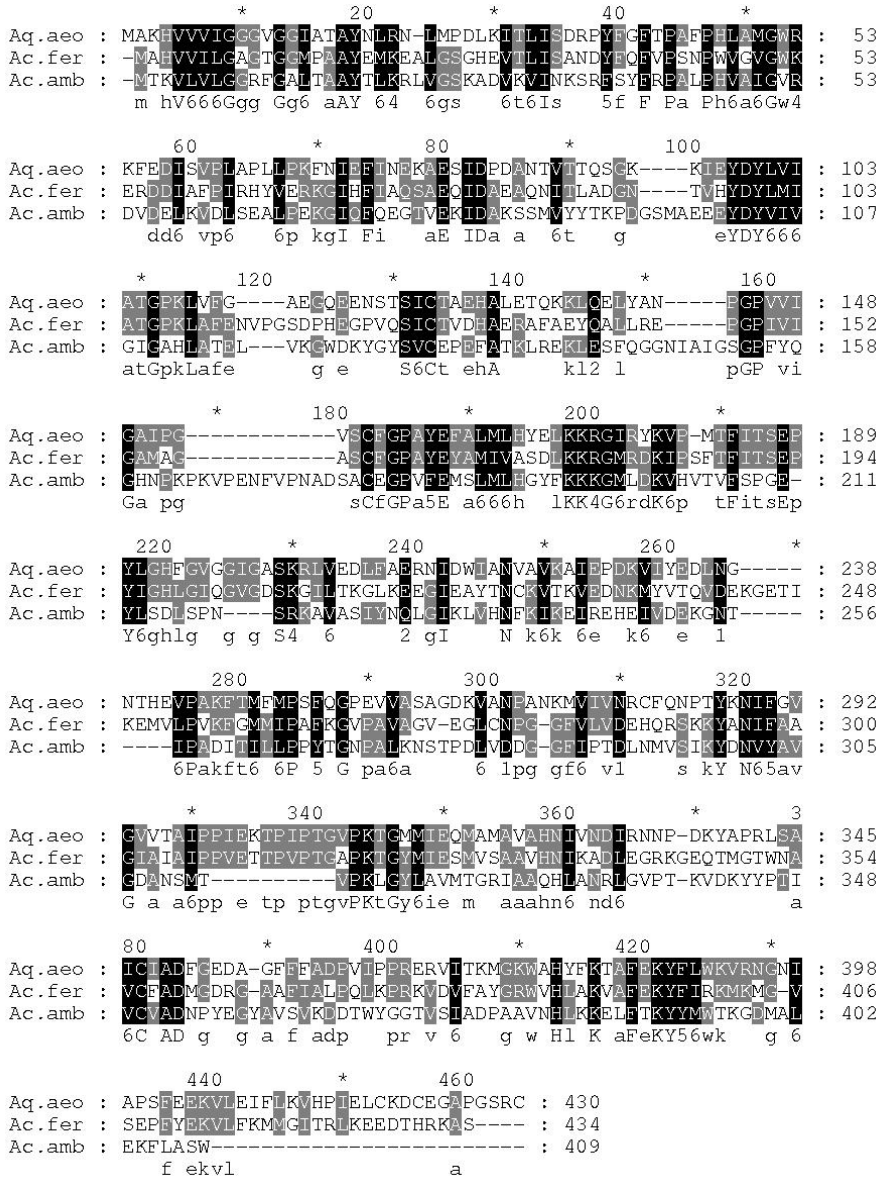


Figure 2.13 - Amino acid sequence alignment of *Acidianus ambivalens*, *Acidithiobacillus ferrooxidans* and *Aquifez aeolicus* SQRs.

The sequences were obtained from the X-ray structures deposited in the Protein Data Bank, are identified as Ac.amb, Ac.fer and Aq.aeo, respectively, and were aligned using the program ClustalW (Chenna *et al.*, 2003).

There is an extra spherical blob of electron density in both $2F_o$-F_c and positive F_o-F_c maps for the two chains of *A. ambivalens* SQR (still visible for chain B in the F_o-F_c map contoured at 6σ), located just above the A conformation of C_{350}. This extra density could not be properly modelled, although we think that this blob of density could correspond to an extra sulfur atom, so that this state would represent an intermediate step towards the polysulfide reaction product (see below). The same feature was observed in the *Aq. aeolicus* and *Ac. ferrooxidans* SQR structures, except for the C_{160}A variant of *Ac. ferrooxidans* SQR, but these were tentatively modelled as sulfide based on occupancy and fit to the spherical positive F_o-F_c electron density peaks. In the C_{160}A mutant of *Ac. ferrooxidans* SQR, there are no additional electron density peaks besides that for the methyl group in the vicinity of A160 C_α atom. Some positive peaks appeared close to the Sγ atom of C356, modelled as a disulfide, but no polysulfide chains were found. Since no catalytic activity was detected for this mutant, it was concluded that the sulfide substrate can interact with C_{356} directly in the absence of the redox active disulfide bridge but the reaction stops there (Cherney *et al.*, 2010).

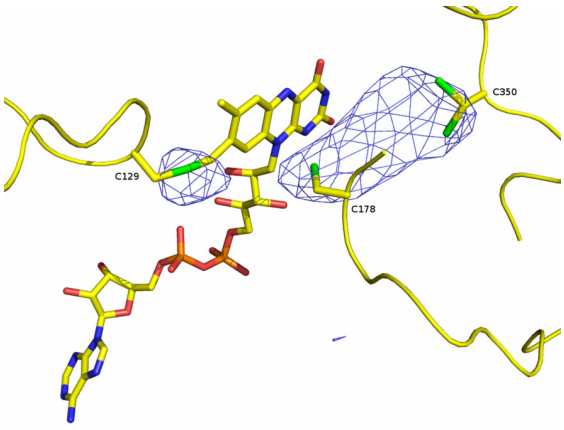


Figure 2.14 – Anomalous difference Fourier map showing the continuous electron density bridging the two redox active site cysteines in *Aciaianus ambivalens*.
The map was calculated with a dataset collected at 1.711 Å wavelength and is contoured at 3.5σ.

Interestingly, the structure of SQR also revealed the presence of two aspartic residues (D_{215} and D_{353}), in the vicinity of C_{350} (Figure 2.9c), with a water molecule closely H-bonded to D_{215} (~2.6 Å), and further away from D_{353} (~3.7 Å), the latter

being highly conserved among the SQRs from Group I (Theissen *et al.*, 2003). The same water molecule is *ca.* 5 Å away from the S_γ atoms of C_{350} and 5.8 Å from the S_2 atom of the trisulfide molecule.

Table 2.3 – Different sulfur adducts found in the three SQR structures known to date

Microorganism	PDB code	C_{XXX} - FAD	C_{XXX} - "S adduct" - C_{XXX} [a]	
A. ambivalens	3H8L	C_{129} - FAD	C_{178} - S_3 - C_{350}	
Aq. aeolicus	3HYV	C_{124} - S - FAD	C_{156} - S - S_8 [b]	C_{347}
			C_{156} - S	C_{347}
			C_{156} - S_7	C_{347}
			C_{156} - S_2	C_{347}
	3HYW	C_{124} - S - FAD	C_{156} - S - S_8 [b]	C_{347}
			C_{156} - S_2	S_4 - C_{347}
			C_{156} - S_7	C_{347}
			C_{156} - S_3	S_2 - C_{347}
	3HYX	C_{124} - S - FAD	C_{156} - S_8 [c]	C_{347}
			C_{156} - S - S_8	C_{347}
Ac. ferrooxidans	3KPG	C_{128}	C_{160} - S_5 - C_{356}	S - C_{356}
	3KPI	C_{128}	C_{160} - S_4 - C_{356}	C_{160} - S
			C_{160} - S_4 - C_{356}	
	3KPK	C_{128}	A_{160}	S - C_{356}

[a] Several adducts per PDB code refers to the different species found in each chain of the asymmetric unit; [b] S_8 in a globular conformation; [c] S_8 in a linear conformation.

PROTEIN CHANNELS AND QUINONE BINDING

The molecular surface representation of SQR showed a channel in the protein, at the *re*-side of FAD, which could be the sulfide entry and/or polysulfide exit pathway (Figure 2.15a). This channel is located between the FAD and the

second Rossmann fold domain, has a minimal diameter of ~5 Å and gives access from the solvent to the catalytic cysteine residues, whereby the most exposed one (C_{350}), is approximately 10 Å buried in the protein interior.

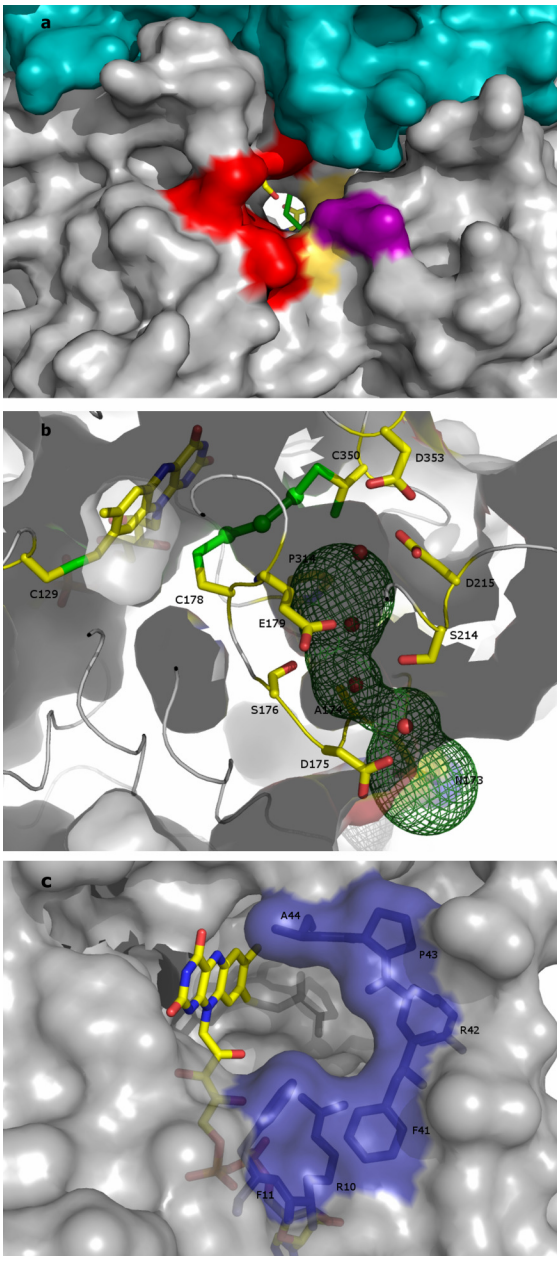


Figure 2.15 – Observed channels in the SQR enzyme from *A. ambivalens*.
a) Molecular surface showing the putative substrate/product channel, with negatively charged residues in red, polar residues in purple and hydrophobic residues in yellow. S_{214} and the sulfur atoms between the redox active cysteine residues are drawn in sticks. Part of chain B is in dark cyan;
b) surface representation of the channel leading access to the redox active site drawn as a green mesh with relevant residues and water molecules displayed;
c) putative quinone binding pocket in surface representation with the relevant hydrophobic residues and FAD displayed in sticks.

The channel is formed mainly by polar and acidic residues, namely N_{173}, S_{176}, S_{214}, D_{175}, D_{215} and E_{179}, but also includes P_{314} and A_{174}, and is filled with several ordered water molecules (Figure 2.15b). Note that the substrate is most probably the neutral, hydrophilic molecule H_2S, taking into account its pK_{a1} of 6.8, and the product a chain of polysulfide or sulfane. It is worth mentioning that the optimum pH for the enzyme activity is around 6.

A possible pathway for proton transfer in *A. ambivalens* SQR, is also located at the *re*-side of FAD and could involve D_{307} and K_{315}, two conserved residues that form a salt bridge and are connected to the surface through a network of water molecules and some protonable groups (H_{160} and K_{163}). This channel is delimited by the extra loop characteristic of *A. ambivalens* SQR and of some other archaeal enzymes from Group I. In FCC, none of these channels are observed most probably because the active site is readily accessible from the solvent.

In *Aq. aeolicus* SQR (Marcia *et al.*, 2009), it was proposed a similar channel to that of *A. ambivalens* SQR for sulfide access to the oxidation site: delimited by the *re*-side of FAD and by the loop S_{155}-P_{159}, the channel, characterized by a positive electrostatic surface potential, runs from the bulk solvent to a cavity that extends for ~15 Å in parallel and 4 Å in perpendicular to the isoalloxazine plane. Similar to *A. ambivalens* SQR, solvent molecules occupy the channel at consecutive sites from the solvent to S_{155} (A_{177} in *A. ambivalens* SQR numbering), and are proposed to mimic the trace of the substrate, H_2S. At the end of the channel sits V_{294}, essential for *Aq. aeolicus* SQR activity (Griesbeck *et al.*, 2002), and strictly conserved in group I SQRs.

The channel proposed for sulfide access in *Ac. ferrooxidans* SQR (Cherney *et al.*, 2010), is somewhat different from those in *Aq. aeolicus* and *A. ambivalens* SQRs. In fact, the channel proposed for sulfide access to the redox active cysteines is "closed" in the structure near C_{128} by the imidazole ring of H_{132} (all *Ac. ferrooxidans* numbering). Hence, the sulfide species bound in the active site pocket has no direct access to this channel or any other close to it. Besides this channel, another three are visible in *Ac. ferrooxidans* SQR which are occupied by FAD (on one side of the protein), and DDM and dUQ (on the "membrane attachement" side of

the protein). One of these, occupied by DDM, is hydrophobic and can be opened to the internal part of the membrane serving as a product release pathway, as similarly proposed for the *Aq. aeolicus* SQR (Marcia *et al.*, 2009). This channel is not visible in *A. ambivalens* SQR since its last 53 a.a. residues are disordered in the structure.

A possible quinone binding site in *A. ambivalens* SQR is the cavity on the *si*-side of FAD (Figure 2.15c). The cavity is delineated by the backbone of residues F_{41} to A_{44} (C_α of A_{44} is only *ca.* 4.7 Å away from N_5 of the isoalloxazine ring), and is flanked by residues R_{10}, F_{11} and F_{41}; it provides a hydrophobic environment suitable for the quinone head to come close and interact with the isoalloxazine ring of FAD. This cavity, accessible from the protein surface, is located just below the proposed in-plane amphipathic helix displaying the shortest path to the membrane plane (see below). Worth mentioning, an additional oblong blob of electron density is visible in the $2F_o$-F_c and positive F_o-F_c maps inside the cavity (closest distance is *ca.* 4.8 Å to N5 of FAD). Attempts to fit and refine caldariella quinone or a DDM molecule were not satisfactory. The quinone head was pushed away from the plane parallel to the isoalloxazine ring and the detergent molecule's aliphatic tail refined only to low occupancy, so nothing was modelled for this extra blob of density. Noteworthy, this extra blob of density sits in a similar location where the quinone molecules were modelled in the *Aq. aeolicus* and *Ac. ferrooxidans* SQR structures (Figure 2.16), corroborating our proposal.

In *Aq. aeolicus* SQR, dUQ was modelled in a $2F_o$-F_c simulated annealed map at 0.8σ and refined to 60 % occupancy (PDB code 3HYW), (Marcia *et al.*, 2009). The aromatic ring of the quinone molecule is stacked between the benzene ring of F_{385} and the side-chain of I_{346}. These residues are strictly conserved in SQRs, but not in FCC, and correspond to F_{389} and V_{349} in *A. ambivalens* SQR, and F_{394} and V_{355} in *Ac. ferrooxidans* SQR. It is interesting to point here that V_{349} in *A. ambivalens* SQR was annotated as an M residue in the sequence databases and it was the X-ray structure solved by us that prompted for this discrepancy (probably due to an error in the sequencing or annotation; M is coded by and AUG codon and V by GUG). Moreover, another potent inhibitor of *A. aeolicus* SQR (Nübel *et al.*, 2000), aurachin C, was found in the inhibitor-bound state structure (PDB code 3HYX),

(Marcia *et al.*, 2009), in a similar position as dUQ confirming the quinone binding site.

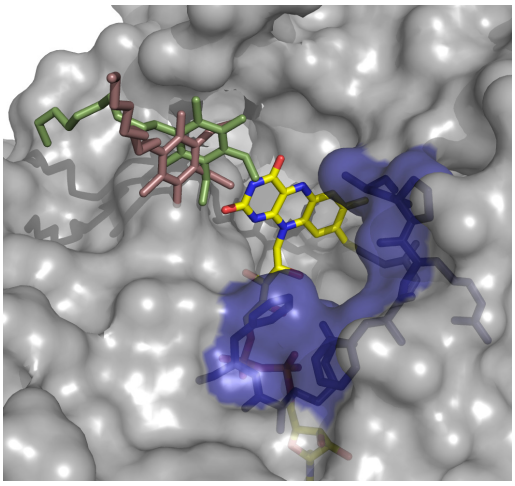


Figure 2.16 – Quinone binding pocket of *A. ambivalens* SQR in surface representation with the relevant hydrophobic residues (blue), and FAD displayed in sticks.
dUQ from *A. aeolicus* (PDB code 3HYW, dark pink), and *Ac. ferrooxidans* (PDB code 3KPG, dark green), are shown in sticks. Residues shown in blue are labelled in Figure 2.15c

Ac. ferrooxidans SQR was cocrystallized with dUQ (PDB code 3KPG), (Cherney *et al.*, 2010), and, although in the same spatial position as the dUQ molecule in *Aq. aeolicus* SQR, there are slight differences in the binding. The aliphatic tail of dUQ in *Ac. ferrooxidans* SQR is "curled" into the enzyme pointing inside the protein, while in *A. aeolicus* points towards the solvent, and the benzoquinone head is closer to the isoalloxazine ring of FAD in *Ac. ferrooxidans* SQR than in *A. aeolicus* (the distance between O4 atom of the quinone and the O2 atom of FAD is less than 3 Å).

MEMBRANE ATTACHMENT

SQR was isolated from the membranes of *A. ambivalens*. The structure herein described does not show structural elements that could be involved in membrane binding. However, the model does not contain the last 53 C-terminal residues for which no electron density is visible, even though crystals of the complete form of the protein were obtained. Analysis of the SQR sequence using TMHMM (http://www.cbs.dtu.dk/services/TMHMM/) and SOSUI (Hirokawa *et*

al., 1998) suggested the inexistence of transmembrane helices, but the AmphipaSeek server (Sapay *et al.*, 2006) predicts an amphipathic helix in-plane membrane anchor within the last 25 a.a. residues. Similar results were obtained with the sequences of other SQRs. This kind of membrane anchoring is supposed to occur via a helix with a large hydrophobic region on one side and a hydrophilic one on the other (Figure 2.17), and has also been observed in other proteins (Nina *et al.*, 2000, Penin *et al.*, 2004). This putative amphipathic helix is most probably located on the *si*-side of the flavin (back of the molecule in relation to Figure 2.9a).

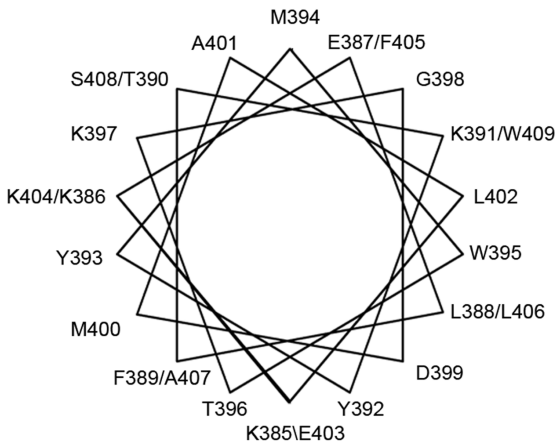


Figure 2.17 – Schematic representation of the putative C-terminal amphipathic helix proposed to be involved in membrane attachment (residues 385-409).

This hypothesis was corroborated by *Ac. ferrooxidans* and *A. aeolicus* SQR structures, where two amphipathic helices mediate membrane attachment. In *A. aeolicus* SQR these two helices (residues 376-412), are also defining the quinone entry channel. It is worth mentioning that *A. aeolicus* SQR sequence is longer than *A. ambivalens* SQR (430 and 409 a.a. residues, respectively), and that residues 413-430 are mediating trimer formation (*A. ambivalens* and *Ac. ferrooxidans*, as well as other FDR enzymes, are proposed to act as dimers). This C-terminal portion of the enzyme is also stabilized by two disulfide bridges (not observed in any other structure), probably needed for rigidity and a stable and site-specific membrane interaction (Marcia *et al.*, 2009).

Cherney and coworkers (Cherney *et al.*, 2010), also propose two amphipathic helices for membrane attachment of the *Ac. ferrooxidans* SQR but the flexibility observed in the last of these two helices (residues 407-427), between the tetragonal (purified in the absence of DDM), and hexagonal forms (purified in the presence of DDM), as well as between different crystals of the same crystal form, make them postulate that this flexible part of the SQR molecule keeps mobile in the membrane environment. This observation is more in accordance to what is observed in *A. ambivalens* SQR.

