Structure to Function Studies in UDP-glucose Dehydrogenases and Nitroreductases

Joana Raquel M. Rocha

Dissertation presented to obtain a Ph.D. degree in Biochemistry,
at the Instituto de Tecnologia Química e Biológica,
Universidade Nova de Lisboa

FCT Fundação para a Ciência e a Tecnologia
MINISTÉRIO DA CIÊNCIA, TECNOLOGIA E ENSINO SUPERIOR

Co-Supervisor: Dr. Edward P. Mitchell

Opponents: Dr. Serge Pérez
Dr. José Trincão
Dr. Arsénio Fialho
Dr. Dolores Roldán
Prof. Dr. Maria Arménia Carrondo

Oeiras, May 2010

Dolores Roldán, Edward Mitchell, Arsénio Fialho, Maria Arménia Carrondo, José Simões, Joana Rocha, Serge Pérez, José Trincão and Carlos Frazão.

10th May 2010

2nd Edition, May 2010

Macromolecular Crystallography Unit
Structural Biology Laboratory
Av. da República
Estação Agronómica Nacional
2780-157 Oeiras
Portugal

Structure to Function Studies in UDP-glucose Dehydrogenases and Nitroreductases

Joana Raquel M. Rocha

Doctoral thesis

J. R. Rocha, Oeiras, Portugal.

The studies described in this thesis were carried out at the Structural Biology Group - Macromolecular Crystallography Unit, Instituto de Tecnlogia Química e Biológica - Universidade Nova de Lisboa, Portugal. They resulted from ongoing scientific collaborations with the Institute for Biotechnology and Bioengineering, Center for Biological and Chemical Engineering, Instituto Superior Técnico – Universidade Técnica de Lisboa, Portugal, the Biochemistry and Molecular Biology Department, University of Córdoba, Spain, and the Structural Biology Group, European Synchrotron Research Facility, France.

The first chapter is an introductory overview of the PhD thesis subjects. The following three chapters correspond to the work performed during the PhD, and are based on articles already published or in preparation for submission in peer reviewed international scientific journals. At the end of each chapter, in the Acknowledgments section, there is a list of all articles' co-authors and their scientific contribution. The last chapter gathers all the relevant conclusions of this PhD thesis work. Appendix I comprise a small chapter on an amidase crystallographic project that delivered results only during the final redaction period of this thesis, and Appendix II serves as an introduction to Macromolecular Crystallography Techniques used in this work.

The author acknowledges the PhD grant BD/24216/2005 from Fundação para a Ciência e Tecnologia. This work has been partially funded by Fundação para a Ciência e a Tecnologia (FCT), Portugal, by grants POCTI/BME/38859/2001, POCI/BIO/58401/2004 and PTDC/QUI/67925/2006, and ESRF.

Table of Contents

Scope of the thesis

The thesis is divided into two parts corresponding to structural studies on two different proteins. The first part concerns the study of two UDP-glucose dehydrogenases (UGDs) from *Sphingomonas elodea* ATCC 31461 and *Burkholderia cepacia* IST 408, both involved in exopolysaccharide production. Their relevance arises because some of these bacterial exopolysaccharides are valuable as established biotechnological products, the former case, whilst others are highly problematic, when used by pathogens in biofilm formation over biological surfaces, as the latter case, namely in the human lungs. The goal of these studies is to increase our knowledge regarding UGDs structural properties, which can potentiate either the design of activity enhancers to respond to the increased demand of useful biofilms, or the design of inhibitors of biofilm production, in order to fight invading pathogens present in several infections. The thesis reports the production and crystallisation of both proteins, the determination of initial phases by single-wavelength anomalous dispersion (SAD) in *S. elodea* crystals using a seleno-methionine isoform, and phasing of *B. cepacia* crystals by molecular replacement (MR) using the *S. elodea* model, as well as the refinement, structural analysis and comparison between the several UGDs structures available during this work. The second part of the thesis concerns the structural characterisation of a nitroreductase from *Rhodobacter capsulatus* B10 (NprA), a flavoenzyme with as yet unknown physiological substrate, which exhibits nitroreductase activity towards several nitroaromatic compounds making it of potential use in bioremediation or cancer therapy. The thesis reports the isolation and crystallisation of the native protein with bound FMN, the structural determination by MR using *Enterobacter cloacae* nitroreductase as search model, the building of the structure model, its refinement and analysis, and a comparison with related proteins.

Abstract

The expression, purification and crystallisation of two uridine-5'-diphospho-glucose dehydrogenases (UGDs) from *Sphingomonas elodea* ATCC 31461 and *Burkholderia cepacia* IST 408 were carried out and their three-dimensional structure determination pursued in presence of different ligands, namely nicotinamide adenine dinocleotide (NAD(H)) and uridine-diphospho-glucuronic acid (UDP-GlcA). *S. elodea* ATCC 31461 UGD, termed UgdG, was expressed, purified and crystallised in the presence of NAD^+, in both native and SeMet derivatised forms. The phase problem was solved by single anomalous dispersion (SAD) using a SeMet derivatised crystal form of UgdG, and this solution was successfully applied by molecular replacement to the crystal form of the native protein, originating a higher resolution structure of UgdG bound to an adenosine-diphospho-ribose molecule, a NAD^+ degradation product. The structural comparison of available UGDs shows conservation of protein-cofactor (visible part) interactions, except for an extra hydrogen bond in UgdG. A 3D phyletic analysis of UGD/GMD family places UgdG in prokaryotic subfamily (group I) closer to eukaryotic members than to the prokaryotic group II proteins. *B. cepacia* IST 408 UGD, termed BceC, was expressed, purified and crystallised in the presence of its final reaction product, UDP-GlcA. Due to their high sequence homology, the phase problem of BceC was addressed by the molecular replacement method using the previously determined UgdG structure as a search model. The refined structure of BceC shows a dimeric protein with one molecule of UDP-glucuronic acid bound per monomer. NAD binding proteins share a glycine-rich signature motif that in the UGD/GMD family includes a conserved tyrosine, GXG***Y***XG. The mutation of this tyrosine in BceC led to almost inactive mutants, thus demonstrating a relevant role within the UGD/GMD enzymatic activity. A comparison of the native and two mutant crystal structures together with enzymatic assays led to the elucidation

of the tyrosine role in the enzymatic mechanism, namely in the final hydrolysis step of the reaction thioester intermediate, where the tyrosine works as final proton conveyer from the aqueous media to the leaving thiolate group. This study also enabled a detailed comparison of human UGD and BceC structures, in order to explore potential medicinal applications. However, the substrate binding pockets of both proteins seem too similar for the design of a potentially selective inhibitor.

The major nitroreductase (NR) of *Rhodobacter capsulatus* B10, termed NprA, was expressed, purified and crystallised in the presence of an *in-vitro* substrate, 2,4-dinitrophenol (DNP). Molecular replacement was used to solve the phase problem, using the *Enterobacter cloacae* nitroreductase structure as search model. The structure revealed a dimeric arrangement with two flavin mononucleotide (FMN) molecules bound at symmetrical crevices of the dimer interface, but with no trace of the substrate. Instead, an acetate molecule, a known inhibitor of Type I nitroreductases, was found bound directly above the cofactor molecule. NprA presents the same overall fold found in known NR structures, in spite of its low sequence identity against family members with available structure. The physiological role of NprA and the mapping of its key catalytic residues remained as yet open questions. Attempts were essayed to obtain NprA complexes with different ligands, but all have remained unsuccessful, eventually due to the rather low reproducibility of NprA crystals. Nevertheless, previous biochemical characterisation has shown that NprA is highly induced by several nitroaromatic compounds, suggesting NprA as a novel candidate for biological remediation or for anti-tumoral therapies.

Sumário

Expressaram-se, purificaram-se e cristalizaram-se as uridina-difosfoglucose desidrogenases (UGDs) de *Sphingomonas elodea* ATCC 31461 e *Burkholderia cepacia* IST 408, e prosseguiu-se com a determinação das suas estruturas tridimensionais na presença de diferentes ligandos, respectivamente do cofactor nicotinamida adenina dinucleotídeo (NAD^+) e de ácido uridina-difosfoglucurónico (UDP-GlcA). A UGD de *S. elodea* ATCC 31461, denominada UgdG, foi expressa, purificada e cristalizada nas formas nativa e derivatizada com selénio-metionina. O problema da fase foi resolvido por dispersão anómala de comprimento de onda único (SAD), utilizando dados de difracção de raios-X do cristal da forma derivatizada. A solução obtida foi de seguida usada com o método de substituição molecular no cristal de UgdG nativa, originando assim uma estrutura a resolução mais elevada de um complexo de UgdG e adenosina-difosforibose, um produto de degradação do NAD^+. A comparação estrutural de diversas UGDs mostrou que se conservam as interacções proteína-cofactor (parte não degradada), mas observa-se uma nova ponte de hidrogénio entre a UgdG e a ribose. Uma analise filética 3D da família UGD/GMD coloca a UgdG no grupo procariótico I, mais próximo do grupo eucariota que dos restantes procariotas do grupo II. A UGD de *B. cepacia* IST 408, designada BceC, foi expressa, purificada e cristalizada na presença do produto final de reacção, o ácido UDP-glucurónico, UDP-GlcA. Devido à elevada homologia nas sequências das duas proteínas o problema da fase em BceC foi resolvido usando a estrutura de UgdG como modelo inicial na substituição molecular. A estrutura refinada exibe uma proteína dimérica com cada monómero de BceC complexado a uma molécula de UDP-GlcA. As proteínas que ligam NAD contêm um motivo rico em glicinas, que na família das UGD/GMDs incluem também uma tirosina conservada, GXG**Y**XG. A mutação desta tirosina em BceC originou mutantes praticamente inactivos, o

que demonstra a importância dessa tirosina na actividade enzimática das UGD/GMDs. A comparação das estruturas nativa e de dois mutantes da BceC conjuntamente com as suas diferentes actividades conduziu à elucidação do papel da tirosina no mecanismo enzimático, nomeadamente no último passo do mecanismo reaccional, a hidrólise do intermediário de tioéster, onde a tirosina facilita a transferência de um protão do meio aquoso para o grupo tiolato em formação. Um dos objectivos deste trabalho consistiu ainda na comparação detalhada entre as estruturas da BceC e da UGD humana, tendo em vista potenciais aplicações à medicina. No entanto as cavidades de ligação do substrato reaccional aparecem como demasiado semelhantes para o design de um potencial inibidor selectivo. A nitroreductase (NR) maioritária de *Rhodobacter capsulatus* B10, denominada NprA, foi expressa, purificada e cristalizada na presença de um substrato *in-vitro,* o 2,4-dinitrofenol. A estrutura da NR de *Enterobacter cloacae* foi usada como modelo para a substituição molecular, permitindo a resolução do problema da fase. A estrutura de NprA revelou-se como um homodímero com duas moléculas de flavina mononucleótido (FMN) ligadas a cada uma das cadeias peptídicas, em bolsos localizados simetricamente na interface dimérica, mas sem vestígios de ligação ao substrato. No entanto, um ião acetato, conhecido inibidor desta família de NRs, foi encontrado ligado à proteína adjacente a cada cofactor. A estrutura de NprA apresenta a topologia das nitroreductases com estrutura conhecida, ainda que a identidade da sua sequência com os vários membros da família seja baixa. Ate à data, a função biológica da NprA e os resíduos cataliticamente relevantes são ainda questões em aberto. Para as aprofundar, tentou-se a obtenção de complexos de NprA com diversos ligandos, até agora infrutíferas, possivelmente devido à fraca reprodutibilidade dos cristais. Ainda assim, uma prévia caracterização bioquímica demonstrou que a NprA é altamente induzida por diversos compostos nitroaromáticos, sugerindo a possibilidade da sua aplicação em biorremediação e no design racional de drogas antitumorais.

Acknowledgements

Many people that provided their personal expertise, valuable assistance and friendship, supported the work here reported. Without them, this work would have not have been possible.

First, I would like to thank my supervisor Dr. Carlos Frazão. He started the project with me and was present all the way through. I am grateful to Carlos for teaching me to be patient and persistent. From Carlos I also learned to never forget how the small details can be very important.

I am very grateful to my co-supervisor Dr. Ed Mitchell, but also to Dr. Ana Maria Gonçalves from ESRF. I'm grateful for the time Ed always had for my research and my doubts, even if it meant late nights, early mornings or weekends, particularly, when collecting data. I thank Dr. Ana Maria Gonçalves for all the patience and the support in the lab when she was back at ITQB and latter, when she moved to the ESRF. I am grateful to both all the help in the last year of my thesis.

I'm also grateful to both group leaders from the Macromolecular Crystallography Group in the ESRF and the Protein Crystallography Laboratory in the ITQB, Dr. Sean McSweeney and Professor Maria Arménia Carrondo, for giving me the opportunity of working in two different and great laboratories.

I thank Dr. Arsénio Fialho, Dr. Ana Teresa Granja and Alma Popescu from Instituto Superior Técnico (IST), for their work on the UgdG and BceC isolation and cloning. Some of the techniques I learn from all were used all the way through the thesis work.

I also thank Dr. Maria Dolores Roldán, Dr. Conrado Moreno-Vivián and Dr. Eva Pérez-Reinado from University of Cordoba (Spain), for their work on the isolation and cloning of NprA but most of all, for the opportunity of working in the lab and the wonderful days spent in such a beautiful city. I am grateful to Dr. Cláudio Soares and Dr. Bruno Vitor from the Modelling Group in ITQB for their help on the NprA-ligand docking studies.

I thank all the people at the Macromolecular Crystallography Unit at ITQB, for many valuable discussions, helping me in laboratory work and some corrections to the text and figures of my thesis. In particular, a warm thanks to Dr. Colin McVey, always available for precious advises.

I also thank all my friends for all the support given during this stage of my life. It would be impossible to mention all the names, but of course they know who they are. Thank you all very much…..

A special thank to Daniele de Sanctis, not only for his help in laboratory work and computing but most of all, for always being there, in the ups and downs of my life.

Of course, I am very grateful to my family, for what I was, what I have conquered and what I have become, for keeping up with my moods and now, enduring the great distance that keeps us apart.

As for my funding I acknowledge a FCT grant SFRH (BD/24216/2005), the ITQB directorate, Protein Crystallography Laboratory in the ITQB and the ESRF.

Index of tables

Illustration Index

Thesis Background

1.1. Bacterial Polysaccharides

Complex carbohydrates comprise the most abundant group of natural products on Earth and in the oceans, as they surround all living cells and play multiple functions in all forms of life (Ruffing & Chen 2006). From their role as energy source or metabolic intermediates, carbohydrates are the major constituents of the shells of insects, arthropodae and also the supporting tissue of plants (Geremia & Rinaudo 2004; Vigetti *et al*. 2006). Moreover, they are present as components of all cell walls, spanning from the world of plants to microbes. In this sense, all organisms have evolved to be able to capture, process, store and sometimes synthesise these carbohydrates. Many bacterial species, fungi and algae are known to produce these compounds. It is estimated that approximately $4x10^{11}$ tons of carbohydrates per year are biosynthesised by plants and bacteria, the majority being produced as polysaccharides, molecules consisting of a large number of one (homopolysaccharide) or more types (heteropolysaccharide) of monosaccharide residues (Broadbent *et al*. 2003; Lidhorst 2007). Because these polymeric compounds are produced by an enormous variety of organisms, including microbes, algae, plants and animals, the diversity of the produced polysaccharides is great (Geremia & Rinaudo 2004). One example of this group of macromolecules is *xanthan* (Figure 1.1.), a polymer produced by the bacterium *Xanthomonas spp*. that confers dessication resistance properties (Blanch *et al*. 2006). Others include the known *fructans,* other bacterial gel polysaccharides, capsule polysaccharides or agar. The unique physiochemical properties of these compounds are mainly determined by their sugar composition and polymerisation (linear or branched polymers), and therefore have been the subject of intensive studies (Vanhooren & Vandamme 1998; Sutherland 2001).

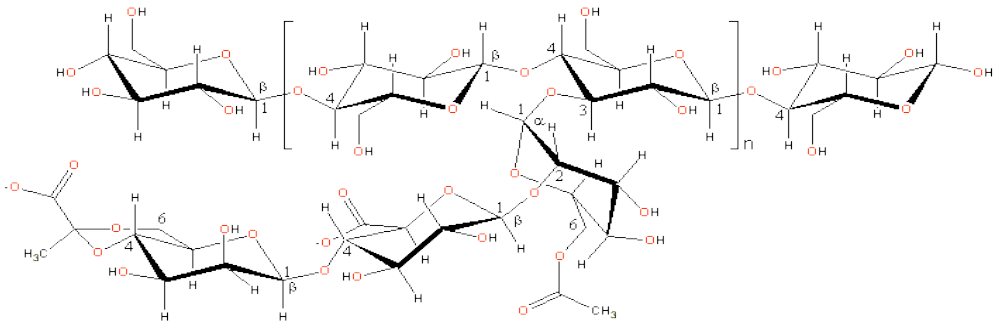


Figure 1.1. *Structure of the polysaccharide xanthan*.

Xanthan gum is an anionic polymer with a β-(1⟶4)-D-glucopyranose glucan (as celullose) backbone with side chains of -(3⟶1)-α-linked D-mannopyranose-(2⟶1)-β-D-glucuronic acid-(4⟶1)-β-D-mannopyranose on alternating residues. Around ~40% of the terminal mannose residues are 4,6-pyruvated and the inner mannose is mostly 6-acetylated (from http://www.lsbu.ac.uk/water/hyxan.html).

The increased demand for natural or modified polymers for various industrial applications in recent years, has led to a renewed interest in these macromolecules. Bacterial polysaccharides have emerged as new industrially important polymeric materials, which are proving economically competitive to some natural gums produced by plants and marine algae (Morin 1998). Both polysaccharides and monosaccharides have a wide application in industry, ranging across the fields of medicine, food industry or crude oil recovery. Some polysaccharides have also been used for the removal of heavy metals from the environment. This is the case of the exopolysaccharides produced by the marine bacteria *Zooglea* sp. and *Enterobacter cloacae*, which has the capacity to chelate chromium, lead and iron ions in solution (Kong *et al*. 1998; Iyer *et al*. 2004). Some of the most relevant bacterial polysaccharides with biotechnological applications are described in Table 1.1. The expression of extracellular polysaccharide (EPS) material is a common feature of several bacteria. Bacterial EPS can be either identical, or closely related to polysaccharides found in eukaryotic species (Sutherland 2004).

Table 1.1. *Examples of bacterial polysaccharide diversity and their industrial application*. (Compiled from De Vuyst & Degeest 1999; Vanhooren & Vandamme 2000; Levander *et al*. 2001; Kumar *et al*. 2007)

Organism	Polysaccharide	Application
Pseudomonas aeruginosa	Alginate	Immobilisation matrix for viable cells and enzymes; coating of roots of seedlings and plants to prevent dessication; microencapsulation matrix for fertilisers, pesticides and nutrients; hypo-allergic wound-healing tissue.
Streptococcus equii and *Streptococcus zooepidemicus*	Hyaluronic acid	Replacer of eye fluid in ophthalmic surgery, in artificial tear-liquid, synovial fluid replica, wound-healing, lotion and moiturising agent in cosmetic industry.
Xanthomonas	Xanthan (E415)	Crude-oil recovery, paints and inks, pesticide and detergent formulations, cosmetics, pharmaceuticals, as a stabiliser and thickner agent in food.
Acetobacter spp.	Cellulose	Temporary artificial skin to heal burns or surgical wounds, nutritional fibers in diet, hollow fibres or membranes in separation technology
Leuconostoc mesenteroids	Dextran	Blood plasma extender or flow improving agent, cholesterol lowering agent, micro-carrier in tissue/cell culture
Clavibacter spp.	Clavan	Potencial in preventing tumour cell colonisation of the lung and the treatment of rheumatoid arthritis, antibody production stimulating agent, in cosmetics as skin moisturising agent.
Lactic acid bacteria	Glucose/galactose rich polymers	Used extensively in dairy product industry, provide beneficial microflora in the gastric tract

The main differences rely on the lack of extensive branching of the bacterial extracellular polymer, as these EPS are usually linear molecules with regular repeating side-chains of not more than tetra- or at most pentasaccharide units (Sutherland 1997). They are synthetised and excreted from the cell either as soluble or insoluble polymers. This EPS coats the outside of the bacterial cell

and as a consequence plays an intimate role in mediating interactions between the bacterium and its surrounding environment (Kumar *et al*. 2007). Bacterial exopolysaccharides produced by several strains have a multitude of applications and occur in two basic forms. In some cases, the polysaccharide may be tightly associated, even covalently bound to the cell surface (Vanhooren & Vandamme 1998), forming a discrete structure termed *capsule* which is usually associated with some pathogenic bacteria, as in the case of group A streptococci or *Pseudomonas aeruginosa* (Hung *et al*. 2007b); or it may be shed in the form of EPS or "slime", which is loosely bound to the cell surface and, as mentioned, in some cases has a wide range of industrial applications. Recent examples are the EPS producers *Lactococcus* and *Streptococcus*, widely used in the dairy product industry. Work has shown that the EPS produced by *Streptococcus thermophilus* can improve the functional properties of low-fat and partly-skimmed mozzarella cheese (Broadbent *et al*. 2003; Savadogo *et al*. 2004; Ruffing & Chen 2006). The precise role of the exopolysaccharide in EPS-producing bacteria is of course dependent upon the ecological niche and the natural environment surrounding the microorganism (Kumar *et al*. 2007). Microorganisms producing EPS are found in various ecological niches (as long as the medium offers a high carbon/nitrogen ratio), such as water effluents from the sugar, paper or food industries (Morin 1998). However, most of the functions attributed to the EPS are related with the cell protection (Kumar *et al*. 2007). The ability of a microorganism to surround itself with an EPS provides it with a protection against dessication in low-moisture environments due to the excellent water-binding properties (Broadbent *et al*. 2003), predation of protozoans and phage attack (Roberts 1996; Jolly *et al*. 2002), osmotic stress (Sutherland 1994). It also allows surface adhesion, affects the diffusion properties into and out of the cells and the penetration of both useful and toxic material, such as metal ions or even antibiotics (Dudman 1977; Sutherland 2001). In the case of pathogenic bacteria the production of the capsular EPS

enables the evasion of phagocytosis as its structure mimics the host cell surface components (Vuong *et al*. 2004; Ventura *et al*. 2006). Exopolysaccharides are also known to play a major role in the formation of cell aggregates and in biofilm adherence on inert and biological surfaces (Vanhooren & Vandamme 1998).

This thesis is centered on the study of the biosynthetic pathway of the exopolysaccharides (EPS) of *Sphingomonas elodea* ATCC 31461, a gram-negative bacterium with the ability to produce *gellan*, and *Burkholderia cepacia* IST408, a bacterium capable of producing *cepacian*. In the first case, the EPS is widely used in different industrial applications, while the second is an example of an EPS produced by an opportunistic pathogenic bacterium capable of colonising cystic fibrosis patient's lungs by adhesion to the mucosa. More specifically, the present work focuses on the expression, purification, biochemical characterisation and three-dimensional structure determination of UDP-glucose dehydrogenase, an enzyme involved in the biosynthetic pathway of EPS formation of both bacteria, with the goal of providing new structural information that can model the enzyme's activity and drive drug design and development (but see below).

1.1.1. *Sphingomonas elodea* and gellan gum

Many gram-negative bacteria synthesise and excrete EPS, which may have different biological roles. As mentioned before, in some cases bacterial EPS are potential or already successful products of biotechnology (Sá-Correia *et al*. 2002; Wang *et al*. 2002). "Sphingans" are structurally related polysaccharides secreted by some species of the genus *Sphingomonas*. This genus is found in a variety of habitats, including many terrestrial and water environments, in plant roots or clinical isolates (White *at al*. 1996). Due to the broad metabolic pathways present in these bacteria, *Sphingomonas* sp. have been widely used in biotechnological applications, ranging from bioremediation to extracellular

polymer production (Pollock 1993), such as the commercial *gellan* gum produced by large-scale fermentation of the strain *Sphingomonas elodea* ATCC 31461, previously known as *Pseudomonas elodea* (Granja *et al*. 2007; Aragão *et al*. 2007).

Gellan gum is a polymer first discovered in the laboratories of Kelco Division of Merck and Co. in 1978 (Figure 1.2; Kaneko & Kang 1979) and it had previously been referred to by the code names S-60 or PS-60. The microorganism was isolated from the *Elodea* plant tissue, and further studies revealed that the gellan gum-producing bacterium was a new strain of the species *Pseudomonas,* and thus baptised as *Pseudomonas elodea*. Successful toxicity trials were completed and gellan gum received approval for use in food in Japan in 1988. In 1994, it was discovered that gellan-producing bacterium was actually a member of the genus *Sphingomonas,* and was thus rebaptised as *Sphingomonas elodea* or *Sphingomonas paucimobilis*. The US FDA approved gellan gum for use as a food additive in 1992 *(*Bajaj *et al*. 2007). Gellan, also known under the food additive code E418, is used as a natural biothickner, stabilising and suspending agent, (Sutherland 1998) approved both in the United States and European Union for applications in food, such as puddings, dessert gels, beverages, dairy products such as yogurt, milk shakes, gelled milk and ice cream, fruit spreads, jams and marmalades, bakery fillings and sauces (Chandrasekaran & Radha 1995; Gonçalves *et al*. 2009). It is also used in fabricated foods, such as restructured meat, fruits and vegetables, and in canned/gelled pet foods (Bajaj *et al*. 2007). In its native form, the polymer is a linear anionic polysaccharide based on a tetrasaccharide repeating unit composed of two molecules of D-glucose (D-Glc), one molecule of L-rhamnose (L-Rha) and another of D-glucuronic acid (D-GlcA), partially sterified with L-glycerate and/or acetate (Figure 1.2.; Omoto *et al.* 1999; Harding *et al*. 2004). Acyl substituents drastically affect the rheology of the formed gels. Deacylation by alkali treatment results in a change from soft and more elastic

thermoreversible gels, to hard, firm and brittle gels (Harding *et al*. 2004; Granja *et al*. 2007).

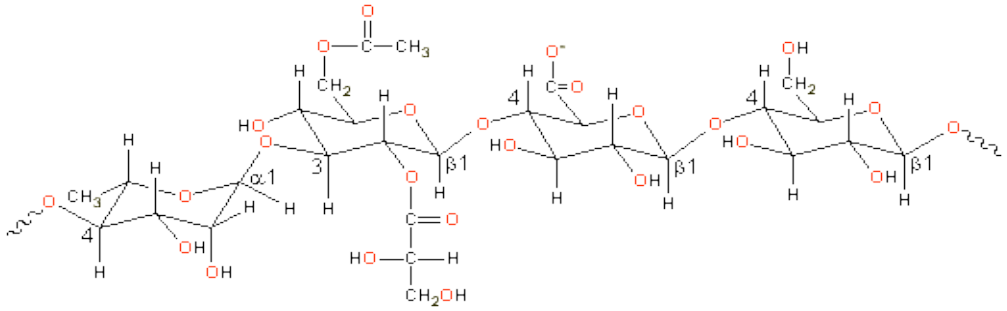


Figure 1.2. *Structure of the tetrasaccharide repeating unit of gellan*.

Gellan gum is a linear polymer based on a tetrasaccharide repeating unit [⟶4)-L-rhamnopyranosyl-(α-1⟶3)-D-glucopyranosyl-(β-1⟶4)-D-glucuronopyranosyl-(β-1⟶4)-D-glucopyranosyl-(β-1⟶] with O(2) L-glyceryl and O(6) acetyl substituents on the 3-linked glucose. It has a high molecular weight, consists of about 50,000 residues and is normally de-esterified by alkali treatment before use in food. (http://www.lsbu.ac.uk/water/hygellan.html).

Compared to other polysaccharides, gellan has important advantages for food applications, such as excellent thermal and acid stabilities, adjustable gel elasticity or rigidity, high transparency, and good flavour release (Fialho *et al*. 2008). However, due to its rheological properties, gellan gum is also exploited in non-food applications such as microbial and plant tissue culture media, solid matrix for DNA gel electrophoresis, the controlled release of certain drugs in the stomach, excipient for nasal and ocular drug delivery (Jansson *et al*. 2005), material for the construction of scaffolds in tissue engineering, and in personal care products such as deodorant gels, sun screens, body lotions, hair conditioners and clear gel toothpaste (Chandrasekaran & Radha 1995; Bajaj *et al*. 2007). The gellan biosynthetic pathway is a multi-step process that initiates with the cytosolic formation of the nucleotide sugars precursors UDP-glucose (UDP-Glc), UDP-glucuronic acid (UDP-GlcA) and TDP-L-Rhamnose (TDP-Rha) (Figure 1.3.), followed by the sequential transfer of the sugar donors to a

lipid carrier, gellan polymerisation and export (Granja *et al*. 2007; Fialho *et al*. 2008). The enzymes involved in TDP-Rha formation belong to the gellan biosynthetic *gel* cluster and are encoded by the *rhsABCD* genes. These genes encode for glucose-1-phosphate thymidylyltransferase (*rhsA*), dTDP-glucose-4, 6-dehydratase (*rhsB*), dTDP-4-keto-L-rhamnose-3,5-epimerase (*rhsC*) and dTDP-L-rhamnose synthase (*rhsD*). The enzymes involved in formatting the other UDP-sugars are phosphoglucomutase (PgmG, EC 5.4.2.2) that catalyses the conversion of D-glucose-6-phosphate into D-glucose-1-phosphate, UDP-D-glucose pyrophosphorylase (UgpG, EC 2.7.7.9) that converts D-glucose-1-phosphate into UDP-D-glucose, and UDP-glucose dehydrogenase (UGD; EC 1.1.1.22) (Sá-Correia *et al*. 2002) forming UDP-D-glucuronic acid from UDP-D-glucose.

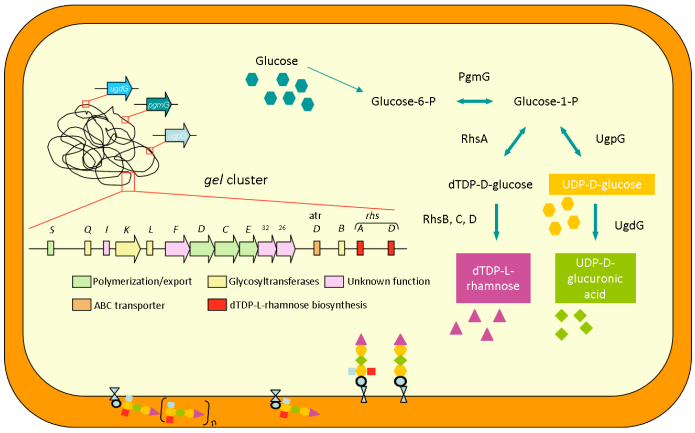


Figure 1.3. *Organisation of the gel cluster and the pathway of gellan biosynthesis*. UDP-D-glucose, UDP-D-glucuronic acid and TDP-L-rhamnose are the activated forms of the sugars units that compose gellan (compiled from Sá-Correia *et al*. 2002; Harding *et al*. 2004).

Interestingly, these three genes are not located in the *gel* cluster of *Sphingomonas elodea* ATCC 31461, but elsewhere in the genome, thus implying a more general role of these proteins in the cell metabolic pathways, and not only in the exopolysaccharide biosynthesis (Silva *et al*. 2005; Fialho *et*

al. 2008). The *gel* cluster also comprises all the genes involved in the polymerisation and export of gellan to the outside of the cell (Sá-Correia *et al*. 2002; Silva *et al*. 2005). The *gelB*, *gelK*, *gelL* and *gelQ* genes encode for glycosyltransferases, the enzymes involved in the tetrasaccharide repeating unit formation, starting with the sequential addition of the first glucose monomer to a lipid carrier, followed by the second, third and fourth sugar monomers, glucuronic acid, glucose and rhamnose, respectively (Sá-Correia *et al*. 2002; Harding *et al*. 2004). The assembly of the repeat unit is completed with the modification of the tetrasaccharide, by the addition of glycerate and acetate groups to the first glucose monomer. This action is carried out by acetyl- and glyceryltransferases, whose genes are also not located in the *gel* cluster, and are elsewhere in the genome of *Sphingomonas elodea* ATCC 31461 (Fialho *et al*. 2008). The *gelC* and *gelE* genes encode for a polysaccharide co-polymerase (PCP) responsible for the polymerisation and extension of the polysaccharide chain. Finally, the export of the polymer is performed via an outermembrane auxiliary (OMA) protein channel encoded by the *gelD* gene.