POSSIBLE MECHANISM FOR SULFIDE OXIDATION

Based on the structure here described, a possible mechanism for the oxidation of sulfide in *A. ambivalens* SQR is hypothesized taking into account: i), the involvement of the two cysteine residues C_{178} and C_{350}; ii), the presence of the three-sulfur bridge between the two cysteine residues; and iii), the possible role of acid-base groups either to increase the nucleophilicity of the hydrogen sulfide molecule or to accept, at least transiently, protons. The oxidative part of the reaction occurs on the *re*-side of the flavin, while the reductive part on the *si*-side, where the primary electron acceptor, the quinone, is proposed to bind. At this stage it would be totally speculative to address the reductive cycle, and thus only the oxidation of hydrogen sulfide will be discussed. Also, a three-dimensional representation for the *R. capsulatus* enzyme was modelled based on the *A. ambivalens* SQR structure, which reveals that what is here proposed for this enzyme is also possible to occur in the *R. capsulatus* SQR, including the putative location of the substrate and proton channels, and the quinone binding cavity. A comparison with the mechanisms proposed for the *Aq. aeolicus* (Marcia *et al.*, 2009), and *Ac. ferrooxidans* (Cherney *et al.*, 2010), SQRs, respectively, is also performed. The mechanisms are essentially the same the difference being the assigned nucleophile in each that is proposed to attack the flavin moiety (see below).

Only two cysteine residues (C_{178} and C_{350} in *A. ambivalens*), are needed for catalysis: the third cysteine residue (C_{129}), is covalently attached to the flavin isoalloxazine ring in *A. ambivalens* SQR and FCC (Chen *et al.*, 1994); it is too far away

from the catalytic site for sulfide oxidation, and, as initially observed by Theissen and coworkers (Theissen *et al.*, 2003) and now corroborated by the analysis of more SQR a.a. sequences, only two cysteine residues are strictly conserved among all SQRs. The *R. capsulatus* SQR model shows that C_{127} (C_{129} in *A. ambivalens* SQR), is certainly involved in the covalent attachment of the FAD. It had also already been mentioned by Griesbeck *et al.* (Griesbeck *et al.*, 2002) that eventually only two cysteine residues would be necessary for the catalytic reaction. The fact that a C_{128}A mutant of *Ac. ferrooxidans* SQR (corresponding to C_{129} in *A. ambivalens* SQR), is fully active corroborates the hypothesis that C_{128} is not involved in the FAD reduction step.

In the "as isolated" *A. ambivalens* SQR crystal structure, the two reactive cysteine residues are linked with three additional sulfurs in between (Figures 2.9c, 2.9d and 2.13; similarly to *Aq. aeolicus* and *Ac. ferrooxidans* SQR structures where different sulfur adducts are visible between the cysteines, Table 2.3). Although it cannot be ascertained at the present stage whether this is an intermediate state of the enzyme, it seems reasonable to propose it, as in most other FDRs, namely in FCC, the catalytic cysteine residues form a disulfide bridge in the enzyme's oxidized state (Chen *et al.*, 1994, Schutz *et al.*, 1999, Giggenbach, 1972). Nevertheless, the possible mechanism described below is also compatible if the state observed in the structure is the resting one. In the present structure, the C_{α}´s of C_{178} and C_{350} are *ca.* 9.3 Å apart; however, those residues are located in loop segments, which can accommodate slight conformational changes so that in the oxidized state a simple disulfide bridge may be formed (Figure 2.18a). The first step of the reaction will be the nucleophilic attack (Figure 2.18b), of the incoming substrate H_2S to the disulfide bridge, being the sulfur incorporated in the nascent polysulfide chain, forming a persulfide at C_{350} and a thiolate anion at C_{178} (Figure 2.18c).

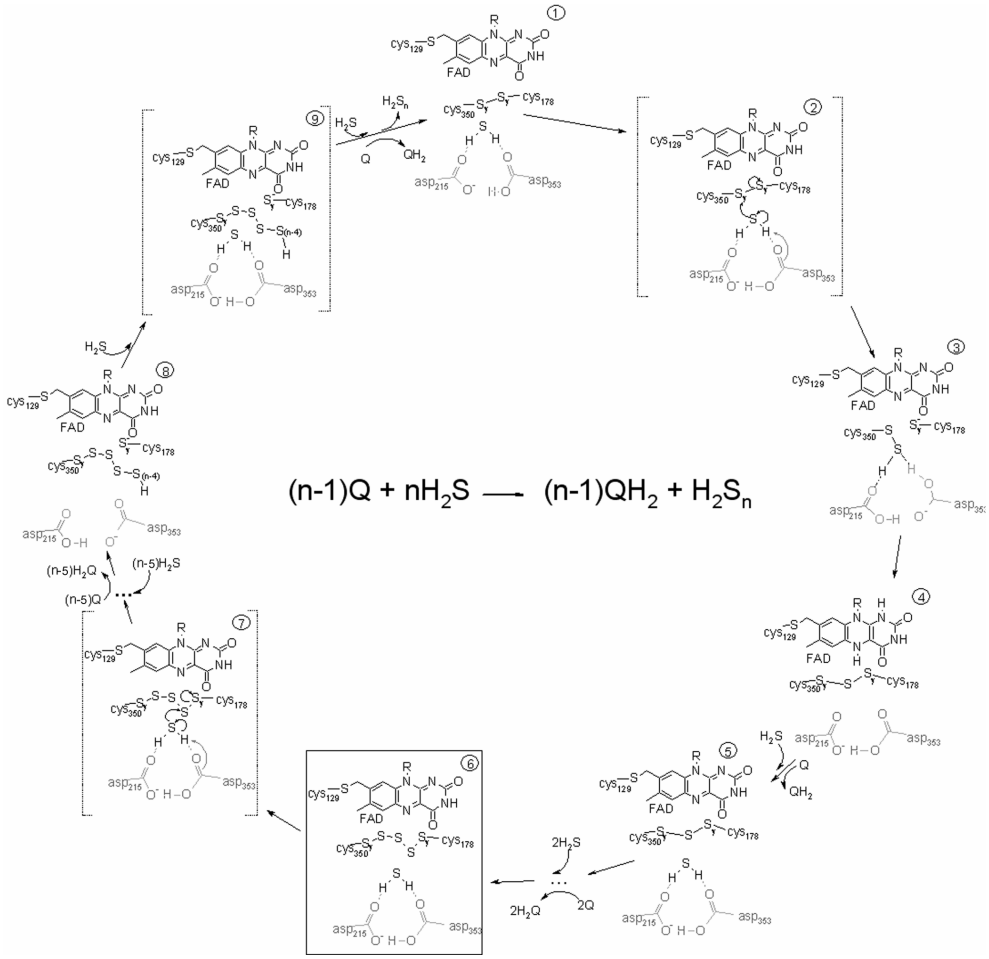


FIGURE 2.18 – Schematic representation of the possible mechanism of *A. ambivalens* SQR reductive half-reaction.

(1) represents a possible native state of the enzyme where the FAD, the two redox active cysteine residues and the two aspartate residues are represented. The H_2S molecule is stabilized by hydrogen bridges with the oxygens from the aspartyls. The attack of an oxygen atom of one of the aspartic residues to a H_2S molecule (2), initiates a cascade of nucleophilic attacks that lead to the break of the disulfide bridge and to the formation of a persulfide at C_{350} and a charge transfer complex with the thiolate of C_{178} as the donor and the oxidized FAD as the acceptor (3 and 8). Then, an active site base abstracts a proton from the persulfide at C_{350}, a trisulfide bridge is established (4), and the flavin may be reduced. State 5), is formed by the arrival of a new sulfide molecule and after electron transfer from the FAD to the quinone. States equivalent to states 1 to 5 are repeated until stereochemical constraints hamper the incorporation of another sulfur atom between the cysteine residues. 6), represents the state observed in the structure. From 7), to 9), (n-4), nucleophilic attacks of new sulfide molecules to until stereochemical reasons promote the release of the polysulfide molecule from the enzyme and the disulfide bridge is restored (1). Square brackets delimit intermediate stages of the mechanism and a black square delimits the state observed in the structure herein described.

The nucleophilicity of the hydrogen sulfide molecule may be increased by the involvement of D_{215} (which is part of the protein channel suggested for substrate conduction), and/or D_{353}; this last aspartic residue is highly conserved among SQRs, but D_{215} is not. However, these residues may be functionally substituted by other a.a. residues: as seen from the a.a. sequence alignments, in SQRs from Group I, one of these aspartic residues (D_{215}), may be substituted by a histidine residue (H_{196}, *R. capsulatus* numbering). Mutants of this residue (Griesbeck *et al.*, 2002) were shown to retain only approximately 40 % of the activity of the wild type enzyme, being the affinity for sulfide decreased, while no change was observed in the K_m for the quinone. In the *R. capsulatus* SQR homology model the imidazole ring of this histidine residue occupies the same spatial position of the carboxylate group from the aspartic residue. This indicates that although not essential for activity, this residue could be involved in SQRs catalytic cycle.

A charge transfer complex between the C_{178} thiolate and FAD_{ox} may occur as proposed before (Griesbeck *et al.*, 2002) (Figure 2.18c). Subsequently, an active site base is suggested to abstract a proton from the persulfide at C_{350} to form a more nucleophilic anion. The nature of this base is still unknown and although it was suggested to be a glutamate residue (E_{165} in *R. capsulatus* which corresponds to E_{184} in *A. ambivalens*), (Griesbeck *et al.*, 2002), the present structure shows that it is not within hydrogen bonding distance of the cysteine residues (S_γ of C_{178} is *ca.* 8.4 Å apart from $O_{\varepsilon 1}$ of E_{184}).

For a trisulfide bridge to be formed, the transfer of two electrons to FAD occurs, possibly through the establishment of covalent adducts with C_{178}. A flavin hydroquinone may be formed (Figure 2.18d), which will be subsequently oxidized by the quinone (Figure 2.18e). Incorporation of two more sulfur atoms to the nascent polysulfide gives rise to an intermediate state with three sulfur atoms between the cysteine residues as observed in the structure here described (Figure 2.18f), and stereochemical constraints will determine the maximum number of sulfur atoms to be incorporated: once it is achieved, two consecutive nucleophilic attacks of two new hydrogen sulfide molecules on the trisulfide, leads to the release of a polysulfide

molecule from the enzyme and the initial disulfide bridge is restored (Figures 2.18f-i).

This same mechanism was also proposed for *Aq. aeolicus* SQR (Marcia *et al.*, 2009), where the formation of a C_{156}-S-S^- disulfide occurs and attacks the C4A atom of FAD. However, Marcia and coworkers, also proposed an alternative mechanism that involves the formation of a C_{124}-S-S^- disulfide that attacks the C8M atom of FAD. Electron transfer would occur during this step leading to the formation of a transient C_{124}-S-S^+. The trisulphide bridge (like in *A. ambivalens* SQR), C_{124}-S-S-S-C_{156} would form after the recombination of the above mentioned cation with the C_{156} thiolate. The role of C_{347} is to take over the growing polysulfide chain and, as a result, to recover the C_{156} thiolate. As mentioned above, the involvement of C_{124} in the oxidation mechanism of sulfide is not possible in FCC and *A. ambivalens* SQR (C_{42} and C_{129}, respectively, are covalently bound to C8M of FAD), as well as in *Ac. ferrooxidans* SQR: although C_{128} is not covalently bound to FAD, a C_{128}A mutant is still fully active (Cherney *et al.*, 2010).

Although, these two mechanisms proposed for *A. ambivalens* and *Aq. aeolicus* SQRs would be perfectly applicable to *Ac. ferrooxidans* SQR, Cherney and coworkers suggest yet another mechanism with the unique feature of the S^- atom of the C_{356}-S-S^- disulfide being the nucleophile that attacks the C4A atom of FAD in the beginning of each cycle (Cherney *et al.*, 2010). The hypothesis is supported by the fact that the distance of the S^- atom of C_{356} disulfide to the C4A atom of FAD is *ca.* 3.2-3.5 Å, much closer than the distances between the other potential nucleophilic sulfur atoms of C_{160} thiol or the C_{160}-S-S^- disulfide and the C4A atom of FAD. Those distances would make such a nucleophilic attack very difficult turning C_{160} in a very unlikely (although, according to authors, not impossible), candidate for the role of the nucleophile (Cherney *et al.*, 2010). Moreover, the fact that an octasulfur ring is found in C_{156} of *Aq. aeolicus* SQR (C_{160} in *Ac. ferrooxidans*), seems to support the proposed mechanism where the polysulfide grows on C_{160} (Cherney *et al.*, 2010).

The proposal of any complex mechanism for sulfide oxidation is hampered by the fact that the reaction product of sulfide oxidation is not known. If what is observed in the structure of the as purified *A. ambivalens* SQR is an intermediate,

then the shortest polysulfide/sulfane product should be S_5^{2-} or S_5H_2. In fact, it has been shown that at pH 6 the most stable form of polysulfide comprises four or five sulfur atoms (Giggenbach, 1972), which can be further substrates to sulfur oxygenase reductase (SOR). However, bigger sulfur adducts (S_7^{2-}, and linear and globular S_8 species), observed in *Aq. aeolicus* and *Ac. ferrooxidans* SQR structures makes it uncertain.

SQR IN *A. AMBIVALENS*, THE LINK BETWEEN SULFUR METABOLISM AND THE AEROBIC RESPIRATORY CHAIN

In aerobically grown *A. ambivalens* the initial step of sulfur metabolism is mediated by a soluble SOR. This enzyme catalyzes the disproportionation of S^0 to sulfite, sulfide and possibly thiosulfate. In this way, not all energy from sulfur is used, since part of it is lost in sulfur reduction to sulfide. However, this energy is recovered by the presence of SQR, which oxidizes sulfide, (producing polysulfides/sulfanes that are again substrates for SOR), and reduces quinones. Thus, this process allows to extract maximum energy from S^0, feeding electrons to the respiratory chain.

The products of sulfur oxidation, sulfite and thiosulfate, are also further metabolized. There are evidences for the existence of a sulfite:acceptor oxidoreductase (SAOR), in the membranes of *A. ambivalens*, although the enzyme has never been isolated (Kletzin *et al.*, 2004), and a membrane bound thiosulfate:quinone oxidoreductase (TQO), has been purified and characterized (Muller *et al.*, 2004). In this way, SOR and SQR create an energetic spiral, allowing to get the maximum of energy from sulfur compounds (Figure 2.19), and, ultimately, the SOR/SQR/TQO/SAOR enzymes enable a full utilization of sulfur for energy conservation in *A. ambivalens* with three direct links of sulfur oxidation to quinone reduction, as the main electron carrier to the quinol:oxygen oxidoreductase, the aa_3 enzyme.

The location of SQRs in respect to either side of the membrane is still unclear. SQR a.a. sequences do not show typical signal peptides for translocation across the cytoplasmic membrane. However, *R. capsulatus* SQR is suggested to be

attached to the periplasmic surface of the cytoplasmic membrane (Schutz *et al.*, 1999). Marcia and coworkers propose that the existence of two disulfide bridges in the C-terminal portion of *Aq. aeolicus* SQR point to the presence of the enzyme in the periplasmic oxidizing environment (Marcia *et al.*, 2009). Moreover, being embedded *ca.* 12 Å in the membrane, and being *Aq. aeolicus* known to accumulate sulfur globules in the cytoplasm, SQR would become an efficient way of cellular uptake of sulfur (Marcia *et al.*, 2009).

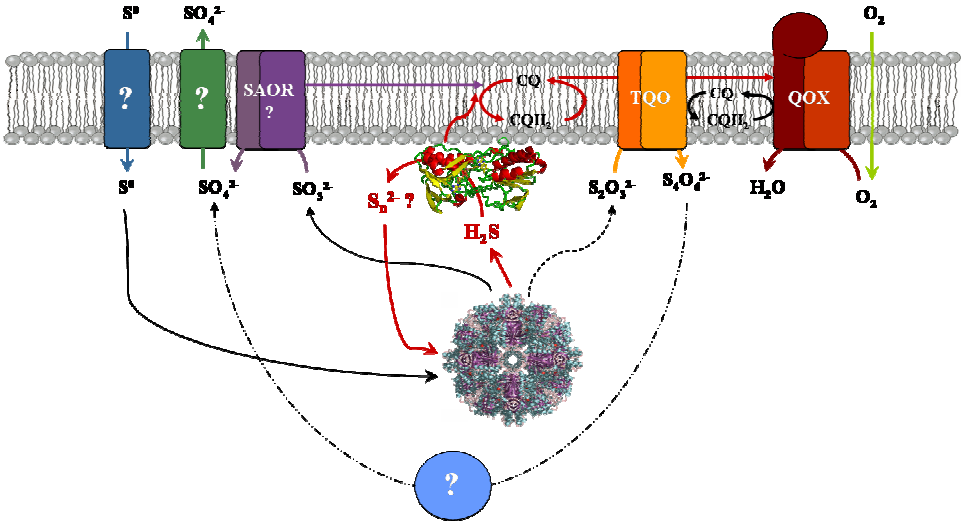


Figure 2.19 – Schematic representation of the sulfur energetic metabolism in *A. ambivalens*. TQO, thiosulfate:quinone oxidoreductase; SAOR, sulfite acceptor oxidoreductase; SQR, sulfide:quinone oxidoreductase is shown attached to the membrane in cartoon representation; QOX, quinol:O_2 oxidoreductase; SOR, sulfur oxygenase reductase is shown in cartoon feeding SQR with H_2S and accepting S_n^{2-}.

ACKNOWLEDGEMENTS

We would like to thank Pedro Matias and Carlos Frazão for help with synchrotron data collection, Bruno Victor for the modelling of *R. capsulatus* SQR, and Cláudio M. Soares for helpful discussions. We also acknowledge Elizabete Pires and Gonçalo da Costa from the Mass Spectrometry service at Instituto de Tecnologia Química e Biológica, Universidade Nova de Lisboa, Oeiras, Portugal.

2.8. REFERENCES

Abrahams, J. P. & A. G. Leslie, (1996) Methods used in the structure determination of bovine mitochondrial F1 ATPase. *Acta Crystallogr D Biol Crystallogr* **52**: 30-42.

Anbar, A. D. & A. H. Knoll, (2002) Proterozoic Ocean Chemistry and Evolution: A Bioinorganic Bridge? *Science* **297**: 1137-1142.

Argyrou, A. & J. S. Blanchard, (2004) Flavoprotein disulfide reductases: advances in chemistry and function. *Prog Nucleic Acid Res Mol Biol* **78**: 89-142.

Arieli, B., E. Padan & Y. Shahak, (1991) Sulfide-induced sulfide-quinone reductase activity in thylakoids of *Oscillatoria limnetica*. *J Biol Chem* **266**: 104-111.

Arieli, B., Y. Shahak, D. Taglicht, G. Hauska & E. Padan, (1994) Purification and characterization of sulfide-quinone reductase, a novel enzyme driving anoxygenic photosynthesis in *Oscillatoria limnetica*. *J Biol Chem* **269**: 5705-5711.

Bandeiras, T. M., C. Salgueiro, A. Kletzin, C. M. Gomes & M. Teixeira, (2002) *Acidianus ambivalens* type-II NADH dehydrogenase: genetic characterisation and identification of the flavin moiety as FMN. *FEBS Lett* **531**: 273-277.

Bricogne, G., E. Blanc, M. Brandl, C. Flensburg, P. Keller, W. Paciorek, P. Roversi, O. S. Smart, C. Vonrhein & T. Womack, (2008) BUSTER-TNT 2.5.1 and autoBUSTER 1.3.1. In: Global Phasing Ltd. Cambridge, pp.

Brito, J. A., T. M. Bandeiras, M. Teixeira, C. Vonrhein & M. Archer, (2006) Crystallisation and preliminary structure determination of a NADH: quinone oxidoreductase from the extremophile *Acidianus ambivalens*. *Biochim Biophys Acta* **1764**: 842-845.

Brune, D. C. & H. G. Trüper, (1986) Noncyclic electron transport in chromatophores from photolithotrophically grown *Rhodobacter sulfidophilus*. *Archives of Microbiology* **145**: 295-301.

Chen, Z. W., M. Koh, G. Van Driessche, J. J. Van Beeumen, R. G. Bartsch, T. E. Meyer, M. A. Cusanovich & F. S. Mathews, (1994) The structure of flavocytochrome c sulfide dehydrogenase from a purple phototrophic bacterium. *Science* **266**: 430-432.

Chenna, R., H. Sugawara, T. Koike, R. Lopez, T. J. Gibson, D. G. Higgins & J. D. Thompson, (2003) Multiple sequence alignment with the Clustal series of programs. *Nucleic Acids Res* **31**: 3497-3500.