Despite the characterisation of many of the gel gene products, there are still some genes with unknown function, some are believed to participate either in regulation mechanisms or as secretion sorting systems by homology with known proteins from EPS operons in other bacteria (Sá-Correia *et al*. 2002; Harding *et al*. 2004). Identification of the genes and elucidation of crucial steps of the gellan biosynthetic pathway indicates some possibilities for the manipulation of gellan production at any of the three levels of its biosynthesis: (*i*) at the level of synthesis of sugar-activated precursors, (*ii*) at the level of the repeat unit assembly or, (*iii*) at polymerisation and export. By modifying the expression of any of the individual genes, or even of a group of these genes, the conversion efficiency and gellan gum yield can be increased. In effect, a few strategies have been applied to the early steps in gellan biosynthesis, for example, the simultaneous overexpression in *Sphingomonas elodea* ATCC

31461 of PgmG and UgdG, which was reported to lead to a 20% increase in gellan yield and EPS viscosity (Sá-Correia *et al.* 2002; Fialho *et al*. 2008).

1.1.2. *Burkholderia cepacia* and cepacian

Until the early 80's *Burkholderia cepacia* was known as the phytopathogen causing the soft rot of onion bulbs, described by Burkholder (Govan & Deretic, 1996). However, during the last 20 years, this bacterium emerged as an opportunistic pathogen in the lungs of immunocompromised individuals, especially cystic fibrosis (CF) patients (Hughes *et al*. 1997; Sist *et al*. 2003). *B. cepacia* is a gram-negative motile rod that belongs to the *B. cepacia* complex (BCC), a group of phenotypically similar (but genetically distinct) environmental bacteria (Goldberg 2007). Although organisms from all genomovars composing the complex have been isolated from patients with CF, *B. cepacia* has been of particular concern, because it is more invasive than other genomovars and possesses several characteristics that render it different from other pathogens (Conway *et al*. 2004; Herashimenka *et al*. 2007). It is highly transmissible, has inherent resistance to multiple antibiotics and antimicrobial peptides (APs) (Ortega *et al*. 2007) and is associated with great virulence. However, little is known about these virulence factors or mechanisms. All these factors combined make the management and treatment of *B. cepacia* infected patients difficult (Jones *et al*. 2001). Because *B. cepacia* is found in lung infections in CF patients, it is often compared with the more common CF pathogen *P. aeruginosa* (Goldberg 2007). *B. cepacia* infection can result in asymptomatic carriage and chronic infection similar to that observed with *P. aeruginosa* infection, but in some cases can result in terminal disease (the so called "*cepacian syndrome*"), that is characterised by a confluent bronchopneumonia and septicemia, which results in death over a period of days, unlike any other CF clinical situation (Hughes *et al*. 1997; Jones *et al*.

2001). A number of investigations carried out on *P. aeruginosa* pointed out to the role of the extracellular polysaccharide alginate in biofilm formation and in the virulence mechanisms of this bacterium (Leid *et al*. 2005). Biofilm formation (Figure 1.4.) is important in CF-associated pulmonary infection and has been well studied in the CF-associated pathogen *Pseudomonas aeruginosa*. Alginate exopolysaccharide (EPS) produced by *P. aeruginosa* makes up its biofilm matrix and mucoid capsule, and protects the cell against the host immune response and antibiotics. It allows adhesion to the host cells, persistence and colonisation of the respiratory tract of patients (Sist *et al*. 2003; Snook *et al*. 2003). The emergence of mucoid *P. aeruginosa* in colonised CF patients is often correlated with the declination of lung function (Conway *et al*. 2004). EPS production also occurs in *B. cepacia* strains. In effect, more than 80% of BCC isolates from CF infected patients were exopolysaccharide producers, indicating that EPS production by these organisms may be important in their pathogenesis (Richau *et al*. 2000; Conway *et al*. 2004).

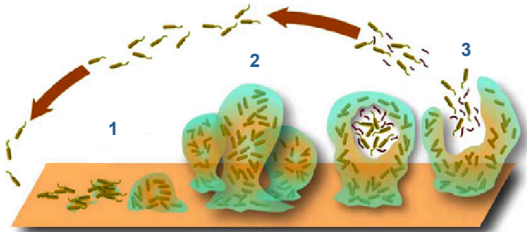


Figure 1.4. *The biofilm life cycle*.

(1) Individual cells populate and attach to the surface, (2) EPS is produced - the biofilm architecture develops and matures (3) Single cells are released from the biofilm and disperse (adapted from the Center for Biofilm Engineering, Montana State University. http://www.erc.montana.edu/biofilmbook).

Several types of EPS have been detected in strains from BCC clinical isolates from all over the world (Herashimenka *et al*. 2007; Zlosnik *et al*. 2008) but curiously, there is one that is most prominent and BCC-specific (Linker *et al*.

2001): a branched acetylated heptasaccharide repeating unit, commonly known as cepacian (Figure 1.5.; Cerantola *et al*. 2000). Although the role of cepacian as a persistence and virulence factor of BCC in CF patients is still not well understood, the production of this EPS is thought to confer enhanced pathogenecity of *B. cepacia* when infecting these patients (Conway *et al*. 2004; Cunha *et al*. 2004).

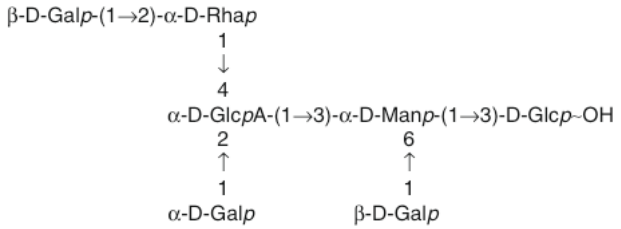


Figure 1.5. *Heptasaccharide repeating unit of the BCC exopolysaccharide, cepacian*. Gal, galactose; Rha, rhamnose; Glc, glucose; GlcA, glucuronic acid; and Man, mannose. The structure includes one acetyl group per repeating unit whose position is unknown (Cescutti *et al.* 2010).

Furthermore, cepacian inhibits neutrophil chemotaxis and scavenges reactive oxygen species in vitro, similarly to the alginate produced by *Pseudomonas aeruginosa* in the CF lung, suggesting that it may protect the producing bacteria from host defenses (Bylund *et al*. 2006). Moreira and co-workers first described the genes involved in cepacian biosynthesis and their physical organisation. Almost all of the genes involved in cepacian formation are located in the same DNA region, in a 16.5 kb operon, the *bce* gene cluster (Figure 1.6.; Moreira *et al*. 2003). At the beginning of the *bce* cluster, *bceA* codes for a bifunctional protein with phosphomannose isomerase (PMI) and GDP-D-mannose pyrophosphorylase (GMP) activities. The resulting product, D-mannose, is incorporated in the cepacian polysaccharide and serves as precursor for GDP-D-rhamnose formation (Richau *et al*. 2000). The *bceC* gene encodes for an UDP-

glucose dehydrogenase (UGD). The reaction product, UDP-glucuronic acid, is the activated donor of D-glucuronic acid for the synthesis of the heptasaccharide repeating unit. The genes *bceB*, *G*, *H*, *J* and *K* are likely to code for glycosyltransferases (GTs) involved in the assembly of the repeating unit. According to the similarity studies preformed by Moreira and coworkers (Moreira *et al*. 2003), *bceB* codes for an undecaprenyl-phosphate glycosyl-1-phosphate transferase (UndPGPT), responsible for the transfer of the first sugar unit into the lipid carrier, which is the first step in the subunit assembly, while the other genes code for all the other glycosyl transferases involved in the transfer of the sugar moieties that compose cepacian.

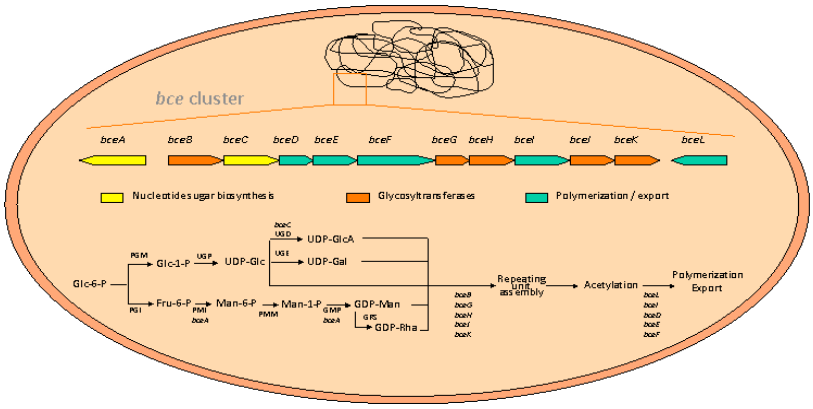


Figure 1.6. *Organisation of the bce gene cluster for cepacian biosynthesis and the pathway for EPS synthesis by B. cepacia* (adapted from Moreira *et al*. 2003).

The genes *bceD*, *E*, *F*, *I* and *L* exhibit homology with genes involved in the polymerisation and export of bacterial surface polysaccharides (Moreira *et al*. 2003). The *bceE* gene encodes an outer-membrane auxiliary (OMA) protein, homologue to the GelD protein of the previously referred *Sphingomonas elodea*. The amino acid sequence analysis of the BceF protein led to the identification of an autophosphorylating protein tyrosine kinase (PTK), belonging to the membrane periplasmic auxiliary (MPA1) protein family, while the same studies identified BceD as a low-molecular mass phosphotyrosine

protein phosphatase (PTPase). The encoded protein BceL is thought to catalyse the translocation of the membrane bound heptasaccharide unit to the periplasmic face of the plasma membrane. Finally, BceI shows similarity to hypothetical proteins from *Burkholderia fungorum* and *Nostoc punctiform* that have been implicated in the polymerisation of the repeating unit in the periplasmic side of the plasma membrane, originating the exopolysaccharide. Genes encoding the enzymes needed for the other nucleotide sugars precursors, such as PGM (with both phosphoglucomutase and phosphomannomutase activities) and UGP (UDP-glucose pyrophosphorylase) are absent from the *bce* locus. Moreover, *B. cepacia* EPS is an acetylated polysaccharide, although the gene coding for the *O*-acetyltransferase enzyme required for acetylation is also not present in the locus. This is consistent with previous reports for other EPS clusters described for gram-negative bacterial species for example, *Sphingomonas elodea* (Moreira *et al*. 2003) referred to earlier.

1.1.3. UDP-Glucose Dehydrogenase (UGD)

To fully understand the biosynthesis of polysaccharides it is necessary to investigate the upstream source of their component sugars. In many cases, the immediate donors of the monomers that compose the polysaccharides are the nucleotide sugars. These are activated forms of monosaccharides, ready to be incorporated into oligo- or polysaccharides. In mammals there are nine different nucleotide sugars, and in plants there are even more but the majority, including the nucleotide-sugars used in bacterial systems, are uridine-diphospho-sugars (UDP-sugars). These compounds may be generated from appropriate monosaccharides and sugar phosphates resulting from metabolism, or they can be generated through the interconversion of existing nucleotide sugars (Figure 1.7.). An example of the latter case is the pathway centered on uridine-diphospho-glucuronic acid (UDP-GlcA; Griffith *et al*. 2004).

UDP-GlcA is the activated donor of the glucuronic acid moiety and it serves many critical roles in a variety of organisms ranging from mammals to bacteria (Campbell *et al*. 2000; Kärkönen & Fry 2006).

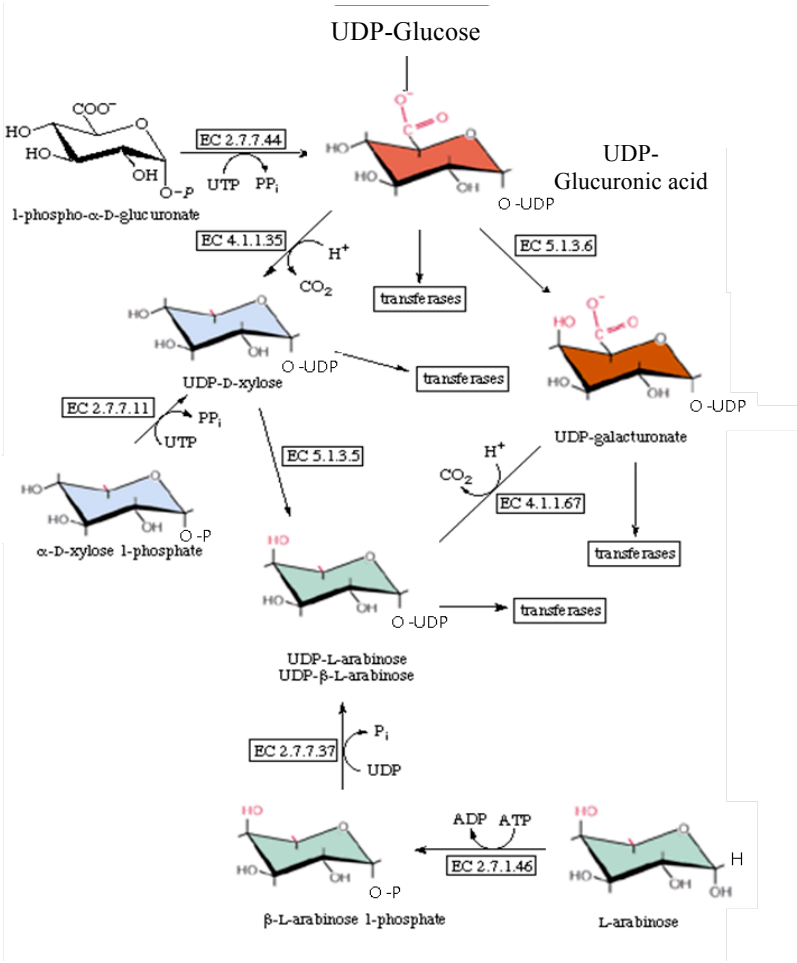


Figure 1.7. *Enzymatic conversion of UDP-glucuronic acid to different sugars*.
(Adapted from http://www.chem.qmul.ac.uk/iubmb/enzyme /reaction/polysacc/UDPAra.html).

In animals, UDP-GlcA is the substrate for UDP-glucuronosyl transferases in liver and gastrointestinal tract, promoting glucuronidation of xenobiotic compounds solubilising these agents (Hempel *et al*. 1994), and it is also essential for biosynthesis of hyaluronan and other glucosaminoglycans (Easley *et al*. 2007; Hung *et al*. 2007b). Matrix polysaccharides in plants may derive

half their mass from monosaccharides generated from UDP-GlcA, as it is the precursor of other nucleotide-sugars like UDP-xylose or UDP-arabinose (Griffith *et al*. 2004; Klinghammer & Tenhaken 2007). In bacteria, UDP-GlcA is used for exopolysaccharide production, being formed exclusively by the action of UDP-glucose dehydrogenase (UGD, EC 1.1.1.22; Wang *et al*. 2003; Moyrand & Janbon 2004). UGD, a sugar nucleotide modifying enzyme, is a member of a small group of NAD^+ dependent four-electron transferring oxidoreductases responsible for the conversion of uridine-diphospho-glucose (UDP-Glc) to UDP-GlcA (Hempel *et al*. 2004). The oxidation is irreversible and gives rise to two moles of NADH per mole of UDP-glucose (Figure 1.8.; De Luca *et al*. 1996).

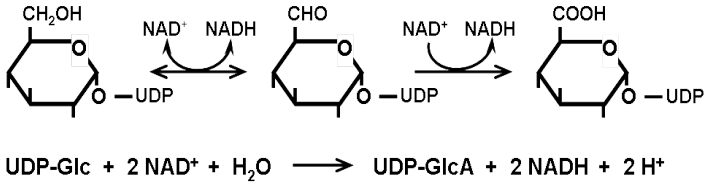


Figure 1.8. *Reaction performed by UDP-glucose dehydrogenase.*
The OH groups of C2, C3 and C4 were omitted for clarity (adapted from Gainey *et al*. 1975).

Other members of this small group include GDP-mannose dehydrogenase and UDP-N-acetylmannosamine dehydrogenase (Campbell *et al*. 2000; Ge *et al*. 2004). Since the discovery of this enzyme in guinea pig and bovine liver more than 50 years ago (Strominger *et al*. 1954), other UGDs have been reported in bacteria (Arrecubieta *et al*. 1996; Campbell *et al*. 1997), plants (Stewart & Copeland, 1998; Kärkönen *et al*. 2005), fungi (Griffith *et al*. 2004), viruses (Landstein *et al*. 1998) and mammals, including the human UDP-glucose dehydrogenase (Sommer *et al*. 2004; Huh *et al*. 2004). The bacterial UGD was described as an active 45 kDa monomer or as a 90 kDa dimer (Campbell *et al*. 1997, 2000), although the degree of oligomerisation varies between organisms.

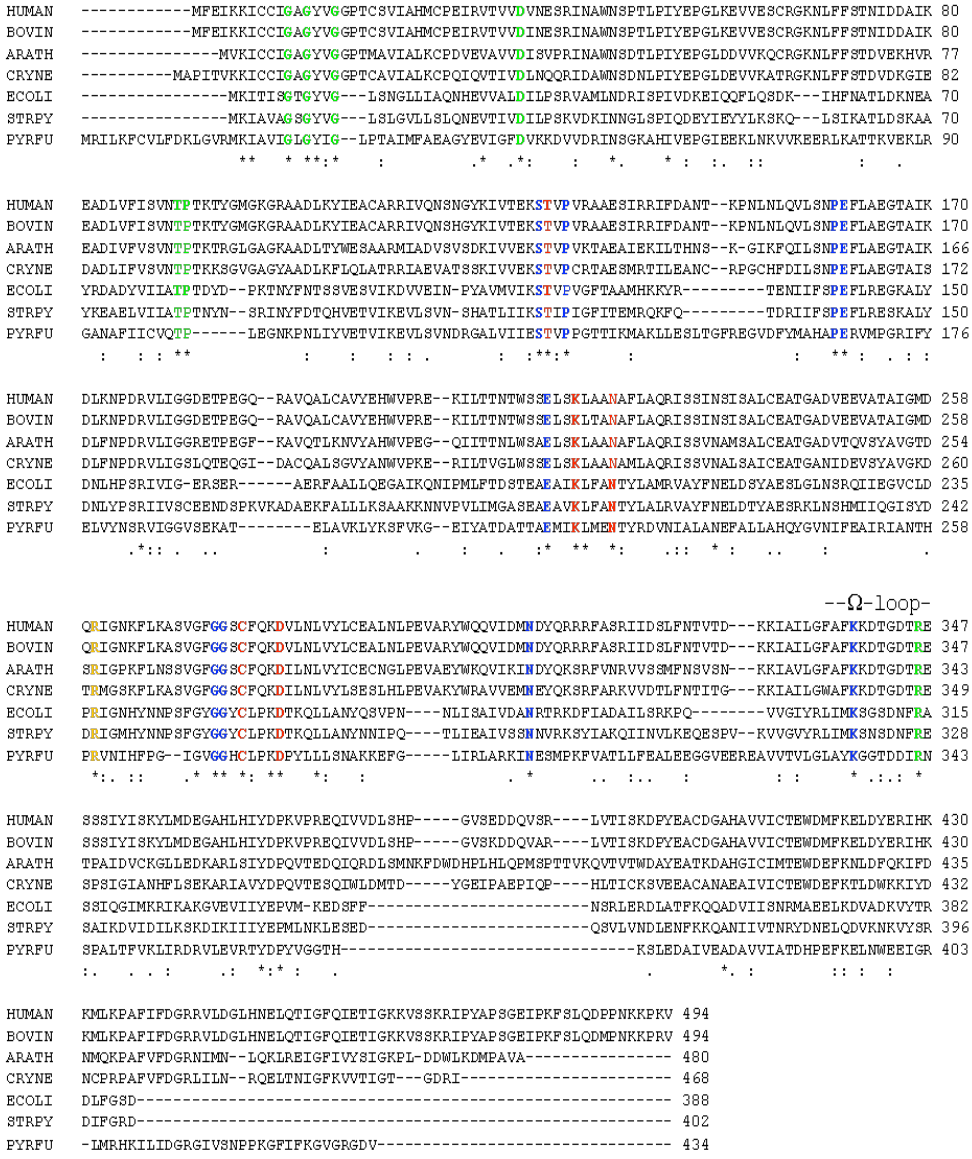


Figure 1.9. *Amino acid sequence alignment of representative UGDs.*

UGDs sequence alignment using ClustalW (Higgins *et al*. 1994) from mammal (human, Spicer *et al*. 1998; bovine, Hempel *et al*. 1994), plant (*A. thaliana*; Kaneko *et al*. 2000), fungi (*C. neoformans*, Moyrand & Janbon 2004), bacteria (*E. coli*, Roman *et al*. 2003; *S. pyogenes,* Dougherty & van de Rijn 1993) and archaea (*P. furiosus;* Robb *et al.* 2001) sources. Functionally conserved residues are highlighted, in particular those involved in NAD^+ binding (green); structural integrity/fold (blue); catalytic activity (red); hydrogen bonding with the UDP-sugar nested at the dimer partner (orange), as described by Campbell and co-workers (Campbell *et al*. 2000).

Most eukaryotic UGDs are described as functional hexamers or "trimers of dimers" of approximately 300 kDa (Feingold & Frazen 1981) with a few exceptions, for example, *Cryptococcus* fungi in which UGD is dimeric (Bar-Peled *et al*. 2004). Regardless of its multimericity and due to high level of similarity in the primary amino acid sequence, UGDs from different organisms are thought to have the same action mechanism, and thus the same residues involved in the two hydrides transfer reaction. A sequence alignment of several UGDs (see Figure 1.9. above) highlights the functional residues of UGD, including those involved in the redox reactions, as described by Campbell and co-workers (Campbell *et al.* 2000). In 2000, Campbell and co-workers described the first three-dimensional structure of the UDP-glucose dehydrogenase from *Streptococcus pyogenes* (Figure 1.10.) elucidating the mechanism of action of this enzyme and some of the residues involved in the oxidation of UDP-glucose to UDP-glucuronic acid (Campbell *et al*. 2000). The exact role of each residue involved has been extensively described later, for both the bacterial (Ge *et al*. 2004) and human (Easley *et al*. 2007) enzymes. The *S. pyogenes* enzyme comprises 402 residues and has two distinct α/β domains, each containing a core β-sheet surround by α-helices. The N-terminal domain (residues 1-196) is a classical NAD^+ binding Rossmann fold domain, and the C-terminal (residues 229-402), a UDP-sugar binding domain also showing a dinucleotide-like binding fold. These domains are connected by a long central α-helix (residues 197-228) that serves as the core for the dimer interface (Campbell *et al*. 2000).

When the present studies started, only the structure of *Streptococcus pyogenes* UDP-glucose dehydrogenase was available (PDB entries 1DLI and 1DLJ). Meanwhile, the UGD structures of *Caenorhabditis elegans*, human and *Porphyromonas gingivalis* were determined and deposited in the PDB (PDB codes 2O3J, 2Q3E/2QG4 and 3GG2, respectively).

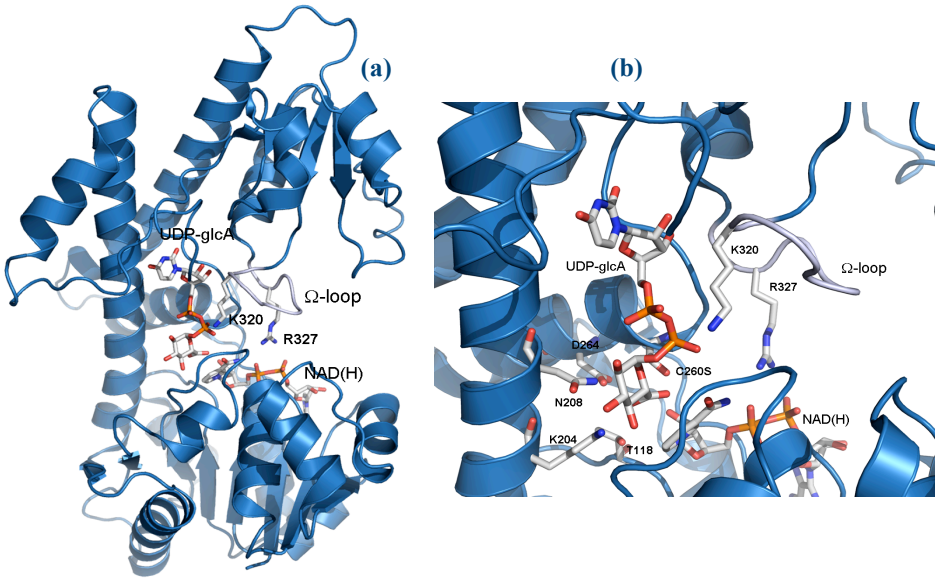


Figure 1.10. T*he three dimensional structure of S. pyogenes UGD (Cys260Ser mutant).* (a) Overall view of *S. pyogenes* UGD monomer (in cartoon representation coloured blue) with bound NADH and UDP-GlcA in stick representation coloured in grey for carbon atoms, red for oxygen, blue for nitrogen and orange for phosphorous. (b) Detailed view of the reaction site with the catalytic and Ω–loop residues as described by Campbell. (PDB code 1DLI; Campbell *et al*. 2000). Residues in stick representation coloured as in a. Images prepared with Pymol (DeLano 2008).

1.1.4. Mechanism of action of UDP-glucose dehydrogenase

This enzyme is thought to follow a "Bi-Uni-Uni-Bi ping pong" mechanism, in which the alcohol binds first and the acid is released last (Campbell *et al*. 1997). The first step of the reaction is the initial oxidation of UDP-glucose and the transfer of one of the C6' hydrides of the sugar moiety to the NAD^+ molecule, to form the first molecule of NADH and an aldehyde intermediate (Figure 1.11.). The covalent catalysis proceeds with a nucleophilic attack of a cysteine residue onto the aldehyde to form a thiohemiacetal. This intermediate is covalently bound to the enzyme and not accessible to aldehyde trapping agents, so that previous studies of this intermediate have failed (Nelsestuen &

Kirkwood 1971; Campbell *et al*. 2000). The thiohemiacetal is then oxidised to a thioester and the second hydride transfer occurs, with the formation and release of a second NADH molecule. The thioester is then irreversibly hydrolysed to UDP-glucuronic acid (Figure 1.11.; Ge *et al*. 1998; Campbell & Tanner 1999). In subsequent work done by Ge and co-workers, mutagenesis studies were performed on the UGD of *Streptococcus pyogenes*, thus revealing the role of some of the residues involved in the two redox reactions, such as the catalytic cysteine or a lysine residue that could serve as key acid/base residue for the first redox reaction (residues 260 and 204 respectively, in the streptococcal UGD) (Campbell *et al*. 2000; Ge *et al*. 2004).

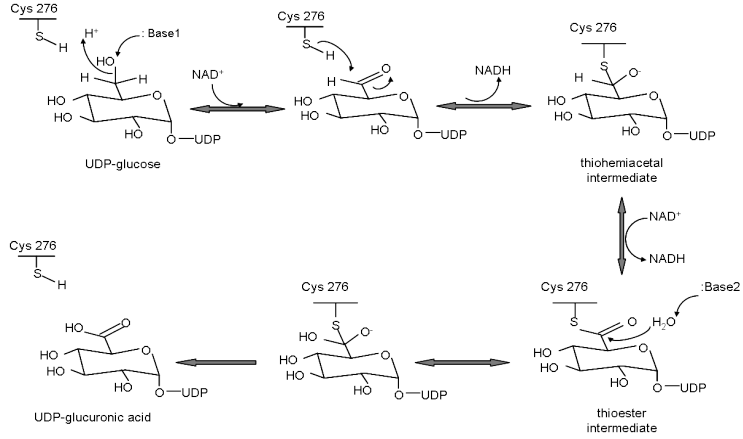


Figure 1.11. *Proposed reaction mechanism for the human UDP-glucose dehydrogenase*. The first oxidation from UDP-Glc to an aldehyde is carried out by a key acid/base residue. The catalytic cysteine attacks the aldehyde intermediate to proceed with the second redox reaction. Residue's numbering according to the human UGD. (Adapted from Easley *et al*. 2007).

Easley carried out the studies on the human UGD and revealed the importance of other residues involved in the reaction mechanism, such as the aspartate 280 (residue 264 on the bacterial enzyme) and a catalytic water molecule, elucidating more clearly how the two reactions occur (Easley *et al*. 2007). According to the results of Easley and others, obtained for the human UGD, the

two-fold oxidation initiates upon the nucleophilic attack of aspartate 280 to a water molecule, which positions and activates this water for catalysis. The activated water can then attack the proton of the C6' hydroxyl and start the first oxidation reaction. The solvent exposed Asp280 attacks the water molecule for a second cycle, which in turn attacks the catalytic Cys276, as this residue has to be deprotonated for the second oxidation and the thiohemiacetal intermediate formation to occur. The positive charge of Lys220 would stabilise the oxyanion whilst the second oxidation and thioester hydrolysis took place (Figure 1.12). This thioester intermediate is then finally hydrolysed, and the final product; UDP-glucuronic acid is released (Ge *et al*. 2004; Easley *et al*. 2007).

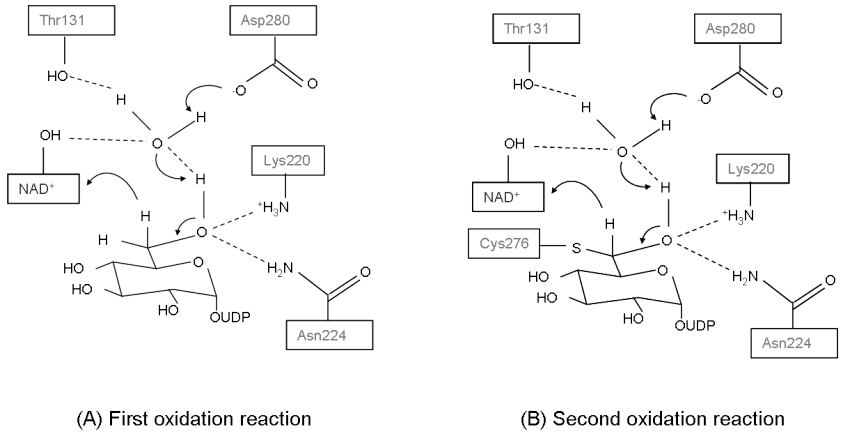


Figure 1.12. *Residues involved in the (A) first and (B) second oxidation reactions of UDP-glucose dehydrogenase*.

Residues numbered according to the human UGD (Ge *et al*. 2004; Easley *et al.* 2007).

In this thesis, the three dimensional structure of the UDP-glucose dehydrogenase from *Sphingomonas elodea* ATCC 31461 (UgdG) and from *Burkholderia cepacia* IST408 (BceC) are presented. The aim of their structural characterisation and of the present studies is to provide new information on the proteins, which may allow the design of potential activators (the case of UgdG) or inhibitors (the case of BceC) and manipulation of the enzyme's activity.

1.2. Bacterial degradation of nitroaromatic compounds

Over the past decades, the massive consumption of natural resources and the development of the industrial synthesis of chemicals have generated a number of environmental problems. This is especially the case with increasing amounts of nitroaromatic compounds and their limited incorporation into biological life cycles (Kulkarni & Chaudari 2007). These xenobiotic compounds are recalcitrant and persist in the environment because they exhibit structural elements or substituents that are rare in nature and limit their biodegradability, thus making them an impairment to survival and causing a huge impact on living organisms and ecosystems (Esteve-Núñez *et al*. 2001). Apart from a few naturally occurring products (Figure 1.13), like chloramphenicol (Smith *et al.* 2007), nitropyoluteorin (Ohmori *et al*. 1978), oxypyrrolnitrin (Kirner *et al*. 1998) and phidolopin, no other natural nitroaromatic compound has been described to date. Nitroheterocyclic and nitroaromatic compounds constitute an enormous range of chemicals that are characterised by the presence of one or more nitro (NO_2) groups on a heterocyclic or aromatic nucleus (Whiteway *et al*. 1998). Only few aromatic compounds, bearing one nitro group as a substituent of the aromatic ring, are produced as secondary metabolites by microorganisms (Higson 1992). The majority of nitroaromatic compounds present in the biosphere, like nitrobenzene, nitrotoluenes, nitrophenols and nitrobenzoates are used and/or produced in chemical industry in the manufacture of dyes, polyurethane foams, herbicides, insecticides, pesticides such as parathion and dinoseb, and solvents (Bryant and Deluca 1991), but also in pharmaceuticals and explosives, like TNT (Dillert *et al*. 1995; Esteve-Núñez *et al*. 2001). They can also be generated during a variety of combustion processes (Chaignon *et al*. 2006). The resulting waste products deriving from all of these activities cause

deleterious effects in biological systems due to their acute toxicity (Kim & Song 2005; Chen *et al*. 2007).

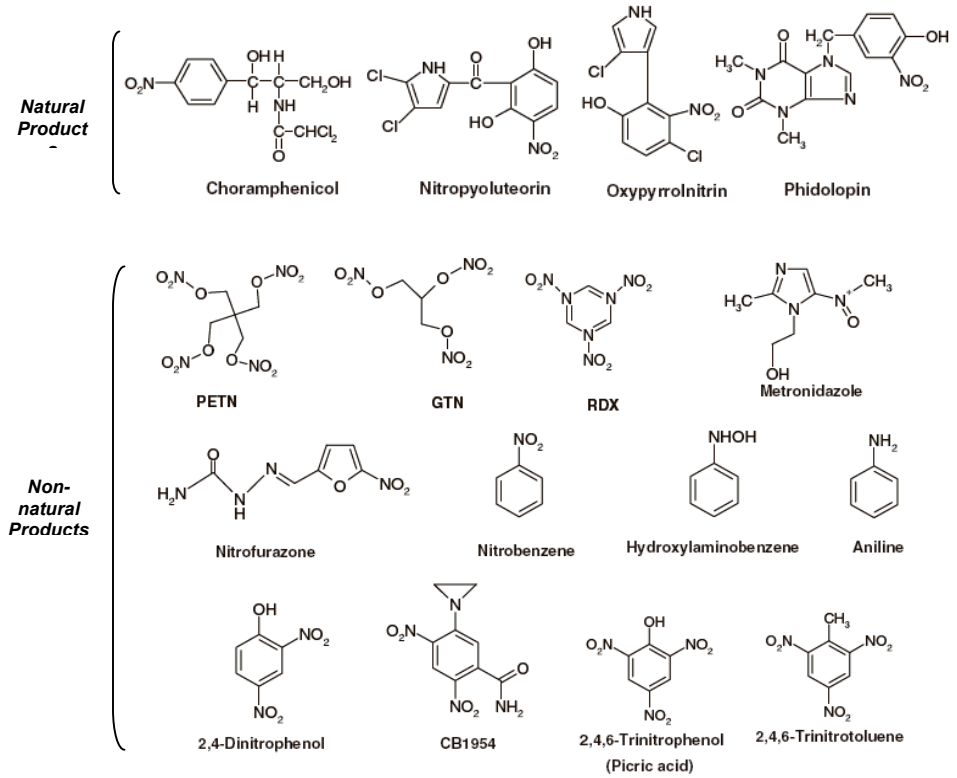


Figure 1.13. *Chemical structure of some of the most relevant natural- and non-natural nitro compounds.*

RDX, hexahydro-1,3,5-trinitro-1,3,5-triazine; GTN, nitroglycerin; PETN, pentaerythritol tetranitrate. (Adapted from Roldán *et al*. 2008).

The stability, persistence and toxicity that make nitroaromatic compounds so valuable to industry render them hazardous when released into the environment (Marvin-Sikkema & Bont 1994). Most of these compounds are (i) considered as poisonous by ingestion, subcutaneous, interperitoneal and intramuscular routes, (ii) exhibit mutagenic and carcinogenic potential (Padda *et al*. 2003; Maeda *et al*. 2007), (iii) and normally decompose and emit toxic fumes of nitrogen oxides (Kulkarni & Chaudari 2007).

2,4,6-Trinitrotoluene (TNT) is a polynitroaromatic compound and one of the most commonly known bulk explosives. TNT is used in military munitions and in civilian mining and quarrying activities (Spain 2000). It was first used on a wide scale during World War I and is still used today. However, the decades of manufacture, manipulation and processing of TNT resulted in high concentrations of contaminants accumulating in soil and water (Funk *et al*. 1996; González-Pérez *et al*. 2007). TNT and its metabolites are known to be toxic, mutagenic and carcinogenic to various organisms, including humans (Honeycutt *et al*. 1996; Esteve-Núñez *et al*. 2001).

2,4-Dinitrophenol (2,4-DNP) is frequently used as a starting material in the production of pesticides, as a polymerisation inhibitor of vinyl aromatics in petrochemical industries (Wang *et al*. 2009), herbicides and dyes. However, it is a very toxic compound for animals, plants and microorganisms because it uncouples the oxidative phosphorylation process in mitochondria (Blasco & Castillo 1997; Kimura *et al*. 2000; Hirooka *et al*. 2005). In living cells, 2,4-DNP can shuttle protons across mitochondria and chloroplast membranes. It defeats the proton gradient across biological membranes, collapsing the proton driving force that the cell uses to produce most of its ATP chemical energy. Instead of producing ATP, the energy of the proton gradient is lost as heat. In this situation, cells counteract the low yields of ATP by further oxidising their reserves of carbohydrates and fat. That was the reason why in the 1930's 2,4-DNP was used in diet pills. The cell's inability to produce ATP is proportional to the dose of 2,4-DNP that is taken. Thus, as the dose increases and energy production is made less efficient, the metabolic rate is increased and more fat is burned. Interestingly, the factor that limits increasing doses is not the lack of energy production, but rather an excessive rise in body temperature due to the heat produced during uncoupling. As a consequence, a 2,4-DNP overdose will cause a fatal fever (Pace & Pace 2002; McFee *et al*. 2002).

Given the increased prevalence of nitroaromatics within our environment, a great deal of interest has focused on the adverse health effects associated with exposure to nitroaromatic compounds, as these are amongst the most potent chemical mutagens and carcinogens identified so far (Bryant and Deluca 1991). Although several cleaning methods, such as carbon adsorption, chemical oxidation, incineration or combustion, are currently in use for the removal of nitroaromatics, none has proved to be sustainable (Kanekar *et al*. 2003; Symons & Bruce 2006; Chen *et al*. 2007; Wang *et al*. 2009). Recently, biological remediation has attracted worldwide attention for the decontamination of nitroaromatic polluted sites (Kulkarni & Chaudhari 2007). The presence of non-natural polynitroaromatic compounds in the environment creates an intense selective pressure, which has led to the evolution of microorganisms capable of degrading nitroaromatic compounds (Blasco & Castillo 1992; Ahmad & Hughes 2002). The incredible versatility inherited by several microbes enabled them to catalyse the mineralisation and/or non-specific transformation of nitroaromatics either by aerobic or anaerobic processes (Kulkarni & Chaudhari 2007). For example, certain strains of *Pseudomonas* (Robertson & Jemba 2005; Kahng *et al*. 2007) and fungi are able to metabolise TNT and use it as nitrogen source (Bayman & Radhkar 1997; Hawari *et al*. 2000). During the last few decades, intensive research has led to the discovery of microbes that are capable of degrading a vast array of toxic compounds. At the same time, an increasing number of nitroaromatic compounds susceptible to microbial degradation has been discovered and a considerable amount of knowledge has been developed about the biodegradation of these compounds (Kulkarni & Chaudhari 2007). Oxygen is the reactive molecule mostly used in the metabolism of many naturally occurring aromatic molecules, through the action of mono- and di-oxygenases. These enzymes give cells the possibility to degrade aromatic compounds and generate many useful intermediates of central metabolic pathways (Spain 1995; Johnson & Spain 2003).

A similar mechanism of oxidative attack of the nitrated aromatic compounds leading to hydroxyl derivatives, via mono or di-oxygenases, followed by the opening of the aromatic ring, is also the expected pathway for the metabolism of mononitroaromatic and even for some dinitroaromatic compounds (Roldán *et al*. 1998; Caballero *et al*. 2005a). However, in the case of polynitrated aromatics, the presence of more nitro groups in the aromatic ring renders the oxidative pathway less efficient. In effect, the deficiency of electrons in the ring, due to the nitro substituents impairs the oxygenase mediated electrophilic attack on the aromatic ring. As a consequence, the reduction of polynitroaromatic compounds is usually favoured both in aerobic or anaerobic conditions (Khan *et al*. 1997; Johnson & Spain 2003; Kim and Song 2005).

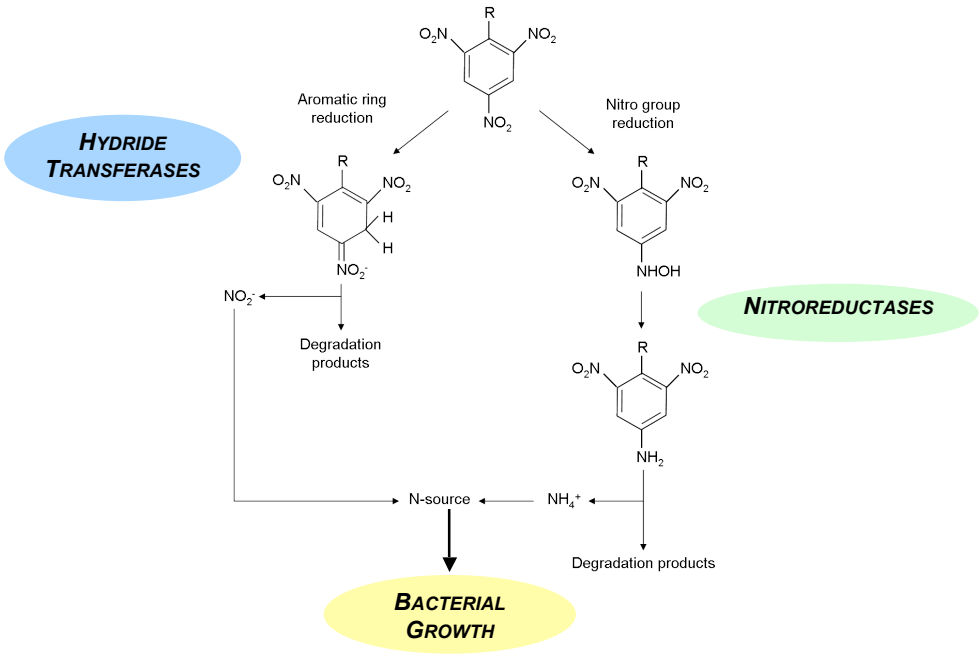


Figure 1.14. *Reductive pathways of polynitroaromatic compounds in bacteria*.
Two pathways are possible: reduction of the aromatic ring through the addition of H^- ions to generate hydride-Meisenheimer complexes catalysed by hydride transferases, or the reduction of the nitro groups (nitroreduction) through the successive addition of electrons, catalysed by nitroreductases. In both cases, an inorganic nitrogen source is generated and it might be used for bacterial growth (adapted from Roldán *et al*. 2008).

Reduction is an initial step in the metabolism of a variety of structurally diverse nitroaromatic compounds, including nitrofurans, nitropyrenes, nitrobenzenes and nitrophenols. This reduction can occur by two dinstinct pathways: (a) the reduction of the aromatic ring by the addition of hydride ions by hydride transferases, forming Hydride-Meisenheimer complexes, which can be further metabolised with the concomitant release of nitrite, or (b) by the reduction of the nitro groups to hydroxylamino or amino compounds, by the actions of nitroreductases, with the release of ammonium if further metabolised (see Figure 1.14. above; Esteve-Núñez *et al*. 2001; Kim and Song 2005).

1.2.1. Bacterial Nitroreductases (NRs)

Nitroreductases belong to the family of flavin-coupled dehydrogenases and reductases that are able to reduce a broad range of nitroaromatic compounds as well as quinones and flavins. These enzymes require reduced nicotinamide adenine dinucleotide (NADH) or reduced NADPH for catalysis (Haynes *et al*. 2002; Chaignon *et al*. 2006; Choi *et al*. 2008). Enzymatic nitroreduction can be performed either through a one- or two-electron mechanism (Figure 1.15.). In the presence of O_2, the one-electron reduction of the nitro moiety leads to superoxide radical formation (Caballero *et al*. 2005b). These enzymes are termed oxygen-sensitive or type II nitroreductases, because these radicals can be easily reoxidised to the parent compounds by O_2, in the so-called "futile cycle", generating superoxide. Thus, these enzymes can mediate the reduction of nitroaromatics only under anaerobic conditions (Watanabe *et al*. 1998; Lee *at al*. 2008). The second reaction is performed by enzymes denominated oxygen-insensitive or type I nitroreductases. In the presence or absence of O_2, they catalyse the sequential reduction of the nitro groups through the addition of electron pairs from NAD(P)H to produce nitroso, hydroxylamine and amino derivatives. However, some of the hydroxylamine intermediates are even more

hazardous than the parent compounds (Koder & Miller 1998; Watanabe *et al*. 1998; Roldán *et al*. 2008).

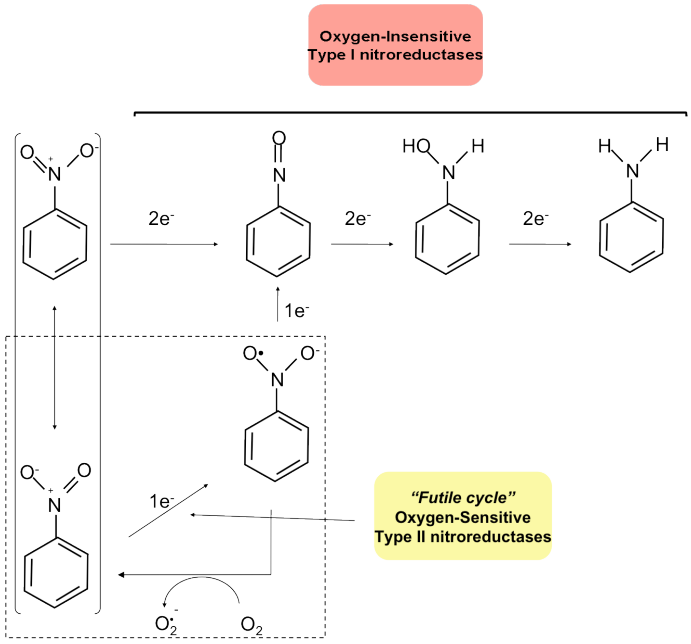


Figure 1.15. *Mechanism of action of Type I and Type II nitroreductases*.
Type II nitroreductases catalyse the $1e^-$ reduction of the nitroaromatic compound forming a nitro anion radical, which in the presence of O_2 will generate superoxide and the parental compound in the so called "futile cycle". Type I nitroreductases catalyse the reduction to a nitroso intermediate, after the $2e^-$ addition to the nitroaromatic compound. Successive enzymatic addition of pairs of electrons leads to the final amino product. The aromatic ring may have several substituents that have been omitted for clarity (adapted from Roldán *et al*. 2008).

Nitroreductases are widely distributed in the bacterial kingdom, but several examples of nitroreductase-like enzymes can also be found in archaea and eukaryotic organisms (Marques de Oliveira *et al*. 2007). Moreover, most bacteria contain several types of nitroreductases, although the best studied belong to the type I oxygen-insensitive enzymes, as several have been purified and/or their genes have been cloned and characterised. Some examples of the most relevant nitroreductases are listed in Table 1.2.

Table 1.2. *Examples of the most relevant bacterial Type I oxygen-insensitive bacterial nitroreductases.* (Adapted from Kulkarni & Chaudhari 2007; Roldán *et al*. 2008).

Bacterial Nitroreductases	Main substrates / Metabolic pathway (if known)	Electron donor(s)	References
Escherichia coli NfsA (major nitroreductase)	Nitrofurazone (antibiotic) and other nitro compounds, chromate	NADPH	Bryant *et al*. 1981; Zenno *et al*. 1996a; Zenno *et al*. 1998.
Escherichia coli NfnB (minor nitroreductase)	Nitrofurazone (antibiotic) and different nitro compounds (including CB1954 and CL-20), chromate	NADPH NADH	Zenno *et al*. 1996b-c; Race *et al*. 2005.
Salmonella enterica NfnB	*p*-nitrophenol, *p*-nitrobenzoate, 1-nitropyrene, 2-aminofluorene, menadione, flavins	NADPH NADH	Watanabe *et al*. 1998; Salamanca-Pinzón *et al*. 2006
Enterobacter cloacae NTR	Nitrofurans, nitrobenzenes, quinones, nitroimidazoles, TNT (reduction to Meisenheimer complex with release of nitrite)	NADPH NADH	Bryant & Deluca 1991; Bryant *et al*. 1991; Haynes *et al*. 2002.
Klebsiella sp. NtrI	2,4-dinitrotoluene, TNT (reduction to hydroxylamino dinitrotoluenes, aminonitrotoluenes and to nitrite via denitration)	NADPH NADH	Kim *et al*. 2003; Kim & Song 2005.
Vibrio fischerii FraseI	FMN, quinones, different nitro compounds	NADPH NADH	Zenno *et al*. 1994.
Vibrio harveyi FRP	FMN, different nitrocompounds, chromate	NADPH	Lei *et al*. 1994.
Bacillus subtilis YwrO	CB1954 (used in antibody-directed enzyme prodrug therapy)	NADPH NADH	Anlezark *et al*. 2002.

Bacillus subtilis YwcG (NfrA1)	Flavins, nitrofurazone, nitrofurantoin	NADPH	Zenno *et al*. 1998b; Chaignon *et al*. 2006.
Staphylococcus aureus NfrA	Flavins, nitrofurazone, nitrofurantoin	NADPH	Streker *et al*. 2005.
Clostridium acetobutylicum NitA	TNT, 2,4-dinitrotoluene	NADH	Kutty & Bennet 2005
Clostridium acetobutylicum NitB	TNT, 2,4-dinitrotoluene	NADPH NADH	Kutty & Bennet 2005
Pseudomonas pseudoalcaligenes NbzA	Nitrobenzene (release of ammonia), TNT (formation of 2,4-dihydroxylamino-6-nitrotoluene), 4-nitrobiphenyl ether	NADPH	Sommerville *et al*. 1995; Fiorella & Spain 1997.
Pseudomonas putida PnrA	TNT, 2,4-dinitrotoluene, 4-nitrotoluene,4-nitrobenzoate, 3,5-dinitroaniline, 2-nitrophenol (release of nitrite, formation of cathecol), 4-nitrophenol	NADPH	Zeyer & Kocher 1988; Caballero *et al*. 2005b; Kulkarni & Chaudhari 2006.

A phylogenetic analysis of several annotated sequences of bacterial nitroreductases shows that the enzymes belonging to the type I class can be further classified into two groups. Group A are usually NADPH-dependent enzymes, of which NfsA from *Escherichia coli* is the main representative (as well as the homologue SnrA from *Salmonella enterica*); and Group B, which are able to use both NADH and NADPH as electron donors, with NfnB from *Escherichia coli* being the representative enzyme (Figure 1.16.). The crystal structure of several nitroreductases has been determined, including the major nitroreductase NfsA from *E. coli* and FRP from *V. harveyi* (pdb codes 1F5V and 1BKJ, respectively) and the minor nitroreductase from *E. coli* NfsB, *E. cloacae* NTR and *V. fischerii* FraseI (pdb codes 1DS7, 1KQB and 1VFR,

respectively) (Tanner *et al*. 1996; Koike *et al*. 1998; Parkinson *et al*. 2000; Kobori *et al*. 2001; Haynes *et al*. 2002).

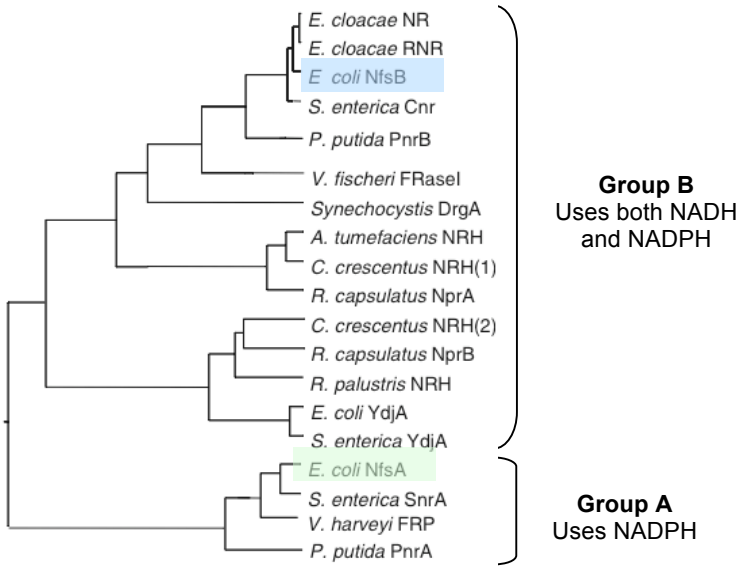


Figure 1.16. *Phylogenetic tree of the Type I oxygen-insensitive bacterial nitroreductases*. Sequence alignment and dendogram were performed with the program ClustalW (Higgins *et al*. 1994). The NfsA from *Escherichia coli* (Zenno *et al*. 1996a), its homologous SnrA from *Salmonella typhimurium* and PnrA from *Pseudomonas putida* JLR11 (Caballero *et al*. 2005b) amongst others, belong to Group A. Group B includes *Enterobacter cloacae* retronitroreductase (RNR) (Koder *et al*. 2001) and nitroreductase (NTR) (Bryant *et al*. 1991), *Salmonella enterica* Cnr, *Pseudomonas putida* JLR11 PnrB (Caballero *et al*. 2005a), *Vibrio fischerii* NAD(P)H-flavin oxidoreductase FraseI (Zenno *et al*. 1994) as well as other uncharacterised nitroreductases (YdjA) from *E. coli* and *Salmonella*.

The analysis of the available three-dimensional structures, together with amino acid comparisons reveals that the overall fold of these proteins is similar, even if the sequence similarity is low between the different members (Zenno *et al*. 1994). Bacterial nitroreductases are globular proteins with an $\alpha+\beta$ fold (Figure 1.17.) including conserved binding domains for the redox centre (FMN) and the electron donor NAD(P)H. Interestingly, nitroreductases lack the canonical Rossmann fold, typical of nucleotide binding proteins (Hecht *et al*. 1995).

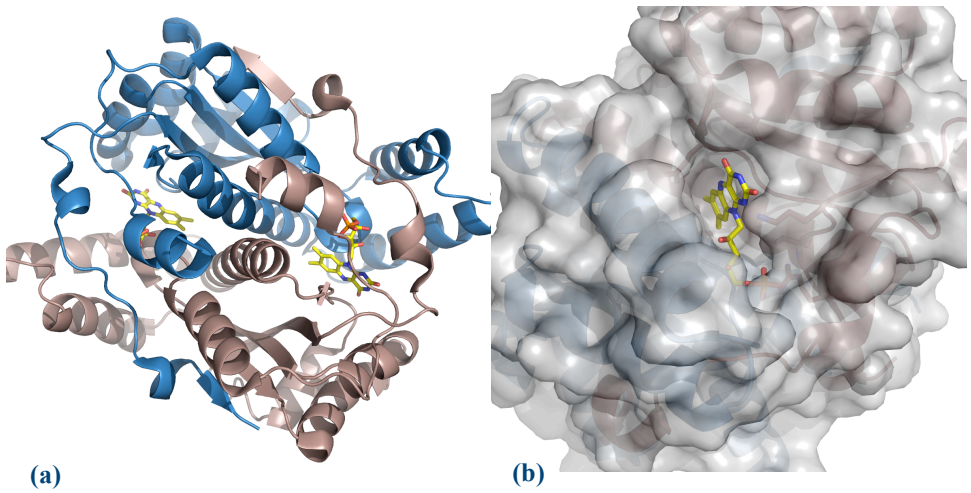


Figure 1.17. *Three-dimensional structure of NfnB, the minor nitroreductase from E. coli*. (a) The dimeric structure of NfnB (PDB code 1DS7; Parkinson *et al*. 2000) with prosthetic groups FMN bound at the dimer interface. Macromolecule in cartoon representation coloured blue and brown for monomers A and B, respectively. (b) FMN binding pocket. Polypeptide represent by its transparent solvent accessible surface and cartoon (coloured as in a). Cofactor in stick representation coloured yellow for carbon, blue for nitrogen, red for oxygen and orange for phosphorous atoms. Figures were produced using Pymol (DeLano 2008).

Nitroreductases are homodimeric enzymes with the two polypeptide chains of 24-30 kDa each, tightly associated, forming an extensive interface with deep pockets where the FMN prosthetic groups are found, establishing hydrogen bonds and hydrophobic interactions with residues from both monomers and therefore, homodimers are likely to be the physiological units of nitroreductases (Koike *et al*. 1998). These interactions are somewhat conserved and involve similar residues in the members of the type I nitroreductase family. Despite the little sequence identity, the core of these proteins is structurally equivalent, possibly indicating a common ancestral flavoprotein (Parkinson *et al*. 2000). The enzymatic reaction follows a "ping-pong bi-bi" mechanism, where the substrate NAD(P)H enters the binding pocket, is oxidised and leaves it in the form of $NAD(P)^+$ before the second substrate enters and binds (Zenno *et al*. 1996b).

1.2.2. Physiological role of bacterial nitroreductases

Although nitroreductases have been well characterised both structurally and biochemically, particularly the minor nitroreductase from *E. coli*, NfnB, the physiological role of these proteins is still an unanswered question. The intensive pressure and selectivity imposed by the growing presence of xenobiotics in the environment has led to the adaptation of microbes and evolution of new metabolic pathways. Such adaptation may encompass mutations at the nitroreductase active sites, decreasing the substrate specificity and allowing the degradation of a wider range of nitroaromatic and nitroheterocyclic compounds (Timmis & Pieper 1999, Roldán *et al*. 2008). Moreover, different physiological roles have been proposed for bacterial nitroreductases. Due to their capacity to reduce flavins, quinones and many types of polynitrated compounds, their role in detoxification has been well accepted. Some nitroreductases seem to have a specific role in the transformation of certain nitroaromatic compounds, as they are located in operons whose genes are involved in particular metabolic pathways. For example, in *Pseudomonas putida* the gene cluster encoding the enzymes for 4-chloronitrobenzene metabolism includes a chlorobenzene nitroreductase, which has a preference for chloronitrobenzene as the substrate (Xiao *et al*. 2006). Another example is the major nitroreductase NfsA from *E. coli* that is highly induced by paraquat, a potent oxidant that generates superoxide. Also, the *ŋfsA* gene is under the control of the SoxRS system, which is involved in the oxidative stress response (Liochev *et al*. 1999, Patterson *et al*. 2002). A third example is FraseI from *Vibrio fischerii* which is involved in the bioluminescent process, catalysing the reduction of the flavin by NAD(P)H, which is required for the luciferase reaction (Lei *et al*. 1994). Interestingly, FraseI and *E. coli* minor nitroreductase, NfnB, are structurally and biochemically very similar. In effect, Zenno and co-workers showed that a single-residue mutation could

transform NfnB into a flavin reductase similar to the FraseI of *V. fischeri* (Zenno *et al*. 1996c). In the same manner, the major nitroreductase from *E. coli*, NfsA, can be converted into a flavin oxidoreductase with similar properties to that of FRP from *V. harveyi* (Zenno *et al*. 1996, 1998a). Recent studies performed with the DrgA protein from the cyanobacterium *Synechocystis* sp. suggest that this enzyme, besides nitroreductase, flavin reductase and quinone reductase activities, may also participate in iron metabolism (Takeda *et al*. 2007). Nevertheless, bacterial type I nitroreductases have received great attention over the last years because of their main role in reducing the toxicity of nitroaromatic compounds, and therefore of their potential use in bioremediation, but also in biomedicine, particularly in cancer therapies (Roldán *et al*. 2008). The ability of these enzymes to also metabolise inert pharmaceutical nitroaromatics into the corresponding highly reactive and cytotoxic hydroxylamine derivatives, allows them to be used as prodrug activators (Johansson *et al*. 2003; Emptage *et al*. 2009). The most widely studied nitroreductase for cancer chemotherapy is the NfnB from *E. coli*, mainly with the dinitrobenzamide prodrug class, and in particular with the 5-aziridinyl-2,4-dinitrobenzamide CB1954. This drug is enzymatically converted by NfnB into the 4-hydroxylamino (highly toxic) and the 2-hydroxylamino (less toxic) derivatives. The 4-hydroxylamino derivative is then converted intracellularly to a DNA-reactive species (Anlezark *et al*. 1992; Green *et al*. 2004). In the antibody-directed enzyme prodrug therapy (ADEPT), NfnB is conjugated with a specific antibody for tumour cells (directing the enzyme to the tumour) and then converts an inert drug, like CB1954, into the cytotoxic hydroxylamine derivative that kills the tumour cells. Two other approaches consist in delivering the enzyme-encoding gene directly to the tumour cells via a tumour-specific vector, either viral-directed (VDEPT) or gene-directed (GDEPT) enzyme prodrug therapy (Green *et al*. 2004). The success of these concepts lies in using an enzyme not normally present in human cells, and a

prodrug that can be converted by the enzyme rapidly and specifically to give a much more cytotoxic product (Anlezark *et al*. 2002). Recently, a nitroreductase from *Bacillus licheniformis* was reported. It is able to convert the CB1954 prodrug but, it is more stable and efficient than NfnB, and thus suitable for use in the "suicide gene therapy", as it produces only the desirable cytotoxic derivative (Emptage *et al*. 2009). The potential application of emergent nitroreductases from many different bacterial species in these antitumoural techniques and, of course, in bioremediation of polynitroaromatic contaminated areas remains unexplored (Roldán *et al*. 2008).

1.2.3. *Rhodobacter capsulatus* and 2,4-dinitrophenol

Rhodobacter capsulatus B10 is a purple nonsulphur aquatic bacterium that exhibits an extensive range of metabolic capacities (Pérez-Reinaldo *et al*. 2005). These include photosynthesis, lithotrophy, and aerobic and anaerobic respiration (Tichi & Tabita 2001). It can also fix nitrogen and synthesise tetrapyrroles, chlorophyls, and vitamin B12. The biochemical versatility together with its genetic characteristics (it possesses one chromosome and an endogenous plasmid of about 130kb (Fonstein & Haselkorn 1993)) allows the bacterium to survive in a variety of habitats, including nitroaromatic contaminated environments. This bacterium is able to use 2,4-dinitrophenol (2,4-DNP) as an electron acceptor, reducing it to 2-amino-4-nitrophenol (2,4-ANP) in anaerobic conditions and in the presence of light (Figure 1.18.; Blasco & Castillo 1992, 1993). 2,4-ANP can be further metabolised under microaerobiosis, through a reaction that releases nitrite into the medium. This is a co-metabolic process, as the photoreduction depends on alternative nitrogen and carbon sources (Pérez-Reinaldo *et al*. 2005, 2008). It is also an inducible pathway that can be reversely inhibited by darkness or ammonium (Blasco & Castillo 1997).

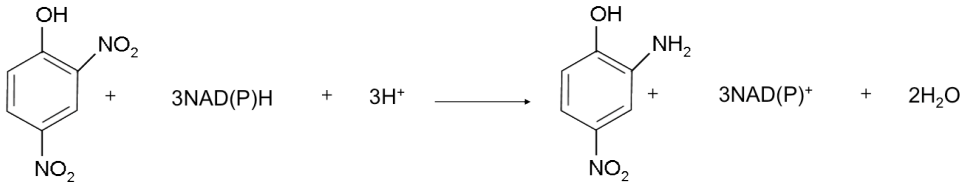


Figure 1.18. *The reduction of 2,4-dinitrophenol to 2-amino-4-nitrophenol*. (Adapted from Blasco & Castillo, 1993).

Two putative genes encoding for the major oxygen-insensitive nitroreductases, involved in the metabolism of 2,4-DNP, were reported in *Rhodobacter capsulatus* B10, *nprA* and *nprB*, although some evidence points to the existence of other non-specific nitroreductases in the bacterium genome (Pérez-Reinado *et al*. 2005). These enzymes show only 14% identity between themselves and, whereas the expression of the NprB protein is constitutive, NprA was shown to be inducible by several aromatic, nitroaromatic and heterocyclic compounds, including benzopyrene, nitrofurazone, the prodrug CB1954, paraquat and of course, 2,4-dinitrophenol (Pérez-Reinado *et al*. 2005, 2008). Recent studies on NprA show that, besides metabolising polynitroaromatic compounds *in vitro*, this enzyme also has high dihydropteridine reductase activity, higher than the activity shown for 2,4-DNP (Pérez-Reinado *et al*. 2008). Dihydropterine reductases (DHPR; EC 1.5.1.34) are involved in the recycling of tetrahydropterine cofactor used, for example, in the hydroxylating reaction of L-phenylalanine to L-tyrosine. Interestingly, N-terminal sequencing of the NfnB and a dihydropteridine reductase from *E. coli* showed that they are similar suggesting homology between the two proteins, and this is the reason why several nitroreductase-like proteins are annotated on databases as putative dihydropteridin reductases (Vasudevan *et al*. 1988, Pérez-Reinado *et al*. 2008). At present, most of the well characterised nitroreductases, at the biochemical and/or structural level are from enterobacteria. The goal of the present studies

was firstly to express, purify and characterise both biochemically and structurally, the inducible type I nitroreductase (NprA) of *Rhodobacter capsulatus* B10, both in the native or ligand-bound forms; and secondly, to map the catalytically important residues, providing new insights into the mechanistic pathway of nitroaromatic reduction. If achieved, these results may help to elucidate the true physiological role of these enzymes.