Cherney, M. M., Y. Zhang, M. Solomonson, J. H. Weiner & M. N. James, (2010) Crystal structure of sulfide:quinone oxidoreductase from *Acidithiobacillus ferrooxidans*: insights into sulfidotrophic respiration and detoxification. *J Mol Biol* **398**: 292-305.

Collaborative Computational Project, N., (1994) The CCP4 suite: programs for protein crystallography. *Acta Crystallogr D Biol Crystallogr* **50**: 760-763.

Cunningham, W. P. & B. W. Saigo, (1999) *Environmental Science: A Global Concern*. McGraw Hill, New York.

DeLano, W., (2002) The PyMOL Molecular Graphics System. In. San Carlos, CA: DeLano Scientific, pp.

Dobbin, L., (1931) *The Collected Papers of Carl Wilhem Scheele*. G. Bell & sons LTD.,York House, London.

Drummond A.J., C. M., Heled J., Kearse N., Moir R., Stones- Havas S., Thierer T., Wilson (2008) Available from http://www.geneious.com/. In., pp.

Ebelmen, J. J., (1845) Sur les produits de la décomposition des espèces minérales de la famille des silicates. *Ann Rev Mines* **12**: 627-654.

Emsley, P. & K. Cowtan, (2004) Coot: model-building tools for molecular graphics. *Acta Crystallogr D Biol Crystallogr* **60**: 2126-2132.

Evans, P. R., (1997) *Proceedings of the CCP4 Study Weekend. Recent Advances In Phasing*: 97-102.

Friedrich, C. G., (1998) Physiology and genetics of sulfur-oxidizing bacteria. *Adv Microb Physiol* **39**: 235-289.

Fuchs, T., H. Huber, S. Burggraf & K. Stetter, (1996) 16S rDNA-based Phylogeny of the Archaeal Order *Sulfolobales* and Reclassification of *Desulfurolobus ambivalens* as *Acidianus ambivalens* comb. nov. *Syst Appl Microbiol* **19**: 56-60.

Giggenbach, W., (1972) Optical Spectra and Equilibrium Distribution of Polysulfide Ions in Aqueous Solution at 20 °C. *Inorg. Chem.* **11**: 1201-1207.

Gomes, C. M., C. Backgren, M. Teixeira, A. Puustinen, M. L. Verkhovskaya, M. Wikstrom & M. I. Verkhovsky, (2001a) Heme-copper oxidases with modified D- and K-pathways are yet efficient proton pumps. *FEBS Lett* **497**: 159-164.

Gomes, C. M., T. M. Bandeiras & M. Teixeira, (2001b) A new type-II NADH dehydrogenase from the archaeon *Acidianus ambivalens*: characterization and in vitro reconstitution of the respiratory chain. *J Bioenerg Biomembr* **33**: 1-8.

Goubern, M., M. Andriamihaja, T. Nubel, F. Blachier & F. Bouillaud, (2007) Sulfide, the first inorganic substrate for human cells. *Faseb J* **21**: 1699-1706.

Greenwood, N. N. & A. n. E. Earnshaw, (1997) *Chemistry of the Elements*. Butterworth-Heinemann, Oxford (UK).

Griesbeck, C., G. Hauska & M. Schütz, (2000) *Biological Sulfide Oxidation: Sulfide-Quinone-Reductase (SQR), the Primary Reaction,* p. 179-203. Research Signpost, Trivandrum, India.

Griesbeck, C., M. Schutz, T. Schodl, S. Bathe, L. Nausch, N. Mederer, M. Vielreicher & G. Hauska, (2002) Mechanism of sulfide-quinone reductase investigated using site-directed mutagenesis and sulfur analysis. *Biochemistry* **41**: 11552-11565.

Hedderich, R., O. Klimmek, A. Kröger, R. Dirmeier, M. Keller & K. O. Stetter, (1998) Anaerobic respiration with elemental sulfur and with disulfides. *FEMS Microbiology Reviews* **22**: 353-381.

Hirokawa, T., S. Boon-Chieng & S. Mitaku, (1998) SOSUI: classification and secondary structure prediction system for membrane proteins. *Bioinformatics* **14**: 378-379.

http://www.cbs.dtu.dk/services/TMHMM/. In., pp.

Hu, H. Y., K. Fujie & K. Urano, (1999) Development of a novel solid phase extraction method for the analysis of bacterial quinones in activated sludge with a higher reliability. *J Biosci Bioeng* **87**: 378-382.

Huber, R., T. Wilharm, D. Huber, A. Trincone, S. Burggraf, H. König, R. Rachel, I. Rockinger, H. Fricke & K. Stetter, (1992) *Aquifex pyrophilus* gen. nov., represents a novel group of marine hyperthermophilic hydrogen-oxidizing bacteria. *Systematic and Applied Microbiology* **15**: 340-351.

Jackson, A. R. W. & J. M. Jackson, (1996) *Environmental Science: The Natural Environment and Human Impact*. Longman, New York.

Jorgensen, B. B., (1982) Ecology of the bacteria of the sulphur cycle with special reference to anoxic-oxic interface environments. *Philos Trans R Soc Lond B Biol Sci* **298**: 543-561.

Kabsch, W., (1993) Automatic processing of rotation diffraction data from crystals of initially unknown symmetry and cell constants. *Journal of Applied Crystallography* **26**: 795-800.

Kasting, F. J., (1993) Earth's early atmosphere. *Science* **259**: 920-926.

Kletzin, A., T. Urich, F. Muller, T. M. Bandeiras & C. M. Gomes, (2004) Dissimilatory oxidation and reduction of elemental sulfur in thermophilic archaea. *J Bioenerg Biomembr* **36**: 77-91.

Knaff, D. B. & B. B. Buchanan, (1975) Cytochrome *b* and photosynthetic sulfur bacteria. *Biochim Biophys Acta* **376**: 549-560.

Krissinel, E. & K. Henrick, (2004) Secondary-structure matching (SSM), a new tool for fast protein structure alignment in three dimensions. *Acta Crystallogr D Biol Crystallogr* **60**: 2256-2268.

Kronneck, P. M. H., (2005) *The Biogeochemical Cycles of the Elements and the Evolution of Life*. Taylor & Francis, Boca Raton (USA).

Laemmli, U. K., (1970) Cleavage of structural proteins during the assembly of the head of bacteriophage T4. *Nature* **227**: 680-685.

Lemos, R. S., C. M. Gomes & M. Teixeira, (2001) *Acidianus ambivalens* Complex II typifies a novel family of succinate dehydrogenases. *Biochem Biophys Res Commun* **281**: 141-150.

Lens, P. N. & J. G. Kuenen, (2001) The biological sulfur cycle: novel opportunities for environmental biotechnology. *Water Sci Technol* **44**: 57-66.

Lloyd, D., (2006) Hydrogen sulfide: clandestine microbial messenger? *Trends Microbiol* **14**: 456-462.

Marcia, M., U. Ermler, G. Peng & H. Michel, (2009) The structure of *Aquifex aeolicus* sulfide:quinone oxidoreductase, a basis to understand sulfide detoxification and respiration. *Proc Natl Acad Sci U S A* **106**: 9625-9630.

Morel, F. M. M. & N. M. Price, (2003) The Biogeochemical Cycles of Trace Metals in the Oceans. *Science* **300**: 944-947.

Muller, F. H., T. M. Bandeiras, T. Urich, M. Teixeira, C. M. Gomes & A. Kletzin, (2004) Coupling of the pathway of sulphur oxidation to dioxygen reduction: characterization of a novel membrane-bound thiosulphate:quinone oxidoreductase. *Mol Microbiol* **53**: 1147-1160.

Murshudov, G. N., A. A. Vagin & E. J. Dodson, (1997) Refinement of macromolecular structures by the maximum-likelihood method. *Acta Crystallogr D Biol Crystallogr* **53**: 240-255.

Nina, M., S. Berneche & B. Roux, (2000) Anchoring of a monotopic membrane protein: the binding of prostaglandin H2 synthase-1 to the surface of a phospholipid bilayer. *Eur Biophys J* **29**: 439-454.

Nübel, T., C. Klughammer, R. Huber, G. Hauska & M. Schütz, (2000) Sulfide:quinone oxidoreductase in membranes of the hyperthermophilic bacterium *Aquifex aeolicus* (VF5). *Archives of Microbiology* **173**: 233-244.

Otwinowski, Z., W. Minor & Charles W. Carter, Jr., (1997) Processing of X-ray diffraction data collected in oscillation mode. In: Methods in Enzymology. Academic Press, pp. 307-326.

Ouml, S. Lkel & M. Grieshaber, (1997) Sulphide oxidation and oxidative phosphorylation in the mitochondria of the lugworm. *J Exp Biol* **200**: 83-92.

Pape, T. & T. R. Schneider, (2004) HKL2MAP: a graphical user interface for macromolecular phasing with SHELX programs. *J. Appl. Cryst.* **37**: 843-844.

Penin, F., V. Brass, N. Appel, S. Ramboarina, R. Montserret, D. Ficheux, H. E. Blum, R. Bartenschlager & D. Moradpour, (2004) Structure and function of the membrane anchor domain of hepatitis C virus nonstructural protein 5A. *J Biol Chem* **279**: 40835-40843.

Powell, M. A. & G. N. Somero, (1986) Hydrogen Sulfide Oxidation Is Coupled to Oxidative Phosphorylation in Mitochondria of Solemya reidi. *Science* **233**: 563-566.

Qu, K., S. W. Lee, J. S. Bian, C. M. Low & P. T. Wong, (2008) Hydrogen sulfide: neurochemistry and neurobiology. *Neurochem Int* **52**: 155-165.

Reinartz, M., J. Tschäpe, T. Brüser, H. G. Trüper & C. Dahl, (1998) Sulfide oxidation in the phototrophic sulfur bacterium *Chromatium vinosum*. *Archives of Microbiology* **170**: 59-68.

Robertson, L. A. & J. G. Kuenen, (1992) *The Colourless Sulfur Bacteria*. Springer.

Sapay, N., Y. Guermeur & G. Deleage, (2006) Prediction of amphipathic in-plane membrane anchors in monotopic proteins using a SVM classifier. *BMC Bioinformatics* **7**: 255.

Schutz, M., I. Maldener, C. Griesbeck & G. Hauska, (1999) Sulfide-quinone reductase from *Rhodobacter capsulatus*: requirement for growth, periplasmic localization, and extension of gene sequence analysis. *J Bacteriol* **181**: 6516-6523.

Schutz, M., Y. Shahak, E. Padan & G. Hauska, (1997) Sulfide-quinone reductase from *Rhodobacter capsulatus*. Purification, cloning, and expression. *J Biol Chem* **272**: 9890-9894.

Shahak, Y., M. Schutz, M. Bronstein, C. Griesbeck, G. Hauska & E. Padan, (1999) *Sulfide-dependent anoxygenic photosynthesis in prokaryotes: Sulfide-quinone reductase (SQR) the initial step*. Plenum Press, New York.

Sheldrick, G. M., (2008) A short history of SHELX. *Acta Crystallogr A* **64**: 112-122.

Shibata, H. & S. Kobayashi, (2006) Characterization of a HMT2-like enzyme for sulfide oxidation from *Pseudomonas putida*. *Can J Microbiol* **52**: 724-730.

Shibata, H., K. Suzuki & S. Kobayashi, (2007) Menaquinone reduction by an HMT2-like sulfide dehydrogenase from Bacillus stearothermophilus. *Can J Microbiol* **53**: 1091-1100.

Smith, P. K., R. I. Krohn, G. T. Hermanson, A. K. Mallia, F. H. Gartner, M. D. Provenzano, E. K. Fujimoto, N. M. Goeke, B. J. Olson & D. C. Klenk, (1985) Measurement of protein using bicinchoninic acid. *Anal Biochem* **150**: 76-85.

Stetter, K. O., (1996) Hyperthermophilic procaryotes. *FEMS Microbiology Reviews* **18**: 149-158.

Stetter, K. O. & G. Gaag, (1983) Reduction of molecular sulphur by methanogenic bacteria. *Nature* **305**: 309-311.

Stetter, K. O., R. Huber, E. Blochl, M. Kurr, R. D. Eden, M. Fielder, H. Cash & I. Vance, (1993) Hyperthermophilic archaea are thriving in deep North Sea and Alaskan oil reservoirs. *Nature* **365**: 743-745.

Susin, S., J. Abian, F. Sanchez-Baeza, M. L. Peleato, A. Abadia, E. Gelpi & J. Abadia, (1993) Riboflavin 3'- and 5'-sulfate, two novel flavins accumulating in the roots of iron-deficient sugar beet (Beta vulgaris). *J Biol Chem* **268**: 20958-20965.

Theissen, U., M. Hoffmeister, M. Grieshaber & W. Martin, (2003) Single eubacterial origin of eukaryotic sulfide:quinone oxidoreductase, a mitochondrial enzyme conserved from the early evolution of eukaryotes during anoxic and sulfidic times. *Mol Biol Evol* **20**: 1564-1574.

Theissen, U. & W. Martin, (2008) Sulfide : quinone oxidoreductase (SQR) from the lugworm *Arenicola marina* shows cyanide- and thioredoxin-dependent activity. *Febs J* **275**: 1131-1139.

Thompson, J. D., T. J. Gibson, F. Plewniak, F. Jeanmougin & D. G. Higgins, (1997) The CLUSTAL_X windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. *Nucleic Acids Res* **25**: 4876-4882.

Urich, T., T. M. Bandeiras, S. S. Leal, R. Rachel, T. Albrecht, P. Zimmermann, C. Scholz, M. Teixeira, C. M. Gomes & A. Kletzin, (2004) The sulphur oxygenase reductase from *Acidianus ambivalens* is a multimeric protein containing a low-potential mononuclear non-haem iron centre. *Biochem J* **381**: 137-146.

van den Ende, F. P., J. Meier & H. van Gemerden, (1997) Syntrophic growth of sulfate-reducing bacteria and colorless sulfur bacteria during oxygen limitation. *FEMS Microbiology Ecology* **23**: 65-80.

Van Driessche, G., M. Koh, Z. W. Chen, F. S. Mathews, T. E. Meyer, R. G. Bartsch, M. A. Cusanovich & J. J. Van Beeumen, (1996) Covalent structure of the flavoprotein subunit of the flavocytochrome *c*: sulfide dehydrogenase from the purple phototrophic bacterium *Chromatium vinosum*. *Protein Sci* **5**: 1753-1764.

Vande Weghe, J. G. & D. W. Ow, (1999) A fission yeast gene for mitochondrial sulfide oxidation. *J Biol Chem* **274**: 13250-13257.

Wakai, S., M. Tsujita, M. Kikumoto, M. A. Manchur, T. Kanao & K. Kamimura, (2007) Purification and characterization of sulfide:quinone oxidoreductase from an acidophilic iron-oxidizing bacterium, *Acidithiobacillus ferrooxidans*. *Biosci Biotechnol Biochem* **71**: 2735-2742.

Wang, R., (2002) Two's company, three's a crowd: can H_2S be the third endogenous gaseous transmitter? *Faseb J* **16**: 1792-1798.

Westhead, D. R., D. C. Hatton & J. M. Thornton, (1998) An atlas of protein topology cartoons available on the World-Wide Web. *Trends Biochem Sci* **23**: 35-36.

Widdel, F. & N. Pfennig, (1991) *The genus Desulfuromonas and other Gram-negative sulfur-reducing eubacteria,* p. 3379-3389. Springer.

Zillig, W., S. Yeats, I. Holz, A. Bock, F. Gropp, M. Rettenberger & S. Lutz, (1985) Plasmid-related anaerobic autotrophy of the novel archaebacterium *Sulfolobus ambivalens*. *Nature* **313**: 789-791.

Zillig, W., S. Yeats, I. Holz, M. Rettenberger, F. Gropp & G. Simon, (1986) *Desulfurolobus ambivalens*, gen. nov., sp. nov., an autotrophic archaebacterium facultatively oxidizing or reducing sulfur. *Syst Appl Microbiol* **8**: 197-203.

Chapter 3

THE CRYSTAL STRUCTURE OF CTP:INOSITOL-1-PHOSPHATE-CYTIDYLYLTRANSFERASE FROM *ARCHAEOGLOBUS FULGIDUS*

The results presented in this chapter were published in

José A. Brito, Nuno Borges, Helena Santos and Margarida Archer (2010). "Production, crystallization and preliminary X-ray analysis of CTP:inositol-1-phosphate cytidylyltransferase from *Archaeoglobus fulgidus*", *Acta Cryst* **F66**, 1463-1465.

José A. Brito, Nuno Borges, Clemens Vonrhein, Helena Santos and Margarida Archer (2011). "Crystal structure of *Archaeoglobus fulgidus* CTP:inositol-1-phosphate cytidylyltransferase, a key enzyme for di-*myo*-inositol-phosphate synthesis in (hyper)thermophiles", *J Bacteriol* **193**, 2177-2185.

NOTE: This work was done in collaboration with the Cell Physiology and NMR Laboratory (ITQB-UNL), headed by Prof. Doctor Helena Santos. NB, cloned, expressed and purified the protein, JAB crystallized the protein and collected diffraction data; JAB and CV processed the diffraction data, solved the phase problem and refined the structure, JAB and MA analyzed the structure, JAB, NB, HS and MA wrote the manuscripts.

3.1. ABSTRACT

Di-*myo*-inositol-1,3'-phosphate is the most widespread organic solute in microrganisms adapted to hot environments. Many Archaea and Bacteria isolated from hot, marine environments accumulate it primarily in response to heat stress. Until now, DIP has never been encountered in organisms with optimal growth temperatures below 60 °C and, hence, the assumption that it plays a role in the thermoprotection of cellular components *in vivo* is often stated (Rodrigues *et al.*, 2007). DIP is part of a class of molecules often referred as compatible solutes or osmolytes, which are of considerable biotechnological interest.

The biosynthesis of DIP involves the activation of inositol to CDP-inositol via the action of a recently discovered CTP:inositol-1-phosphate cytidylyltransferase (IPCT) activity. In most cases, IPCT is part of a bifunctional enzyme comprising two domains: a cytoplasmic domain with IPCT activity and a membrane domain with di-*myo*-inositol-phosphate-phosphate synthase (DIPPS) activity catalysing the synthesis of di-*myo*-inositol-1,3'-phosphate-1'-phosphate from CDP-inositol and L-*myo*-inositol phosphate.

Herein we describe the X-ray structure determination of the IPCT cytoplasmic domain of the bifunctional IPCT/DIPPS enzyme from the hyperthermophilic archaeon *Archaeoglobus fulgidus* DSMZ 7324 in its apo-form to 1.9 Å resolution and bound to citrate to 2.4 Å resolution. The enzyme exhibited apparent K_m values of 0.9 and 0.6 mM for inositol-1-phosphate (inositol-1P) and CTP, respectively. The optimal temperature for catalysis was in the range of 90 – 95 ºC and V_{max} determined at 90 ºC was 62.9 μmol.min^{-1}.mg of protein^{-1}. The structure of IPCT is composed of a central seven-stranded mixed β-sheet, of which six β-strands are parallel, surrounded by six α-helices, a fold reminiscent of the dinucleotide-binding Rossmann fold. The enzyme shares structural homology with other pyrophosphorylases showing the canonical motif G-X-G-T-(R/S)-X_4-P-K. CTP, inositol-1P and CDP-inositol were docked into the catalytic site which provided insights into the binding mode and high specificity of the enzyme for CTP.

3.2. OSMOADAPTATION AND OSMOREGULATION

Archaeal organisms, like all microrganisms living in aqueous environments, face alterations in the water activity (a_w), due to the fluctuations in the concentrations of dissolved salts and sugars (Santos & da Costa, 2002). The water activity a_w is defined as the quotient of the vapour pressure (p) of water in a given substance or media, by the vapour pressure (p_0) of pure water at the same temperature:

$$a_w = \frac{p}{p_o} \qquad \text{(Eq. 3.1)}$$

In general terms, it defines the amount of water that is available for microbial growth. Pure water has an a_w of 1.0 and the lower this value "the less water is available" for growth; an a_w of 0.5 is considered to be the lowest threshold for microbial proliferation.

The majority of plants and bacteria maintain an osmotic pressure in the cytoplasm higher than that of the surrounding environment:

$$p_{cytopl} > p_{environ} \qquad \text{(Eq. 3.2)}$$

This outward-directed pressure, called turgor, is essential for cell division and growth, and must be assured at all times (Kempf & Bremer, 1998). Changes in the environment osmolarity, most often caused by alterations in the concentration of external Na^+ ions, will trigger the flux of water across the cytoplasmic membrane leading to alterations in the cell volume (Roberts, 2000). This changes the concentrations of intracellular metabolites and can lead to the inhibition of a variety of cellular processes (Csonka, 1989, Csonka & Hanson, 1991); in the case of reduced external Na^+, the influx of water can even lead to cell lysis.