UDP-glucose dehydrogenase from *Burkholderia cepacia* IST 408 (BceC)

2. Cloning, expression, purification, crystallisation and preliminary crystallographic studies of BceC, an UDP-glucose dehydrogenase from *Burkholderia cepacia* IST 408

The work described here is based on the published refereed paper but now includes a more detailed description of the materials and methods, as well as further figures. Reference: Rocha, J; Popescu, A.; Sá-Correia, I. Fialho, A.; and Frazão, C. (2010b) "Cloning, expression, purification, crystallisation and preliminary crystallographic studies of BceC, an UDP-glucose dehydrogenase from *Burkholderia cepacia* IST 408." *Acta Crystallographica Section* F66: 269-271.

2.1. Abstract

Burkholderia cepacia is an opportunistic pathogen in cystic fibrosis patients. The production of its virulent polysaccharide biofilms requires UDP-glucose dehydrogenase (UGD) enzymatic activity. This enzyme is responsible for the NAD^+-dependent two-fold oxidation of UDP-glucose into UDP-glucuronic acid, a biofilm precursor. The 470 amino-acid residue *B. cepacia* IST 408 UGD, usually known as BceC, was cloned, expressed, purified and crystallised. Its preliminary crystallographic analysis at 2.09 Å resolution and a molecular replacement study are reported herein. Crystals belong to orthorhombic space group $P2_12_12_1$ and contain four molecules in the asymmetric unit.

2.2. Introduction

Bacteria belonging to the *Burkholderia cepacia* complex (BCC), a group of seventeen closely related species, have emerged as highly problematic

opportunistic human pathogens in patients with cystic fibrosis (CF). BCC members often express a mucoid phenotype, associated with the production of large amounts of exopolysaccharides (EPS), suggesting a possible role of EPS in BCC persistence in the CF airways (Cunha *et al*. 2003; Sousa *et al*. 2007). Since the sugar composition of the EPS produced by different members of the BCC complex is somewhat similar, this type of extracellular polysaccharide was termed as cepacian (Sist *et al*. 2003). Cepacian is composed of a branched acetylated heptasaccharide repeat-unit with D-glucose, D-rhamnose, D-mannose, D-galactose and D-glucuronic acid in the ratio 1:1:1:3:1, respectively (Cescutti *et al*. 2000). The pathway leading to the nucleotide sugar precursors necessary for cepacian biosynthesis has been proposed for the mucoid CF clinical isolate *B. cepacia* IST 408 (Richau *et al*. 2000), and the cepacian cluster of genes (named *bce*) directing its biosynthesis has been identified (Moreira *et al*. 2003).
Our interest in *B. cepacia* IST 408 uridine-5'-diphosphoglucose (UDP-glucose) dehydrogenase (encoded by the cepacian gene called *bceC*) arises from its pivotal role in providing the UDP-glucuronic acid (UDP-GlcA) precursor necessary for cepacian biosynthesis. UDP-glucose dehydrogenase (UGD; EC 1.1.1.22) catalyses the two-fold NAD^+-dependent oxidation of UDP-glucose (UDP-Glc) to UDP-glucuronic acid (UDP-GlcA) (Campbell *et al*. 1997; Campbell & Tanner 1999). UGDs have been the subject of several studies because they are present not only in bacteria, but also in multiple prokaryotic and eukaryotic cells, where UDP-GlcA is the activated donor of D-glucuronic acid, which can be converted into other compounds and serves in many critical roles in different metabolic pathways, ranging from mammals to bacteria (Campbell *et al*. 2000; Hwang & Horowitz 2002; Griffith *et al*. 2004; Sommer *et al*. 2004; Stewart & Copeland 2006). At the time of writing, only four X-ray crystal structures of UGD enzymes are publicly available, namely from the prokaryotic *Streptococcus pyogenes* (Campbell *et al*. 2000) with PDB entries 1DLI and 1DLJ; UGD from *Porphyromonas gingivalis* (Bonanno *et al*.; not

published) with PDB entry 3GG2, and two eukaryotic representatives from *Caenorhabditis elegans* PDB code 2O3J, not published, and the human UGD, PDB codes 2QG4 and 2Q3E, also not published. The determination of several 3D structures of UGDs allows the identification of the conserved stereochemical features within this family of enzymes, which may lead to a deeper understanding of its reaction mechanism.

2.3. Materials and Methods

2.3.1. Cloning and expression of *bceC* gene from *B. cepacia* IST 408

The *bceC* gene from *Burkholderia cepacia* IST 408 was amplified by polymerase chain reaction (PCR) using its genomic DNA as the template, *Pwo* DNA polymerase (Roche Diagnostics, Mannheim, Germany), and the oligonucleotides BPCRbceC (5'-GGGGGATCCATGAATCTGACTAT-3') and HPCRbceC (5'-GGGAAGCTTGAAACGGGTTAC-3') designed based on the *bceC* nucleotide sequence. The PCR product (1,413 bp) was digested with *Bam*HI and *Hind*III (recognition sites underlined) and cloned into the cloning vector pWH844 (Schirmer *et al*. 1997), generating pBceC. This plasmid carries the *bceC* gene preceded by a sequence coding for six histidines, a glycine and a serine residue present for purification purposes. The DNA insert cloned in pBceC was sequenced to confirm the fidelity of DNA amplification. The nucleotide sequence of the *bceC* gene from *B. cepacia* IST408 has been deposited in the GenBank database under the accession number GQ451909.

Overexpression of the *bceC* gene was carried out by cultivation of *Escherichia coli* SURE (Stratagene) transformants harbouring the plasmid pBceC, in 1000 ml LB medium supplemented with 100 $\mu g \cdot ml^{-1}$ of ampicillin at 310 K, until an

OD_{600nm} of 0.6 was reached. The cells were then induced with 0.3 m*M* isopropyl β-D-thiogalactoside (IPTG) for 6h at 301 K, harvested by centrifugation (10 000*g*, 30 min, 277 K) and the obtained pellets were stored at 253 K.

2.3.2. Protein purification

Cells were resuspended in 20 m*M* Na_2HPO_4/NaH_2PO_4 buffer pH 7.4, 20 m*M* imidazole, 1 *M* NaCl and disrupted in a French press. Crude cell extract was obtained by centrifugation at 27 000 *g* for 40 min at 277 K and the supernatant was applied onto a 5 ml HisTrap column (GE Healthcare), pre-equilibrated with 20 m*M* Na_2HPO_4/NaH_2PO_4 buffer pH 7.4, 20 m*M* imidazole, 1 *M* NaCl, and connected to an ÄKTA Explorer Instrument (GE Healthcare) according to the manufacturer's recommendation. The column was washed with buffer *A* (20 m*M* Na_2HPO_4/NaH_2PO_4 buffer pH 7.4, 20 m*M* imidazole, 1 *M* NaCl) to remove any unbound protein, and a concentration gradient of imidazole (20-500 m*M*) was applied. The BceC protein eluted at approximately 250 m*M* imidazole in a symmetrical chromatography peak (Figure 2.1.). The eluted fractions were immediately pooled and buffer-exchanged with 25 m*M* Tris-HCl pH 8.3, 50 m*M* NaCl, 2.5 m*M* dithiothreitol (DTT), 0.25 m*M* UDP-GlcA and 0.5 m*M* oxidised nicotinamide adenine dinucleotide (NAD^+), using a PD10 desalting column (GE Healthcare). The composition and purity of the eluted fractions were confirmed by SDS-Page analysis and only homogenous fractions were used thereafter. The protein migrates as a single polypeptide with an estimated molecular mass of approximately 52.6 kDa (51.3 kDa from the native protein, plus 1.3 kDa corresponding to the 6xHis-tag).

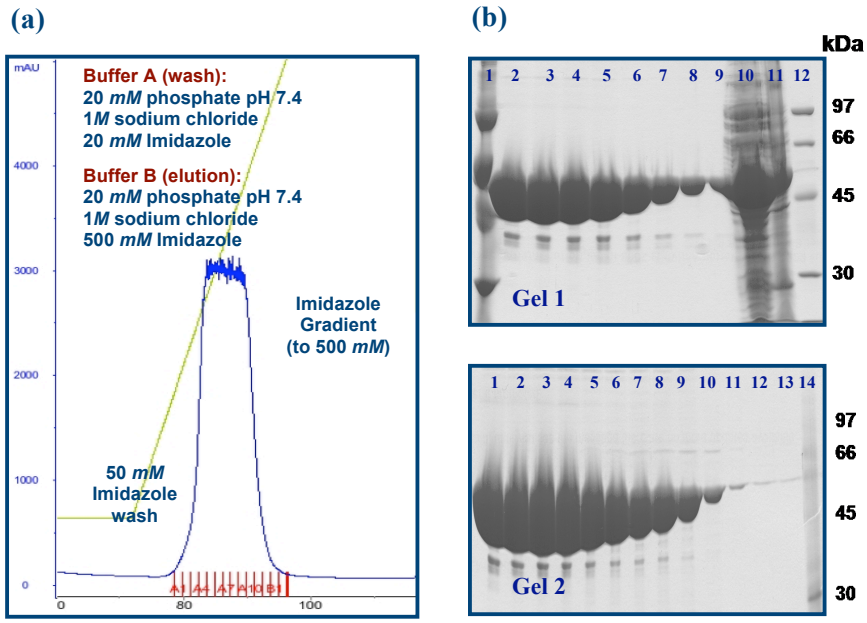


Figure 2.1. *The purification of the BceC native protein*.

(a) Affinity chromatogram for the purification of the his-tagged BceC protein. Protein detected by the absorbance at 280nm (y coordinates; blue line). BceC elutes in approximately 20 ml (x coordinates) almost pure in a symmetrical peak at approximately 250 m*M* of imidazole concentration (green line). (b) SDS-Page analysis of the HisTrap purified fractions of BceC using the Comassie blue staining. Fractions containing BceC were pooled together and the protein was concentrated up 10 mg·ml^{-1}. *Legend Gel1*: 1/12. Sigma MW marker (Sigma Aldrich); 2-9. Peak eluted fractions; 10-11. Soluble BceC. *Legend Gel2*: 1-13. Peak eluted fractions (cont.); 14. Sigma MW marker (Sigma Aldrich).

The protein was concentrated in a Vivapore 10/20 concentrator (Vivascience Ltd, UK) to approximately 10 mg·ml^{-1} before storage at 193 K. The protein concentration was determined by Bradford's method (Bradford, 1976) using bovine serum albumin fraction V (Sigma, France) as the standard, and by direct measurement of absorbance at 280 nm in a NanoDrop ND-1000 Spectrophotometer using an extinction coefficient of 38120 M^{-1}·cm^{-1}.

2.3.3. Protein crystallisation

Protein crystallisation screens were performed with a "Honeybee" Microsys 4000 XL Cartesian™ Dispensing Systems robot (Genomic Solutions, USA) using the vapour-diffusion method. 768 sitting drops consisting of 100 nl of protein solution at 10 $mg.ml^{-1}$ plus 100 nl of precipitant solution were equilibrated against 100 µl of precipitant solutions from crystallisation screens Classics, PEGs, MbClass, MbClass II, pHClear, pHClear II, Ammonium sulfate and MPD from Qiagen Canada Inc. (Montreal, Canada). A crystalline precipitate was found in various drops from the different screens, but only MbClass II and pHClear II screens generated visible single crystals. In particular, solution F7 of MbClass II (200 m*M* $(NH_4)_2SO_4$, 100 m*M* $NaCH_3CO_2$ pH 4.6 and 12% (w/v) PEG 4K) and solution D8 of pHClear II (100 m*M* citric acid pH 5.0 and 30% (v/v) isopropanol) gave the most regularly shaped crystals. Manual reproduction and further optimisation of these conditions was tried in order to decrease nucleation and promote further crystal growth. Finally, only the F7 condition of MbClass II showed reproducible results at the microlitre scale, and its crystals were improved by changing parameters such as the ratio of protein *vs.* precipitant volumes and by the use of additives from the 96 solutions of the Additive Screen, Hampton Research (Aliso Viejo, USA). Crystals were allowed to grow up to 24h, as longer times led to higher instability during cryoprotection and lower resolution in diffraction experiments. The best crystals (Figure 2.2.) were obtained at 293 K from sitting drops of 1 µl of protein solution 5 $mg{\cdot}ml^{-1}$ (see above *section 2.3.2.*) and 1 µl of precipitant solution 200 m*M* $(NH_4)_2SO_4$, 100 m*M* $NaCH_3CO_2$ pH 4.5, 11% (w/v) PEG 4K and 50 m*M* NaF, against 500 µl of precipitant solution in the well. Cryoprotection of crystals was achieved in a two-step procedure *i*) by transferring them into a washing (and stabilising) solution of mother liquor containing 15% (w/v) PEG 4K to remove precipitated protein surrounding

crystals, followed by *ii*) a quick soak in a cryoprotective solution (the same as step *i*) supplemented with 25% (v/v) glycerol.

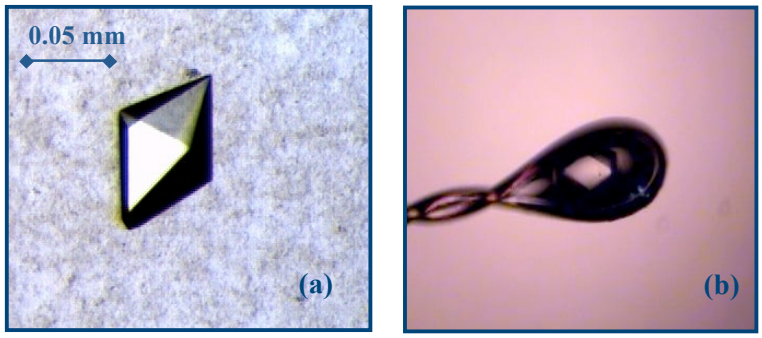


Figure 2.2. *Photographs of BceC crystals.*

(a) Native *B. cepacia* BceC forms octahedral-like crystals that reach an approximate size of 0.05, 0.05, and 0.07 mm within 24 h of growing time. (b) A cryo-cooled crystal mounted on a loop at the beam line ID14-1 of ESRF (Grenoble, France) ready for data collection.

2.3.4. X-ray diffraction analysis and phase problem solution

Cryoprotected crystals were flash cooled in liquid nitrogen and diffraction data collected at the European Synchrotron Radiation Facility (ESRF), Grenoble, France, at station ID14-1 on an ADSC Q210 CCD detector.

Diffraction images were processed with *MOSFLM* (Leslie 1992) to reflection intensities, scaled and merged together with *SCALA,* and reduced to structure factor amplitudes using *TRUNCATE* from the CCP4 Suite (Collaborative Computational Project, Number 4, 1994) (see Table 2.1 for crystal and diffraction data and statistics). A molecular replacement search by *PHASER* (McCoy *et al*. 2007) using a preliminary model at 3.4 Å resolution of the bacterial UGD from *Sphingomonas elodea* (see section *4.3.4*.) with 42 % sequence identity as the target structure led to final translation function Z-scores of 12.0, 9.9, 9.6 and 5.1 for each of the four molecules and final log likelihood gain (LLG) of 293.

Table 2.1. *Crystal and diffraction data statistics for a native crystal of BceC.*

Crystal	***BceC_native***
ESRF beam line	ID14-1
Wavelength (Å)	0.934
Resolution (Å)	40.23-2.09 (2.20-2.09)
Space group	$P2_12_12_1$
Unit cell dimensions	
a (Å)	97.49
b (Å)	109.08
c (Å)	187.69
No. of measured reflections	442772 (61050)
No. of unique reflections	118686 (17111)
Reflection redundancy	3.7 (3.6)
$\langle I \rangle / \langle \sigma(I) \rangle$	11.8 (4.2)
R_{pim} [a]	0.028 (0.101)
R_{rim} [b]	0.054 (0.195)
R_{sym} [c]	0.046 (0.166)
Completeness (%)	99.8 (99.7)
Mosaicity (º)	0.26
V_M ($Å^3Da^{-1}$)	2.36
Solvent content (%)	47.9
Wilson B ($Å^2$)	24.2
Estimated no. of molecules in ASU	4

Data within parentheses refer to the outer resolution shell; [a]$Rpim = \Sigma_h [1/(N-1)]^{1/2} \Sigma_i |I_i(h) - \langle I(h) \rangle| / \Sigma_h \Sigma_i I_i(h)$, where N is the data redundancy, I is the observed intensity and $\langle I \rangle$ is the average intensity of multiple observations from symmetry-related reflections. It is an indicator of the precision of the final merged and averaged data-set; [b]$Rrim = Rmeas = \Sigma_h [N/(N-1)]^{1/2} \Sigma_i |I_i(h) - \langle I(h) \rangle| / \Sigma_h \Sigma_i I_i(h)$,, where N is the data redundancy, I is the observed intensity and $\langle I \rangle$ is the average intensity of multiple observations of symmetry-related reflections. It is an indicator of the average spread of the individual measurements; [c]$Rsym = \Sigma_h \Sigma_i |I_i(h) - \langle I(h) \rangle / | \Sigma_h \Sigma_i I_i(h)$, where I is the observed intensity and $\langle I \rangle$ is the average intensity of multiple observations from symmetry-related reflections.

2.3.5. *In vitro* oligomerisation analysis of BceC

An estimation of the number of molecules in the asymmetric unit (a.u.) (Matthews 1968) taking into account the resolution dependent distribution of the Matthews coefficient in the PDB (Kantardjieff & Rupp 2003), indicated 4 molecules in the a.u. with a probability of 72 % (3 or 5 molecules would correspond to probabilities of 13 or 14 %, respectively). To access the oligomeric state of BceC (both in solution and crystal structure) several techniques were used. Size exclusion chromatography was carried out for molecular weight determination, by comparison with molecular weight standards (Sigma-Aldrich) ran under the same conditions. A Superdex 200 10/300 GL column (GeHealthcare) was used with a flow rate of 0.25 ml·min^{-1}. The BceC protein elutes (both in the presence or absence of substrates) in a single chromatography peak at 13.4 ml of elution volume, corresponding to an estimated molecular weight of about 100 kDa (according to the calibration curve).

(a)

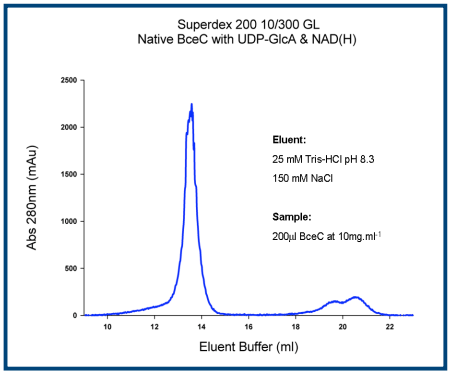


Figure 2.3. *Determination of the oligomeric state of BceC in solution using size exclusion chromatography*.

Experiments were performed on a Superdex 200 10/300 GL column (GE Healthcare) and a protein concentration of 10 mg·ml^{-1}. (a) BceC with bound substrates UDP-GlcA (0.25 m*M*) and NAD(H) (0.2 m*M*).

(b)

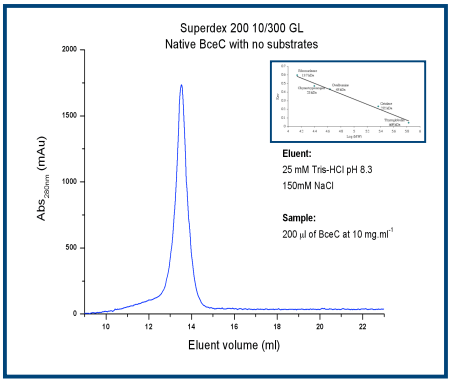


Figure 2.3. (cont). *Determination of the oligomeric state of BceC in solution using size exclusion chromatography*.

(b) BceC without bound substrates. In both cases, the protein elutes at 13.4 ml corresponding to a calculated molecular weight of approximately 100 kDa (according to a calibration curve). Molecular weight standards (Biorad): thyroglobulin, 669 kDa (9 ml); catalase, 232 kDa (12 ml); ovalbumin 43 kDa (15.15 ml); chymotrypsinogen A 25 kDa (15.7 ml) and ribonuclease A 13.7 kDa (17.7 ml). The calibration curve is shown as insert.

These results suggest that BceC is a homodimeric species under the experimental conditions used (see Figure 2.3. a, b above).

Dynamic Light Scattering (DLS) assays were performed with BceC (2 mg·ml^{-1}) at 20°C in both presence and absence of bound substrates. The experimental conditions used were the same as the size exclusion chromatography assays (Figure 2.3.). A single peak with an average particle diameter of about 8.7 nm was obtained. This corresponds to a single species in solution with an estimated molecular weight of approximately 100 kDa. These results suggest that, in these experimental conditions, BceC is a dimer regardless of the presence or absence of bound ligands (Figure 2.4. a, b).

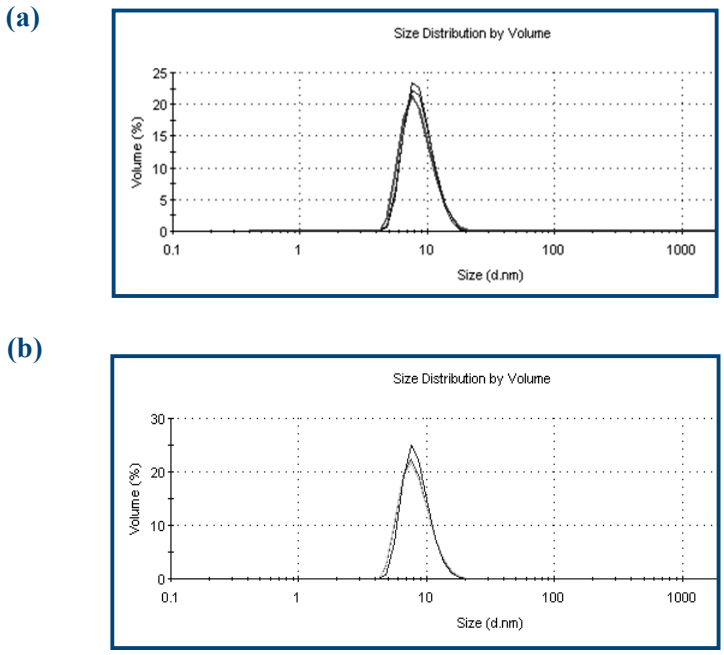


Figure 2.4. *BceC oligomeric state determination using DLS.*

(a) DLS curves for native BceC with no ligands bound. (b) DLS studies for BceC with UDP-GlcA (0.25 m*M*) and NAD(H) (0.5 m*M*) bound. Assays were performed in triplicate, using a solution of 25 m*M* Tris-HCl pH 8.3 supplemented with 150 m*M* NaCl, as the eluent buffer (same as chromatographic curves on Figure 2.3.). The assay was performed also for BceC with bound UDP-Glc or UDP-GlcA alone (data not shown), exhibiting the same results. In all cases, the average particle diameter is around 8.7 nm, which corresponds to an estimated molecular weight of approximately 100 kDa suggesting a homodimer of BceC (2 x 52.6 kDa).

The pH dependent oligomerisation state of BceC (over a range of 4-9) was also evaluated, both in the presence or absence of bound ligands (Figure 2.5. a, b). In the pH 6-9 ranges, a single peak with an average particle diameter of about 8.7 nm is shown. Again, this corresponds to a species with a molecular weight of approximately 100 kDa. In the acidic 4-5 range, a single peak is also shown but with a much higher diameter, indicating severe aggregation.

These results show BceC as a stable dimeric species over pH 6-9, with or without bound ligands.

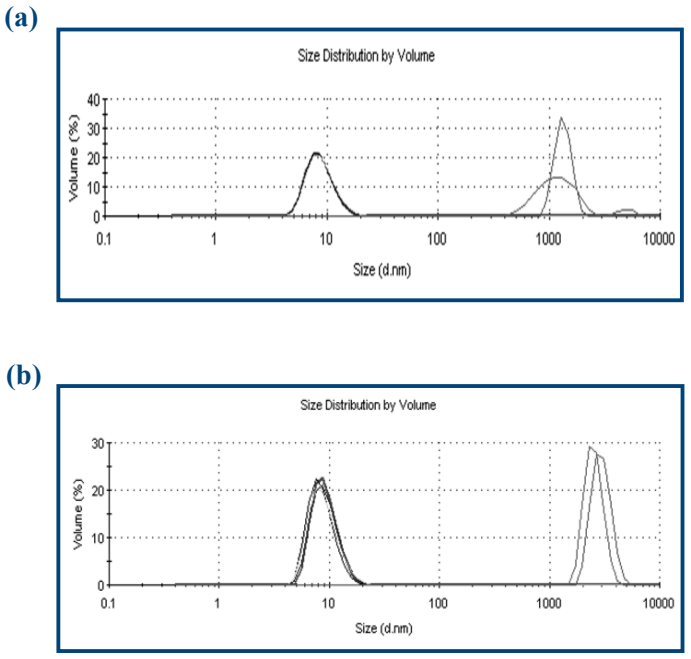


Figure 2.5. *BceC oligomeric state determination using DLS, over a pH range 4-9.*
(a) Dynamic light scattering (DLS) curves for native BceC with no ligands bound at different pHs ranging from 4-9. (b) DLS curves for BceC with bound UDP-GlcA and NAD(H) at pH ranges 4-9. All assays were performed in triplicate. The assay was performed also for BceC with bound UDP-Glc or UDP-GlcA alone (data not shown), exhibiting the same results. In pH range 6-9, the average particle diameter is around 8.7 nm, which corresponds to an estimated molecular weight of approximately 100 kDa suggesting that, in solution, BceC is a homodimeric species (2 x 52.6 kDa). At pH range 4-5, both native and ligand bound forms show severe aggregation, judging by the excessive average particle diameter (1000 or 3000 nm, respectively).

Taking this into consideration, the Matthews Coefficient (Vm= 2.36 Å^3/Da) may indicate the presence of two BceC dimers in the asymmetric unit with approximately 48 % solvent content.

2.4. Results

Uridine-diphospho-glucose dehydrogenase from *Burkholderia cepacia* IST 408 termed BceC, was cloned and expressed. Extensive crystallisation screens were performed. Octahedral-shaped crystals were obtained and diffraction data were collected and characterised. The best crystal diffracted up to 2.09 Å resolution and belonged to the orthorhombic space group $P2_12_12_1$. The phase problem was solved by molecular replacement using a preliminary model of the homologous structure of *Sphingomonas elodea* enzyme, UgdG. The BceC crystal contains four molecules in the asymmetric unit, which are arranged as a pair of dimers, corroborating the experimental results obtained in solution with other biochemical techniques.

2.5. Acknowledgements

The authors would like to thank the staff at ESRF in Grenoble, for the use of synchrotron radiation. Joana Rocha (JR) acknowledges the PhD fellowship SFRH/BD/24216/2005, from Fundação para a Ciência e Tecnologia (FCT). This work was funded by FCT grants POCTI/BME/38859/2001, POCI/BIO/58401/2004 and PTDC/QUI/67925/2006. Alma Popescu (AP) at Instituto Superior Técnico (IST) did the cloning of the *bceC* gene. The protein expression and purification was done by JR at Instituto de Tecnologia Química e Biológica (ITQB) following AP experience and protocols. JR at ITQB did the crystallisation, data collection, phase determination and structural solution of the BceC protein.

3. Structure of *Burkholderia cepacia* IST 408 UDP-glucose dehydrogenase (BceC) and the role of Tyr10 in the intermediate thioester hydrolysis

The work described in the following sections is based on a paper prepared for submission to a referee evaluated scientific journal, but here including a more detailed description of the materials and methods, as well as further figures and results. Reference: Rocha, J.; Popescu, A.; Sá-Correia, I.; Fialho, A. and Frazão, C. (2010c) "Structure of *Burkholderia cepacia* IST 408 UDP-glucose dehydrogenase (BceC) and the role of Tyr10 in the intermediate thioester hydrolysis."

3.1. Abstract

Members of the *Burkholderia cepacia* complex (BCC) are serious respiratory pathogens in immunocompromised individuals and in patients with cystic fibrosis (CF). They are exceptionally resistant to many antimicrobial agents, have the capacity of spreading between patients and lead to declining lung function, with necrotising pneumonia. BCC members often express a mucoid phenotype, associated with the secretion of an exopolysaccharide (EPS), known as cepacian. There is much evidence supporting the fact that cepacian is a major virulence factor of BCC bacteria. Uridine-5'-diphosphoglucose dehydrogenase (UGD) is the enzyme responsible for the NAD-dependent two fold oxidation of UDP-glucose (UDP-Glc) to UDP-glucuronic acid (UDP-GlcA), which is a key step in cepacian biosynthesis. UGD is therefore a logical target for the development of inhibitors or blockers of cepacian biosynthesis, preventing BCC lung colonisation and avoiding chronical infections in CF patients. Extensive mutagenic studies have been performed on the active site of UDP-glucose dehydrogenases and several crystallographic structures have been produced in

order to elucidate the mechanism of action in this family of sugar-nucleotide modifying enzymes. The UGD family is built up of two Rossmann-like domains connected by a whole alpha-region. This subdomain, similarly to other dehydrogenases, harbours the catalytic centre and is the main contributor to homodimerisation. The UGD family contains a strictly conserved tyrosine residue (Y10 in BceC) within the glycine-rich motif (GXG*Y*XG) of its first Rossmann-like domain. Several BceC Y10 mutants were produced and revealed severely impaired dehydrogenase activity. Crystal structures of native BceC and its Y10K and Y10S mutants were obtained as complexes with the reaction product UDP-GlcA at 1.75, 2.8 and 1.70 Å resolution, respectively. In order to evaluate potential selective inhibitors of BceC *vs.* the human UGD, a superposition of the UGD structures from both sources was examined, but showed a very high structural homology. Additionally, the activity studies of BceC mutants together with the examination of Tyr10 environment, led to its identification as the key catalytic residue in the final hydrolytic step of nucleotide-sugar 6-dehydrogenases.

3.2. Introduction

The BCC consists of at least 17 species that have been isolated from sputum of CF patients. Over the last 15 years, members of the BCC have emerged as highly problematic opportunistic human pathogens in immunocompromised individuals and in patients with cystic fibrosis (CF) (Sá-Correia *et al.* 2002; Moreira *et al*. 2003). CF patients are likely to see their disease accelerating, as BCC pulmonary colonisation leads to a rapid decline of the lung function and in some cases the development of the "cepacian syndrome", a fatal condition that is characterised by necrotising pneumonia with fever and bacteremia (Govan & Deretic 1996; Jones *et al*. 2001). Treatment of *B. cepacia* (a member of BCC)

infected patients has proved to be of great difficulty as this bacterium is intrinsically resistant to multiple antibiotics including the potent antipseudomonal compounds available, is associated with great virulence and it is highly transmissible. Despite all of the intensive studies carried out on *B. cepacia*, little is known about the pathogenic mechanisms of this bacterium, although it has been shown to possess many of the virulence factors known to play a role in *P. aeruginosa* CF infection (Mahenthiralingam *et al.* 2002). In the CF lung, *P. aeruginosa* secretes abundant amounts of the exopolysaccharide (EPS) *alginate* which is believed to afford protection from the host immune system response and antibiotic therapy and promotes persistence of the bacteria in the respiratory tract (Snook *et al.* 2003). Interestingly, it was observed that a vast range of the BCC isolates amongst the CF patients often express a mucoid phenotype associated with the production of EPS. Cepacian is the main EPS produced by BCC isolates and is composed of a branched acetylated heptasaccharide repeat unit with D-glucose, D-rhamnose, D-mannose, D-galactose and D-glucuronic acid in the ratio 1:1:1:3:1, respectively (Sist *et al*. 2003). The cepacian biosynthesis is a multi-step process directed by genes belonging to the cepacian gene cluster *bce* (Moreira *et al*. 2003). A potential strategy for combating BCC infection is to block *cepacian* biosynthesis, through the application of specific inhibitors for the biosynthetic enzymes. Because UDP-glucose dehydrogenase is a key enzyme required for the synthesis of the cepacian polysaccharide precursor UDP-glucuronic acid, it is a logical target for this approach (Snook *et al.*, 2003, Hung *et al.*, 2007).

UDP-glucose dehydrogenase (UGD) (EC 1.1.1.22) catalyses two successive NAD-dependent oxidations of UDP-glucose, producing UDP-glucuronic acid (Campbell *et al*. 2000; Easley *et al*. 2007). Due to the striking similarity in the primary amino acid sequence of the active site in the UGDs from different species, it has been proposed that the reaction mechanism is probably the same for all the species ranging from bacteria, to animals and plants (Hempel *et al.*,

1994, Roman *et al.*, 2003, Ge *et al.*, 2004). Intensive studies on the reaction mechanism have therefore been carried out, both on the bacterial and human enzymes (Campbell *et al*. 2000, Ge *et al*. 2004, Sommer *et al*. 2004, Easley *et al*. 2007). In this work, the structure of *B. cepacia* IST 408 UDP-glucose dehydrogenase (BceC) was determined and its biochemical characterisation was performed. Mutagenic studies were also performed on a strictly conserved residue amongst all UGDs, Tyr10, which is present in the Rossmann fold motif GXG**Y**XG, and not in the protein's active site. Studies on several Y10 mutants, described here, show that this residue plays an important role in the enzymatic activity, as the mutations proved to produce an inactive protein.

3.3. Experimental Procedures

3.3.1. DNA manipulation and generation of Y10 mutants

Point mutants of the BceC protein were generated from the wild-type pBceC construct using a protocol based on the Quickchange site-directed mutagenesis kit (Stratagene, La Jolla, CA) (see section *2.3.1.*). Based on the nucleotide sequence of *bcec,* the oligonucleotides Y10S_up (5' – ATC ATC GGC AGC GGT TCC GTA GGT CTT GTC ACC – 3'), Y10S_rev (5' – GGT GAC AAG ACC TAC GGA ACC GCT GCC GAT GAT – 3'), Y10K_up (5' – CT ATC ATC GGC AGC GGT AAG GTA GGT CTT GTC ACC GG – 3'), Y10K_rev (5' – CC GGT GAC AAG ACC TAC CTT ACC GCT GCC GAT GAT AG – 3'), Y10F_up (5' - ATC GGC AGC GGT TTC GTA GGT CTT GTC -3') and Y10F_rev (5' - GAC AAG ACC TAC GAA ACC GCT GCC GAT -3') were designed for mutants Y10S, Y10K and Y10F, respectively (mutated codons are underlined). The mutated sequences were confirmed by N-terminal sequencing, and the plasmids harbouring the Y10 mutants were transformed into

Escherichia coli SURE strain (Stratagene, La Jolla, CA). The oligonucleotides were purchased from MWG Biotech. Restriction enzymes were from Gibco BRL and DNA polymerases from Roche Diagnostics (Mannheim, Germany).

3.3.2. Protein purification and crystallisation

BceC wild type and Y10 mutants were expressed and purified as previously described (see *sections 2.3.1.* and *2.3.2.*). Briefly, the cells harbouring the recombinant constructs of pBceC, pBceC_y10f, pBceC_y10s and pBceC_y10k were grown at 37º C up to an OD_{600nm} of approximately 0.6 in LB medium supplemented with 100 $\mu g \cdot ml^{-1}$ ampicillin, and induced at 28º C for about 6 hours by addition of IPTG to a final concentration of 0.3 m*M*. Cells were harvested by centrifugation (7 000*g*, 40 min, 4º C), resuspended in phosphate buffer 20 m*M*, pH 7.4, containing sodium chloride 1 *M,* imidazole 20 m*M* and an EDTA-free protease inhibitor cocktail (Roche Diagnostics, Mannheim, Germany), and disrupted on a French press. Nucleic acids and cell debris were removed by centrifugation, 17 000*g* for 40 min at 4º C, and the supernatant was applied onto a 5 ml nickel-chelating column (HisTrap FF, Ge-Healthcare), previously equilibrated in the same phosphate buffer. Proteins were eluted using a linear imidazole gradient up to 500 m*M* (Figure 3.1. a, b, c), pure fractions were pooled and the buffer exchanged to Tris-HCl 25 m*M* pH 8.3, sodium chloride 50 m*M*, dithiothreitol (DTT) 2.5 m*M*, UDP-glucuronic acid 0.25 m*M*, and nicotinamide adenine dinucleotide (NAD^+) 0.5 m*M*. All proteins were stored at -80 ºC with final concentration of approximately 10 $mg \cdot ml^{-1}$. UDP-glucuronic acid and NAD^+ were obtained from Sigma. Wild-type crystals of BceC were obtained as described elsewhere (see *section 2.3.3.*). A sitting-drop, vapour diffusion, crystallisation screen around those conditions also provided crystals for mutant proteins Y10S and Y10K.

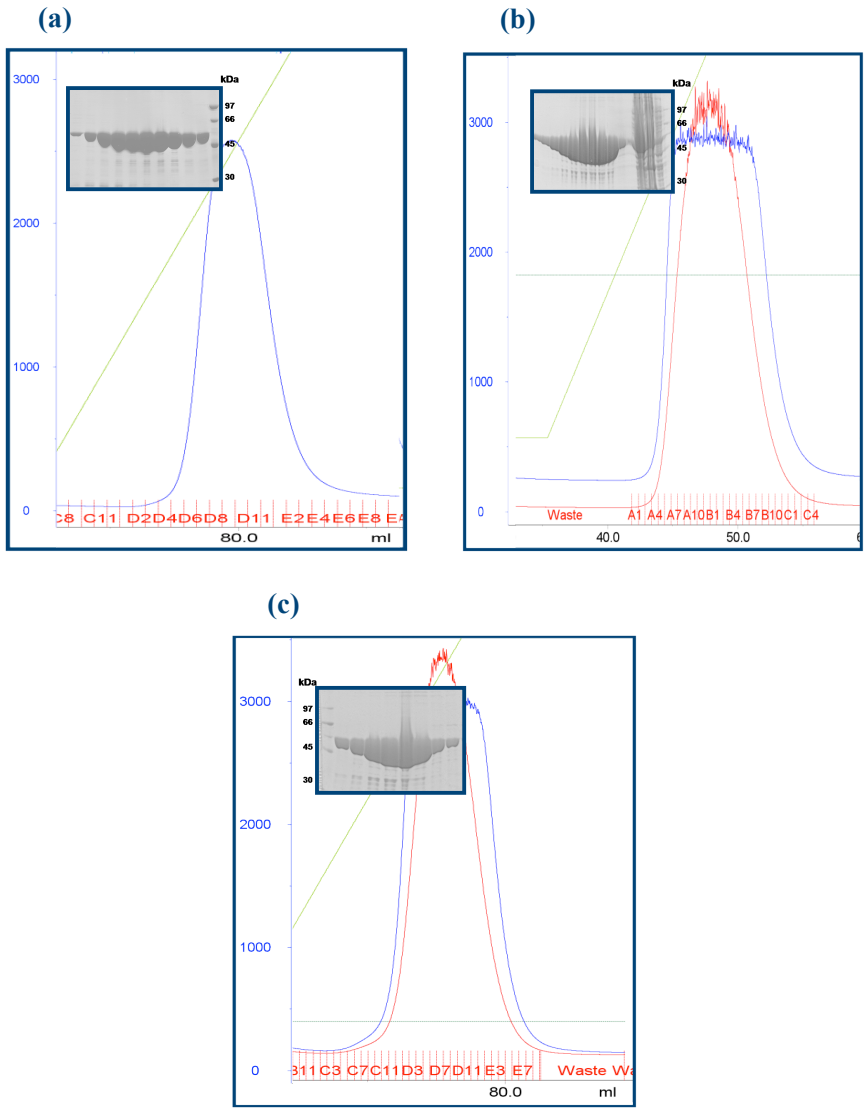


Figure 3.1. *Purification of the Y10 BceC mutants.*

All mutants were purified on a HisTrap FF Affinity chromatography column (GE Healthcare) using the native BceC protein purification protocol. Mutant proteins exhibit the behaviour of native BceC and elute at approximately 250 m*M* of imidazole concentration (green line; see *section 2.*). Protein and DNA contents were monitored by UV absorbance (mAU) at 280 nm (blue curve) and 254 nm (red curve), respectively. (a) Chromatogram and SDS-PAGE gel analysis (insert) of the Y10F BceC mutant. (b) Chromatogram and SDS-PAGE gel analysis (insert) of the Y10K BceC mutant. (c) Chromatogram and SDS-PAGE gel analysis (insert) of the Y10S BceC mutant. The SDS-PAGE gels (inserts) show only the peak fractions for each of the purifications.

In spite of more extensive crystallisation screens, no crystals were obtained for the Y10F mutant. Optimised Y10S and Y10K mutant crystals were obtained by equilibration of 1 µl of mutant proteins solution and 1 µl of a precipitant solution of sodium acetate buffer 100 m*M* pH 4.5, ammonium sulphate 200 m*M* and PEG 4K 11-14% (w/v). Octahedral shaped crystals appeared and grew within 24 hours, reaching dimensions up to 0.06-0.1 mm (Figure 3.2.).

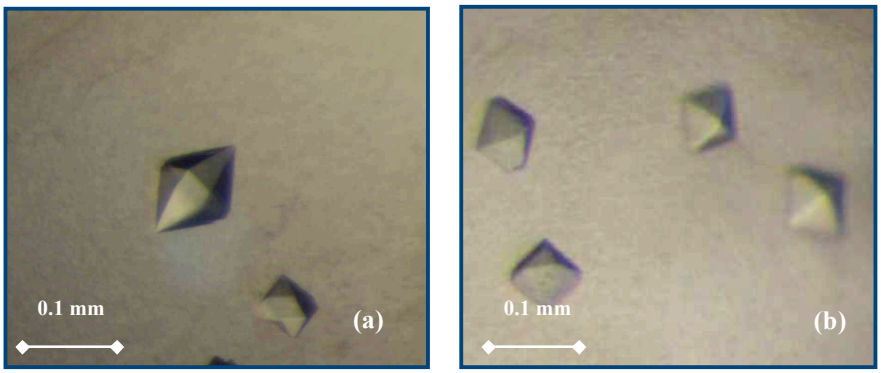


Figure 3.2. *Crystals of the Y10 BceC mutants.*
The crystals were obtained for mutants Y10K and Y10S using a crystallisation solution equal (or very similar) to that used for the native BceC protein. The BceC mutant Y10F did not crystallise, even though an extensive screening of different crystallisation conditions was performed. (a) Crystals of the Y10K mutant; (b) Crystals of the Y10S mutant.

3.3.3. Standard enzymatic activity

BceC catalyses the 1:1 conversion of UDP-glucose to UDP-glucuronic acid, while reducing two molecules of NAD^+ to NADH. For the screening of UDP-glucose dehydrogenase enzymatic activity of both native and mutant BceC proteins, the change in absorbance at 340 nm that accompanies the NAD^+ reduction to NADH was monitored. The assays were performed for 60 minutes

at 30 °C in 1 ml reaction volume containing Tris-HCl 100 m*M* pH 8.7, $MgCl_2$ 10 m*M*, DTT 2 m*M*, UDP-glucose 2 m*M* and NAD^+ 1.5 m*M*. The reaction started with the addition of 3 μg of pure protein to the cuvette. Control assays were performed with reaction mixtures lacking the enzyme. All assays were performed in triplicate and specific activities were calculated from the measured absorbance values at 340 nm using 6220 $M^{-1}cm^{-1}$ as the molar extinction coefficient of NADH. UDP-Glucose was obtained from Sigma (Sigma-Aldrich).

3.3.4. Diffraction data collection and processing

Native crystals were cryoprotected by transferring them directly into a drop of a solution of mother liquor including glycerol 25% (v/v). Cryoprotection of the Y10S and Y10K mutant crystals required consecutive transfer between drops of solutions of mother liquor including glycerol at increasing concentrations, up to 25% (v/v). Crystals were then rapidly plunged into liquid nitrogen and stored for diffraction data collection. Diffraction data were measured at beam line ID14-1 of the European Synchrotron Radiation Facility (ESRF), Grenoble, France. Diffraction images were processed with *MOSFLM* (Leslie 1992), intensities were merged and scaled with *SCALA* (Evans 2006), and the corresponding magnitudes calculated with *TRUNCATE* (French 1978), from the CCP4 suite (The CCP4 suite: programs for protein crystallography 1994).

3.3.5. Structural solution and refinement

Native BceC initial phases were obtained by molecular replacement using *PHASER* (McCoy *et al*. 2007) and a preliminary model of UDP-glucose dehydrogenase from *S. elodea* UgdG (*sections 4.* and *5.*) as the search structure. *RESOLVE* prime-and-switch was used to remove the bias from the search

structure, and to further improve the phases (Terwilliger 2004b). *ARPWARP* extended the docked model (Perrakis *et al*. 2001), which was iteratively completed and corrected on a graphics workstation using *COOT* (Emsley & Cowtan 2004) and structurally refined with *PHENIX.REFINE* (Afonine *et al*. 2005). The phases for the mutants were obtained from the BceC native structure upon truncation of Tyr10 into Ala, using a rigid-body refinement with *REFMAC5* (Murshudov *et al*. 1997) in the case of the Y10S mutant and a molecular replacement step with *PHASER* (McCoy *et al*. 2007) in case of Y10K. Models were iteratively refined with *PHENIX.REFINE* (Afonine *et al*. 2005), and edited and fitted to their electron density and completed with *COOT* (Emsley & Cowtan 2004). Model completion of disordered loops was facilitated using non-crystallographic symmetry (NCS) averaged maps within *COOT*.
The progress of the refinement was monitored by the use of R_{free}, where 5% of the reflections were constantly set aside for cross-validation.

3.3.6. Structural analysis

Detection and selection of translation-libration-screw domains in the protein chains were performed with the *TLMSD* server (http://skuld.bmsc.washington.edu/~tlsmd; Painter & Merrit 2006). Molecular stereochemistry was checked within *COOT* (Emsley & Cowtan 2004), intermolecular contacts and crystallographic packing were checked with *PISA* (Krissinel & Henrick 2007) and structural comparisons amongst the models were performed with *SSM* (Krissinel & Henrick 2004). Images were prepared with *PYMOL* (Delano 2008).

3.4. Results and Discussion

3.4.1. Crystal characterisation and diffraction data processing of native BceC and its Y10 mutants

Upon successive crystal optimisation and data collection sessions at ESRF, data sets of native BceC and its Y10S and Y10K mutants reached final resolutions of 1.75, 1.70 and 2.80 Å, respectively (Figure 3.3.). All data were integrated with *MOSFLM* (Leslie 1992) and the intensities scaled, merged and reduced to structure factors magnitudes using *SCALA* (Evans 2006). All crystals belong to space group $P2_12_12_1$, show similar cell dimensions, and contain 4 molecules in the asymmetric unit (*a.u.*), corresponding to calculated solvent contents in the range 47.5-49.5 % and Matthews coefficients within 2.34-2.41 $Å^3Da^{-1}$ (see Table 3.1. for further details).

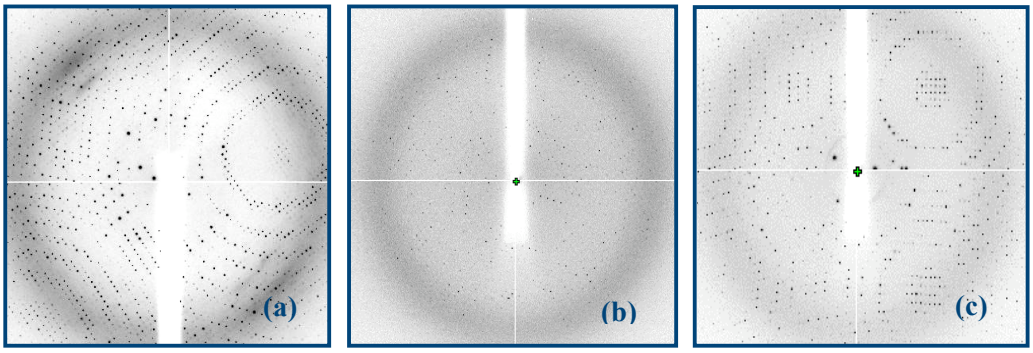


Figure 3.3. *Diffraction images from the crystals of native and Y10 BceC mutants.* Diffraction patterns of (a) native BceC crystal; (b) Y10K mutant; (c) Y10S mutant. Crystals of native and mutant BceC were tested and collect at the ESRF (Grenoble, France).

Table 3.1. *Data collection and crystallographic statistics for native BceC and Y10 mutants.*

Crystal	*Native*	*Y10S*	*Y10K*
ESRF beam line	ID14-1	ID14-1	ID14-1
Wavelength (Å)	0.9334	0.9334	0.9334
Resolution (Å)	47.56 – 1.75 (1.84 – 1.75)	47.40 – 1.70 (1.79 - 1.70)	57.83 – 2.80 (2.95 - 2.80)
Space-group	$P2_12_12_1$	$P2_12_12_1$	$P2_12_12_1$
a (Å)	97.63	97.36	101.62
b (Å)	108.93	108.62	109.25
c (Å)	187.71	187.46	183.44
No. measured refs.	750258 (109058)	803863 (115360)	206820 (30278)
No. unique refs.	197721 (29063)	217359 (31490)	50832 (7331)
Redundancy	3.8 (3.8)	3.7 (3.7)	4.1(4.1)
I/σ(I)	6.7 (2.0)	7.6 (2.0)	2.8 (1.7)
R_{pim} [a]	0.038 (0.188)	0.044 (0.283)	0.119 (0.241)
R_{rim} [b]	0.076 (0.369)	0.087 (0.554)	0.249 (0.501)
R_{sym} [c]	0.066 (0.315)	0.074 (0.474)	0.217 (0.436)
Completeness (%)	98.4 (99.7)	99.8 (100.0)	99.8 (100.0)
Mosaicity (º)	0.26	0.25	0.8
V_M ($Å^3Da^{-1}$)	2.43	2.34	2.41
Solvent content (%)	49.5	47.5	48.9
No. molecules in *a.u.*	4	4	4
Wilson B ($Å^2$)	18.31	17.66	42.22

Data within parentheses refer to the outer resolution shell; [a] $R_{pim} = \Sigma_h [1/(N-1)]^{1/2} \Sigma_i |I_i(h) - \langle I(h) \rangle| / \Sigma_h \Sigma_i I_i(h)$, where N is the data redundancy, I is the observed intensity and $\langle I \rangle$ is the average intensity of multiple observations from symmetry-related reflections. It is an indicator of the precision of the final merged and averaged data-set; [b] $R_{rim} = R_{meas} = \Sigma_h [N/(N-1)]^{1/2} \Sigma_i |I_i(h) - \langle I(h) \rangle| / \Sigma_h \Sigma_i I_i(h)$, where N is the data redundancy, I is the observed intensity and $\langle I \rangle$ is the average intensity of multiple observations of symmetry-related reflections. It is an indicator of the average spread of the individual measurements; [c] $R_{sym} = \Sigma_h \Sigma_i |I_i(h) - \langle I(h) \rangle| / |\Sigma_h \Sigma_i I_i(h)$, where I is the observed intensity and $\langle I \rangle$ is the average intensity of multiple observations from symmetry-related reflections.

3.4.2. Phase determination and structural refinement of native BceC

Initial molecular replacement attempts of native BceC using the PDB entry 1DLI, the structure of UDP-glucose dehydrogenase from *Streptococcus pyogenes* (Campbell *et al*. 2000) as the model, were not successful, most probably due to their too low homology as their sequence identity is about 23 %. However, once a preliminary model of UDP-glucose dehydrogenase from *Sphingomonas elodea* was obtained in our laboratory (see *section 4*.), with a sequence identity of 42 % against BceC, the molecular replacement solution was readily obtained (see *section 2*.). *RESOLVE*, with automatically implemented four-fold NCS averaging and the "prime and switch" procedure (Terwilliger 2004b) was used to improve the phases of the native BceC molecular replacement solution. Readily interpretable electron density maps were produced with a *figure of merit* of 0.48, which included well defined densities for an activated sugar UDP-glucuronic acid molecule nestled in each BceC monomer, in a model with 1358 out of 1912 possible residues in the *a.u*. Nevertheless, only 349 residues were docked in sequence. Using the 1.75 Å resolution diffraction dataset of native BceC, obtained in the meantime, and the program *ARPWARP* (Perrakis *et al*. 2001), the four chains model was extended to 1604 residues, of which 1530 were docked in the sequence. The protein and solvent structure were examined at a graphics workstation against their σ_A maps (Read 1986), and the model improved accordingly. As result of a translation, libration and screw motion (TLS) determination of the models using *TLMSD* (Painter & Merrit 2006), four homologous sequential domains were assigned to each chain and used both for TLS refinement and to define NCS restraints among chains, in atomic positional and isotropic *a.d.p.* refinement. Steps of structure refinement and editing were iteratively repeated until convergence of R_{work}/R_{free} to 16.3/19.5 %. Using all data the final R factor is 16.7 %.

3.4.3. Phase problem solution and structural refinement of Y10K and Y10S BceC mutants

As native BceC and its Y10S mutant produced isomorphous crystals, cell dimensions differ only up to 0.3 %, a rigid-body refinement of the four chains BceC model produced suitable phases for the Y10S mutant. For the non-isomorphous Y10K mutant crystal (cell lengths differ up to 4% in the ***a*** direction) phases were obtained by molecular replacement using the refined structure of native BceC as the search model. For the two mutants, iterative molecular refinement, inspection and edition of the structures proceeded similarly to that described for the native case. Their R_{work}/R_{free} are 16.8/19.8 % and 22.8/26.63 %, and final R-values of 16.6 and 21.7 %, for the Y10S and Y10K cases, respectively.

3.4.4. Crystal contents of native BceC and Y10 mutants

While the two residues of the linker between the His-tag (see *section 2.3.1.*) and the first methionine of the protein could be modelled as Gly-1 and Ser0 (except for Y10K mutant where only the serine was visible), the 6 histidine residues of the tag and the last 10 residues (up to 15 in the case of the Y10K mutant) of the C-termini were not visible in the electron density maps and therefore omitted from the models. Several main-chain residues are not entirely visible in the electron density in all the three structures but they were modelled according to their NCS consensual positions. They belong mainly to loops exposed to the solvent and their refined *a.d.p.* values reflect their disordered behaviour. Similarly, some side-chains, mainly from hydrophilic residues, were modelled in alternate conformations (see Table 3.2 for further details). Globally, 1841, 1841 and 1796 out of 1912 (four chains each with 8+470) residues were modelled in the native BceC, Y10S and Y10K mutant structures, respectively.

A total of 1415, 1311 and 8 solvent molecules were modelled as waters in the native, Y10S and Y10K structures, respectively. Additionally, each protein chain also encompasses a docked UDP-glucuronic acid molecule, and whenever a persistent positive (mFo-DFc) density (Read 1986), was found in a chemically suitable environment, acetate, sulphate, glycerol or Tris molecules were modelled, in particular for the higher resolution crystals of native and Y10S mutant (see Table 3.2. for further details).

3.4.5. BceC monomer topology

A topology analysis of the BceC monomer (dimensions of approximately 75, 45 and 35 Å) using *TOPS* (Westhead *et al*. 1999) shows two αβα sandwich type dinucleotide binding Rossmann-like domains, at the N- and C-terminal sections of the protein (residues 1-205 and 305-459, respectively) linked together by an entirely α-helical sub-domain (residues 206-304) (Figure 3.4.). We succeeded to co-crystallise BceC only with the reaction product UDP-GlcA, but in other UGD structures (Campbell *et al*. 2000; Kavanagh *et al*. 2007 a, b, unpublished, PDB entries 1DLJ and 2Q3E) the active site could be mapped with the two enzyme substrates of (di)-nucleotide type, the oxidising nicotinamide adenine dinucleotide (NAD^+) at the nucleotide binding N-domain (NBD), and nucleotide-sugars analogues at the C-domain, hence denominated as the sugar binding domain (SBD) (Campbell *et al*. 2000) (but see below). Each αβα domain includes two consecutive Rossmann-fold motifs built up of three parallel beta strands interspersed by (one or two) alpha helices interconnected by variable sized loops. In the NBD (β1 to β8), and after the last strand (β6) of the canonical dinucleotide binding Rossmann-like fold, one finds a 2x(αβ) motif (α7-β7-α8-β8), in which the two β-strands extend the molecule central twisted β-sheet in an anti-parallel fashion, *via* two main-chain hydrogen bonds connecting strands β6 and β7.

Table 3.2. *Refinement and model quality statistics for native and Y10 mutants BceC.*

Parameter	*Native*	*Y10S*	*Y10K*
R_{work} (%)	16.3	16.8	22.8
R_{free} (%)	19.5	19.8	26.3
R (%)	16.7	16.6	22.2
[a] Protein chain <*a.d.p.*> (Å^2)	22, 24, 30, 23 (25)	22, 24, 31, 22 (24)	30, 26, 29, 25 (28)
[a] Glucuronic acid <*a.d.p.*> (Å^2)	12, 11, 12, 12 (12)	12, 11, 14, 12 (12)	25, 25, 25, 17 (23)
No. of residues per chain	461, 460, 460, 460	461, 460, 460, 460	449, 447, 451, 449
No. of solvent waters	1415	1311	8
No. of solvent sulphates	16	14	6
No. of solvent acetates	4	5	--
No. of solvent glycerol	9	6	--
No. of solvent Tris	--	1	--
No. atoms modelled in alternating conformations	341	333	0
Bond distances rmsd (Å)	0.005	0.005	0.008
Bond angles rmsd (º)	1.0	1.0	1.3
Residues distribution in preferred, allowed, or outlier regions of Ramachandran diagram (%)	96.8, 2.9, 0.3	97.1, 2.7, 0.2	95.8, 3.9, 0.3

[a]Values for each of the four crystallographically independent molecules (chains) followed by their average in parentheses.

Such a 2x(αβ) motif is also found in the available UGD structures, but whilst in UGDs from *Caenorhabdtis (C.) elegans* (Zhang *et al*. 2006, unpublished, PDB entry 2O3J) or from *Homo (H.) sapiens* (Kavanagh *et al*. 2007a, b, unpublished, PDB entries 2QG4 and 2QE3), or from *Porphyromonas gingivalis* (Bonnano *et al*. 2009, unpublished, PDB entry 3GG2) there are two main-chain hydrogen bonds linking this motif with the central sheet, in *Streptococcus (S.) pyogenes* UGD one only finds a single hydrogen bond, and thus its 2x(αβ) motif was assigned as a small and independent additional β-sheet (Campbell *et al*. 2000).

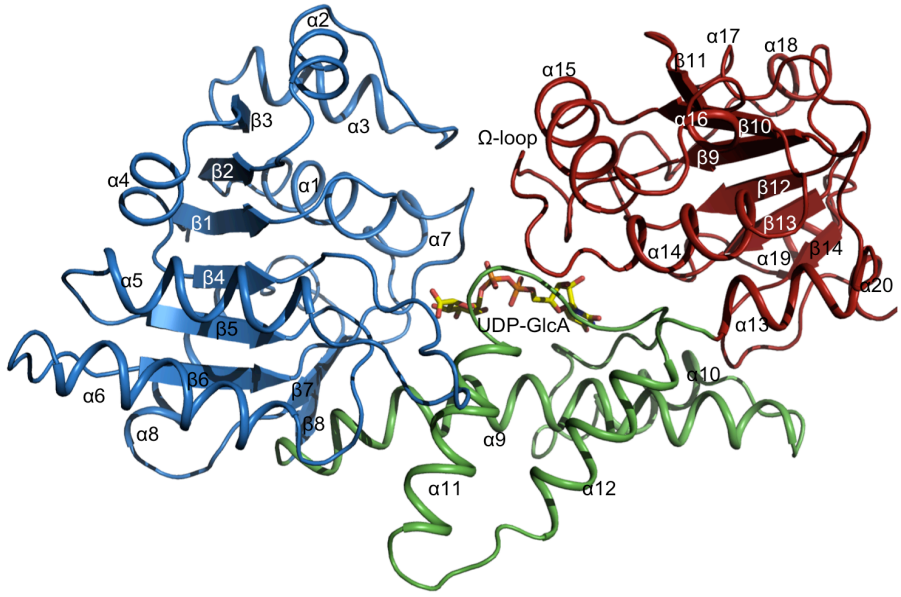


Figure 3.4. *The three-dimensional crystal structure of BceC monomer.*
Cartoon representation of UDP-glucose dehydrogenase from *B. cepacia* IST408 (BceC). The BceC monomer has dimensions of approximately 75, 45 and 35 Å. The structure consists of two αβα sandwich dinucleotide binding Rossmann fold domains, at the N- (blue) and C-terminal (red) sections, connected by a α-helical sub-domain (green), responsible for the homodimer formation. UDP-GlcA in stick representation coloured yellow for carbons, red for oxygen, blue for nitrogen and orange for phosphorus. Secondary structure elements classified numerically.

The intermediate α sub-domain (α9 to α12) is located around the homodimer two-fold non-crystallographic symmetry rotation axis. It is therefore implicated

in BceC homodimerisation and, as it also includes the catalytic centre (see below), it will be assigned here as the dimerisation and catalytic sub-domain (DCSD). It is composed of two pairs of helices connected by a 20 residue long loop, between α10 and α11, which surrounds the nucleotide-sugar substrate. The DCSD is initiated by a 30 residue α-helix (α9), a conserved helix that connects two Rossmann-like folds in NAD-dependent dehydrogenases (Kutzenko *et al*. 1998). In this context it is relevant to note that the fit of NBDs from UGDs and GDP-mannose dehydrogenase (GMD, Snook *et al*. 2003) also leads to the superposition of this long interconnecting helix (not shown), although the second Rossmann-like domain of GMD is swapped and overlaps the first Rossmann-like domain of the dimer mate in UGDs (Snook *et al*. 2003). The DCSD of UGDs/GMDs has been classified as belonging to the 6-phosphogluconate dehydrogenase C-terminal domain-like *super family* in SCOP 1.73 (Murzin *et al*. 1995; Andreeva *et al*. 2008), or *clan* in PFAM (Finn *et al*. 2008), being characterised as a multi-helical common core formed around two long anti-parallel helices related by (pseudo) twofold symmetry. Additionally, these two databanks define 12 or 6 families of domains homologous to the UGDs/GMDs DCSD, respectively. A search for the closest homologue domains in the PDB (Berman *et al*. 2000) using DALI (Holm *et al*. 2008) and the coordinates of BceC residues within the range 207-302, led to the structures of NDP-N-acetyl-D-galactosaminuronic acid dehydrogenase from *Methanosarcina* (*M.*) *mazei* (Malashkevich *et al*. 2009, unpublished, PDB entry 3G79), in β-hydroxyacid dehydrogenases (Papagrigoriou *et al*. 2006, unpublished, PDB entry 2GF2; Reitz *et al*. 2008, PDB entry 3CKY), and in cytokine-like nuclear factor N-pac protein (Tickle *et al*. 2007, unpublished, PDB entry 2UYY), for cut-offs in the DALI Z-score down to 3 and in rmsd up to 3 Å, for the 96 Cα atoms used. Whilst the NDP-N-acetyl-D-galactosaminuronic acid dehydrogenase is clearly homologous to UGD structures, with their two Rossmann-like dinucleotide binding domains linked

by the DCSD, the two β-hydroxyacid dehydrogenases and the cytokine-like nuclear factor N-pac protein structures form another set of similarly folded protein that contain only one Rossmann-like dinucleotide binding domain followed by an entirely α-domain, similar to UGD's DCSD. These structures diverge from UGDs mainly by lacking the second Rossmann-like fold, by having different paths in the DCSD intermediate 20 residue long loop and in the DCSD fourth helix, which is shortened to the first three turns. A packing analysis of these structures confirms that they all homodimerise *via* interactions between their DCSDs. It is noteworthy that the catalytic residues in all the mentioned enzymes are provided by their DCSD.

The C-terminal αβα domain (α13 to β14) has the internal six stranded pleated β-sheet surrounded by eight helices. In contrast to a low 12% sequence identity it shows a rather high 3D structural similarity against the NBD domain, with a rmsd of 1.8 Å between 96 pairs of Cα atoms (out of 154 residues) upon fit of the two domains using a superposition radius cut-off of 1.75 Å. Such a strong homology between the two Rossmann-like domains has already been noted in *S. pyogenes* UGD (Campbell *et al*. 2000), as well as in general to NAD dependent dehydrogenases (Kutzenko *et al*. 1998) or to other related enzymes (Campbell *et al*. 2000). The observation of proteins containing remarkable chain-fold similarities without apparent sequence conservation has been usually attributed to a gene duplication event at an ancestor form of the protein, which upon evolution retained the functional tertiary structure but lost sequence conservation (Lang *et al*. 2000).

3.4.6. Quaternary structure of BceC

Oligomeric structures of nucleotide sugar dehydrogenases may vary, *e.g.* the bovine and human enzymes occur as hexamers (Franzen *et al*. 1980) while BceC displays a dimeric form (Figure 3.5. a, b), as deduced by the *ca.* 100 kDa

single particles found in solution according to dynamic light scattering (DLS) or gel filtration chromatography assays (see *section 2*.). They correspond to approximately two times the expected molecular mass of the 470 residue protein monomer and thus show that the dimer is the functional unit of BceC, similarly to *E. coli* (Sieberth *et al*. 1995) and to *P. aeruginosa* (Hung *et al*. 2007) UGDs. The analysis of the crystal packing corroborates such a dimeric arrangement, with the four molecules in the asymmetric unit associated in two dimers. The dimer formation gives an occlusion of about $2.6x10^3$ $Å^2$ that corresponds to 13% of otherwise solvent accessible surface. Each of the remaining crystal packing interacting regions does not exceed 3.5% of the monomers solvent accessible areas.