In Archaea, to avoid lysis under low-osmolarity, or dehydration under high-osmolarity growth conditions (efflux of water from the cytoplasm to the surrounding environment), and to adjust to fluctuations in the a_w's of the media and,

concomitantly, of "cytoplasmic water", cells possess active mechanisms that allow a rapid and effective adaptation to these changes. The response to altered osmotic pressure has three main components: the recognition of osmotic imbalance by an osmosensor, the accumulation of solutes in response to the imposed pressure difference and the stabilization of macromolecules under the new intracellular conditions (Fig. 3.1).

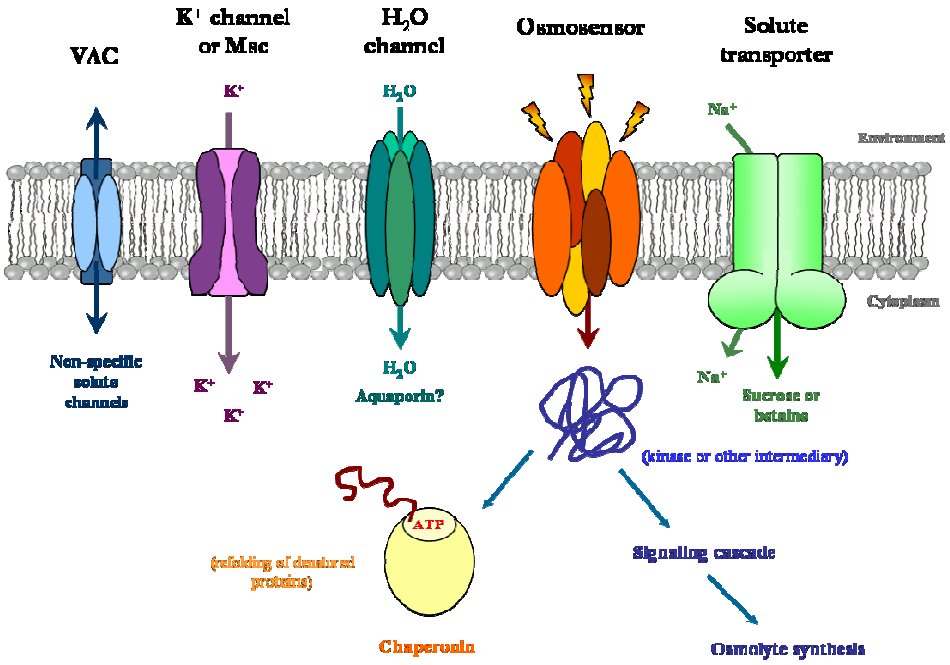


Figure 3.1 – Schematic of the different components of osmotic stress mechanisms that may be operational in Archaea [adapted from (Roberts, 2000)].
VAC, Volume-Activated Channels; Msc, MechanoSensitive ion Channels.

Volume-Activated Channels (VAC's), are a class of transmembrane proteins that seem to play a critical role in the hypoosmotic response. When cells swell due to hypoosmotic shock, the release of cytoplasmic solutes can induce an efflux of water and a return to the original cell volume. VAC's are the channels that respond to these volume/pressure alterations and appear to serve as conductors for expulsion of a multitude of solutes (Roberts, 2000).

MechanoSensitive ion Channels (Msc's), are a family of channels that are gated by membrane tension and thought to be primary biosensors for osmoregulation in bacteria (Blount & Moe, 1999). These channels can lead the internalization of K^+ ions upon hyperosmotic shock., the so-called "salt-in-cytoplasm" strategy (Roberts, 2000).

An immediate response of cells to osmotic imbalance is the movement of water molecules from the low solute concentration environment to the high solute concentration environment until the a_w is the same on both sides of the membrane. Although water can diffuse across the lipid bilayer, water fluxes have been shown to occur in mammalian and plant cells through defined channels, namely aquaporins. This large family of membrane channels involved in osmoregulation have already been shown to exist also in Bacteria [*e.g.*, the *aqpZ* gene in *Escherichia* (*E.*) *coli* codes for an aquaporin shown to mediate large water fluxes in response to sudden changes in extracellular osmolarity (Delamarche *et al.*, 1999), PDB code 1RC2 (Savage *et al.*, 2003)], and Archaea (*e.g.*, AqpM from *Methanothermobacter marburgensis* (Kozono *et al.*, 2003), PDB code 2F2B (Lee *et al.*, 2005), and AqpM from *Arc. fulgidus*, PDB code 3NE2, no citation available), being involved in mediating water fluxes (Roberts, 2000).

Cells can also possess membrane attached "osmosensors" that undergo conformational changes in response to alterations in extracellular a_w. Direct osmosensing could be achieved by this membrane component in response to changes in its interactions with particular solutes, hydration changes or macromolecular crowding (Wood, 1999). Otherwise, it can detect and respond to changes in membrane structure. These conformational changes are coupled to long time-scale adaptations usually the accumulation of low molecular weight solutes (either by internalization or *de novo* synthesis of these compounds), or the promotion of correct protein folding by chaperonins or other macromolecular complexes (changes in a_w can have a pernicious effect on protein stability and folding) (Roberts, 2000).

Solute accumulation strategies are ubiquitous responses observed throughout all three domains of Life. Some of these solutes, like glycine betaine or

sucrose, are rarely synthesized *de novo* and are actively transported into cells. Several of these transporters have been identified in (hyper)thermophilic Archaea, *e.g. Sulfolobus solfataricus*, *S. tokodaii*, *Pyrococcus furiosus* or *Methanococcus jannaschii* (vide TransportDB (Ren *et al.*, 2004)), namely transporters of sugars (sucrose, glucose and others), and aminoacids (proline, glycine). One of the most successful strategies for solutes accumulation in (hyper)thermophilic Archaea and Bacteria involves the utilization of polyol phosphodiesters, namely di-glycerol-phosphate and di-*myo*-inositol-phosphate (Empadinhas & da Costa, 2006).

3.3. COMPATIBLE SOLUTES

The term "compatible solute" was firstly introduced by A. D. Brown in 1976 to describe "small organic compounds used for osmotic adjustment that do not interfere with cell function" (Brown, 1976). These compatible solutes (CS), or osmolytes, can be amino acids (glutamate, proline, glycine), and their derivatives, sugars (sucrose, trehalose, mannose, glucose), and their derivatives, polyols, betaines and ectoines that can be taken from the environment or synthesized *de novo* (Empadinhas & da Costa, 2006, da Costa *et al.*, 1998). Implicit in its definition is the fact that compatible solutes protect proteins and other cell components from osmotic-induced dehydration. However, the role of compatible solutes goes beyond osmotic adjustment alone, to the protection of cells and cell components from freezing, desiccation, high temperature and oxygen radicals (Santos & da Costa, 2002).

It has also been inferred recently that CS may be involved in the rescue of folding defects and exert a chemical chaperoning function in neurodegenerative pathologies and metabolic disorders (Muller *et al.*, 2005, Leandro & Gomes, 2008). Several diseases, like Alzheimer, Parkinson and Huntington disease, cystic fibrosis and retinitis pigmentosa are caused by defects in protein trafficking, folding and/or aggregation. These diseases are referred to as Conformational Disorders and in recent years several CS are being used to study its effects in the rescue of folding

defects (Table 3.1). Besides this potential clinical application, CS are of considerable biotechnological interest and are used in pharmaceutical formulations, vaccine stabilizers, cosmetics, and skin treatment and protection (Muller *et al.*, 2005).

Table 3.1 - Human proteins involved in misfolding disorders rescued by CS (adapted from (Leandro & Gomes, 2008)

Disease	Protein involved	Localization	Mechanism	Molecule
Cystic fibrosis	CFTR	Membrane	Mistrafficking	Glycerol
Huntington	Huntingtin	Cytoplasm	Aggregation	Trehalose
Homocystinuria	CBS	Cytoplasm	Misfolding	TMAO, glycerol, sorbitol and proline
Phenylketonuria	PAH	Cytoplasm	Misfolding	Glycerol
Nephrogenic diabetes insipidus type II	AQP2	Membrane	Mistrafficking	Glycerol, TMAO and DMSO
Albinism	Tyrosinase	Membrane	Mistrafficking	Tyrosine
Parkinson	α-synuclein	Cytoplasm	Misfolding/aggregation	TMAO
Creutzfeld-Jacob	Prion	Several cellular locali	Aggregation	Acridine-based analogue
Machado-Joseph	Ataxin-3	Nucleus/cytoplasm	Aggregation	Glycerol, TMAO and DMSO

(CFTR) cystic fibrosis transmembrane conductance regulator; (CBS) cystathionine-β-synthase; (PAH) phenylalanine hydroxylase; (AQP2) aquaporin-2.

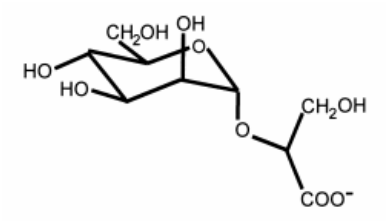
Thermophile and hyperthermophile microrganisms accumulate CS that have not been found, or are rarely encountered, in mesophilic microrganisms. These are generally negatively charged molecules, as are most of the CS from Archaea, while other microrganisms accumulate neutral or zwitterionic molecules. Most of the CS

found in hyperthermophilic Archaea (Fig. 3.2), and thermophilic Bacteria, have their negative charge due to carboxylate, phosphate or sulphate groups (Roberts, 2005, Santos & da Costa, 2002). This anionic nature of the solutes might be explained by the fact that they can contribute to the balance of the high intracellular concentration of inorganic cations, namely K^+ (Muller *et al.*, 2005).

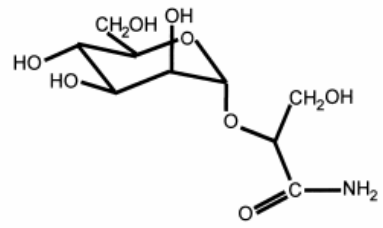
These unusual osmolytes can be α- or β-amino acids, like α-glutamate or N^{ε}-acetyl-β-lysine, mannosylglycerate (MG), di-*myo*-inositol-1,3'-phosphate (DIP), di-glycerol-phosphate (DGP), or di-mannosyl-di-*myo*-inositol-1,3'-phosphate (MDIP), among others (Fig. 3.2). MG, DIP and DGP are hallmark solutes of archaeal organism living in very hot environments. Members of the genera *Aeropyrum*, *Stetteria*, *Pyrococcus*, *Thermococcus* and *Archaeoglobus* have ben shown to accumulate MG at the same time as DIP (Muller *et al.*, 2005). *Pyrolobus fumarii* and *Pyrodictium occultum* have been shown also to accumulate DIP as *Methanococcus igneus* and *Palaeococcus ferrophilus* accumulate MG but not DIP. In the light of these findings it has been suggested that these CS are specifically associated with life at high temperatures (Santos & da Costa, 2002).

3.4. DI-*MYO*-INOSITOL-1,3'-PHOSPHATE (DIP)

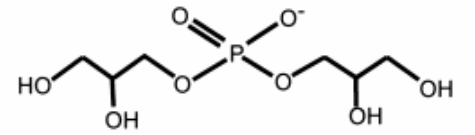
DIP was first identified by Scholz and co-workers in the archaeon *Pyrococcus woesei* (Scholz *et al.*, 1992), where they showed that the potassium salt of DIP provided optimum enzyme stabilization when the activity of glyceraldehyde-3-phosphate dehydrogenase from that organism was tested at 105 ^{o}C under anaerobic conditions. DIP was later encountered in many other hyperthermophiles including members of the genera *Methanotorris*, *Thermococcus*, *Thermotoga*, *Aquifex*, *Pyrodictium*, *Aeropyrum*, *Archaeoglobus*, *Stetteria* and *Pyrolobus*, and accumulates in response to salinity or temperature stress (Goncalves *et al.*, 2003, Santos & da Costa, 2002).



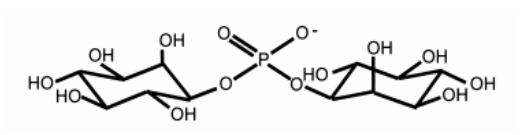
Mannosylglycerate (MG)



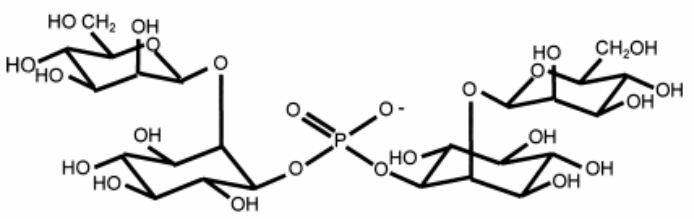
Mannosylglyceramide



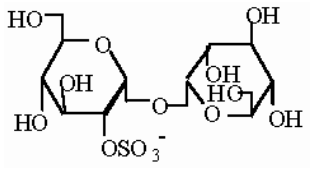
Di-glycerol-phosphate (DGP)



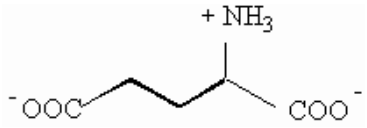
Di-*myo*-inositol-1,3'-phosphate (DIP)



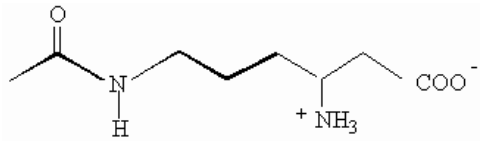
Di-mannosyl-di-*myo*-inositol-1,3'-phosphate (MDIP)



2-sulfotrehalose



α-glutamate



N$^{\varepsilon}$-acetyl-β-lysine

Figure 3.2 – Compatible solutes found only in (hyper)thermophiles (adapted from (Santos & da Costa, 2002, Muller *et al.*, 2005, Roesser & Muller, 2001).

In fact, DIP is the most widespread small-molecular weight solute being highly restricted to hyperthermophilic Archaea and Bacteria, and has never been found in mesophilic organisms (Santos & da Costa, 2002). In *Methanococcus* (*M.*) *igneus* and *Pyrococcus* (*P.*) *furiosus*, intracellular DIP increases with increasing extracellular concentrations of NaCl as well as at supraoptimal temperatures (> 80 °C for *M. igneus* and 98 to 101 °C for *P. furiosus*) (Ciulla *et al.*, 1994, Martins & Santos, 1995). The unusual intracellular high concentration of K^+ ions in *P. woesei* and its extreme optimal growth temperature (100 to 104 °C), led to the suggestion that DIP could also act as a counterion of K^+ with a possible thermostabilizing function (Zwickl *et al.*, 1990).

Until some years ago, the only known pathway for inositol biosynthesis in all organisms was the conversion of D-glucose-6-phosphate to L-*myo*-inositol-1-phosphate (inositol-1P), via L-*myo*-inositol-1-phosphate synthase and hydrolysis of inositol-1P to *myo*-inositol via inositol-1P phosphatase (Loewus, 1990, Loewus *et al.*, 1990). A possible pathway would then be to use *myo*-inositol and inositol-1P for the synthesis of DIP where the activation of inositol-1P with CTP and condensation of the resultant CDP-inositol with another molecule of *myo*-inositol would generate DIP (Chen *et al.*, 1998). This was demonstrated by Chen *et al.* (Chen *et al.*, 1998), where in crude protein extracts of *M. igneus*, the DIP biosynthesis occurred in a single step from activated CDP-inositol and *myo*-inositol via a DIP synthase activity. This last step in the pathway of DIP had striking similarities to the already know pathway to the biosynthesis of phosphatidylserine: CDP-inositol is attacked by a *myo*-inositol hydroxyl group with the assistance of a general base provided by the DIP synthase enzyme (Chen *et al.*, 1998).

A few years later, Rodionov *et al.* (Rodionov *et al.*, 2007), in *Thermotoga* (*T.*) *maritima*, and Borges *et al.* (Borges *et al.*, 2006), in *Archaeoglobus* (*Arc.*) *fulgidus* proposed almost simultaneously a different pathway proceeding in two steps with the occurrence of a phosphorylated residue that is subsequently dephosphorylated. Borges and co-workers (Borges *et al.*, 2006) proposed the DIP biosynthesis in *Arc. fulgidus* to proceed from glucose-6-phosphate in the following way: (i) glucose-6-phosphate is converted into inositol-1P by L-*myo*-inositol-1-phosphate

synthase; (ii) inositol-1P is activated to CDP-inositol at the expense of CTP; (iii) CDP-inositol is coupled with another molecule of inositol-1P to yield a phosphorylated intermediate, di-*myo*-inositol-1,1'-phosphate-3-phosphate (DIPP); (iv) finally, DIPP is dephosphorylated into DIP by the action of a phosphatase. Rodrigues and colleagues (Rodrigues *et al.*, 2007) later proposed a different stereochemistry to DIP based on ^{13}C-NMR results and discovered that the enzyme that catalyzes the second and third steps in the pathway is only one being a bifunctional CTP:L-*myo*-inositol-1-phosphate cytidylylytransferase/di-*myo*-inositol-1,3'-phosphate-1-phosphate synthase (IPCT/DIPPS) (Fig. 3.3).

3.5. CTP:INOSITOL-1-PHOSPHATE CYTIDYLYLTRANSFERASE FROM *ARCHAEOGLOBUS FULGIDUS*

Although the pathways proposed by Borges *et al.* (Borges *et al.*, 2006), and Rodionov *et al.* (Rodionov *et al.*, 2007), were essentially the same, the most striking difference between them is that there was only one gene, AF0263, coding for the L-*myo*-inositol-1-phosphate cytidylyltransferase (IPCT), and the di-*myo*-inositol-1,3'-phosphate-1-phosphate synthase (DIPPS), activities in *Arc. fulgidus*, whereas in *T. maritima* those activities were encoded by two separate genes, TM1418 and TM1419. The existence of these genes encoding for the key activities in DIP biosynthesis, i.e., IPCT and DIPPS, was firmly identified in several microorganisms (*Arc. fulgidus*, *P. furiosus*, *Thermococcus kodakarensis*, *Aquifex aeolicus* and *Rubrobacter xylanophilus*) by functional expression in *E. coli* (Rodrigues *et al.*, 2007). In most organisms known to accumulate DIP, the two activities are present in a single polypeptide chain, constituting the bifunctional enzyme IPCT/DIPPS; however, in *T. maritima* the two activities are encoded by separate, consecutive genes and the same gene organization was predicted in other members of hyperthermophilic Bacteria and in a few Archaea (Rodionov *et al.*, 2007).

The bifunctional enzyme from *Arc. fulgidus* coding for the IPCT/DIPPS activities was found by a functional genomics approach to the microorganism's

genome and one gene, AF0263, encoding a bifunctional protein with high homology to nucleotidyltransferases (N-terminal domain), and to CDP-alcohol phosphatidyltransferases (C-terminal domain), was found. This hypothesis was substantiated when attempts were made to purify the protein from *Arc. fulgidus* cell extracts based on the DIPPS activity and this activity was not separated from the IPCT activity even after four chromatographic steps (Rodrigues *et al.*, 2007).

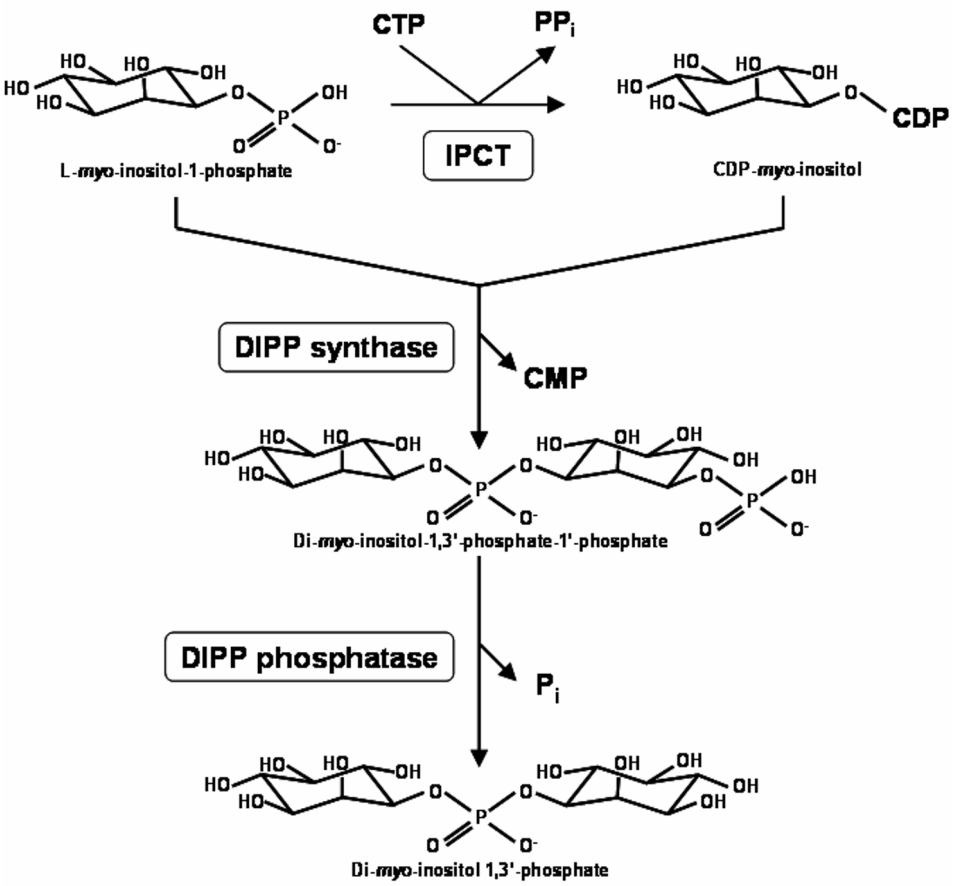


Figure 3.3 - Pathway for di-*myo*-inositol phosphate synthesis.
Enzymes: CTP:L-*myo*-inositol-1-phosphate cytidylyltransferase (IPCT); di-*myo*-inositol-1,3'-phosphate-1'-phosphate synthase (DIPP synthase); di-*myo*-inositol-1,3'-phosphate-1'-phosphate phosphatase (DIPP phosphatase).