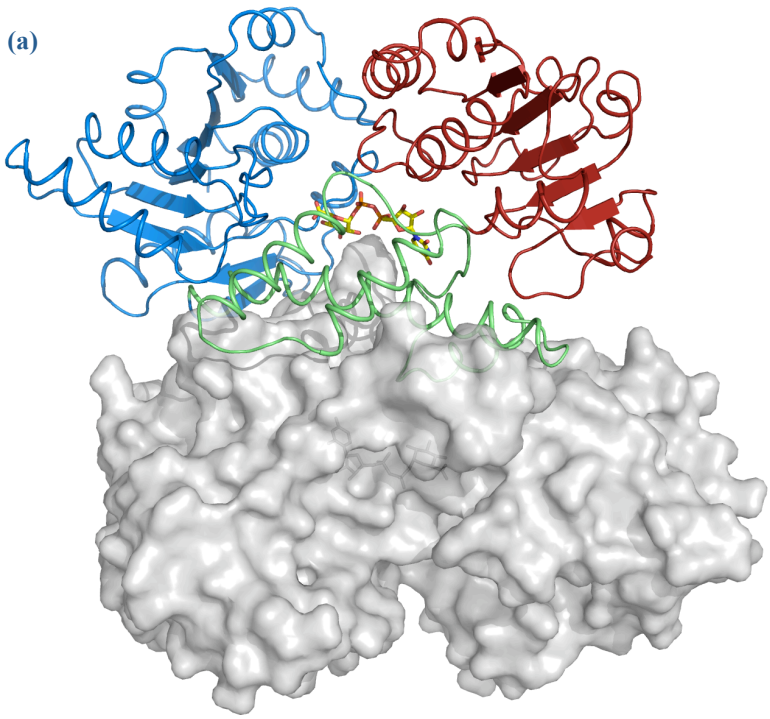


Figure 3.5. *The BceC dimer structure*.

(a) "Front" view of the BceC dimer. One monomer is displayed as cartoon representation with a docked UDP-GlcA molecule in stick representation (both coloured as in Figure 3.4.), and the second monomer is represented by its solvent accessible surface (grey).

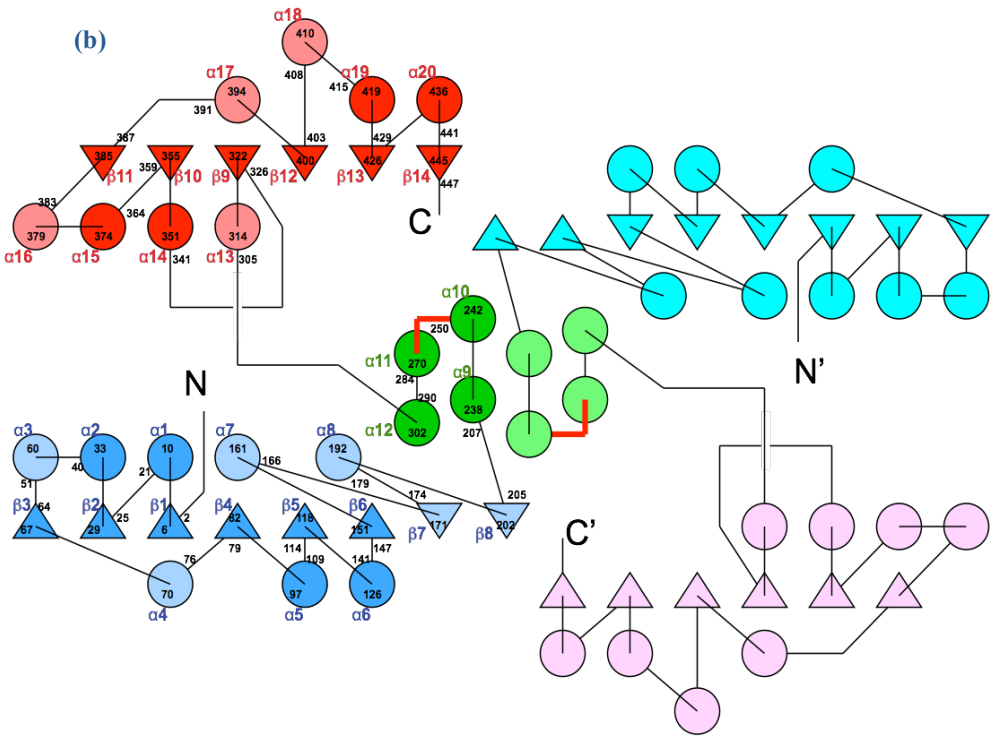


Figure 3.5. (cont.) *The BceC dimer structure.*

(b) Topology of the BceC dimer, with triangles representing β-strands and circles representing α-helices. In the monomer on the left, dark blue and dark red represent the features common to both the dinucleotide binding Rossmann-like domains, in contrast to the different, light blue and light red, secondary structure elements. The dimerisation sub-domain is shown in green. The monomer on the right is colour coded in cyan, pink and light green for the N-terminal, C-terminal and intermediate subdomain, respectively. Secondary structure elements were labelled numerically both for α-helical or β-strand elements. Residues delimiting BceC secondary structure elements are numbered black. Catalytic loop represented by the red line. Image prepared with Tops (Westhead *et al*. 1999).

As described above, dimerisation involves several inter-helical interactions between the DCSDs of the dimer partners, with particular relevance to an anti-parallel coupling of the two mates long central helices α9. The dimerisation interface involves sixty-five residues (of the one hundred residues that compose each DCSD) and involves *ca.* 20 hydrogen bond interactions.

3.4.7. The cofactor Binding Pocket

The three dimensional structures presented here, were obtained only upon addition of both the reaction product UDP-GlcA and the cofactor NAD^+, to the protein solution, even if no cofactor molecule could be successfully positioned and refined in any of the structures. In an attempt to compare the BceC cofactor-binding pocket with those of the human and *S. pyogenes* UGDs, a superposition of their 3D models was performed (Figure 3.6.).

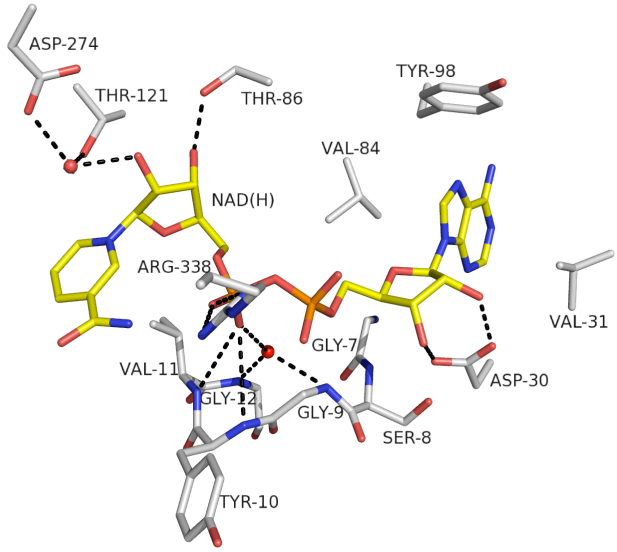


Figure 3.6. *BceC cofactor binding pocket with a docked NAD^+/NAD(H) model.* Superposition of BceC against *S. pyogenes* and human UGD-NAD(H) complexes led to a model of NAD(H) docked in BceC (atoms as sticks colored blue for nitrogen, red for oxygen, orange for phosphorus and white or yellow for carbon atoms in BceC or NAD(H) model, respectively). The black dashed lines illustrate hydrogen bond network between BceC and its cofactor model, defining a stretch of residues that in the UDP-glucose/GDP-mannose family is characterised by the GXGXXG sequence (Bellamacina 1996; Snook *et al.* 2003).

In BceC, the cofactor lies in a solvent accessible cleft near the carboxy ends of strands β1, β4 and β5 and is shown in an extended conformation, similarly to

several NAD^+-enzyme complexes (Bellamacina 1996; Snook *et al*. 2003), with most of its atoms exposed to the solvent. This is in agreement with a "Bi-Uni-Uni-Bi ping pong" mechanism (Ordman & Kirkwood, 1977; Ge *et al*. 2004), where the UDP-Glc binds first and the UDP-GlcA is released last, with the cofactor molecules being readily reduced and released to the medium, without causing major structural changes in the protein domains. Several conserved residues in the UDP-glucose/GDP-mannose family are involved in cofactor-protein interactions. At the beginning of the Rossmann fold a stretch of 30-35 amino acid residues shows a high degree of similarity, reminiscent of different types of dehydrogenases such as alcohol or aldehyde dehydrogenases, being denominated as the *fingerprint* region (Bellamancina 1996). The side-chain of Asp30, a conserved residue in this region of the dinucleotide-binding domain, establishes hydrogen bonds with the adenine ribose O2 and O3 hydroxyls. Also located in the fingerprint region is the glycine-rich consensus sequence, GXGXXG, also known as the phosphate-binding loop, where main-chain amides hydrogen bond, either directly or via bridging water molecules, to the pyrophosphate moiety. Interestingly, conservation in this UGD/GMD family of proteins seems to go beyond the three glycine residues, as the fourth position in this motif is strictly conserved (Tyr10 in BceC, see below). Another important residue interacting with the pyrophosphate moiety is the strictly conserved Arg338, located in a Ω-loop formed by residues 328-340 that includes 5 internal hydrogen bonds, which was pointed by Campbell as responsible for sequestering the reaction intermediates (Campbell *et al*. 2000). Ω-loops are non-regular secondary structures that account for about one quarter of all residues in globular proteins (Leszczymski & Rose, 1986; Fetrow 1995). They are characterised by a polypeptide chain that follows a 3D loop-shaped course (Leszczymski & Rose, 1986), and although they do not contain repeating backbone dihedral angles or regular patterns of hydrogen bonds, they have been shown to play significant roles in protein function and stability (Fetrow 1995).

3.4.8. The substrate/product binding pocket

In the substrate-binding pocket, where the reaction product was found, six residues are within van der Waals distances to UDP-GlcA, which establishes hydrogen bonds with further eleven residues and ten solvent water molecules. The UDP moiety is flanked by a coiled stretch comprising residues 259-267. The uracyl and Phe259 rings show a π-edge stacking similar to that observed on Phe265 and Tyr249, in the homologous proteins from *H. sapiens* ((Kavanagh *et al*. 2007a), unpublished, PDB entry 2QE3) and *S. pyogenes* (Campbell *et al*. 2000) UGDs, respectively. A list of BceC hydrogen bond interactions in the substrate binding pocket is presented in Table 3.3., together with those found in the UGD structures of *H. sapiens*, and in the thoroughly described *S. pyogenes* structure (Campbell *et al*. 2000). Almost all N or O atoms of the substrate are engaged in hydrogen bonding interactions with the enzyme. Main-chain interactions are prevalent with uracyl or ribose moieties, while the interactions with glucose, usually in multiple hydrogen bonds, occur mainly with residue side-chains (Figure 3.7.). Also noteworthy is the higher residue conservation in the vicinity of the active site, when compared with the remaining substrate binding pocket. Regarding the enzyme interactions with the glucuronate moiety one finds Arg254 from the dimer mate extending its side-chain into the partner's active site, in particular with its guanidinium group forming hydrogen bonds with the O2' and O3' hydroxyls of the sugar. Additionally, the Lys214 ε-amino group forms a hydrogen bond with O4' and, together with Asn218 side-chain amide, also hydrogen bonds to O61 of the carboxylic group, which O62 is in turn involved in hydrogen bonds with the sulfydryl of catalytic Cys270 and further with the hydroxyl of Tyr10 and the carboxylic of Glu154.

Table 3.3. *Hydrogen bonding interactions within the UGDs substrate's binding pocket.* BceC *(this structure), H. sapiens* (Kavanagh *et al*. 2007, not published, PDB entry 2Q3E) and *S. pyogenes* (Campbell *et al*. 2000, PDB entry 1DLJ) UGDs are represented. The substrate is divided into uracyl, ribose, diphosphate and glucose/glucuronic acid sections (first column). Equivalent atom labels of reaction substrate/product (second column) appear in the residue's rows (columns 3-5) to which they form hydrogen bonds (defined up to 3.2 Å initial cut-off distance, but homologous interactions up to 3.4 Å were included). Conserved residues are highlighted in bold; residues bridged to the reaction substrate/product *via* a solvent water molecule are presented in italic within parenthesis. Hydrogen-bonding distance values (Å) are within parentheses. Contacts with the active cysteine (or its serine mutant) are marked with *.

		BceC	***H. sapiens* UGD**	***S. pyogenes* UGD**
Uracyl	O4, O4, O4'	Tyr261_N (3.0)	Lys267_N (3.2)	Asn251_N (2.9) (*R401_NE*)
	N3	Tyr261_O (3.0)	Lys267_O (2.8)	Asn251_O (2.9)
	O2	*(Asn229_ND2)* Arg431_NH1 (3.2)	Ser269_OG (2.8)	Ser253_OG (2.8)
Ribose	O2*, O2C, O2D	(*Glu406_OE1, 2*)	Arg442_NH1 (2.9) Arg442_NH2 (3.1) (*Glu416_OE2*)	Asp402_OXT (2.7)
	O3*, O3C, O3D	**Gly**267_N (3.1), Phe330_O (3.0)	**Gly**273_N (2.9), Phe338_O (2.7)	**Gly**257_N (2.9), Met319_O (2.8)
Diphosphate	O1A	(***Lys****331_NZ*)	**Lys**339_NZ (2.7) (*Arg260_NH1*)	**Lys**320_NZ (3.2) (*Asp402_OD1*)
	O2A			Tyr249_OH (2.7)
	O3A	**Lys**331_NZ (3.3)	**Lys**339_NZ (3.2)	**Lys**320_NZ (3.4)
	O1B	(***Glu158****_OE1*) (***Gly267****_O*) (***Cys****270_N*) (*Phe_271_N*)	(***Gly****273_O*)	(***Glu****145_OE1*) (***Gly****275_O*) (****Ser****260_N*) (*Leu261_N*)
	O2B	**Glu**158_N (3.3) **Lys**331_NZ (2.8)	**Glu**165_N (2.9)	**Glu**145_N (2.9), **Lys**320_NZ (2.8)
Glucose	O2'	**Arg**254_NH1 (2.9)	**Arg**260_NH1 (2.8)	**Arg**244_NH1 (3.1)
	O3'	(***Phe****155_O*) **Arg**254_NH2 (2.9) (***Lys214****_NZ*) (***Asn****218_OD1*)	**Phe**162_O (2.7), **Arg**260_NH2 (3.0) (***Asn****224_OD1*)	**Phe**142_O (2.70) **Arg**244_NH2 (3.2) (***Lys****204_NZ*) (***Asn****208_OD1*)
	O4'	**Leu**156_O (2.7) **Lys**214_NZ (3.0)	**Leu**163_O(2.5) **Lys**220_NZ (3.1)	**Leu**143_O(2.6) **Lys**204_NZ, (3.0)
	O61, O6', O'P	**Lys**214_NZ (2.8), **Lys**218_ND2 (3.0) (***Thr****121_OG1*) (***Asp****274_OD2*) ***Cys**270_SG (3.4)	**Lys**220_NZ (2.9), **Lys**224_ND2 (2.9) (***Thr****131_OG1*) (***Asp****280_OD2*) ***Cys**276_SG (3.6)	**Lys**204_NZ (2.8) **Lys**208_ND2 (2.9) (***Thr****118_OG1*) (***Asp****264_OD2*) ***Ser**260_OG (3.4)
	O62, O'Q	**Tyr**10_OH (2.5) **Glu**154_OE1 (2.6) ***Cys**270_SG (3.2)	N/A	***Ser**260_OG (2.3)

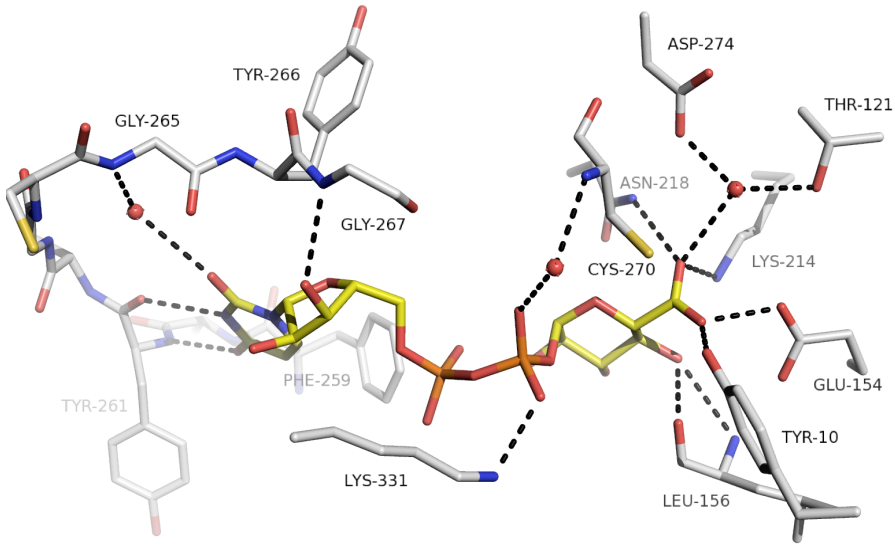


Figure 3.7. *The substrate/product binding pocket of BceC.*

Several BceC residues interact via main-chain or water mediated bonds with the UDP-GlcA (by analogy the substrate UDP-Glc). However, whilst the sugar crevice shows a myriad of strictly conserved residues and hydrogen bonds, including the catalytic amino acids, the pyrophosphate and uridine binding pockets show many main-chain or water mediated interactions, hence a lower sequence conservation when compared for example, with the co-factor NAD^+ binding-pocket. Residues and UDP-GlcA in stick representation coloured as in Figure 3.6. Distances of the hydrogen bonds were removed for clarity. For further details see Table 3.3. above.

These two last interactions are not observed either in *S. pyogenes* or in human UGDs. However, these structures were obtained as ternary complexes including both reactants, while in the present study no NAD^+ could be found. Superposition of the three structures shows that the actual side-chain conformations of Tyr10 and Glu154 in BceC would produce severe clashes with the nicotinamide carboxamide group of NAD^+, if docked into the enzyme. In contrast to the cofactor cleft, the substrate-binding pocket is buried deeper in

the protein, requiring protein structural changes for access and release of reactant and product. The side-chain of Lys331, a conserved residue belonging to the Ω-loop, hydrogen bonds with O3A atom of the pyrophosphate moiety. As described above, the Ω-loop includes another strictly conserved residue, Arg338, which by analogy with the *H. sapiens* or *S. pyogenes* structures should interact with the pyrophosphate from NAD^+. The two salt bridges with both reactants are thought to ensure the loop's "closed" conformation, whereby substrate intermediates are locked in place during the different reaction steps. The O1B atom of the pyrophosphate is hydrogen bonded to a solvent water molecule that in turn forms several hydrogen bonds in the immediate vicinity of the catalytic site, namely with main-chain amide from Gly268, Cys270 and Phe271, and thus helps to set the catalytic positioning of substrate and Cys270.
One aim of this study was to accomplish a detailed comparison of human and BceC structures in order to explore potential medicinal applications. However the substrate binding pockets of both proteins revealed to be too similar for the design of a potentially selective inhibitor.

3.4.9. The UDP/GMD-6-dehydrogenase catalytic mechanism

The alcohol to carboxylic acid transformation catalysed by the UGD/GMD family involves a four electrons oxidation of the pyranose C6 atom, and corresponds to a global replacement of methylene by an acyl group. The reaction includes two successive steps of hydride transfer to NAD^+ and a final hydrolysis of a thioester intermediate, which allows the incorporation of a second oxygen atom into C6. The elucidation of the reaction mechanism (see Figure 3.11.) has been successively fine-tuned since the early scheme proposed by Campbell and co-workers for *S. pyogenes* UGD (Campbell *et al*. 1997; Ge *et al*. 1998; Campbell & Tanner 1999; Easley *et al*. 2007). By analogy with the mechanism described by Easley and collaborators, the reaction in BceC is

initiated by the activation of a structural water molecule (Ge *et al.* 2004; Easley *et al.* 2007). The water is deprotonated by the carboxylate group of Asp274, to which it is hydrogen bonded. The general base capacity of Asp274 is enhanced by the close proximity of the ε-amino group of Lys273 (at hydrogen bonding distance) (Sommer *et al.* 2004). As both Asp274 carboxylate and Lys273 ε-amino groups are solvent accessible, they may function as an exit route for the two protons that will be extracted from UDP-Glc during the overall reaction. The generated hydroxide is kept in place by a hydrogen bond with the side-chain of Thr121 (Ge *et al.* 2004), whose positioning is assured by hydrogen bonds with Ser120 and Pro123 residues. Upon deprotonation of the UDP-Glc hydroxyl at O6 position by the hydroxide, the intermediate oxyanion is stabilised by electrostatic interactions with the nearby ε-amino group of Lys214 at hydrogen bonding distance (Campbell *et al.* 2000), until a first hydride at C6 is transferred to NAD^+. The resulting aldehyde intermediate (Campbell *et al.* 1997) remains sequestered by the enzyme (Nelsestuen & Kirkwood 1971; Ge *et al*. 2004). With the increase of the C6 oxidation state it becomes susceptible to nucleophilic attack, due to the electron withdrawing effect of the now double bonded oxygen. The catalytic water/hydroxide is also neighboring the sulfydryl group of Cys270 and thus may also extract its proton. The formed nucleophilic thiolate readily attacks the acyl C6 leading to the formation of a covalently bonded thiohemiacetal intermediate (Ge *et al.* 2004; Sommer *et al.* 2004; Chaikuad *et al.* 2009, not published). The resulting oxyanion is stabilised once more by electrostatic interactions with the nearby ε-amino group of Lys214 (Campbell *et al.* 2000) until transfer of the second hydride at C6 to a freshly docked NAD^+, and the global oxidation of C6 is accomplished with the thioester formation. The final hydrolysis step involves a second nucleophilic attack to C6, most likely by the catalytic water/hydroxyl, which will disrupt the sulfur-carbon bond and generate the final carboxylic group at C6. This is the rate-limiting step of the reaction (Ordman & Kirkwood 1997) and requires the

stabilisation of the forming thiolate. However, as the ε-amino group of Lys214 is located at the opposite side of C6, it is too far away to be invoked for catalysis, and the relevant catalytic agent has until now remained elusive.

3.4.10. UGD/GMDs final hydrolysis step is catalysed by the conserved Y of their GXGYXG motif

Sequence alignment within the UDP-Glucose and GDP-Mannose dehydrogenase family (Granja *et al*. 2007) shows the conservation of a tyrosine residue (Figure 3.8.) at the second "X" position of the Rossmann fold glycine rich motif (Bellamacina, 1996).

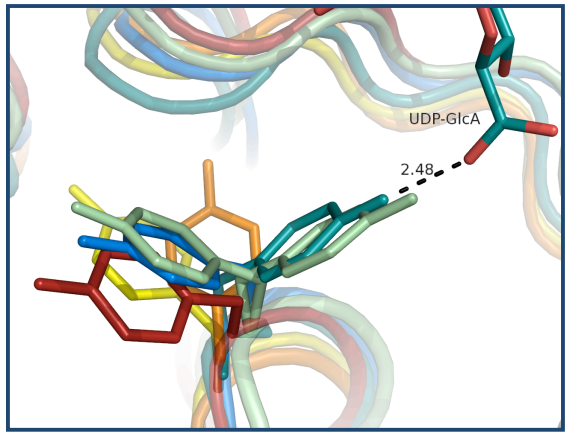


Figure 3.8. *The conserved tyrosine residue at the Rossmann fold phosphate-binding motif.* A 3D superposition of UGD structures shows that the side-chain of Tyr10 (BceC residue numbering) populates a variety of Chi1 torsional conformations. Noteworthy, only for the BceC and *P. gingivalis* UGD structures does Tyr10 assume a *gauche (+)* conformation. This conformation is not possible for the other structures, because they all contain a ligand in the cofactor-binding pocket, which would imply severe clashes with such a Tyr conformation. UGDs colouring scheme as follows (PDB entries within parentheses): BceC in dark green; *S. elodea* UgdG in orange (Rocha *et al*. 2009a); *S. pyogenes* in blue (1DLI); *C. elegans* UGD in yellow (2O3J); *H. sapiens* UGD in red (2QE3); and *P. gingivalis* UGD (3GG2) in light green.

In the native BceC structure the hydroxyl group of the conserved tyrosine residue establishes a hydrogen bond with the carboxylic group of UDP-GlcA similarly to that found in *P.gingivalis* UGD (PDB code 3GG2; Bonnano *et al*. 2009, not published). We proceeded to carry out mutagenesis studies on BceC whereby the Tyr10 residue was mutated into a phenylalanine, a serine or a lysine and followed their activity under saturating concentrations of substrate and cofactor at pH 8.7 (see *section 3.3.3.*). The mutants show severely impaired activity (Figure 3.9.) although a latent activity was nevertheless detected. Interestingly, while studying *B. subtilis* UGD, Mijakovic and co-workers have also pointed out Y10 as critically important, as the Y10F mutation on this UGD led to an accentuated decrease in the enzyme activity (Mijakovic *et al*. 2004).

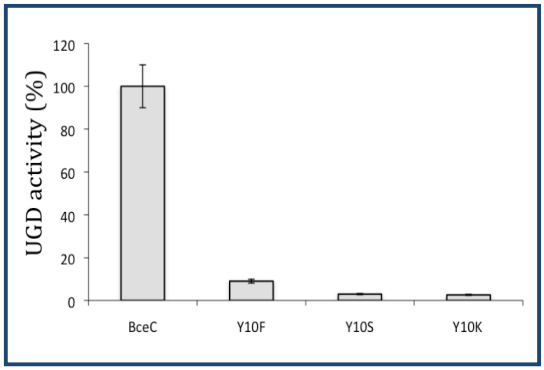


Figure 3.9. *Graphical representation of the native and Y10 BceC mutant's activity*.
The UDP-glucose dehydrogenase activity was measured as the conversion of NAD^+ to NADH, in the presence of saturating conditions of substrate and cofactor and plotted as the mean ± standard deviation of three independent determinations.

Comparison of BceC with the Y10S/K mutant three-dimensional structures shows no differences apart from the referred mutation (Figure 3.10.). In BceC, the hydroxyl group of Tyr10 is involved in hydrogen bonds with the carboxylate of UDP-GlcA, two water molecules and the thiol group of Cys270.

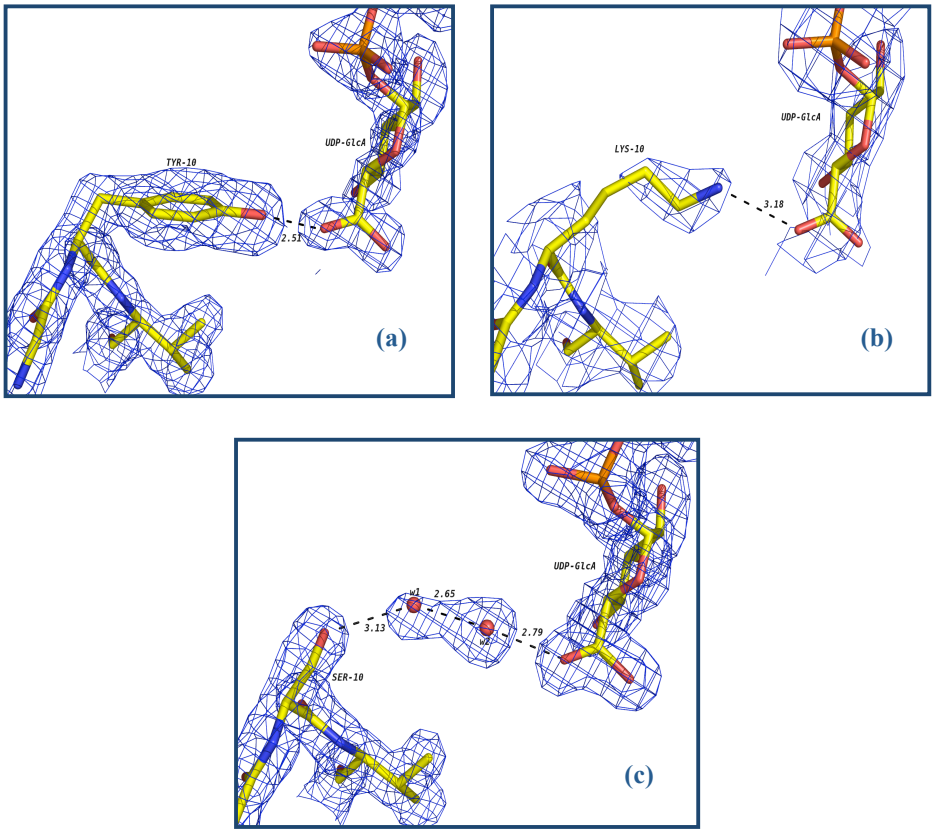


Figure 3.10. *Structures of native BceC and its mutants Y10K and Y10S.*
(a) In native BceC Tyr10 hydroxyl is involved in a hydrogen bond with the UDP-GlcA carboxylate; (b) In the Y10K mutant a salt bridge is formed between the Lys10 ε-amino group and UDP-GlcA; (c) In the Y10S mutant two solvent water molecules are involved in a hydrogen bonding network between Ser10 Oγ and UDP-GlcA carboxylate. Residues are in stick representation coloured yellow for carbon, blue for nitrogen, red for oxygen and orange for phosphorus atoms. Electron density 2mFo – DFc maps represented in blue mesh contoured at 1.0σ. Hydrogen bonds depicted in dashed black lines with corresponding distances in Å.

The pratical loss of activity upon Tyr10 mutations implies that Tyr10 is involved in the final hydrolysis step of the enzymatic mechanism, as it is well positioned to function as stabiliser of the thiolate leaving group, after the nucleophilic attack by the catalytic hydroxide to acyl C6 of the thioester intermediate.

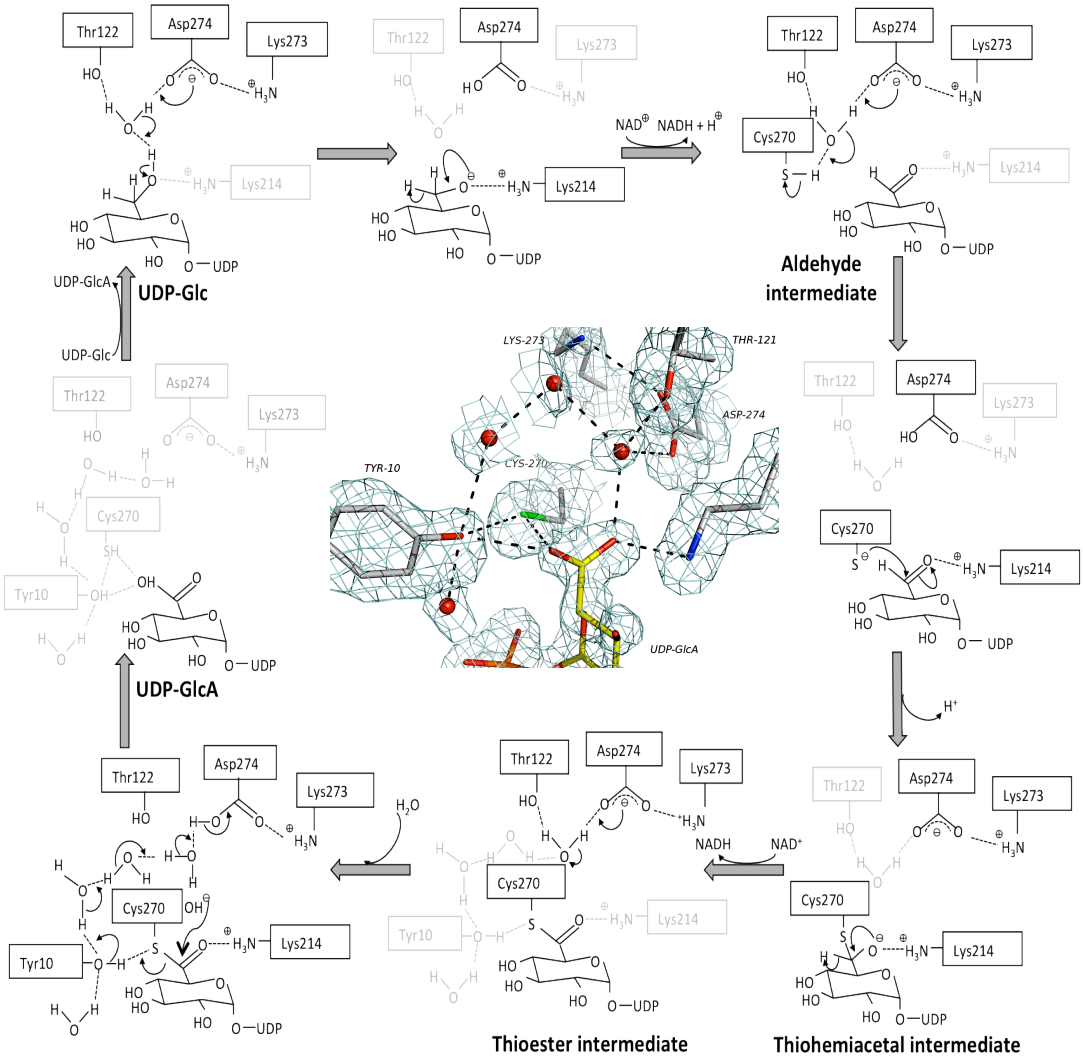


Figure 3.11. *Mechanism of BceC enzymatic conversion of UDP-Glc to UDP-GlcA*.