Biochemical characterization of the enzyme in protein extracts led to the conclusion that the IPCT domain was absolutely specific for CTP and L-*myo*-inositol-1-phosphate, while other nucleotide donors (ATP, GTP and UTP), and other alcohol acceptors (*myo*-inositol, D-*myo*-inositol-1-phosphate, glycerol and DL-glycerol-3-phosphate), were not substrates to the enzyme. The DIPPS domain recognized CDP-L-*myo*-inositol, CDP-D-*myo*-inositol and CDP-glycerol as alcohol donors but L-*myo*-inositol-1-phosphate was the only alcohol acceptor used by the enzyme. It was also able to catalyze the synthesis of glycerol-phospho-*myo*-inositol-1-phosphate (GPIP), when CDP-glycerol was provided in combination with inositol-1P. Glycerol-phospho-*myo*-inositol (GPI), is a structural chimera of DIP and DGP (diglycerol-phosphate), only identified to date in the genus *Aquifex* and in *Arc. fulgidus*. While DIP increases consistently in response to elevated temperature and DGP to osmotic stress, the level of GPI seems to respond to a combination of both stresses (Borges *et al.*, 2006).

More importantly, it was predicted that the bifunctional enzyme IPCT/DIPPS combines a cytosolic domain (IPCT) with a membranar domain (DIPPS). Therefore, the determination of the structure/function relationship of the bifunctional enzyme emerges as an attractive challenge. The attempts to crystallize the whole functional enzyme were unsuccessful thus far, and hence we set out to determine the structure of IPCT, the cytoplasmic domain, as a first step towards understanding the full catalytic process. IPCT is an intrinsically interesting enzyme, being the sole nucleotidyltransferase known to use inositol-1P as substrate. Enzymes classified in the nucleoside triphosphate-transferase family (PF00483) typically transfer nucleoside monophosphate (NMP) from nucleoside triphosphates (NTP) to an acceptor phosphoryl group belonging to a small molecule such as phosphocholine, hexose-1-phosphate or ribitol-5-phosphate (Kwak *et al.*, 2002, Kim *et al.*, 2010, Baur *et al.*, 2009). This activity leads to release of pyrophosphate and production of a nucleoside diphospho-acceptor that is subsequently utilized by glycosyltransferases in a myriad of reactions of vital importance for the cellular functions.

Herein, we describe the first X-ray structure of the IPCT domain of the IPCT/DIPPS bifunctional enzyme from *Archaeoglobus fulgidus* DSMZ 7324. CTP, inositol-1P and CDP-inositol were docked to the structure which gave insights into the binding mode and substrate specificity of the enzyme.

3.6. MATERIALS AND METHODS

MATERIALS

CTP was purchased from Sigma-Aldrich (St. Louis, MO). Inositol-1P was produced enzymatically as described in (Rodrigues *et al.*, 2007). DGP was supplied by bitop AG (Witten, Germany).

PROTEIN PRODUCTION AND BIOCHEMICAL CHARACTERIZATION

In order to define the IPCT domain in the bifunctional IPCT/DIPPS from *Arc. fulgidus* (AF0263), the amino acid sequence of the AF0263 gene (436 a.a.) was aligned with the monofunctional IPCT and DIPPS sequences from *Hyperthermus butylicus*, *Aeropyrum pernix*, *Thermotoga* (*T.*) *maritima*, and *T. petrophila* using the program ClustalW (Larkin *et al.*, 2007). On the basis of this analysis, the IPCT cytoplasmic domain in the *Arc. fulgidus* bifunctional IPCT/DIPPS was identified as the N-terminal polypeptide starting at M-I-N-V-D and going up to residue 232. This domain was cloned and expressed in *E. coli* (Rodrigues *et al.*, 2007). The chromosomal DNA from *Arc. fulgidus* was isolated according to Ramakrishnan and Adams (1995). The IPCT domain was amplified by PCR using *Pfu* DNA polymerase (Fermentas). The set of primers used for the amplification were IPCT-Fwd (5' CGG**CATATG**ATAAATGTTGACGGAG-3', the *Nde*I restriction site is shown in bold) and IPCT-Rev (5'-CGG**CTCGAG**TTACAAGGCTCTATTAG-3', the *Xho*I restriction site is shown in bold). The PCR product was cloned in the pET19b plasmid (Novagen) between the *Nde*I and *Xho*I sites; the pET19b plasmid carries an N-terminal ten histidines tag sequence followed by an enterokinase site and three

cloning sites (*Nde*I, *Xho*I and *Bam*HI). The correct nucleotide sequence of the construction was confirmed by DNA sequencing (AGOWA, Germany). *E. coli* BL21(DE3)pLys cells, bearing the construction, were grown at 37 ºC in LB medium supplemented with ampicillin (100 μg.mL^{-1}), and chloramphenicol (34 μg.mL^{-1}), to an Abs_{600nm} of 0.6 and induced with 1 mM IPTG for 4 h.

Cells were collected by centrifugation, suspended in 20 mM Tris-HCl pH 7.6 plus 10 mM $MgCl_2$ and disrupted in a French press. IPCT was purified from heat treated cell extracts (30 min at 70 ºC) in a His-Trap column and the histidine-tag was cleaved with enterokinase (18 h at 37 ºC). Subsequently, the enterokinase treated sample was applied onto a second His-Trap column. The flow-through fraction was dialyzed overnight against Tris-HCl (50 mM, pH 7.6) and concentrated in a Centricon (MW 10,000). A last polishing step was performed applying IPCT onto a Superdex G75 gel filtration column. Protein concentration was estimated by the Bradford method (Bradford, 1976), and purity of the final preparation was evaluated by 10% SDS-PAGE (Laemmli, 1970). The molecular mass of IPCT was estimated by gel filtration using a Superdex S200 column equilibrated with 20 mM Tris-HCl pH 7.5 and 150 mM NaCl. Lysozyme (14.3 kDa), superoxide dismutase (32.6 kDa), albumin (66.8 kDa), and alcohol dehydrogenase (150 kDa), were used as standards.

The IPCT activity was determined in a reaction mixture (final volume of 400 μL), containing 50 mM Bis/Tris/Propane, 10 mM $MgCl_2$, 5 mM CTP, and 5 mM inositol-1P. The mixture was pre-incubated for 2 min at the same temperatures of the assay in 2 mL glass tubes. The reactions were initiated by addition of IPCT (5 μg), and stopped at different time points (0, 15 and 30 seconds) by immersion in liquid nitrogen. Afterwards, 10 μL of EDTA (0.5 M, pH 8) and 100 μL of ^2H_2O were added to the reaction mixture. The amount of CDP-inositol produced was quantified by ^{31}P-NMR spectroscopy using DGP as an internal concentration standard. Spectra were acquired on a Bruker DRX500 spectrometer (Bruker, Rheinstetten, Germany). The effect of Mg^{2+} (0, 10 and 20 mM) on IPCT activity was studied. The temperature profile for activity was determined between 60 ºC and 115 ºC at $pH^{25\ ºC}$ 8.0 ($pH^{25\ ºC}$ means that pH values were measured at 25 ºC and no temperature correction was applied). The pH profile for activity was

determined at 90 ºC using 50 mM Bis/Tris/Propane; pH values were measured at 25 ºC but the values displayed ($pH^{90\ ºC}$) are corrected for the temperature of 90 ºC using the conversion factor for Bis/Tris/Propane ($\Delta pKa/\Delta T$ = -0.015). Kinetic parameters (V_{max} and K_m) were determined at 90 ºC and $pH^{90\ ºC}$ of 7.0 in reaction mixtures containing 4 mM CTP and 0-7.5 mM inositol-1P or 0-4 mM CTP and 7.5 mM inositol-1P. The actual CTP and inositol-1P concentrations of the stock solutions were determined by NMR. Kinetic parameters were obtained by fitting the data to Michaelis-Menten equations using the software Origin 5.0 Professional (Microcal Software Inc., MA). The data included six concentration values of each substrate.

CRYSTALLIZATION AND DATA COLLECTION

Protein sample was concentrated using an Amicon YM10 concentrator device (Millipore) to about 12 $mg.mL^{-1}$ in 10 mM HEPES buffer pH 7.5. Initial crystallization screenings were performed by vapour-diffusion using the Cartesian Robot Dispensing System ("mini-bee" Genomic Solutions) and the following commercially available screens: Classics I and II Suites (Qiagen) and SaltRx (Hampton Research). The flat-bottom 96-well plates (Greiner Bio-One) harbored sitting-drops containing 100 nL of protein sample mixed with 100 nL of precipitant solution equilibrated against a reservoir of 90 µL precipitant. The sealed plates were then incubated at room temperature (20 ºC).

Initial crystalline material appeared in several conditions of the different screens which mainly contained salts (i.e. sodium/potassium phosphate, and sodium malonate, citrate or formate) as precipitants at low pH values (4.5 - 6). Optimization of these conditions included grid screens surrounding the crystallization hits using the hanging-drop vapour-diffusion technique. Drops consisted of 1.5 µL protein plus 1.5 µL precipitant solutions equilibrated against 500 µL reservoir solution in 24-well EasyXtal tool plates (Qiagen), at room temperature (20 ºC). The optimized crystallization condition yielding diffraction-quality crystals of IPCT consisted of 1.2-1.4 M trisodium citrate dihydrate, 100 mM HEPES pH 6.8-7.2 and 10 mM $MgCl_2$.

Single crystals with approximate dimensions of 0.15×0.15×0.15 mm^3 (Fig. 3.4), were used for data collection at cryogenic temperature using the crystallization buffer/mother liquor supplemented with 10% glycerol as cryoprotectant and flash-cooled at 100 K in a nitrogen-gas steam (Oxford Cryosystems 700).

The X-ray diffraction images were collected by ADSC Q4 and ADSC Q315R CCD detectors on ESRF beamlines ID14-2 and ID23-1 (ESRF, Grenoble), respectively. Data were indexed and processed using XDS (Kabsch, 1993) and scaled with SCALA from the CCP4 program package (Evans, 1997). Crystals belonged to the orthorhombic space group $P2_12_12$ with unit-cell parameters a = 154.7 Å, b = 83.9 Å and c = 127.7 Å. Data statistics are summarized in Table 3.2. Calculation of Matthews coefficients (Matthews, 1968) yielded a V_m of 2.62 $Å^3.Da^{-1}$ for a solvent content of approximately 53% with 6 molecules in the asymmetric unit, or 2.25 $Å^3.Da^{-1}$ for 45% solvent content and seven molecules in the asymmetric unit. Thus, the protein content in the crystal asymmetric unit of IPCT was not clear at this point.

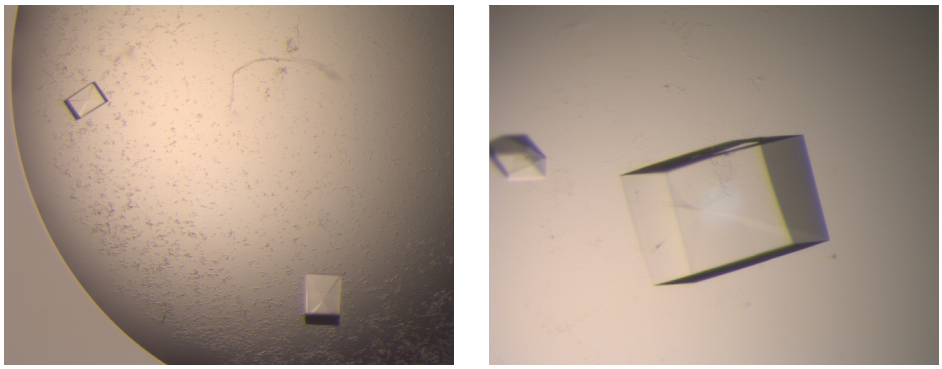


Figure 3.4 - IPCT crystals grown in 1.2 M trisodium citrate dihydrate, 100 mM HEPES pH 7.2 and 10 mM $MgCl_2$.
The crystals took approximately two weeks to appear and grew to dimensions of 0.15×0.15×0.15 mm^3.

A self-rotation function calculated with the program POLARRFN (Evans, 1997) from CCP4 indicated the existence of three two-fold non-crystallographic

symmetry axes (Fig. 3.5). Two of these were approximately perpendicular to the crystallographic c-axis while the third one is either coincident or slightly angled relative to it, since no strong non-origin peaks were observed in a native Patterson map (not shown).

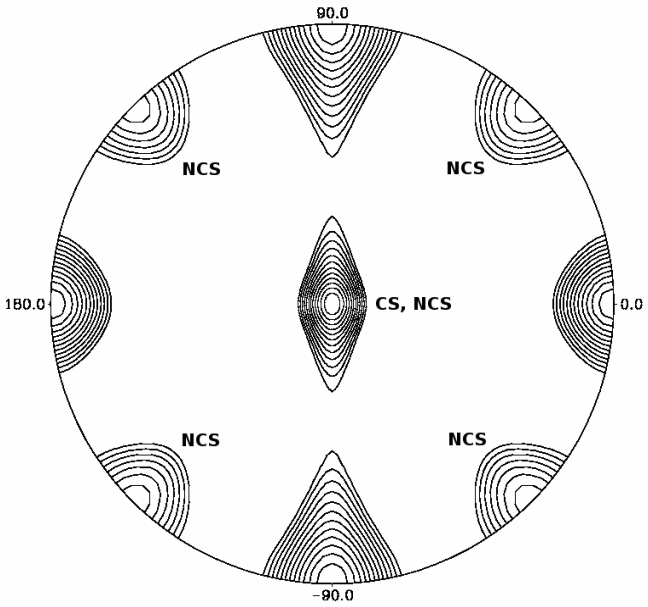


Figure 3.5 - Self-rotation Patterson-function plot for IPCT in the resolution range $3.0 \leq d \leq 15$ Å with integration range $5.0 \leq R \leq 15$ Å.

Peaks in the section shown at $\kappa = 180^{o}$ give the orientation of the crystallographic (CS) and noncrystallographic (NCS) twofold rotation axes in the crystal structure. The plot is scaled to a maximum value of 100 and the contours are drawn at five-unit intervals starting at 40. The figure was prepared with the programs POLARRFN, NPO and XPLOT84DRIVER (Evans, 1997).

IPCT was also crystallized using sodium malonate as precipitant. Crystals were obtained at 20 ^{o}C in hanging-drops using 1.5 μL of protein sample and 1.5 μL of reservoir solution consisting of 1.8 M sodium malonate pH 6.5-7.0. These crystals were flash-cooled directly in liquid nitrogen prior to data collection. Data from these crystals were measured at ESRF beamline ID23-1 using an ADSC Q315r CCD detector, indexed and processed using XDS (Kabsch, 1993) and scaled with SCALA from the CCP4 program package (Evans, 1997). IPCT crystals in this crystallization

condition belonged to space group P2 with 12 or 13 molecules in the asymmetric unit corresponding to V_m values of 2.47 and 2.28 $Å^3.Da^{-1}$ and solvent contents of 50.2 and 46.1%, respectively (Matthews, 1968).

STRUCTURE SOLUTION AND REFINEMENT

Search for potential derivatives was performed by soaking freshly prepared IPCT crystals crystallized in citrate with crystallization solution containing heavy-atom compound salts (K_2Cl_4Pt, $K(CN)_2Au$ or $HgCl_2$ - JBScreen Heavy, Jena Biosciences) in concentrations ranging from 0.5 to 5 mM and incubation times from minutes to several days, at room temperature. In addition, a quick soak was also done with 1 M KI for 30 seconds prior to crystal cryo-cooling.

Experimental phases were determined by the Single Isomorphous Replacement with Anomalous Scattering (SIRAS) method using a mercury derivative (5 mM $HgCl_2$ for 4 hours). SHELXC and SHELXD (Sheldrick, 2008), both within HKL2MAP (Krissinel & Henrick, 2004), were used to prepare the experimental data and to locate the heavy atom sites, respectively. SHELXD indicated clear anomalous signal up to 5 Å but SHELXE did not produce an interpretable map. A few positive hits were found out of 1000 trials, the best solution showing Correlation Coefficients - CCall of 17.98% and CCweak of 9.62%, and PATFOM 1.37; a total of 10 Hg sites were found with occupancies higher than ca. 0.4. This initial set of 10 sites was then refined in SHARP (de la Fortelle & Bricogne, 1997) and subsequent analysis of log-likelihood gradient maps revealed 2 additional sites, giving a total of 12 heavy atom positions. The calculated phases had a mean figure of merit of 0.29 to 2.8 Å resolution and were improved through a single cycle of solvent flipping with SOLOMON (Abrahams & Leslie, 1996) and extended to the resolution of the native dataset (2.4 Å). The resulting map was used to build an initial atomic model using the program BUCCANEER (Cowtan, 2006), which built 1168 residues in 61 chains (the longest having 104 residues). Upon examination of the model, it became clear that one non-crystallographic symmetry (NCS) operator related two dimers of the enzyme and this information was used for NCS averaging with DM (Cowtan, 1994). This further improved the maps which allowed BUCCANEER to proceed with

model building, reducing the number of chains and sequencing more side-chains. Iterative model building and crystallographic refinement were performed with the programs COOT (Emsley *et al.*) and BUSTER-TNT (Bricogne *et al.*, 2009). The malonate-crystallized IPCT was solved by molecular replacement with PHASER (McCoy *et al.*, 2007), using one monomer of the citrate-crystallized IPCT, without solvent and other ligands, as search model for 12 molecules in the asymmetric unit. The model was built into the electron density maps with COOT (Emsley *et al.*) and refinement was carried out with BUSTER-TNT (Bricogne *et al.*, 2009).

The homologous 3D structures were superimposed using the secondary structure matching algorithm (SSM) (Krissinel & Henrick, 2004) in COOT and CTP, inositol-1P and CDP-inositol were manually fitted followed by structure idealization with REFMAC5 (Murshudov *et al.*, 1997), visualization in COOT and subsequent analysis in PyMol (DeLano, 2002). All structural figures were drawn using PyMol (DeLano, 2002).

The coordinates of the malonate-IPCT (apo form) and the citrate-IPCT of *Arc. fulgidus* DSMZ 7324 have been deposited in the Protein Data Bank with accession codes 2XME and 2XMH, respectively.

3.7. RESULTS AND DISCUSSION

BIOCHEMICAL CHARACTERIZATION OF IPCT FROM *ARC. FULGIDUS*

The biochemical and kinetic properties of IPCT from *Arc. fulgidus* were investigated using CTP and inositol-1P as substrates. The presence of Mg^{2+} was absolutely required for activity and maximal activity was obtained with 20 mM Mg^{2+}. The activity of the enzyme was undetectable at temperatures below 60 ºC, and maximal activity was reached between 90 and 95 ºC (Fig. 3.6a). Therefore, the temperature for optimal catalysis is around 10 ºC above the temperature for optimal growth of the host organism, a feature commonly found in other enzymes from hyperthermophiles (Borges *et al.*, 2004, Rodrigues *et al.*, 2009).

The pH profile for *Arc. fulgidus* IPCT showed a maximum of activity between 6.5 and 7.5 (Fig. 3.6b). Kinetic parameters of the enzyme were determined under optimal conditions (Table 1).