By analogy with the mechanistic characterisation in *S. pyogenes* and human UGDs, successive deprotonations of the catalytic water initiate the processes leading to two hydride transfers from methylene C6 of UDP-Glc, generating aldehyde and thioester intermediates. That catalytic water is also proprerly positioned to perform a nucleophilic attack to the thioester. Tyr10, located at the opposite side relative to the acyl group, functions as a proton conveyer from the aqueous media to the forming thiolate, thus catalysing the hydrolysis of the thioester with formation of UDP-glucuronic acid (insert). The active and passive residues in each step of the catalytic mechanism are highlighted in black and grey, respectively.

Tyr10 may therefore work as a proton conveyer from the aqueous hydrogen bonding proton-wire system to the forming thiolate (see Figure 3.11. above). All three mutants lack this capacity at residue 10. In Y10F there is no hydroxyl group. In Y10S an Oγ hydroxyl exists, but the side-chain is too short for an effective proton transfer to thiolate. In Y10K at biological pH the ε-amino group is in the protonated ammonium form, and therefore there is no electron doublet free to convey protons from the aqueous media to the forming thiolate. In spite of their similar pKa's (above 10) tyrosine and lysine show contrasting catalytic capacities, because the conveyance of aqueous protons by tyrosine is not dependent upon its general acid-base properties – due to the two hydrogen bonded water molecules a proton always remains attached to the Tyr10 hydroxyl, throughout the catalytic process.

3.5. Acknowledgements

The authors would like to thank the staff at ESRF in Grenoble, for the use of synchrotron radiation. Joana Rocha (JR) acknowledges the PhD fellowship SFRH/BD/24216/2005, from Fundação para a Ciência e Tecnologia (FCT). Alma Popescu (AP) at Instituto Superior Técnico (IST) did the cloning of native BceC and y10f_bceC genes. JR, following AP experience and protocols, performed the cloning of the y10k_bceC and y10s_bceC mutants, and the three proteins expression and purification, as well as the crystallisation, data collection, phase determination, structural solution and analysis.
This work was funded by FCT grants POCTI/BME/38859/2001, POCI/BIO/58401/2004 and PTDC/QI/67925/23006.

UDP-glucose dehydrogenase from *Sphingomonas elodea* ATCC 31461 (UgdG)

4. Cloning, expression, purification, crystallisation and preliminary crystallographic studies of UgdG, an UDP-glucose dehydrogenase from *Sphingomonas elodea* ATCC 31461

The work described here is based on the published refereed paper but now including a more detailed description of the materials and methods, as well as new figures. Reference: Rocha, J.; Granja, A. T.; Sá-Correia, I. Fialho, A. and Frazão, C. (2010a) "Cloning, expression, purification, crystallisation and preliminary crystallographic studies of UgdG, an UDP-glucose dehydrogenase from *Sphingomonas elodea* ATCC 31461". *Acta Crystallographica* F66, 69-72.

4.1. Abstract

Gellan gum, a commercial gelling agent produced by *Sphingomonas elodea* ATCC 31461, is a high-value exopolysaccharide (EPS). UDP-glucose dehydrogenase (UGD; EC 1.1.1.22) is responsible for the NAD dependent twofold oxidation of UDP-glucose into UDP-glucuronic acid, one of the key components for gellan biosynthesis.

S. elodea ATCC 31461 UGD, termed UgdG, was cloned, expressed, purified and crystallised in both native and SeMet derivatised forms, in hexagonal and tetragonal space groups, and diffracting X-rays up to 2.40 and 3.40 Å resolution, respectively. Experimental phases were obtained for the tetragonal SeMet derivatised crystal form from a single anomalous dispersion (SAD) experiment. This structure was successfully used as a molecular replacement probe for the hexagonal crystal form of the native protein.

4.2. Introduction

Bacterial exopolysaccharides (EPS) are biotechnology products of high interest due to their rheological properties. This is the case of gellan gum; a multifunctional gelling agent produced in high yields by the non-pathogenic bacteria strain *Sphingomonas elodea* ATCC 31461. It has approval in the US and EU for food use as a gelling, stabilizing and suspending agent, either alone or in combination with other hydrocolloids. In its native form, gellan is a linear high molecular weight anionic EPS based on a tetrasaccharide repeat unit composed of two molecules of D-glucose, one of L-rhamnose and one of D-glucuronic acid. The native gellan is partially esterified with acyl substituents (1 mole of glycerate and 0.5 mol of acetate) per repeat unit (Jay *et al*. 1998). Gellan and gellan-like polymers have unique characteristics and find many applications, particularly in the food, pharmaceutical and biomedical fields (Sá-Correia *et al*. 2002; Fialho *et al*. 2008). Due to the broadness of its commercial applications, the biosynthetic pathway leading to the production of gellan has gained a significant amount of interest (Aragão *et al*. 2007, Sá-Correia *et al*. 2002). Gellan gum biosynthesis is a multi-step process that starts with the production of nucleotide-sugar precursors uridine-5'-diphosphoglucose (UDP-glucose or UDP-Glc), UDP-glucuronic acid (UDP-GlcA) and deoxythymidine-diphospho-L-rhamnose (dTDP-L-Rha), followed by formation of the tetrasaccahride repeating unit, gellan polymerisation and export (Moreira *et al*. 2004; Fialho *et* al. 2008). Three gene clusters necessary for gellan production have been described (Harding *et al*. 2004) but interestingly, genes encoding the enzymes for the UDP-Glc and UDP-GlcA synthesis (*pgmG*, *ugpG* and *ugdG*) are located at different positions of the *S. elodea* genome, perhaps indicating a broader function of these enzymes (Harding *et* al. 2004, Silva *et* al. 2005, Granja *et al*. 2007; Aragão *et al*. 2007). The identification of key genes like

those mentioned above, and the elucidation of the crucial steps in the gellan pathway indicate that possibilities now exist for exerting control in gellan production by acting, for example, on the expression levels/activity of these genes/enzymes. Our interest in the *S. elodea* ATCC 31461 UDP-glucose dehydrogenase, termed UgdG (EC 1.1.1.22), derives from its crucial role in providing UDP-GlcA, one of the precursors for gellan biosynthesis. In this work we report the crystallisation and phase problem determination of UgdG from *S. elodea* ATCC 31461.

4.3. Materials and Methods

4.3.1. Cloning and Expression of native and SeMet derivatised UgdG

The complete sequence of the *ugdG* gene from *S. elodea* ATCC 31461 was amplified by polymerase chain reaction (PCR; Granja *et al.* 2007) using the cosmid pCG8 as the template DNA, *Pwo* DNA polymerase (Roche Diagnostics, Mannheim, Germany), and the oligonucleotides UGD1 (5'-AAAGGATCCGTGCGCATTGCGATG-3') and UGD2 (5'-AAAAAGCTTTCAGTCGCGGCTCG-3'), which were designed based on the *ugdG* nucleotide sequence. The PCR product (1,320 bp) was digested with *Bam*HI and *Hind*III (recognition sites underlined in the primers used) and cloned into the vector pWH844 (Schirmer *et al.*, 1997), generating pUgdG. This plasmid carries the *ugdG* gene preceded by a sequence coding for six histidine residues present for purification purposes. The DNA insert cloned in pUgdG was sequenced to confirm the fidelity of DNA amplification. Over-expression of the *ugdG* gene was carried out by cultivation of *Escherichia coli* BL21 transformants harbouring the plasmid pUgdG in 600ml LB medium

supplemented with 100 μg·ml^{-1} of ampicillin at 310 K, until an OD_{600nm} of 0.6 was reached. Expression was then induced with 0.3m*M* isopropyl *β*-D-thiogalactoside (IPTG) for 3 h, harvested by centrifugation (10 000 *g*, 30 min, 277 K) and the obtained pellets stored at 253 K. For the production of the SeMet-UgdG derivatised protein (Molecular Dimensions protocol), methionine-auxotrophic strain *E. coli* B834 (DE3) carrying pUgdG was cultivated in 600 ml of LB medium (supplemented with 100 μg·ml^{-1} of ampicillin) at 310 K, until an OD_{600nm} of 0.6 was reached. Cells were harvested by centrifugation (10 000 *g*, 10 min, 277 K) and gently resuspended in SelenoMet Medium Base (Molecular Dimensions, UK). This process was repeated three times to remove any traces of LB medium containing methionine. Finally, cells were resuspended in SelenoMet medium and 250x concentrated seleno-methionine solution was added (Molecular Dimensions, UK). Induction occurred for 3 h after the addition of 0.3 m*M* IPTG, and cells were harvested by centrifugation (10 000 *g*, 30 min, 277 K). The supernatant was discarded and the pellet stored at 253 K.

4.3.2. Purification of native and SeMet derivatised UgdG

The following purification protocol was used for both the native and SeMet derivative UgdG proteins. Cells were resuspended in 20 m*M* phosphate pH 7.4, 20 m*M* imidazole, 1 *M* NaCl and disrupted in a French press.

Crude cell extract was obtained by centrifugation (27 000 *g* for 40 min at 277 K), and the supernatant applied onto a 5 ml HisTrap column (GE Healthcare) pre-equilibrated with 20 m*M* phosphate pH 7.4, 20 m*M* imidazole and 1 *M* NaCl, connected to an ÄKTA Explorer Instrument (GE Healthcare) according to the manufacturer's recommendation.

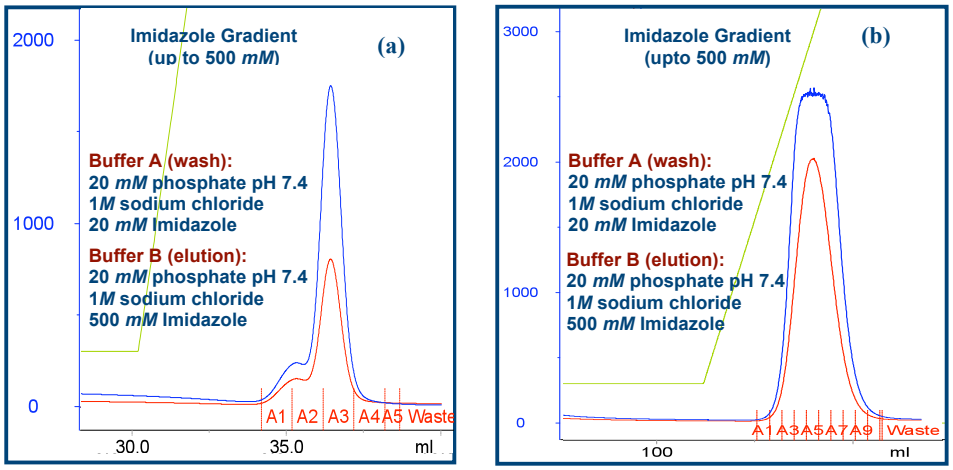


Figure 4.1. *The purification of the native and SeMet derivatised UgdG.*

(a) Affinity chromatogram for the purification of the His-tagged native UgdG protein. (b) Affinity chromatogram for the purification of the His-tagged SeMet derivatised UgdG. Protein detected by UV absorbance (mAU) at 280nm and 254 nm (y coordinates; blue and red lines). UgdG elutes almost pure in a symmetrical peak at approximately 250 mM of imidazole concentration (green line). The experimental conditions were the same for both protein forms.

The column was washed with buffer *A* (20 m*M* phosphate pH 7.4, 20 m*M* imidazole, 1 *M* NaCl) to remove any unbound protein, and a gradient of imidazole 20-500 m*M* was applied. His_6-UgdG recombinant protein was eluted at approximately 250 m*M* imidazole in a symmetrical chromatography peak (see above Figure 4.1.a). Imidazole was removed from the eluted fractions of the UgdG protein, replacing the buffer with 25m*M* Tris-HCl pH 8.3, 50 m*M* NaCl, 2.5 m*M* dithiothreitol (DTT) supplemented with 1 m*M* of oxidised nicotinamide adenine dinucleotide (NAD^+), or 0.5 m*M* UDP-GlcA and 1 m*M* NAD^+, using a PD10 desalting column (GE Healthcare). In the case of SeMet-UgdG derivatised protein, the eluted fractions were pooled together and buffer-exchanged with 25 m*M* Tris-HCl pH 8.3, 50 m*M* NaCl, 2.5 m*M* DTT and 1 m*M* NAD^+, using a PD10 desalting column (GE Healthcare; see Figure 4.1. b).

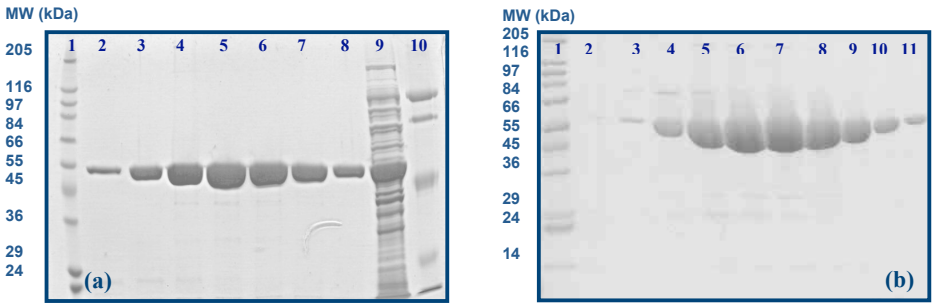


Figure 4.2. *SDS-Page analysis for the purifications of native and SeMet derivatised UgdG.* (a) SDS-Page gel for the purification of the His-tagged native UgdG using Coomassie blue staining. The protein migrates as a single polypeptide with an estimated molecular mass of approximately 48.5 kDa (47.2 kDa of native protein plus 1.3 kDa of His6-tag). *Legend:* 1. Wide range Sigma Marker; 2-8. Native UgdG eluted fractions; 9. Soluble fraction; 10: Low range Sigma Marker. (b) SDS-Page gel for the purification of the His-tagged SeMet derivatised UgdG using Coomassie blue staining. The protein migrates to a slightly higher molecular weight than the native UgdG due to the Selenium atoms. *Legend:* 1. Wide range Sigma Marker; 2-11. SeMet derivatised UgdG eluted fractions. Fractions of both UgdG forms were pooled, buffer exchanged and concentrated up 10 $mg \cdot ml^{-1}$ and 5.5 $mg \cdot ml^{-1}$ for native and derivative, respectively.

Both UgdG native protein and SeMet-UgdG derivative protein were concentrated in Vivapore 10/20 concentrators (Vivascience Ltd, UK) to 4.5 $mg \cdot ml^{-1}$ and 5.5 $mg \cdot ml^{-1}$ respectively, before storage at 193 K. The composition and purity of both proteins were confirmed to be homogenous by SDS-polyacrylamide gel electrophoresis (see Figure 4.2. above). The derivatised protein migrates to a slightly higher molecular weight due to the presence of selenium methionines that substitute the sulphur methionines of native UgdG. Protein concentration was determined by Bradford's method (Bradford, 1976) and by measurement of absorbance at 280 nm in a NanoDrop ND-1000 Spectrophotometer using an extinction coefficient of 18490 $M^{-1} \cdot cm^{-1}$.

4.3.3. Crystallisation of native and SeMet derivatised UgdG

An initial crystallisation screen of native UgdG in the presence of NAD^+ was performed in batch mode on an Oryx6 crystallisation robot (Douglas Instruments), using 96 solutions from the Classic Screen from Qiagen Canada Inc. (Montreal, Canada). Drops consisting of 0.5 µl of protein solution at 4.5 mg·ml^{-1} in 25 m*M* Tris-HCl pH 8.3, 50 m*M* NaCl, 2.5 m*M* DTT and 1 m*M* NAD^+ plus 0.5 µl of precipitant solution were left equilibrating at 293 K under paraffin oil. Crystals formed within 24 hours for conditions H5 (200 m*M* Li_2SO_4, 100 m*M* Tris-HCl pH 8.5 and 30% (w/v) PEG 4K) and A7 (100 m*M* Na-citrate pH 5.6, 20% (v/v) isopropanol and 20% (w/v) PEG 4K). Crystals appearing in condition H5 had dimensions up to 0.3, 0.15, 0.01 mm (thin plates), but could not be reproduced. Other, mostly hexagonal shaped prismatic crystals appeared in condition A7, were reproducible and could be further optimised (Figure 4.3. a) adding 50 *mM* sodium fluoride as additive to the mother liquor. Cryoprotection of these crystals was achieved by passing crystals through a solution of their mother liquor supplemented with 25% (v/v) glycerol. Crystallisation screens on the SeMet derivatised protein were performed using a "Honeybee" Microsys 4000 XL Cartesian™ Dispensing Systems robot (Genomic Solutions, USA) using the vapour-diffusion method. Drops consisting of 100 nl of 5.5 mg·ml^{-1} protein solution in 25 m*M* Tris-HCl pH 8.3, 50 m*M* NaCl, 2.5 m*M* DTT, 1 m*M* NAD^+ and 100 nl of precipitant solution were equilibrated against 100 µl of 672 conditions of precipitant solutions from the crystallisation screens Classics, PEG, MbClass, MbClass II, pHClear, pHClear II and MPD screens from Qiagen Canada Inc. (Montreal, Canada). Crystalline precipitate was found in several conditions from different screens. From those, solution A7 from PEG screen (100 m*M* MES pH 6.5 and 40% (v/v) PEG 200) produced the most promising crystallites.

Upon optimisation, octahedral-like crystals appeared within 24-48 h, and grew to approximate dimensions of 0.05, 0.05, 0.08 mm (Figure 4.3. b) in drops also containing other precipitated material.

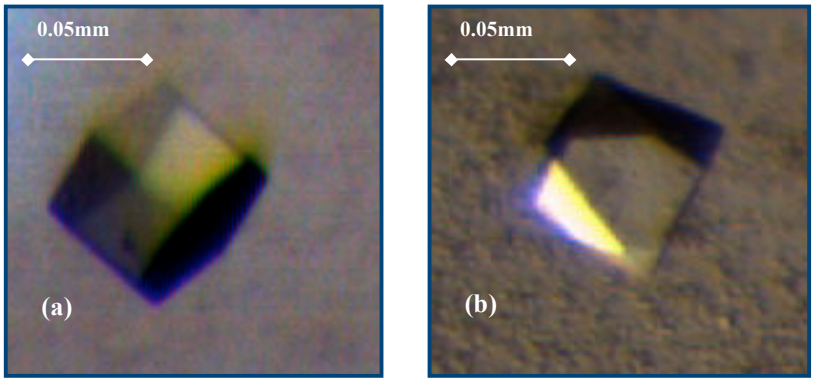


Figure 4.3. *Crystals of native and SeMet derivatised UgdG.*
(a) Native UgdG in hexagonal space-group (dimensions *ca*. 0.04, 0.04 and 0.06 mm). (b) SeMet-UgdG derivative crystal in tetragonal space-group (crystal dimensions *ca*. 0.05, 0.05 and 0.08 mm).

Crystals had to be "washed" with mother liquor to remove this precipitate prior to cryoprotection, which was achieved by quick soaking in mother liquor augmented with 20% (v/v) PEG 400. Crystals were then flash cooled and stored in liquid nitrogen for diffraction data collection.

4.3.4. X-ray diffraction analysis and phase problem solution

Cryoprotected crystals under a nitrogen stream at 100 K were measured using the European Synchrotron Radiation Facility (ESRF), (Grenoble, France), at stations ID23-1 (UgdG) and ID14-3 (SeMet-UgdG). The diffraction images from native protein crystals (Figure 4.4.a) were processed with *MOSFLM* (Leslie 1992) and the corresponding intensities were scaled and merged together using

SCALA (Evans 2006) from the CCP4 suite (The CCP4 suite: programs for protein crystallography, 1994).

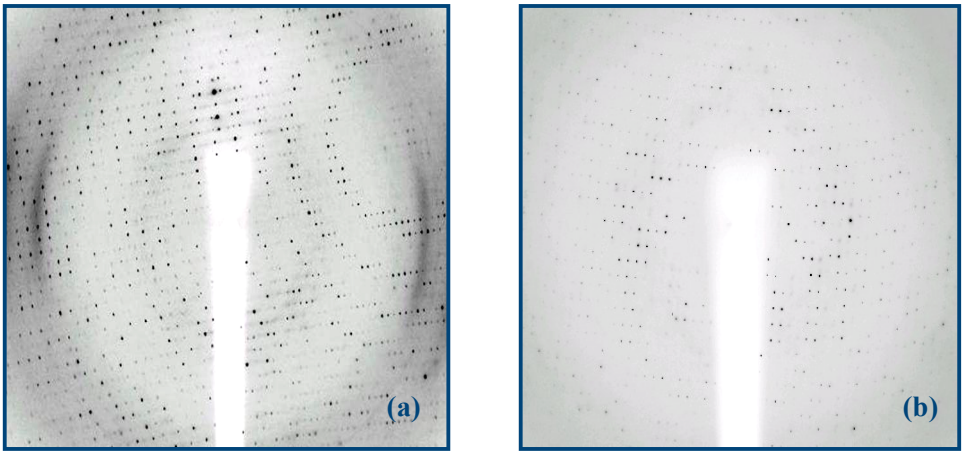


Figure 4.4. *X-ray diffraction images for Sphingomonas elodea UgdG.*
(a) Image for native UgdG; (b) Image for UgdG SeMet derivatised form. Data sets were collected at the ESRF beamlines ID23-1 and ID14-3 for native and SeMet derivatised UgdG crystals, respectively. UgdG native crystals diffracted to 2.4 Å resolution, whilst the Se derivatives diffracted only to 3.4 Å resolution.

The diffraction images of the SeMet derivatised protein (Figure 4.4.b) were integrated with *DENZO*, and intensities were scaled and merged together with *SCALEPACK* (Otwinowski & Minor 1997). Final magnitudes from both crystals were calculated with *TRUNCATE* (French & Wilson 1978). Crystallographic data and processing statistics are presented in Table 4.1.

Table 4.1. *Crystal and diffraction data statistics for the native and SeMet derivatised UgdG. crystals.*

UgdG crystal	Native	SeMet Derivative
ESRF Beamline	ID23-1	ID14-3
Wavelength (Å)	1.074	0.931
Resolution (Å)	45.98–2.40 (2.53–2.40)	48.76–3.40 (3.52–3.40)
Space-group	P622	$P4_32_12$
a = b (Å)	167.4	109.0
c (Å)	84.3	175.8
No. reflections	335219 (48456)	197981 (12797)
No. unique	27740 (3957)	15086 (1444)
Redundancy	12.1 (12.2)	13.1 (13.5)
(I/σ(I))	5.2 (1.7)	25.1 (3.8)
R_{pim} [a]	0.029 (0.129)	0.023 (0.204)
R_{rim} [b]	0.102 (0.460)	0.085 (0.763)
R_{sym} [c]	0.098 (0.441)	0.082 (0.729)
R_{anom} [d]	N/A	0.038 (0.189)
Completeness (%)	100.0 (100.0)	99.2 (99.5)
Mosaicity (º)	0.28	0.65
V_M (Å^{3}Da^{-1})	3.62	2.65
Solvent content	0.66	0.53
No. of molecules in *a.u.*	1	2
Wilson B (Å^2)	44.4	85.0

Data within parentheses refer to outer resolution shells. [a] $R_{pim} = \Sigma_h [1/(N-1)]^{1/2} \Sigma_i |I_i(h) - \langle I(h) \rangle| / \Sigma_h \Sigma_i I_i(h)$, where N is the data redundancy, I is the observed intensity and $\langle I \rangle$ is the average intensity of multiple observations from symmetry-related reflections. It is an indicator of the precision of the final merged and averaged data set (Weiss 2001). [b] $R_{rim} = R_{meas} = \Sigma_h [N/(N-1)]^{1/2} \Sigma_i |I_i(h) - \langle I(h) \rangle| / \Sigma_h \Sigma_i I_i(h)$, where N is the data redundancy, I is the observed intensity and $\langle I \rangle$ is the average intensity of multiple observations of symmetry-related reflections. It is an indicator of the average spread of the individual measurements (Weiss 2001). [c] $R_{sym} = \Sigma_h \Sigma_i |I_i(h) - \langle I(h) \rangle / | \Sigma_h \Sigma_i I_i(h)$, where I is the observed intensity and $\langle I \rangle$ is the average intensity of multiple observations from symmetry-related reflections. [d] $R_{anom} = \Sigma_h |I(h) - I(\hbar) / | \Sigma_h I(h)$, expresses the magnitude of the anomalous signal.

Assuming two molecules in the *a.u.* and according to the resolution dependent distribution of the Matthews coefficient, the tetragonal form of *S. elodea* UgdG crystal has a calculated solvent content of 53 % and a V_M of 2.65 $Å^3Da^{-1}$ with an estimated probability of 89 % (Matthews 1968; Kantardjieff & Rupp 2003). Initial attempts to obtain phases by molecular replacement using the, at the time, available UGD model from *Streptococcus pyogenes* (PDB entry 1DLJ (Campbell *et al*. 2000), with a sequence identity against UgdG of 23 %) were unsuccessful. Single anomalous dispersion (SAD) phasing was thus used with the diffraction data from the *S. elodea* UgdG seleno-methionine derivatised tetragonal crystal form. The automated protocol of *SOLVE* (Terwilliger 2004) located 18 selenium sites, as expected from two molecules in the *a.u.*, and calculated initial experimental phases and electron density maps. The selenium atom sub-structure allowed the determination of the NCS operator within the *a.u.*, which programs *DM* (Cowtan & Main 1998) and *RESOLVE* (Terwilliger 2004) used to improve the experimental phases via NCS averaging of the electron density, up to a figure of merit of 0.83. *RESOLVE* (Terwilliger 2004) and *FFEAR* (Cowtan 1998) were further used to localise protein chain fragments, mainly as poly-Ala chains. Their inspection with Xfit (McRee 1999) on a stereo-graphics workstation taking into account the NCS operator, allowed their two-fold degeneracy to be resolved and an initial model of *S. elodea* UgdG monomer, with 346 out of 438 residues (albeit including 318 poly-Ala residues), was accordingly produced (Figure 4.5.). Molecular replacement was attempted once more on the 2.4 Å native data, but this time using the preliminary UgdG monomeric model derived from the Se-methionine data as search probe. *PHASER* (McCoy *et al.* 2005) was successful and located the monomer position in the hexagonal *a.u.* (final z-scores of RFZ = 3.5 and TFZ = 9.1). The validity was corroborated by examination of omit maps (Figure 4.5.). The model completion and structural refinement of the UDP-glucose dehydrogenase from *S. elodea* ATCC 31461 is in progress.

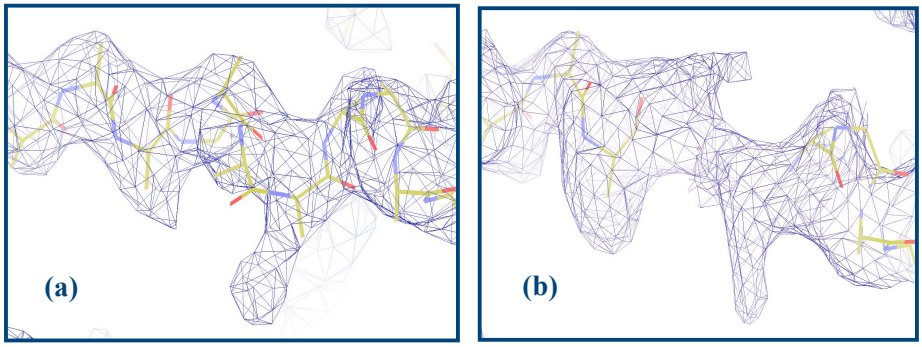


Figure 4.5. *Electron density map sections showing a helical motif of S. elodea native and SeMet derivatised UgdG.*

(a) Experimental SeMet derivative σ_A map (Read, 1986) obtained by *RESOLVE* (Terwilliger 2004) after a two-fold NCS averaging at 3.4 Å resolution, and contoured at 2σ (dark blue mesh); (b) Molecular replacement solution by *PHASER* (McCoy *et al*. 2005), after a σ_A omit map (Read 1986) calculation, at 2.4 Å resolution and contoured at 1.5σ (dark blue mesh). Four residues of the helical motif were omitted from the model in the map calculation using *REFMAC5* (Murshudov *et al*. 1997). Bonds in stick representation, with carbons in yellow, nitrogen in blue and oxygen in red.

4.4. Results

Uridine-diphospho-glucose dehydrogenase from *Sphingomonas elodea* ATCC 31461 was cloned and expressed. Extensive crystallisation screens were performed in order to address the phase problem. Well diffracting hexagonal-shaped crystals were obtained and diffraction data were collected and characterised. The best diffracting crystal belongs to the hexagonal space group P622 and diffracted up to 2.4 Å resolution. The phase problem was solved by single anomalous dispersion (SAD) using the SeMet derivatised protein. The UgdG SeMet crystals diffracted up to 3.4 Å resolution and belong to the tetragonal $P4_32_12$ space group, with a dimer in the asymmetric unit. The native crystal contains one molecule of UgdG in the asymmetric unit, but the dimeric

protein can be generated by the crystal's symmetry, corroborating the experimental results obtained in solution with other biochemical techniques indicating that UgdG is a dimeric protein under physiological conditions.

4.5. Acknowledgements

The authors gratefully acknowledge ESRF, Grenoble, for provision of synchrotron radiation. JR is a recipient of PhD fellowship from Fundação para a Ciência e Tecnologia (FCT), Portugal, SFRH/BD/24216/2005. Joana Rocha (JR) would like to thank Ana Teresa Granja (ATG) at Instituto Superior Técnico (IST) for the cloning of the *ugdG* gene. The protein expression and purification was done by JR at Instituto de Tecnologia Química e Biológica (ITQB) following ATG experience and protocols, further optimising and making the necessary adjustments for large scale preparation. JR at ITQB performed the isolation of UgdG selenised form, but also the crystallisation, data collection, phase solution and structural determination of both proteins. This work was partially supported by FCT grants POCTI/BME/38859/2001, POCI/BIO/58401/2004 and PTDC/QUI/67925/2006.

5. The three dimensional structure of the *Sphingomonas elodea* ATCC 31461 Uridine-5'-diphosphoglucose dehydrogenase (UgdG)

The work described in the following sections is based on paper prepared for submission to a referee evaluated journal but now including a more detailed description of the materials and methods, as well as new figures and results. Reference: Rocha, J.; Fialho, A.; Sá-Correia, I. and Frazão, C. "The three dimensional structure of the *Sphingomonas elodea* ATCC 31461 Uridine-5'-diphosphoglucose dehydrogenase (UgdG) (2010d).

5.1. Abstract

Gellan is an exopolysaccharide (EPS) produced by the non-pathogenic bacterium *Sphingomonas elodea* ATCC 31461, which is used on a wide range of biotechnological applications. Uridine-5'-diphosphoglucose dehydrogenase (UGD) is the enzyme responsible for the NAD-dependent two fold oxidation of UDP-glucose (UDP-Glc) to UDP-glucuronic acid (UDP-GlcA), which is a key step in *gellan gum* biosynthesis. UDP-GlcA is the precursor for the glucuronate moiety that constitutes this EPS. Here we report the X-ray crystal structure of *S. elodea* ATCC 31461 UGD, termed UgdG, in complex with its cofactor NAD(H). UgdG exhibits two dinucleotide binding Rossmann fold domains at the N- and C-terminal regions connected by a 4-helical bundle subdomain, responsible for homodimerisation and where the active site and catalytic residues are located. Its structural comparison with the available UGDs shows a strong conservation, both in primary and tertiary structures. A phyletic analysis based on the 3D structures shows that prokaryotic UGDs split into two distinct

groups, with UgdG belonging to group I, which is closely related to the eukaryotic members.