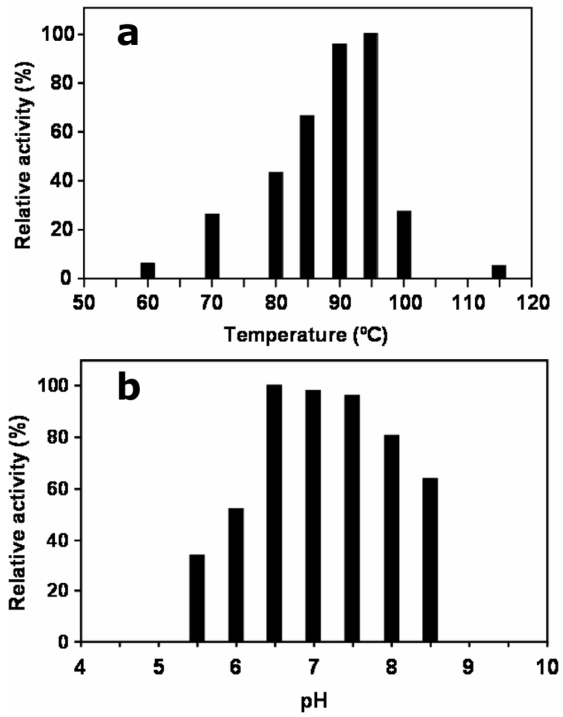


Figure 3.6 - Temperature (a) and pH (b) profiles of IPCT from *Arc. fulgidus*.
The enzyme activity was determined between 60 ºC and 115 ºC, and between $pH^{90\ ºC}$ 5.5 and 8.5 in Bis/Tris/propane buffer.

A typical Michaelis-Menten kinetics was observed for the two substrates. The V_{max} (62.9 μmol.min^{-1}.mg^{-1}) was 2.5-fold lower than that of the *T. maritima* IPCT (Rodionov *et al.*, 2007). In respect to substrate affinity, the truncated *Arc. fulgidus* IPCT showed K_m values of 0.87 mM for CTP and 0.58 mM for inositol-1P which are in the same range of magnitude as those of the homofunctional enzyme from *T. maritima* (Rodionov *et al.*, 2007).

TABLE 3.3 - Kinetic properties of the recombinant IPCT of *Arc. fulgidus* and effect of Mg^{2+}. K_m and V_{max} values were determined at 90 °C and $pH^{90\,°C}$ 7.0.

Parameters	IPCT
K_m (mM)	
Inositol-1P	0.87 ± 0.09
CTP	0.58 ± 0.08
V_{max} (µmol.min^{-1}.mg of protein^{-1})	62.9 ± 1.4
Mg^{2+} (mM)	
0	0[a]
10	67[a]
20	100[a]

[a] Expressed as percentages of the maximum activity.

Next page: **TABLE 3.4 - Data collection and refinement statistics (values between parentheses refer to the highest resolution shell)**

$$^{*}\mathrm{R}_{merge} = \sum_{hkl} \sum_{i} \left| I_i(hkl) - \overline{I(hkl)} \right| \Big/ \sum_{\mathrm{hkl}} \sum_{\mathrm{i}} I_i(hkl)$$

$$^{\#}\mathrm{R}_{pim} = \sum_{\mathrm{hkl}} \left[1 \Big/ (N-1) \right]^{1/2} \sum_{i} \left| I_i(hkl) - \overline{I(hkl)} \right| \Big/ \sum_{\mathrm{hkl}} \sum_{\mathrm{i}} I_i(hkl)$$

Calculated with the program SCALA, R_{merge} and R_{pim} are indicators of the precision of the final merged and averaged data-set, where I_i *(h k l)* is the observed intensity of the i^{th} measurement, $\overline{I(hkl)}$ is the average intensity of multiple observations of symmetry-related reflections and *N* is redundancy.

	IPCT:$HgCl_2$	**Citrate-IPCT***	**Malonate-IPCT (apo)**
PDB ID code	-	2XMH	2XME
Space group	P 2_1 2_1 2	P 2_1 2_1 2	P 1 2 1
Unit cell			
a, b , c (Å)	153.6, 83.5, 127.8	154.7, 83.9, 127.7	86.03, 127.55, 141.48
α, *β*, *γ* (º)	90.00, 90.00, 90.00	90.00, 90.00, 90.00	90.00, 90.56, 90.00
Data collection			
Beamline	ESRF ID14-2	ESRF ID23-1	ESRF ID14-2
Wavelength (Å)	0.933	1.067	0.933
Resolution range (Å)	49.3-2.9 (3.10-2.90)	48.28-2.40 (2.53-2.40)	142.49-1.89 (1.99-1.89)
Total no. of observations	116728 (9974)	224153 (32846)	941436 (58133)
No. of unique observations	32022 (2686)	64618 (9409)	219870 (29484)
R_{merge}* (%)	8.0 (59.7)	5.0 (42.6)	12.6 (48.9)
R_{pim}# (%)		2.8 (26.2)	6.6 (45.3)
Completeness (%)	97.4 (99.2)	98.8 (99.5)	90.3 (83.0)
Mean < I / σ(I) >	12.2 (2.6)	13.4 (2.3)	10.1 (1.5)
Refinement statistics			
Resolution range (Å)	-	35.14-2.40 (2.46-2.40)	27.67-1.89 (1.94-1.89)
R_{work} / R_{free} (%)	-	23.3 / 27.5	20.6 / 23.7
Geometry			
rmsd bonds (Å)	-	0.01	0.01
rmsd angles (º)	-	1.23	1.10
Ramachandran statistics			
Residues in most favoured regions (%)	-	94.9	98.3
Residues in allowed regions (%)	-	5.0	1.6
Residues in disallowed regions (%)	-	0.1	0.1
Mean B factors			
Protein ($Å^2$)	-	73.1	45.9
Solvent ($Å^2$)	-	60.9	41.0

* Used as "native" for SIRAS phasing

MODEL QUALITY

IPCT crystallized in space group $P2_12_12$ (a = 154.9 Å, b = 83.9 Å and c = 127.9 Å) with 6 molecules in the asymmetric unit using citrate as precipitant at pH 8.2 (V_m is 2.6 $Å^3.Da^{-1}$ and solvent content is ~53% (Matthews, 1968)). The structure was solved by SIRAS based on a Hg-derivative and refined to an R_{factor} of 23.3% and R_{free} of 27.0% at 2.4 Å resolution. The final IPCT model comprises 1183 amino acid residues, 204 water and 4 citrate molecules. Data collection and refinement statistics are shown in Table 3.4.

No clear electron density was visible at both termini (16 to 19 amino acid residues in the N-terminus and 10 to 12 residues in the C-terminus, depending on the chain, were not built in the model), and for a flexible loop composed of residues 27-37. Attempts to soak or co-crystallize IPCT with its substrates were performed but no well defined electron density was observed in the calculated maps for either inositol-1P and/or CTP nearby the active site, instead a citrate molecule was observed (Fig. 3.7a).

We then searched for another crystallization condition without citrate and optimized one containing malonate as precipitant at pH ~7. In this condition, IPCT crystals belonged to space group P2 with 12 molecules in the asymmetric unit corresponding to a V_m value of 2.5 $Å^3.Da^{-1}$ and a solvent content of ~50% (Matthews, 1968). Soaking and co-crystallization experiments were also unsuccessful in this condition since no clear electron density for CTP and/or inositol-1P was visible in the active site. The final model comprises 2419 amino acid residues, 1171 water and 8 glycerol molecules, and was refined to 1.9 Å resolution with an R_{factor} of 20.6% and R_{free} of 23.7%. The electron density maps are clearly defined, apart from the N and C termini of all subunits (similarly to the citrate-IPCT, chains start at residues 16-19 and end up at 220-222 out of 232), and a loop. This loop (residues 27 to 37) is disordered in most chains and contains gaps in the final model, except in chains E and F. Interestingly, these two loops are quite close to each other in the crystallographic dimer (*e.g.* distance between the C_α atoms of G_{34E} and G_{33F} is only 6.6 Å), similar to the other pairs: A-B, C-D, and so on. The final models of citrate-

and malonate-IPCT are almost identical, yielding rmsd values between 0.24-0.69 Å for Cα superposition of the different chains (Holm & Sander, 1996). Subsequent descriptions will refer to chains E and/or F of the malonate-IPCT structure, unless otherwise stated.

The Ramachandran plot, as assessed with RAMPAGE (Lovell *et al.*, 2003), shows that all non-glycine amino acid residues lie within allowed regions, with the exception of R_{91A} and D_{109E} for citrate-IPCT. PROCHECK (Laskowski *et al.*, 1993) analysis showed no bad contacts and a final model with a good geometry and stereochemistry.

OVERALL FOLD OF IPCT AND RELATED STRUCTURES

Each monomer of IPCT has overall dimensions of ~45×45×40 $Å^3$ and is organized into two domains, a core domain and a sugar-binding domain. The core domain (residues 16-135 and 173-222), consists of a central seven-stranded mixed β-sheet (β_1-β_7), with order 3214657, where β_6 is antiparallel to the rest, surrounded by six helices, a fold reminiscent of the dinucleotide-binding Rossmann fold (28). At one of its ends, the central β-sheet is topped by a two-stranded β-sheet (β_{5a} and β_{5b}), that participates in a 30 residue long stretch connecting strands β_5 and β_6. One face of the central β-sheet is packed against helices α_1, α_2 and α_4, and the other face, which binds the nucleotide, stacks against helices α_3, α_5 and α_6 (this helix being typical of nucleotide binding proteins). The sugar-binding domain (residues 136-172), comprises a short antiparallel three-stranded β-sheet, which sits against the exposed face of the central β-sheet (where the nucleotide binds), with several conserved residues involved in the sugar-phosphate moiety recognition (*e.g.*, L_{24}, G_{174}, and W_{216}, IPCT numbering). A topology diagram drawn with the program TOPS is presented in Figure 3.7b (35).

The citrate-IPCT crystallographic structure contains 6 molecules in the asymmetric unit, whereas the malonate one has 12 molecules. Despite the different packing in both crystal forms, the dimeric arrangement between two molecules (*e.g.*, chains A-B, C-D or E-F) is quite similar (Figs. 3.7c and 3.7d).

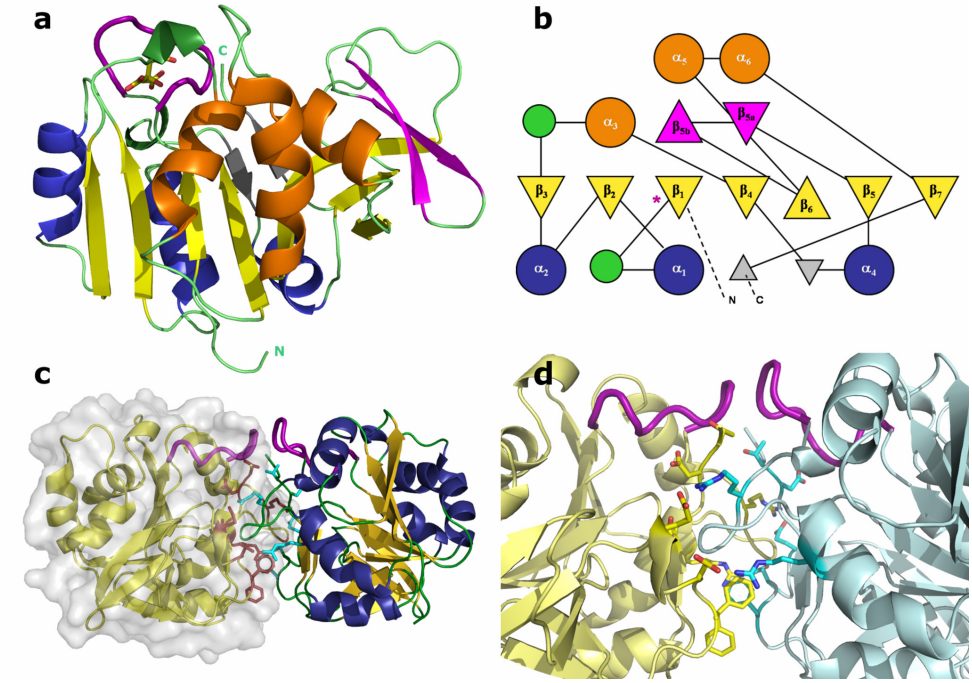


Figure 3.7 – Overall fold of IPCT, topology diagram and dimer interface formed between two neighbouring monomers.

(a) Cartoon representation of IPCT monomer; color scheme according to topology diagram in (b); the active site loop is shown as a thicker purple loop herein and after, and citrate is drawn in sticks (carbon - yellow and oxygen - red); (b) TOPS-generated topology diagram of IPCT. Relevant secondary structure elements are numbered and the location of the active site loop is pointed with an asterisk; (c) Cartoon and surface representation of IPCT dimmer; chain F is drawn yellow with the molecular surface, residues at the interface are coloured in magenta (chain F) and cyan (chain E); (d) Zoomed view at the dimer interface, residues establishing side-chain interactions are displayed (atom colour code: carbon is yellow for chain F and cyan for chain E, oxygen is red and nitrogen is dark blue for both chains).

According to PISA (18), a dissociation energy (ΔGdiss) of 6.4 kcal.mol-1 is predicted for this dimeric assembly in solution. The interface is mainly hydrophilic with an area around 1110 Å^2 corresponding to *ca.* 11% of the total solvent-accessible area of each monomer. Most residues that contribute to the interface stabilization are located in loops between β_1-α_1, β_5-β_{5a} and at the C-terminal loop. The interface is

stabilized by 14 H-bonds, 2 of which are salt-bridges. Several hydrogen bonds are established between main-chain atoms of both chains, namely between R_{41}-D_{220}, G_{43}-D_{218}, and D_{144}-G_{43}.

Additional stabilization is achieved by side-chain/main-chain interactions between R_{50}-F_{142}, T_{221}-L_{39} and D_{218}-G_{44}. Moreover, $R_{41E/F}(NH_2)$ is H-bonded to $E_{147F/E}(OE_2)$. Molecular mass determinations by gel filtration indicate that there is a mixture of monomers and dimers in solution, with prevalence of the monomeric form. At this stage we think that the relevance of the interface interactions in the crystal structure for the physiological assembly of the enzyme can not be fully assessed before the structure of the whole IPCT/DIPPS bifunctional protein is characterized.

Comparison of one molecule of IPCT (monomer) with other structures in the Protein Data Bank (PDB) using the DALI server (14), reveals that the highest matches are found with the N-terminal domain of the bifunctional N-acetylglucosamine-1-phosphate uridylyltransferases (GlmU) from different sources (PDB codes: 1G97, 2OI6, 1HM9, 1HV9, 1FWY and 2V0I) and with glucose-1-phosphate thymidylyl-/uridylyl-transferases (PDB codes: 2PA4, 1FXO, 2E3D and 2UX8), showing Z scores between 20 and ~18, and rms deviations ranging from 2.5 to 2.8 Å for ~190 superimposed C_α atoms. Other cytidylyltransferases (CTP:phosphocholine-, 3-deoxy-manno-octulosonate- and D-glucose-1-phosphate-), with PDB codes 1JYL, 1GQC, and 1TZF, respectively, have Z values from 17.7 down to ~14.5. Interestingly, GlmU is also a bifunctional enzyme, containing an homologous nucleotidyltransferase domain at the N-terminus but with an acetyltransferase domain at the C-terminus, which adopts a left-handed parallel β-helix structure LβH). The biological assembly of the bifunctional GlmU is a trimer, whereas most members of other nucleotidyltransferases are dimers or tetramers. These structural-related enzymes share between 15% and 22% of sequence identity with IPCT and show conservation in the core region, consisting of a mixed β-sheet of seven strands with the same order and orientation as IPCT, a fold typical of nucleotide-diphospho-sugar transferases (SCOP Superfamily 53448).

Interestingly, the 32 best hits retrieved in a blast search using the sequence of *Arc. fulgidus* IPCT belong to hyperthermophilic or thermophilic organisms, probably reflecting the specialized function of IPCT in DIP synthesis (Fig. 3.8); the best match relates to proteins of *Thermococcus* spp. that display sequence identities between 49% and 56%.

As expected, the *Arc. fulgidus*' protein appears in a cluster dominated by members of the Euryarchaeota; curiously, three representatives of the domain Bacteria also cluster in this group, which is clearly separated from a second major group comprising bacteria of the order Thermotogales. The closest mesophilic homologues share *ca.* 30% sequence identity with *Arc. fulgidus* IPCT and belong to the bacterial genera *Gluconacetobacter* and *Granulibacter*. Whether these mesophilic counterparts are able to use inositol-1P and CTP remains elusive.

THE ACTIVE SITE

A citrate molecule is observed in the active site pocket in three out of six molecules of IPCT crystallized using citrate as precipitant (chains A, D and E, Fig. 3.7a). The citrate is establishing hydrogen bonds with the main-chain nitrogen atoms of G27, L28, K37 and R68. Its presence in only some chains might be related with the position of the disordered loop comprising G27-K37. This loop is not visible in most chains of both citrate- and malonate-IPCT structures and, if visible, it adopts different conformations (Fig. 3.9).

This loop encompasses the signature sequence G-X-G-T-(R/S)-X_4-P-K of nucleotidyltransferases and, in IPCT, it starts at G27 and has the sequence G-L-G-T-R-L-G-G-V-P-K. In the apo-RmlA structure (PDB 1FZW), this loop is also disordered but becomes traceable in the thymidine containing complexes (PDB 1G2V and 1G0R) (Blankenfeldt *et al.*, 2000).

Attempts to obtain structures of IPCT in complex with substrates CTP, inositol-1P and CTP:inositol-1P, either by co-crystallization or soaking experiments, did not succeed in both crystal forms. However, due to the core fold conservation among sugar nucleotidyltransferase structures, and based on the DALI results, we were able to identify the catalytic region of IPCT.

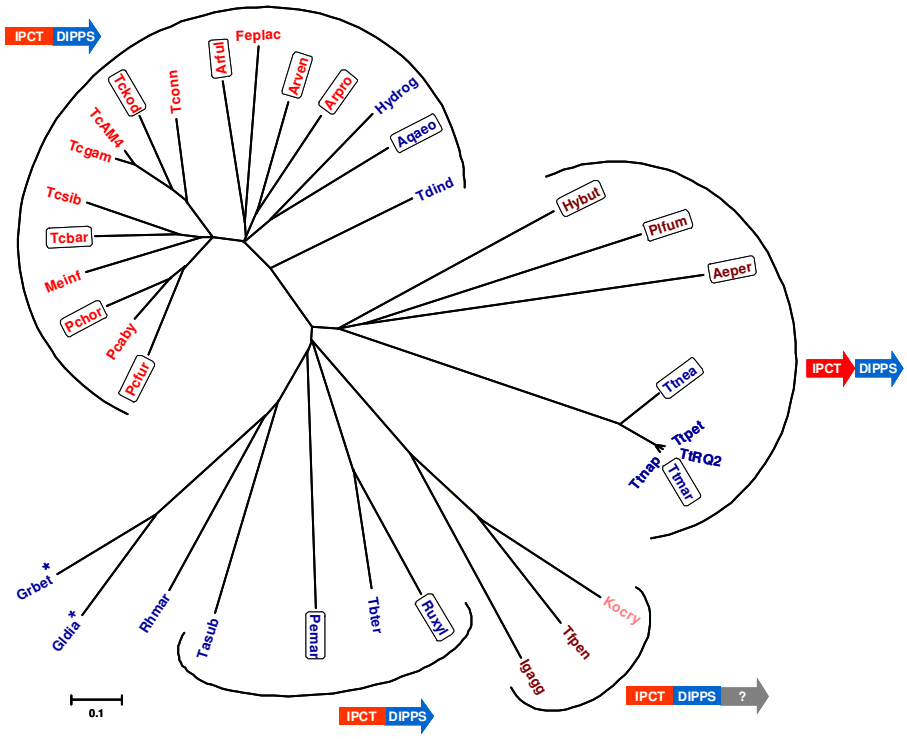


Fig. 3.8. Unrooted phylogenetic tree based on available homologues of *A. fulgidus* IPCT. Homologues of IPCT were retrieved from the existing protein databases (GenBank) using BLAST. Amino acid sequences of bifunctional IPCT/DIPPS were artificially separated. MEGA 4.1 software (Tamura *et al.*, 2007) was used for sequence alignment and to generate the phylogenetic tree using the neighbour-joining method. Bar, 0.1 change per site. The bacterial IPCT are depicted in blue, the euryarchaeotal in red, crenarchaeotal in brown, and korarchaeotal in pink. Mesophilic homologues are labelled with an asterisk. The empty box designates the organisms known to accumulate DIP. IPCT DIPPS and IPCT DIPPS indicate that the IPCT and DIPPS genes are fused and separated, respectively. The genes containing IPCT, DIPPS and an extra domain are indicated by IPCT DIPPS ?. Abbreviations: Arful, *Archaeoglobus fulgidus*; Arpro, *Archaeoglobus profundus*; Arven, *Archaeoglobus veneficus*; Tconn, *Thermococcus onnurineus* NA1; Tcsib, *Thermococcus sibiricus* MM739; Tcgam, *Thermococcus gammatolerans* EJ3; Tckod, *Thermococcus kodakarensis* KOD1; TcAM4, *Thermococcus* sp. AM4, Tcbar, *Thermococcus barophilus* MP; Pchor, *Pyrococcus horikoshii* OT3; Pcaby, *Pyrococcus abyssi*; Pcfur, *Pyrococcus furiosus* DSM 3638; Feplac, *Ferroglobus placidus* DSM 10642; Meinf, *Methanocaldococcus infernus* ME; Plfum, *Pyrolobus fumarii*; Aeper, *Aeropyrum pernix* K1; Hybut, *Hyperthermus butylicus* DSM 5456; Igagg, *Ignisphaera aggregans* DSM 17230; Kocry, Candidatus *Korarchaeum cryptofilum* OPF8; Hydrog, *Hydrogenivirga* sp. 128-5-R1-1; Ttnap, *Thermotoga naphthophila* RKU-10; Ttnea, *Thermotoga neapolitana* DSM 4359; Ttmar, *Thermotoga maritima*; Ttpet, *Thermotoga petrophila*; TtRQ2, *Thermotoga* sp. RQ2; Aqaeo, *Aquifex aeolicus* VF5; Ruxyl, *Rubrobacter xylanophilus* DSM 9941; Pemar, *Persephonella marina* EX-H1; Tbter, *Thermobaculum terrenum* ATCC BAA-798; Tasub, *Thermaerobacter subterraneus* DSM 13965; Tdind, *Thermodesulfatator indicus*; Tfpen, *Thermofilum pendens* Hrk 5; Rhmar, *Rhodothermus marinus* DSM 4252; Grbet, *Granulibacter bethesdensis* CGDNIH1; Gldia, *Gluconacetobacter diazotrophicus* PAl 5.

The most homologous structure to IPCT, besides GlmU, is glucose-1-phosphate thymidylyltransferase (RmlA) from *Pseudomonas* (*P.*) *aeruginosa* (rmsd of 2.57 Å for 188 C_α atoms) (Blankenfeldt *et al.*, 2000). RmlA is involved in the synthesis of dTDP-L-rhamnose, an important component in the cell wall of many microorganisms. Superposition of various structures of RmlA apo- and complexes (dTTP, glucose-1-phosphate and dTDP-D-glucose) shows no relevant structural changes upon ligand binding, as also observed with other nucleotidyltransferase structures (Fig. 3.10a). We identified the relevant catalytic residues of IPCT based on various structures of RmlA. Moreover, this structural analysis was extended to include other homologous nucleotidyltransferases, such as CTP:phosphocoline cytidylyltransferase from *Streptococcus pneumoniae* (PDB 1JYL) (Kwak *et al.*, 2002), and α-D-glucose-1-phosphate cytidylyltransferase from *Salmonella typhii* (PDB 1TZF) (Koropatkin & Holden, 2004), leading to similar results. For the sake of simplicity, we will refer to the comparison with RmlA throughout the manuscript.

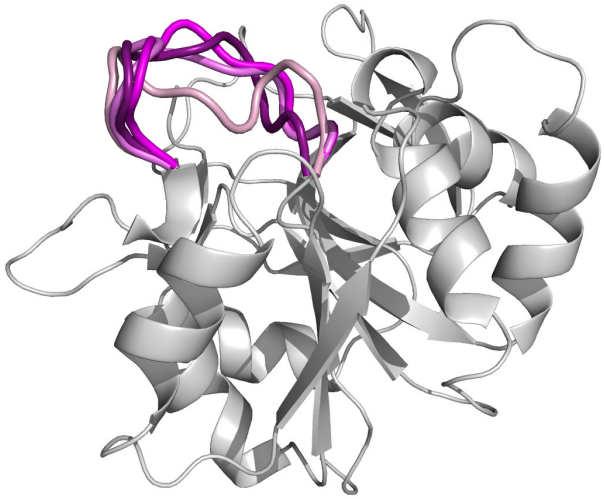


Figure 3.9 - Cartoon representation of IPCT highlighting the mobility of the loop G27-K37. This cartoon shows the several conformations the loop G27-K37 can adopt (chains D and E from the citrate-IPCT structure and chains E and F from the malonate-IPCT structure in different shades of pink); for the sake of clarity, only on complete chain is visible (chain D from the citrate-IPCT, in grey), and the citrate molecules are not shown.

The active site of IPCT is thus located in a pocket formed by the sugar- and nucleotide-binding domains of the enzyme (Fig. 3.10b). The pocket entrance comprises a series of polar or charged residues: K_{37}, R_{68}, E_{93}, K_{163}, E_{194}, S_{198} and D_{220}. The cavity is formed by residues from strands β_1, β_2, β_4 and β_6 (namely L_{24}, A_{25}, V_{65}, A_{66}, M_{115}-H_{118}, D_{172}, G_{174} and F_{176}), helices α_3 and α_6 (residues G_{95}, N_{96} and V_{201}), and residues P_{38}, S_{99} and W_{216}.

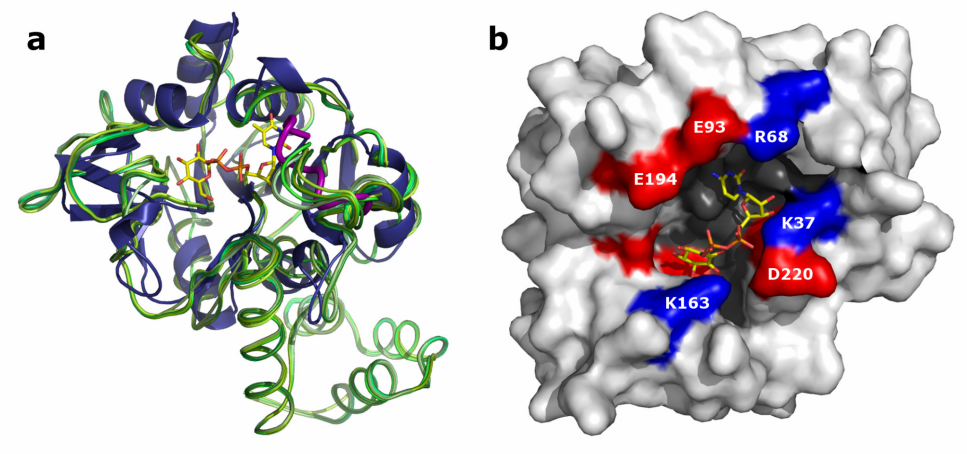


Figure 3.10 - (a) Superposition of IPCT (cartoon in blue) with several structures of glucose-1-phosphate thymidylyltransferase (RmlA) from *P. aeruginosa*: apo form (1FZW, ribbon in dark green), and in complex with different ligands (PDB codes 1FXO, 1G0R, 1G2V, 1G1L, 1G23 and 1G3L, ribbon in different shades of green). dTDP-D-glucose (from 1G1L) is shown in sticks (C- yellow, O- red, N- blue and P-orange); (b) IPCT active site pocket with fitted CDP-inositol; residues at the entrance are labelled.

The binding sites of nucleotides are highly conserved comprising a Rossmann fold motif ($\alpha\beta\alpha\beta\alpha$). In IPCT, the CTP should be nested in a groove composed by the N-termini of helices α_2 and α_5, and the C-termini of strands β_1, β_2 and β_4. Based on the structure of RmlA in complex with dTTP (PDB 1G2V) (Blankenfeldt *et al.*, 2000) we have fitted a CTP molecule into the active site of IPCT (Fig. 3.12a). In RmlA, several conserved amino acid residues, namely G_{10}, T_{14}, R_{15}, K_{25} (RmlA numbering), are involved in the stabilization of the nucleotide. In IPCT,

the phosphate moiety may be stabilized by T_{30}, R_{31} and K_{37} (corresponding to T_{14}, R_{15} and K_{25} in RmlA, Figure 3.11), of which T_{30} and R_{31} are located in the disordered loop.

Furthermore, A_{26} and D_{117} (G_{10} and D_{110} in RmlA), may establish several H-bonds to the ribose and pyrimidine ring, respectively. Q_{82}, a catalytically important residue in RmlA (replaced by H_{89} in IPCT, Figure 3.12), along with G_{10} (A_{26} in IPCT) and G_{87} (G_{95} in IPCT), provide specificity for thymidine, while a tight loop formed by Q_{82}-G_{87} is responsible for the specificity of pyrimidine over purine bases (Blankenfeldt *et al.*, 2000).

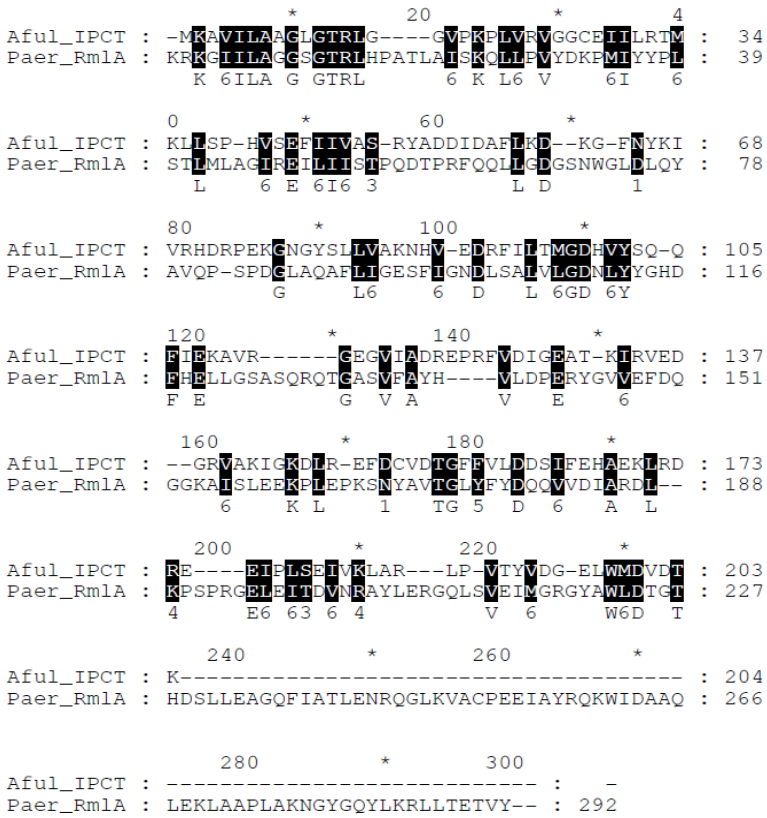


Figure 3.11 - Amino acid sequence alignment of *Archaeoglobus fulgidus* IPCT and *Pseudomonas aeruginosa* RmlA.

The sequences were obtained from the X-ray structures deposited in the Protein Data Bank, are identified as Aful_IPCT and Paer_RmlA, respectively, and were aligned using the program ClustalW (Chenna *et al.*, 2003).

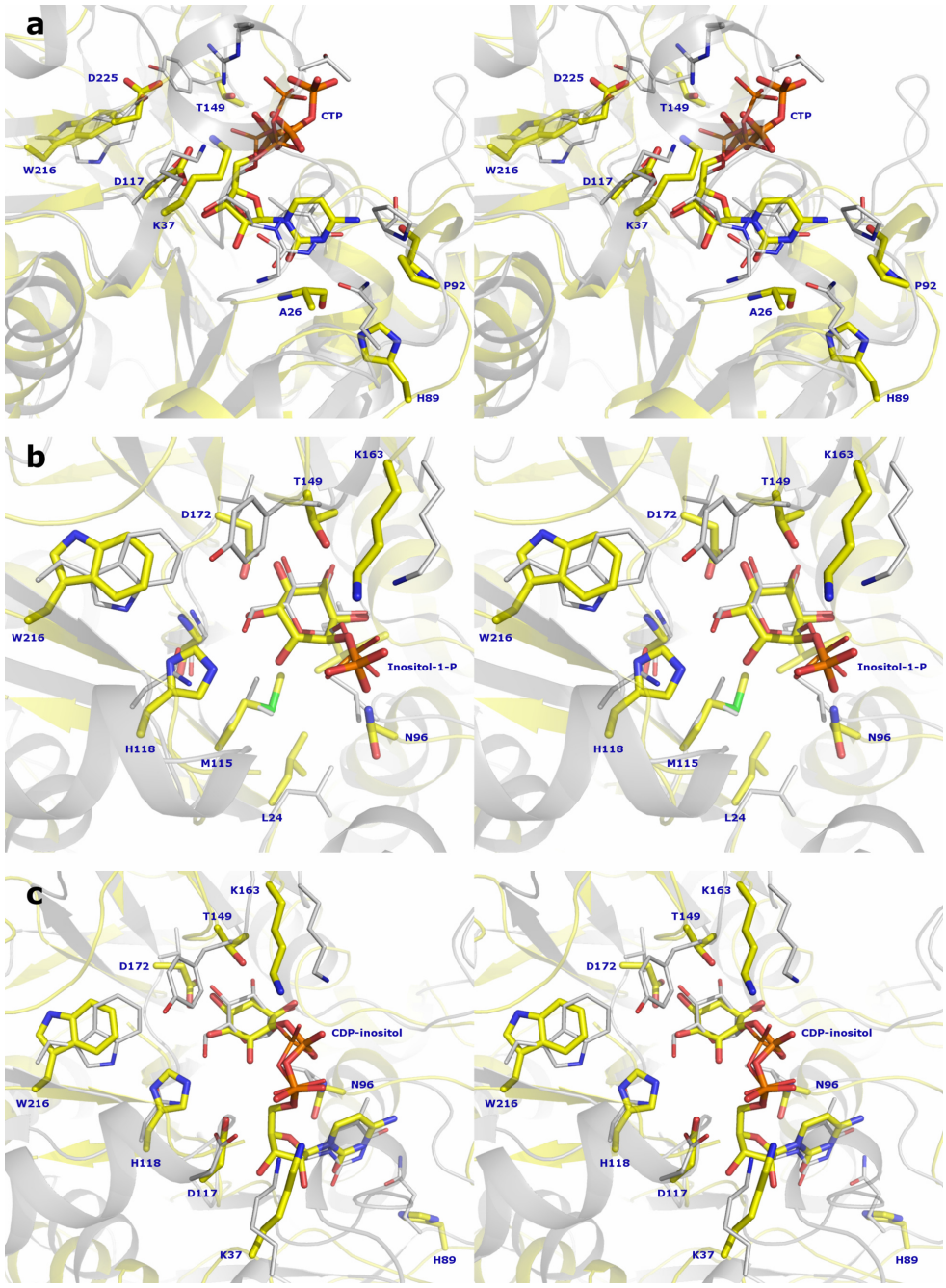


Figure 3.12 - Stereo view of structural superposition of (a) IPCT fitted with CTP and RmlA with dTTP (1G2V); (b) IPCT fitted with inositol-1P and RmlA with G-1P (1G23); (c) IPCT structure fitted with CDP-inositol and RmlA with dTDP-D-glucose (1G1L); IPCT ribbon is drawn in yellow and RmlA in grey with ligands shown in sticks (same atom color code as in figure 4, with phosphate atoms in orange, sulphur atoms in green and carbon atoms of RmlA in grey).

IPCT is highly specific for CTP as no activity was detected using UTP, ATP and GTP as nucleotide donors (Rodrigues *et al.*, 2007). The structural basis for this substrate preference becomes apparent from our CTP-fitted model: the NH_2 amino group of cytidine is able to form an H-bond with the main-chain oxygen of P_{92}, which is not possible with UTP (NH2 substituted by a carbonyl group). By analogy with RmlA, the tight loop (H_{89}-G_{95} in IPCT) is proposed to be responsible for the specificity of cytosine over purine bases.

Inositol-1P was docked into IPCT by superposition with the structure of RmlA bound to glucose-1-phosphate (PDB code 1G23) (Blankenfeldt *et al.*, 2000), where the ligand is stabilized by several hydrogen bonds and van der Waals interactions (Blankenfeldt *et al.*, 2000). The same type of interactions is expected in the inositol-1P fitted structure of IPCT since most of the residues involved in substrate binding are conserved among nucleotidyltransferases (Fig. 3.12b).

The side-chains of H_{118}, T_{149} and D_{172} (N_{111}, Y_{145} and V_{172} in RmlA), can establish H-bonds with the inositol ring, and N_{96} and K_{163} (L_{88} and K_{162} in RmlA), with the phosphate moiety. Additionally, several hydrophobic residues are present in the surroundings of the ligand, such as M_{115}, G_{174}, L_{197} and W_{216} in IPCT (L_{108}, G_{174}, I_{199} and W_{223} in RmlA). The structure of IPCT fitted with CDP-inositol (Fig. 3.12c) reveals protein interactions quite similar to those found in inositol-1P and CTP docked models and hence will not be subjected to further discussion.

Mg^{2+} is known to be essential for the catalytic reaction of nucleotidyltransferases but its role is not fully understood. Blakenfeldt *et al.* (Blankenfeldt *et al.*, 2000) proposed that this cation could be coordinating the β- and γ-phosphates of dTTP in a chelated manner, leaving the active site as a complex with pyrophosphate. However, several structures, such as glucose-1-phosphate uridylyltransferase from *E. coli* (Thoden & Holden, 2007), and glucose-1-phosphate cytidylyltransferase from *Salmonella typhi* (Koropatkin & Holden, 2004), show the magnesium bound to the final product of each enzyme, specifically to the α- and β-phosphoryl oxygens of the NDP-sugar product. A more direct role of Mg^{2+} could be to accelerate catalysis by positioning the phosphate oxygen of the

sugar-phosphate adjacent to the α-phosphate of CTP to align the incoming second substrate for in-line nucleophilic attack, as suggested for CTP:phosphocholine cytidylyltransferase of *Streptococcus pneumoniae* (Kwak *et al.*, 2002).

CONCLUDING REMARKS

The crystallization of IPCT was a crucial step toward the elucidation of the catalytic mechanism of this unique enzyme able to catalyze the synthesis of CDP-inositol. The occurrence of this metabolite was first disclosed by Borges and co-workers (Borges *et al.*, 2006).

The first 3D-structure of CTP:inositol-1-phosphate cytidylyltransferase (IPCT) revealed an overall architecture similar to the dinucleotide-binding Rossmann fold and a fairly good match with the structures of glucose-1-phosphate thymidylyl-/uridylyl-transferases. It is surprising that inositol-1P, a novel substrate for nucleotidyltransferases, does not imply a distinctive structural layout for catalysis. On the other hand, it is curious that IPCT is structurally less related to known cytidylyltransferases, namely those recognizing other polyol-phosphates such as glycerol-, ribitol- or methylerythritol-phosphate.

A blast in the genome databases with the *Arc. fulgidus* IPCT sequence clearly shows a strong correlation of the closest matches with thermo- or hyperthermophilic host organisms. This feature underlies the commitment of CDP-inositol to the synthesis of DIP, a compatible solute thus far exclusive to (hyper)thermophiles. This apparently specialized role of CDP-inositol is an intriguing finding that challenges the usual, multitasking character of most metabolites.

ACKNOWLEDGEMENTS

We thank the ESRF for financial and technical support for data collection. The authors acknowledge M. Henriques and I. Pacheco for technical assistance and Pedro Matias for help during data collection. The NMR spectrometers are part of the National NMR Network (REDE/1517/RMN/2005), supported by Programa Operacional Ciência e Inovação and Fundação para a Ciência e a Tecnologia.

3.8. REFERENCES

Abrahams, J. P. & A. G. Leslie, (1996) Methods used in the structure determination of bovine mitochondrial F1 ATPase. *Acta Crystallogr D Biol Crystallogr* **52**: 30-42.

Baur, S., J. Marles-Wright, S. Buckenmaier, R. J. Lewis & W. Vollmer, (2009) Synthesis of CDP-activated ribitol for teichoic acid precursors in *Streptococcus pneumoniae*. *J Bacteriol* **191**: 1200-1210.

Blankenfeldt, W., M. Asuncion, J. S. Lam & J. H. Naismith, (2000) The structural basis of the catalytic mechanism and regulation of glucose-1-phosphate thymidylyltransferase (RmlA). *Embo J* **19**: 6652-6663.

Blount, P. & P. C. Moe, (1999) Bacterial mechanosensitive channels: integrating physiology, structure and function. *Trends Microbiol* **7**: 420-424.

Borges, N., L. G. Goncalves, M. V. Rodrigues, F. Siopa, R. Ventura, C. Maycock, P. Lamosa & H. Santos, (2006) Biosynthetic pathways of inositol and glycerol phosphodiesters used by the hyperthermophile *Archaeoglobus fulgidus* in stress adaptation. *J Bacteriol* **188**: 8128-8135.

Borges, N., J. D. Marugg, N. Empadinhas, M. S. da Costa & H. Santos, (2004) Specialized roles of the two pathways for the synthesis of mannosylglycerate in osmoadaptation and thermoadaptation of *Rhodothermus marinus*. *J Biol Chem* **279**: 9892-9898.

Bradford, M. M., (1976) A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding. *Anal Biochem* **72**: 248-254.

Bricogne, G., E. Blanc, M. Brandl, C. Flensburg, P. Keller, W. Paciorek, P. Roversi, O. S. Smart, C. Vonrhein & T. Womack, (2009) BUSTER, version 2.8.0. In: Global Phasing Ltd. Cambridge, United Kingdom, pp.

Brown, A. D., (1976) Microbial water stress. *Bacteriol Rev* **40**: 803-846.

Chen, L., E. T. Spiliotis & M. F. Roberts, (1998) Biosynthesis of di-*myo*-inositol-1,1'-phosphate, a novel osmolyte in hyperthermophilic archaea. *J Bacteriol* **180**: 3785-3792.

Chenna, R., H. Sugawara, T. Koike, R. Lopez, T. J. Gibson, D. G. Higgins & J. D. Thompson, (2003) Multiple sequence alignment with the Clustal series of programs. *Nucleic Acids Res* **31**: 3497-3500.

Ciulla, R. A., S. Burggraf, K. O. Stetter & M. F. Roberts, (1994) Occurrence and Role of Di-*myo*-Inositol-1,1'-Phosphate in *Methanococcus igneus*. *Appl Environ Microbiol* **60**: 3660-3664.

Cowtan, K., (1994) 'dm': An automated procedure for phase improvement by density modification. *Joint CCP4 and ESF-EACBM Newsletter on Protein Crystallography* **31**: 34-38.

Cowtan, K., (2006) The Buccaneer software for automated model building. 1. Tracing protein chains. *Acta Crystallogr D Biol Crystallogr* **62**: 1002-1011.

Csonka, L. N., (1989) Physiological and genetic responses of bacteria to osmotic stress. *Microbiol Rev* **53**: 121-147.

Csonka, L. N. & A. D. Hanson, (1991) Prokaryotic osmoregulation: genetics and physiology. *Annu Rev Microbiol* **45**: 569-606.

da Costa, M. S., H. Santos & E. A. Galinski, (1998) An overview of the role and diversity of compatible solutes in Bacteria and Archaea. *Adv Biochem Eng Biotechnol* **61**: 117-153.

de la Fortelle, E. & G. Bricogne, (1997) Maximum-likelihood heavy-atom parameter refinement for multiple isomorphous replacement and multiwavelength anomalous diffraction methods. *Methods Enzymol* **276**: 472-494.

Delamarche, C., D. Thomas, J. P. Rolland, A. Froger, J. Gouranton, M. Svelto, P. Agre & G. Calamita, (1999) Visualization of AqpZ-mediated water permeability in *Escherichia coli* by cryoelectron microscopy. *J Bacteriol* **181**: 4193-4197.

DeLano, W., (2002) The PyMOL Molecular Graphics System. In. San Carlos, CA: DeLano Scientific, pp.

Empadinhas, N. & M. S. da Costa, (2006) Diversity and biosynthesis of compatible solutes in hyperthermophiles. *Int Microbiol* **9**: 199-206.

Emsley, P., B. Lohkamp, W. G. Scott & K. Cowtan, Features and development of Coot. *Acta Crystallogr D Biol Crystallogr* **66**: 486-501.

Evans, P. R., (1997) *Proceedings of the CCP4 Study Weekend. Recent Advances In Phasing*: 97-102.

Goncalves, L. G., R. Huber, M. S. da Costa & H. Santos, (2003) A variant of the hyperthermophile *Archaeoglobus fulgidus* adapted to grow at high salinity. *FEMS Microbiol Lett* **218**: 239-244.

Holm, L. & C. Sander, (1996) Mapping the protein universe. *Science* **273**: 595-603.

Kabsch, W., (1993) Automatic processing of rotation diffraction data from crystals of initially unknown symmetry and cell constants. *Journal of Applied Crystallography* **26**: 795-800.

Kempf, B. & E. Bremer, (1998) Uptake and synthesis of compatible solutes as microbial stress responses to high-osmolality environments. *Arch Microbiol* **170**: 319-330.

Kim, H., J. Choi, T. Kim, N. K. Lokanath, S. C. Ha, S. W. Suh, H. Y. Hwang & K. K. Kim, (2010) Structural basis for the reaction mechanism of UDP-glucose pyrophosphorylase. *Mol Cells* **29**: 397-405.

Koropatkin, N. M. & H. M. Holden, (2004) Molecular structure of alpha-D-glucose-1-phosphate cytidylyltransferase from *Salmonella typhi*. *J Biol Chem* **279**: 44023-44029.

Kozono, D., X. Ding, I. Iwasaki, X. Meng, Y. Kamagata, P. Agre & Y. Kitagawa, (2003) Functional expression and characterization of an archaeal aquaporin. AqpM from *Methanothermobacter marburgensis*. *J Biol Chem* **278**: 10649-10656.

Krissinel, E. & K. Henrick, (2004) Secondary-structure matching (SSM), a new tool for fast protein structure alignment in three dimensions. *Acta Crystallogr D Biol Crystallogr* **60**: 2256-2268.

Kwak, B. Y., Y. M. Zhang, M. Yun, R. J. Heath, C. O. Rock, S. Jackowski & H. W. Park, (2002) Structure and mechanism of CTP:phosphocholine cytidylyltransferase (LicC) from *Streptococcus pneumoniae*. *J Biol Chem* **277**: 4343-4350.

Laemmli, U. K., (1970) Cleavage of structural proteins during the assembly of the head of bacteriophage T4. *Nature* **227**: 680-685.

Laskowski, R. A., M. W. MacArthur, D. S. Moss & J. M. Thornton, (1993) PROCHECK: a program to check the stereochemical quality of protein structures. *Journal of Applied Crystallography* **26**: 283-291.

Leandro, P. & C. M. Gomes, (2008) Protein misfolding in conformational disorders: rescue of folding defects and chemical chaperoning. *Mini Rev Med Chem* **8**: 901-911.

Lee, J. K., D. Kozono, J. Remis, Y. Kitagawa, P. Agre & R. M. Stroud, (2005) Structural basis for conductance by the archaeal aquaporin AqpM at 1.68 A. *Proc Natl Acad Sci U S A* **102**: 18932-18937.

Loewus, F., (1990) *Inositol biosynthesis*. Wiley-Liss, Inc., New York, N. Y.

Loewus, F., J. D. Everard & K. A. Young, (1990) *Inositol metabolism: precursor role and breakdown*. Wiley-Liss, Inc., New York, N. Y.

Lovell, S. C., I. W. Davis, W. B. Arendall, 3rd, P. I. de Bakker, J. M. Word, M. G. Prisant, J. S. Richardson & D. C. Richardson, (2003) Structure validation by Calpha geometry: phi,psi and Cbeta deviation. *Proteins* **50**: 437-450.

Martins, L. O. & H. Santos, (1995) Accumulation of Mannosylglycerate and Di-*myo*-Inositol-Phosphate by *Pyrococcus furiosus* in Response to Salinity and Temperature. *Appl Environ Microbiol* **61**: 3299-3303.

Matthews, B. W., (1968) Solvent content of protein crystals. *J Mol Biol* **33**: 491-497.

McCoy, A. J., R. W. Grosse-Kunstleve, P. D. Adams, M. D. Winn, L. C. Storoni & R. J. Read, (2007) Phaser crystallographic software. *J Appl Crystallogr* **40**: 658-674.

Muller, V., R. Spanheimer & H. Santos, (2005) Stress response by solute accumulation in archaea. *Curr Opin Microbiol* **8**: 729-736.

Murshudov, G. N., A. A. Vagin & E. J. Dodson, (1997) Refinement of macromolecular structures by the maximum-likelihood method. *Acta Crystallogr D Biol Crystallogr* **53**: 240-255.

Ren, Q., K. H. Kang & I. T. Paulsen, (2004) TransportDB: a relational database of cellular membrane transport systems. *Nucleic Acids Res* **32**: D284-288.

Roberts, M. F., (2000) Osmoadaptation and osmoregulation in archaea. *Front Biosci* **5**: D796-812.

Roberts, M. F., (2005) Organic compatible solutes of halotolerant and halophilic microorganisms. *Saline Systems* **1**: 5.

Rodionov, D. A., O. V. Kurnasov, B. Stec, Y. Wang, M. F. Roberts & A. L. Osterman, (2007) Genomic identification and in vitro reconstitution of a complete biosynthetic pathway for the osmolyte di-*myo*-inositol-phosphate. *Proc Natl Acad Sci U S A* **104**: 4279-4284.

Rodrigues, M. V., N. Borges, C. P. Almeida, P. Lamosa & H. Santos, (2009) A unique beta-1,2-mannosyltransferase of *Thermotoga maritima* that uses di-*myo*-inositol phosphate as the mannosyl acceptor. *J Bacteriol* **191**: 6105-6115.

Rodrigues, M. V., N. Borges, M. Henriques, P. Lamosa, R. Ventura, C. Fernandes, N. Empadinhas, C. Maycock, M. S. da Costa & H. Santos, (2007) Bifunctional CTP:inositol-1-phosphate cytidylyltransferase/CDP-inositol:inositol-1-phosphate transferase, the key enzyme for di-*myo*-inositol-phosphate synthesis in several (hyper)thermophiles. *J Bacteriol* **189**: 5405-5412.

Roesser, M. & V. Muller, (2001) Osmoadaptation in bacteria and archaea: common principles and differences. *Environ Microbiol* **3**: 743-754.

Santos, H. & M. S. da Costa, (2002) Compatible solutes of organisms that live in hot saline environments. *Environ Microbiol* **4**: 501-509.

Savage, D. F., P. F. Egea, Y. Robles-Colmenares, J. D. O. C. Iii & R. M. Stroud, (2003) Architecture and Selectivity in Aquaporins: 2.5 Å X-Ray Structure of Aquaporin Z. *PLoS Biol* **1**: e72.

Scholz, S., J. Sonnenbichler, W. Schafer & R. Hensel, (1992) Di-*myo*-inositol-1,1'-phosphate: a new inositol phosphate isolated from *Pyrococcus woesei*. *FEBS Lett* **306**: 239-242.

Sheldrick, G. M., (2008) A short history of SHELX. *Acta Crystallogr A* **64**: 112-122.

Tamura, K., J. Dudley, M. Nei & S. Kumar, (2007) MEGA4: Molecular Evolutionary Genetics Analysis (MEGA) software version 4.0. *Mol Biol Evol* **24**: 1596-1599.

Thoden, J. B. & H. M. Holden, (2007) The molecular architecture of glucose-1-phosphate uridylyltransferase. *Protein Sci* **16**: 432-440.

Wood, J. M., (1999) Osmosensing by bacteria: signals and membrane-based sensors. *Microbiol Mol Biol Rev* **63**: 230-262.

Zwickl, P., S. Fabry, C. Bogedain, A. Haas & R. Hensel, (1990) Glyceraldehyde-3-phosphate dehydrogenase from the hyperthermophilic archaebacterium *Pyrococcus woesei*: characterization of the enzyme, cloning and sequencing of the gene, and expression in Escherichia coli. *J Bacteriol* **172**: 4329-4338.

Chapter 4

CONCLUDING REMARKS

In the beginning of this third millennium, technology has already surpassed the dreams of many. We have not yet reached the point of having space ships travelling around the Universe (who knows what they would find?), and Moon-based facilities, like it was thought to happen by the end of the 1900s and presented in the plot of "Space: 1999" series. However, who is to say that our children or grandchildren will not witness those, so far, unthinkable wonders?

Great opportunities for exploration await us as we unravel mysteries and gather knowledge and answers in biochemistry and systems biology, genetics and biomedicine, astrobiology and chemistry of the elements. The work presented in this thesis was a small contribution to the present knowledge and a step less needed for the forthcoming generations to engage. It gathered a few answers in structural biology and biochemistry of each presented enzyme while opening new perspectives and listed a new set of questions. The main conclusions were drawn in each presented chapter and here only some concluding remarks and future perspectives will be referred.

The work on the **sulfide:quinone oxidoreductase** (SQR) from *A. ambivalens* (Brito *et al.*, 2009), was a landmark in the sulfur metabolism field since it was the first SQR structure to be solved. Unfortunately, the structure did not had the impact we anticipated since questions like "how does membrane attachment occurs?" and "how does quinone binds?" could not be definitely answered due to incompleteness of the crystallographic model. However, the two later structures of SQRs, from *Aq. aeolicus* (Marcia *et al.*, 2009), and *Ac. ferrooxidans* (Cherney *et al.*, 2010), corroborated some of our proposals, namely the way that membrane attachment occurs, through two C-terminus amphipathic helices, and the quinone binding pocket. Moreover, the proposed mechanisms for the reductive half-reaction for those two enzymes, which were solved in complex with quinones, quinone analogues and, for the *Ac. ferrooxidans* enzyme, a $C_{160}A$ mutant, are in line with what we proposed for the *A. ambivalens* enzyme.

In *A. ambivalens* SQR, two cysteine residues (C_{178} and C_{350}), are involved in sulfide oxidation and are responsible for the transfer of electrons to the quinone pool. A third cysteine residue (C_{129}), is not involved in the oxidation mechanism and

covalently binds the cofactor FAD. In *Aq. aeolicus*, this residue corresponds to C_{124}, which is too far away from the FAD and does not covalently binds to it. Instead, a labile cysteine persulfide bond is observed that opened the question of this third cysteine residue being also involved in the sulfide oxidation mechanism. Since there are SQRs with only two cysteines (corresponding to the C-terminal cysteines, C_{178} and C_{350} in *A. ambivalens*), or even with only one, like the *Sulfolobus tokodaii* homologue, it would be interesting to characterize and solve the structure of such enzymes.

Our Laboratory may start a collaboration with the Laboratory of Doctor Thomas Hanson (College of Marine and Earth Studies, University of Delaware, USA), who recently functionally analysed three SQR homologues in *Chlorobaculum* (*C.*) *tepidum* (CT0117, CT0876 and CT1087), for their ability to catalyze sulfide-dependent ubiquinone reduction and if they are required for growth on sulfide (Chan *et al.*, 2009). In this work, Hanson and coworkers showed that, depending on the growth conditions (*e.g.*, concentration of sulfide in the growth media), the three transcripts show different expression levels engaging the cells different metabolic responses (Chan *et al.*, 2009). It will be very interesting to solve the structures of these three enzymes and try to infer from them what differences trigger these different behaviours when growing on sulfide.

When SQR from *A. ambivalens* was first isolated it was characterized as a type-2 NADH:quinone oxidoreductase having a FMN cofactor covalently bound to the protein through a histydinyl residue (Gomes *et al.*, 2001). The crystal structure presented here showed that the cofactor was indeed an FAD covalently bound to the enzyme through a cysteinyl residue. Moreover the NADH:quinone oxidoreductase activity was due to the exposure of the FAD to the solvent which lead to an unspecific reaction with NADH. Work on a true type-2 NADH:quinone oxidoreductase from *A. ambivalens* is currently undergoing in our Laboratory.

The thermostability of a protein results from versatile stabilization strategies that correspond to subtle structural modifications (Stelter *et al.*, 2008), i.e., quite minor changes in the structure of an enzyme can lead to large changes in thermostability (Vieille & Zeikus, 2001). Although there is no single universal

mechanism that promotes thermostability there are some common features that have been identified as contributing to thermostability (Li *et al.*, 2005). Among these are: an increase in hydrogen bonds, additional or improved electrostatic interactions caused by salt bridges or networks thereof, optimized hydrophobic interactions, increased compactness and packing densities, increased polar compared with non-polar surface areas, improved binding of metal ions, replacement of energetically unfavourable conformations by G, truncation of solvent-exposed loops, a higher number of Pro's and β-branched amino acids in loops, association to oligomers and reductions of the content of the thermally labile amino acids N, Q, C and M (Sterner & Liebl, 2001).

Looking to the primary sequence of the **CTP:inositol-1-phosphate cytidylyltranferase** (IPCT) from *Arc. fulgidus*, two striking features immediately arise: first, the high content of hydrophobic residues (103 out of the 232 residues of the construct used in this study are either A, F, I, L, M, P, V or W), and, second, the high percentage of β-branched amino acids (ca. 21% of the residues are I, T or V). All of these residues are spread all over the structure of the protein but they are mostly found in the core of the structure interacting with each other and more so on β-strands. These two features are in agreement with what Sterner *et al.* (Sterner & Liebl, 2001) proposed as stabilization characteristics for (hyper)thermophilic enzymes. Moreover, the percentage of thermally labile amino acids in IPCT is as low as *ca.* 6% of the proteins content also in agreement with Sterner *et al.* (Sterner & Liebl, 2001).

Regarding the increased polar compared with non-polar surface areas, there is a cluster of polar, negatively charged amino acids where the ligands (CTP and L-*myo*-inositol-1-phosphate) should bind, surrounded by several smaller clusters of positively charged amino acids. The only region of the protein which does not seem to be charged is the N-terminus where the first 15-17 residues (depending on the chain), of the protein are not visible in the electron density maps. However, these residues are mainly hydrophobic (I_2, V_4, L_9, I_{11}, F_{12}, A_{13} and I_{16}) with only two negatively charged a.a. (D_5 and E_7), three positively charged a.a. (K_{10}, R_{15} and K_{17}), and four polar a.a. (N_3, G_6, Y_8 and G_{14}).

The work on the IPCT domain of the bifunctional enzyme IPCT/DIPPS, is an important step towards the final goal of understanding the full catalytic route for DIP synthesis. DIP, being an hallmark of marine microorganisms that thrive optimally in very hot environments, was observed to increase its pool preferentially in response to heat stress, leading to the view that it protects proteins, and other cellular macromolecules, against the detrimental effects of high temperature (Santos *et al.*, 2007), with potential application in the field of biotechnology. Work on the bifunctional IPCT/DIPPS enzyme is currently ongoing in our Laboratory with promising results. The fact that the DIPPS domain is a membranar domain with three membrane spanning helices increases the difficulty of the project; however, the full-length IPCT/DIPPS enzyme has been crystallized and optimization trials are underway.

In this work, two *de novo* structures were determined with implications in the sulfur metabolism and compatible solutes metabolic pathways fields. We believe that the ongoing work in our Laboratory will answer some of the questions that remain open but we are very pleased with this initial step towards the elucidation of the structure function relationship on these enzymes based on the X-ray structure solution of the sulfide:quinone oxidoreductase from *Acidianus ambivalens* and the CTP:inositol-1-phosphate cytiylyltransferase from *Archaeoglobus fulgidus*.

ITQB-UNL | Av. da República, 2780-157 Oeiras, Portugal
Tel (+351) 214 469 100 | Fax (+351) 214 411 277

www.itqb.unl.pt

4.1. REFERENCES

Brito, J. A., F. L. Sousa, M. Stelter, T. M. Bandeiras, C. Vonrhein, M. Teixeira, M. M. Pereira & M. Archer, (2009) Structural and functional insights into sulfide:quinone oxidoreductase. *Biochemistry* **48**: 5613-5622.

Chan, L. K., R. M. Morgan-Kiss & T. E. Hanson, (2009) Functional analysis of three sulfide:quinone oxidoreductase homologs in *Chlorobaculum tepidum*. *J Bacteriol* **191**: 1026-1034.

Cherney, M. M., Y. Zhang, M. Solomonson, J. H. Weiner & M. N. James, (2010) Crystal structure of sulfide:quinone oxidoreductase from *Acidithiobacillus ferrooxidans*: insights into sulfidotrophic respiration and detoxification. *J Mol Biol* **398**: 292-305.

Gomes, C. M., T. M. Bandeiras & M. Teixeira, (2001) A new type-II NADH dehydrogenase from the archaeon *Acidianus ambivalens*: characterization and in vitro reconstitution of the respiratory chain. *J Bioenerg Biomembr* **33**: 1-8.

Li, W. F., X. X. Zhou & P. Lu, (2005) Structural features of thermozymes. *Biotechnol Adv* **23**: 271-281.

Marcia, M., U. Ermler, G. Peng & H. Michel, (2009) The structure of *Aquifex aeolicus* sulfide:quinone oxidoreductase, a basis to understand sulfide detoxification and respiration. *Proc Natl Acad Sci U S A* **106**: 9625-9630.

Santos, H., P. Lamosa, T. Q. Faria, N. Borges & C. Neves, (2007) *The physiological role, biosynthesis and mode of action of compatible solutes from (hyper)thermophiles*. ASM Press, Washington, D.C.

Stelter, M., A. M. Melo, M. M. Pereira, C. M. Gomes, G. O. Hreggvidsson, S. Hjorleifsdottir, L. M. Saraiva, M. Teixeira & M. Archer, (2008) A novel type of monoheme cytochrome *c*: biochemical and structural characterization at 1.23 A resolution of *Rhodothermus marinus* cytochrome *c*. *Biochemistry* **47**: 11953-11963.

Sterner, R. & W. Liebl, (2001) Thermophilic adaptation of proteins. *Crit Rev Biochem Mol Biol* **36**: 39-106.

Vieille, C. & G. J. Zeikus, (2001) Hyperthermophilic enzymes: sources, uses, and molecular mechanisms for thermostability. *Microbiol Mol Biol Rev* **65**: 1-43.