5.2. Introduction

The increased demand for natural polymers for various industrial applications in recent years has led to a renewed interest in exopolysaccharide (EPS) production by microorganisms (Chi & Fang, 2005; Kumar *et al*. 2007). Many gram-negative bacteria synthetise and secrete extracellular polymers which may have diverse biological roles, such as virulence factors in plant and animal pathogens, signalling molecules in bacteria-plant interactions and protection against predation (from protozoans) or osmosis (Sutherland 2001). Interestingly, some of these bacterial polymers are potential or already accepted products of biotechnology (Sá-Correia *et al*. 2002), which are gradually proving economically at par with natural gums produced by algae and plants. In effect, the advantages of microbial EPS over plants or marine microalgae polysaccharides are their novel functions, constant chemical and physical properties and stable supply (Sutherland 1997; Broadbent *et al*. 2003; Chi & Fang 2005). Gellan is a high molecular weight EPS produced in high yields by the non-pathogenic bacterium *Sphingomonas elodea* ATCC 31461, formerly known as *Pseudomonas elodea* (Harding *et al*. 2004; Silva *et al*. 2005). The polymer is composed of a tetrasaccharide repeating unit containing two molecules of D-glucose, one of L-rhamnose and one of D-glucuronic acid. In the native form, it has also O-acetyl and L-glyceryl substituents in one of the D-glucose monomers (Sá-Correia *et al*. 2002). The rheological properties of gellan vary considerably with several external factors such as temperature or the presence of cations in solution, but more importantly, the degree of acetylation of the polymer. Chemical deacylation of gellan causes the conversion of a soft,

elastic and thermoreversible gel to a harder, firm and brittle gel (Sutherland 1997; Jay *et al*. 1998). When compared to other polysaccharides, gellan presents several advantages such as excellent thermal and acid stability, adjustable elasticity and rigidity; high transparency and good flavor release (Fialho *et al*. 2008).

Due to these characteristics, gellan has been used in different biotechnological applications ranging from bioremediation to food industry applications. Also known as the food additive code E418, gellan is being used as a gelling, stabilising and suspending agent in dessert gels, icings, sauces, puddings and microwavable food, but also in several personal care products, such as lotions, creams and toothpaste (Sutherland 2001). In the biomedical field, gellan also found its place being used as an excipient for nasal, ocular, gastric and colonic drug delivery (Jansson *et al*. 2005). It has also proved to be a suitable material for the construction of 3D scaffolds for tissue engineering as well as a substitute for agar in bacterial and plant tissue culture media (Ciardelli *et al*. 2005; Smith *et al*. 2007). Regarding the environmental field, gellan is used in the degradation of gasoline or for the encapsulation of bacteria for bioaugmentation of contaminated aquifers (Moslemy *et al*. 2003; Moslemy *et al*. 2004). It is one of the few bacterial gums with gelling properties. There are approximately 900 worldwide patents over the applications of gellan, with the number increasing each year, suggesting that gellan has an attractive commercial market as it is proving to be one of the most important commercialised bacterial exopolysaccharides. The characterisation of the gellan biosynthetic genes and the genetic engineering of the pathway, together with the comparison of the yield and the properties of the polymers produced, are therefore of great importance, and will allow the understanding of crucial aspects in the regulation as well as the major bottlenecks of the gellan metabollic pathway (Fialho *et al*. 2008). The metabolic pathway of gellan production is a multi-step process that initiates with the intracellular formation of the nucleotide sugar-precursors,

followed by the formation of the repeating unit and ultimately, gellan polymerisation and export (Sá-Correia *et al*. 2002; Harding *et al*. 2004).

One of the enzymes involved in this complex process is uridine-diphospho-glucose dehydrogenase (UGD; EC 1.1.1.22), which catalyses the NAD-dependent two fold oxidation of UDP-Glc to UDP-GlcA (Granja *et al*. 2007). UGD belongs to a small family of nucleotide-sugar modifying enzymes widely spread in nature (Ge *et al*. 2004). It can be found in the mamallian liver where UDP-GlcA is the substrate for several glucuronosyl transferases involved in xenobiotic solubilisation, allowing the excretion of these agents out of the organism (Tukey *et al*. 2000), or the production of hyaluronic acid, a matrix polysaccharide present in human connective tissues (Spicer *et al*. 1998; Huh *et al*. 2004). In plants, UDP-GlcA is the precursor for other activated sugars such as uridine-diphospho-arabinose (UDP-Ara) or uridine-diphospho-xylose (UDP-Xyl), which constitute the plants matrix polysaccharides (Kärkönen 2005; Klinghammer & Tenhaken 2007). In bacteria, UDP-GlcA is used for the production of extracellular polymers that play different roles according to the microorganism and the ecological niche surrounding it (Broadbent *et al*. 2003; Kumar *et al*. 2007).

Our interest in the *S. elodea* ATCC 31461 UDP-glucose dehydrogenase, termed UgdG, arises from its pivotal role in the gellan gum biosynthetic pathway. It is the only bacterial source of the glucuronate moiety that composes the tetrasaccharide repeating unit of the polymer. UgdG is therefore a perfect target for enzyme manipulation and the design of activity enhancers. The knowledge of its three-dimensional structure is essential to enable this.

5.3. Experimental Procedures

5.3.1. Protein purification and crystallisation

UgdG wild type and SeMet derivatised proteins were expressed and purified as previously described in detail (see *sections 4.3.1.* and *4.3.2.*). Briefly, cells harbouring the plasmid constructs for native or SeMet derivatised UgdG expression, were grown in LB supplemented with ampicillin, harvested, and resuspended in sodium chloride (1 *M*) and phosphate buffer (20 m*M*) at pH 7.4. After disruption, the cell free extracts were applied onto a nickel-chelating column (Histrap FF, GeHealthcare). Proteins were eluted using a linear imidazole gradient and the pure fractions were pooled and buffer exchanged to a suitable storage buffer. Both proteins were confirmed to be homogenous and of high purity by SDS-polyacrylamide gel electrophoresis, and their concentration determined both by Bradford's method (Bradford, 1976) or direct measurement of their absorbance at 280 nm, using a extinction coefficient of 18490 $M^{-1}{\cdot}cm^{-1}$. The proteins were stored at -80º C with final concentrations of approximately 4.5 and 5.5 $mg{\cdot}ml^{-1}$ for native and SeMet derivatised proteins, respectively. Wild-type crystals were firstly obtained in a microbatch experiment (see *section 4.3.3.*) but the availability of a "Honeybee" Microsys 4000 XL Cartesian™ Dispensing Systems robot (Genomic Solutions, USA) allowed the screening of new conditions using vapour diffusion techniques. New crystals of native UgdG were obtained with solution 56 of the PEG II screen from Qiagen Canada Inc. (100 m*M* MES pH 6.0, 10% (w/v) PEG 6000, 1*M* LiCl). SeMet derivatised crystals were obtained as described previously (see *section 4.3.3.*). Cryoprotection of UgdG native crystal was achieved by a quick soak in a solution of their mother liquor including 25% glycerol, in order to minimise the risk of crystal damage in X-rays. Crystals were then flash cooled and stored in liquid nitrogen for diffraction data collection.

5.3.2. Diffraction data collection and processing

Native UgdG diffraction data were measured at beam line ID23-1 of the European Synchrotron Radiation Facility (ESRF), Grenoble, France. The diffraction images were processed with *MOSFLM* (Leslie 1992), intensities were scaled and merged with *SCALA* (Evans, 2006) and final magnitudes were calculated with *TRUNCATE* (French & Wilson 1978) from the *CCP4* suite (The CCP4 suite: programs for protein crystallography, 1994).

5.3.3. Structural solution and refinement

Native crystal UgdG phases were obtained by molecular replacement using *PHASER* (McCoy *et al*. 2005) and the preliminary model of the seleno-methionine derivative UgdG (see *section 4.3.4.*) as the search structure. Model editing and completion against electron density maps at a graphics workstation and iterative structural refinement of the UgdG monomer were performed with *COOT* (Emsley & Cowtan 2004) and *REFMAC5* (Murshudov *et al*. 1997), respectively. The progress of the refinement was monitored by the use of R_{free}, where 5% of the reflections were constantly set aside for cross-validation.

5.3.4. Structural analysis

Molecular stereochemistry was evaluated using the program *PROCHECK* (Laskowsky *et al*. 1993). Inter-molecular contacts and crystallographic packing were checked with *PISA* (Krissinel & Henrick 2007), and structural superposition of homologous models were produced with *SSM* (Krissinel & Henrick 2004) and *MODELLER* (Sali & Blundell 1993). Images were produced with *PYMOL* (Delano 2008). 3D-structural comparisons were performed with *MAMMOTH* (Lupyian *et al.* 2005), a multiple structure sequence-independent

alignment program based on "all-against-all" pairwise comparison of protein models. Briefly, an overall alignment based on models coordinates is performed. The models are divided into heptapeptides, and for each successive Cα pair, a unit vector is computed. These unit vectors are then grouped at a same origin, the root mean square distance between all pairs of heptapeptides is calculated and a similarity matrix is generated. The latter is then used for 3D alignment based on backbone (local) similarity. A similarity score is computed and output as a dendogram file. The obtained dendogram was fed into *PHYLIP* (Felsenstein 1993) to create a structural-based phyletic relationship tree.

5.4. Results and Discussion

5.4.1. Crystallographic data

UgdG was produced, purified and crystallised as described elsewhere (see *section 4.3.*). Crystals were further optimised with the new crystallisation conditions (see Experimental Procedures) leading to improved diffraction data up to 2.37 Å resolution. They belong to space-group P622 with cell dimensions a=b=167.6 Å, c= 84.8 Å, containing one molecule in the asymmetric unit (*a.u.*), which corresponds to a Matthews coefficient of 3.6 $Å^3Da^{-1}$ and a calculated solvent content of 66 %. Crystallographic statistics are presented in Table 5.1.

5.4.2. Phase problem solution and structural refinement of UgdG

Since initial molecular replacement attempts of native UgdG using the, at that time, available structure of UGD from *S. pyogenes* (Campbell *et al*. 2000) as the search structure were not successful (sequence identity 23 %), seleno-methionine derivatised UgdG (SeMet-UgdG) was produced and crystallised.

Table 5.1. Crystallographic parameters and data-collection statistics for native UgdG crystals.

Crystal	***UgdG***
ESRF Beamline	ID14-3
Wavelength (Å)	0.931
Resolution (Å)	59.55–2.37 (2.50–2.37)
Space-group	P622
a, *b*, *c* (Å)	167.6, 167.6, 84.8
No. reflections	329626 (48366)
No. unique	29012 (4161)
Redundancy	11.4 (11.6)
(I/σ(I))	6.6 (1.9)
R_{pim} [a]	0.024 (0.124)
R_{rim} [b]	0.082 (0.429)
R_{svm} [c]	0.078 (0.411)
Completeness (%)	100.0 (100.0)
Mosaicity (º)	0.27
V_M (Å^{3}Da^{-1})	3.64
Solvent content (%)	66.2
No. of molecules in *a.u.*	1
Wilson B (Å^2)	42.7

Data within parentheses refer to outer resolution shells. [a] $R_{pim} = \Sigma_h [1/(N-1)]^{1/2} \Sigma_i |I_i(h) - \langle I(h) \rangle| / \Sigma_h \Sigma_i I_i(h)$, where N is the data redundancy, I is the observed intensity and $\langle I \rangle$ is the average intensity of multiple observations from symmetry-related reflections. It is an indicator of the precision of the final merged and averaged data-set (Weiss, 2001). [b] $R_{rim} = R_{meas} = \Sigma_h [N/(N-1)]^{1/2} \Sigma_i |I_i(h) - \langle I(h) \rangle| / \Sigma_h \Sigma_i I_i(h)$, where N is the data redundancy, I is the observed intensity and $\langle I \rangle$ is the average intensity of multiple observations of symmetry-related reflections. It is an indicator of the average spread of the individual measurements (Weiss, 2001). [c] $R_{sym} = \Sigma_h \Sigma_i |I_i(h) - \langle I(h) \rangle| / |\Sigma_h \Sigma_i I_i(h)$, where I is the observed intensity and $\langle I \rangle$ is the average intensity of multiple observations from symmetry-related reflections.

The crystals belonging to tetragonal space-group $P4_32_12$, diffracted up to 3.4 Å resolution, and their phases were obtained by single anomalous dispersion methods (SAD), leading to an incomplete, mostly poly-Ala truncated model (only 346 out of 438 residues). Using this preliminary SeMet-UgdG model, a molecular replacement solution for the initial diffraction data set to 2.4 Å resolution of the hexagonal native crystals was readily obtained by *PHASER* (McCoy *et al.* 2007), with final Z-scores RFZ = 3.5 and TFZ = 9.1 (*section 4.3.4.*). Readily interpretable electron density maps were produced, which included density of the NAD(H) cofactor nestled into the UgdG monomer. The model was improved through successive cycles of manual building and refinement using the later obtained higher resolution data (2.37 Å), until convergence of R_{work}/R_{free} to 18.2/22.6 %. The last refinement cycle included the use of three TLS groups, determined with the *TLSMD* server (Painter & Merrit 2006). The final R factor deduced for all data is 18.4 %. Further refinement statistics are presented in Table 5.2.

5.4.3. UgdG crystal contents

While the two residues of the linker between the His-tag and the first valine residue of the UgdG protein could be modelled as Gly-1 and Ser0, the 6 histidines of the N-terminal tag and residues 90-92 and 191-192 were not visible in the electron density maps due to crystallographic disorder, and therefore omitted from the models. Globally, the native structure of UgdG contains 433 out of the 440 (2+438) residues, 125 solvent water molecules and two chloride solvent species.

Table 5.2. *Refinement statistics and stereochemistry assessment for native UgdG*.

A stereochemical analysis of the final UgdG model, revealed 91.3%, 8.7%, 0.0% and 0.0% of the residues are located in the most favourable, additional allowed, generously allowed and disallowed regions of the Ramachandran plot, respectively.

Parameter	***UgdG***
R_{work} (%)	18.2
R_{free} (%)	22.6
R (%)	18.4
Number of residues	433
Solvent water molecules	125
Other ligand molecules (NAD(H), chloride, glycerol)	4
Average protein atoms <*a.d.p.*> ($Å^2$)	43.2
Average main-chain <*a.d.p.*> ($Å^2$)	42.0
Average side-chain <*a.d.p.*> ($Å^2$)	44.4
Nicotinamide adenine dinucleotide [NAD(H)] <*a.d.p.*> ($Å^2$)	46.8
Average solvent <*a.d.p.*> ($Å^2$)	42.0
Deviations from the ideal	
Bond distances rmsd (Å)	0.014
Bond angles rmsd (º)	1.51

These were introduced whenever the initially modelled waters did not compensate for persistent positive m|Fo|-D|Fc| density (Read 1986), were refined to too low atomic displacement parameters (*a.d.p.'s*) when compared with neighbouring atoms, or had a positively charged close environment, since 1 *M* LiCl was present in the crystallisation media. Additionally, the protein molecule also exhibits a docked cofactor molecule modelled as adenine-diphospho-ribose with occupancy of 70%, as the residual σ_A density maps (Read, 1986) presented no evident density that could account for the nicotinamide moiety of the cocrystallised NAD^+ cofactor molecule.

5.4.4. UgdG monomeric topology

The single UgdG monomer in the crystal *a.u.* has dimensions of approximately 70, 40 and 38 Å, and is composed of two domains with an α/β topology connected by an α-helical bundle domain (Figure 5.1.). Both N- and C-terminal exhibit a three-layer αβα sandwich-like domain with a central β-sheet flanked on both sides by α-helices, typical of a dinucleotide-binding fold. The N-terminal domain (residues 1-199) shows a six-stranded parallel β-sheet (β1-β6) interspersed with seven α-helices (α1-α7), a canonical Rossmann fold that serves as the nicotinamide adenine dinucleotide (NAD(H)) binding domain, followed by an anti-parallel β-α-β extension (β7-α8-β8), that leads straight into the central domain. This is an entirely α-helical region (residues 200-307), composed of a four-helical bundle (α9-α12) and it is responsible for the homodimerisation of UgdG. This structural arrangement is present in several dehydrogenases whose homodimerisation is also produced *via* interactions between the central helical domains of the two interacting monomers. The C-terminal region of the UgdG molecule (residues 308-438) also exhibits a dinucleotide binding like Rossmann fold, with a β-sheet core (β9 to β14) flanked on both sides by several α-helices (α13 to α17).

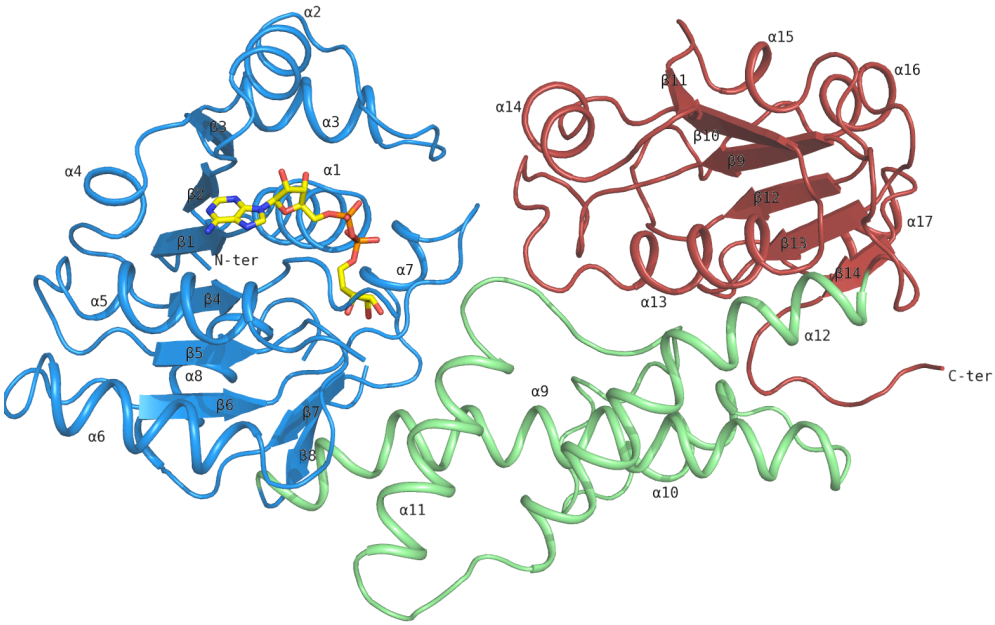


Figure 5.1. *Three-dimensional structure of the UgdG monomer with an incomplete cofactor NAD(H) bound.*

The UDP-glucose dehydrogenase from *S. elodea* ATCC 31461 (UgdG) exhibits an α/β fold found in other nucleotide-sugar dehydrogenases and comprises two αβα sandwich type dinucleotide binding Rossmann-like domains, at the N- (blue) and C-terminal (red) parts, connected by an α-helical sub-domain (green). Secondary structure details were classified numerically for both α-helical or for β-strands elements. Cofactor molecule in stick representation coloured yellow for carbon, blue for nitrogen, red for oxygen and orange for phosphorus.

The N- and C-terminal domains show a striking similarity when superimposed, except for some missing α-helical elements on the C-terminal domain (Figure 5.2), even if their sequence identity is low (*ca*. 17%). The presence of two similar domains in one protein was first described for bovine liver rhodanese (Ploegman *et al*. 1978) and has been found in many protein structures (Lang *et al*. 2000).

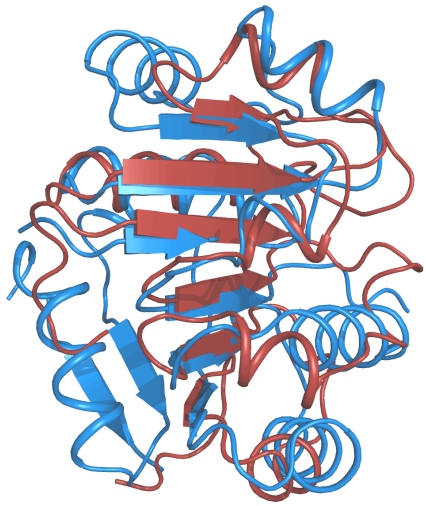


Figure 5.2. *The superposition of the N-and C-terminal domains of UgdG.*
The two dinucleotide-binding Rossmann-like domains, the N- (blue) and C-terminal (red) exhibit a striking similarity when superposed, with an rmsd of 2.7 $Å^2$ over 101 alpha carbons (Krissinel & Henrick 2007).

5.4.5. The interface of the UgdG dimer

An analysis of the crystal packing contacts strongly suggests that UgdG forms dimeric species, as each molecule uses 2650 $Å^2$ (13.4% of its otherwise 19718 $Å^2$ solvent accessible surface) to form an interface with its closest crystal neighbour (Figure 5.3. a, b, c). The dimeric mates are related by a crystallographic two-fold rotation axis, which brings the N-domain of one monomer closer to the C-domain of the other. This oligomerisation is consistent to what has been described for other members of this family (Campbell *et al*. 2000; Snook *et al*. 2003; Hung *et al*. 2007), although in some cases a higher quaternary arrangement is also observed. This is the case of the most studied eukaryotic UGDs, such as bovine and human, where a "trimer of dimers" was found to be the predominant arrangement although the role of subunit association is not clear, as it is not necessary for activity (Easley *et al*. 2007).

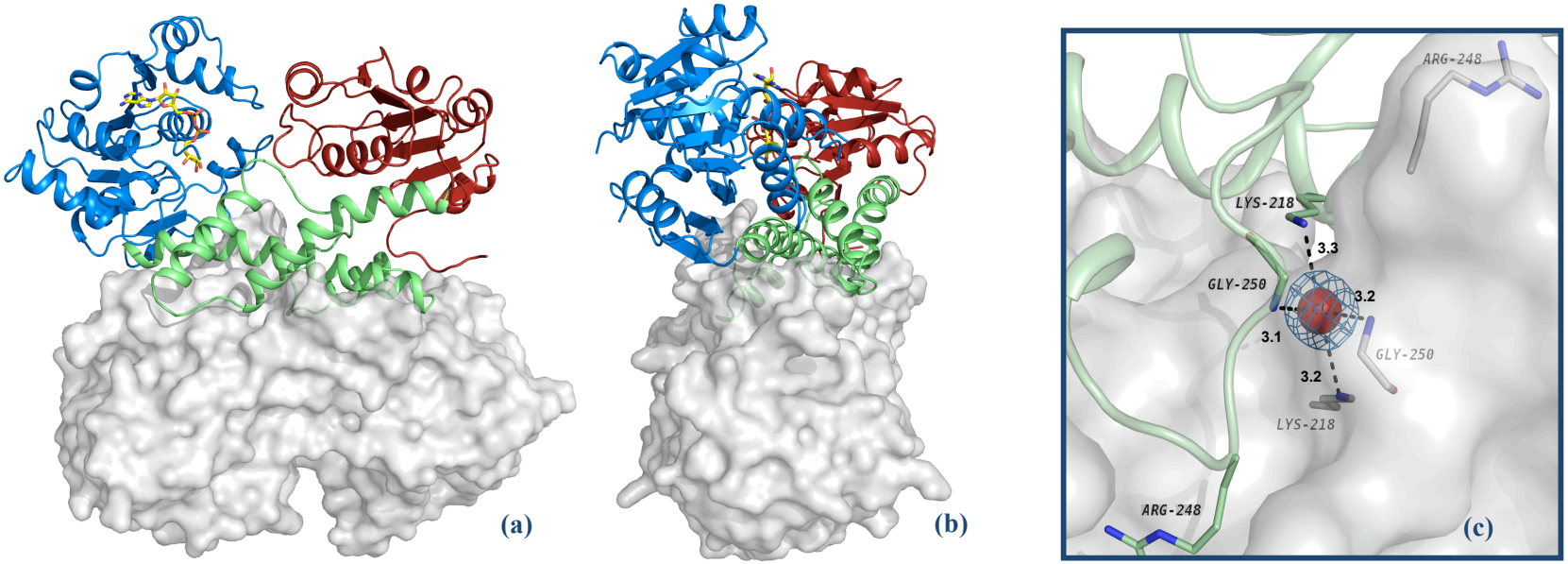


Figure 5.3. *The quaternary arrangement and the dimer interface of UgdG.*

(a) and (b) Perpendicular views of the UgdG dimer with one monomer shown as a cartoon with the N- and C-terminal domains in blue and red, respectively, and the middle subdomain in green. The other monomer is represented by its solvent accessible surface coloured (in grey). The (incomplete) cofactor molecule is represented in sticks coloured yellow for carbon, blue for nitrogen, red for oxygen and orange for phosphorus. (c) Zoomed view of the dimeric interface showing a chloride ion (red sphere) on a crystallographic two-fold axis. Residues side-chains are presented in sticks with atoms coloured green or grey for carbon in each monomer, and in blue for nitrogen and red for oxygen atoms. A 2mFo-DFc electron density contoured at 1σ is shown as a blue mesh around the chloride ion. Hydrogen bonds represented as black dashed lines with distances in Å.

Additionally, the rat and the *Cryptococcus* fungi UGDs have been described as tetrameric and dimeric, respectively (Sivaswami *et al*. 1972; Bar-Pelled *et al*. 2004). On the other hand, prokaryotic UGDs seem to be dimeric (Sieberth *et al*. 1995), even though the UGD monomer of *Streptococcus pyogenes* was also described as active (Campbell *et al*. 1997). Interestingly, the UgdG dimer interface shows a chloride ion lying exactly at the two-fold symmetry axis that relates the two monomers. This is likely to be a crystallisation artefact, as the precipitant solution contained 1 *M* lithium chloride (see *section 4.3.3.*) and it has not been described for any of the nucleotide (NDP)-sugar dehydrogenase members of the family. The ion interacts with the main-chain and side-chain amines of Gly250 and Lys218, respectively, of the symmetrically related monomers (see Figure 5.3. c). The interface is built mainly by interactions between the helical subdomains of both monomers, a scheme verified for other dehydrogenases and NAD-binding proteins (Kutzenko *et al*. 1998).

Table 5.3. *Hydrogen bonding network at the dimeric interface of UgdG.*
Residues involved and the respective bond distances were defined using PISA (Krissinel & Henrick 2007).

Hydrogen bonds at the UgdG dimeric interface		
Chain A	***Distance (Å)***	***Chain B***
Asn247 [N]	2.7	Asp166 [OD2]
Asp235 [N]	3.0	Thr203 [OG1]
Ile285 [N]	3.2	Glu224 [OE1]
Arg284 [N]	2.8	Glu224 [OE2]
Arg202 [NH1]	2.8	Val232 [O]
Arg167 [NH2]	2.7	Gly242 [O]
Lys218 [NZ]	3.1	Arg248 [O]
Asp166 [OD2]	2.7	Asn247 [N]
Thr203 [OG1]	3.0	Asp235 [N]
Glu224 [OE1]	3.2	Ile285 [N]
Glu224 [OE2]	2.8	Arg284 [N]
Val232 [O]	2.8	Arg202 [NH1]
Gly242 [O]	2.7	Arg167 [NH2]
Arg248 [O]	3.1	Lys218 [NZ]

Dimerisation occurs upon the close packing of both central helices (α9) of the two subdomains, through an extensive network of hydrogen bonds and van der Waals interactions, even if the primary structure conservation is very low in this region. Amongst these interactions, 14 are hydrogen bonds (see Table 5.3. above), the remainder being van der Waals contacts. The interface involves interactions between 63 of the about hundred residues that compose the subdomain, but also a few N-terminal residues, such as the strictly conserved Phe152 and Arg167. Several aromatic residues are involved in the interface formation, including Phe214, Phe221, Phe253 and Phe265 (Figure 5.4.). The majority of these residues are not strictly conserved but some aromatic character conservation is observed, when compared to other UGDs.

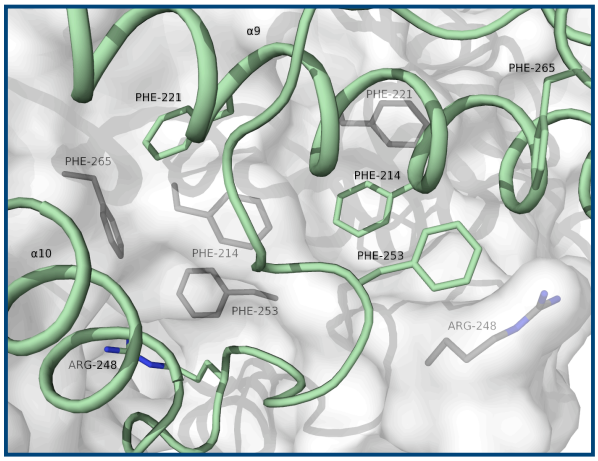


Figure 5.4. *The hydrophobic character of the dimer interface*.

The UgdG dimeric interface has a strong hydrophobic character with conservation of aromatic residues. Monomer A is in cartoon representation and monomer B in its transparent solvent accessible surface. Residues represented in sticks coloured as in Figure 5.3.

For example, *S. pyogenes* UGD has tyrosine residues at these equivalent positions (Campbell *et al*. 2000), while human UGD also has phenylalanine residues except for a serine on position 232, equivalent to Phe221 on UgdG.

This might contribute to the higher similarity of UgdG and eukaryotic UGDs, as pointed by Granja and co-workers (Granja *et al*. 2007). Arg248, one of the few strictly conserved residues present at the interface, was proposed by Campbell (Campbell *et al*. 2000) to be determinant of substrate specificity since the residue's side-chain contacts with the active site and with the substrate from the partner. Such an interaction is not detected in UgdG due to the absence of the substrate in the respective binding pocket. Other strictly conserved residues involved in dimerisation are Leu215, Asp 246, Ile249 and Leu272 (Leu227, Asp258, Ile261 and Leu284 for human, and Leu211, Asp242, Ile 245 and Leu268 for *S. pyogenes* UGDs).

5.4.6. The NAD(H) cofactor binding pocket of UgdG

Despite all of the efforts to produce a UgdG-NAD(H) complex with a complete cofactor bound, one observes a lack of density at the nicotinamide ring of the co-crystallised NAD(H). Therefore, only its adenosine, diphosphate and ribose groups could be modelled into the electron density map (Figure 5.5.). Such a phenomenon was also observed in other NAD-binding proteins, and has been justified by the possibility of different conformations of the ring within the crystal, or by a possible hydrolysis of the glycosidic bond, under the experimental conditions, originating several subproducts (Margolis *et al*. 1976; Miksic & Brown 1978; Bellamacina 1996). In UGDs, an inspection of available NAD(H) complexes seems to indicate that cofactor degradation is pH dependent, originating the adenosine-diphospho-ribose (ADP-ribose) subproduct. In *S. pyogenes* and human UGD-NAD(H) structures (Campbell et al 2000; Kavanagh *et al*. 2007), where the cofactor molecule is complete, crystallisation occurred at pH 7.8 and 6.5, respectively. In the case of GMD, where an incomplete cofactor molecule was found bound to the protein (Snook *et al*. 2003), crystallisation occurred at an acidic pH 4.6, similar to that of UgdG

(pH 5.6 or 6; see *sections 4.3.3.* and *5.3.1*). Moreover, the 2mFo-DFc and mFo-DFc electron density maps of UgdG show the change in the chirality of the NAD(H)-ribose at C1D, suggesting the substitution of the N1N amine group by a hydroxyl group, leading to ADP-ribose, and release of the nicotinamide moiety to the media. Also, NAD^+ binding is a fast and labile state, as the cofactor must exchange rapidly to allow enzyme turnover, and full oxidation of the alcohol to acid. This helps to explain why the cofactor could not be refined to its full occupation, but to a value of 70% (see *section 5.4.3.*). The visible part of the cofactor adopts an extended conformation in a cleft within the N-terminal domain, nearby the N-extremities of β1 and β4. A comparison with available UGD-NAD(H) complexes (Campbell *et al.* 2000; Kavanagh *et al.* 2007 not published) shows that the hydrogen bonding network between the protein and the visible part of the cofactor is maintained, and that if the side-chain of Glu151 had adopted a Chi1 *trans* conformation, as in ternary complexes of human and *S. pyogenes* UGDs, there would be room available for the (not visible) nicotinamide ring of NAD (Figure 5.5.). Instead, it adopts a *gauche (+)* conformation also found in the BceC structure without any docked cofactor (*section 3.*). Modelling the fragment as ADP-ribose allows an extra hydrogen bond between UgdG and the truncated cofactor fragment (see Figure 5.5.). In other UGD/GMD structures (Campbell *et al.* 2000; Snook *et al.* 2003), the nicotinamide ring is found binding in a *syn* orientation; with its C4 carbon (where the hydride transfer occurs) located near the catalytic Cys264 as required for catalysis, and its carboxamide group pointing to the side-chain of Glu155 (not shown). A detailed description of the hydrogen bonding interactions between the cofactor and the already characterised UGDs is given on Table 5.4. The adenine moiety lies on a patch formed by several hydrophobic residues, some conserved like Val84 and Tyr99, and other non-conserved residues such as Lys31 and Ala103 (substituting Val31 and Val102

in GMD). All of these residues are within van der Waals distance from the base (Figure 5.5.; Snook *et al*. 2003).

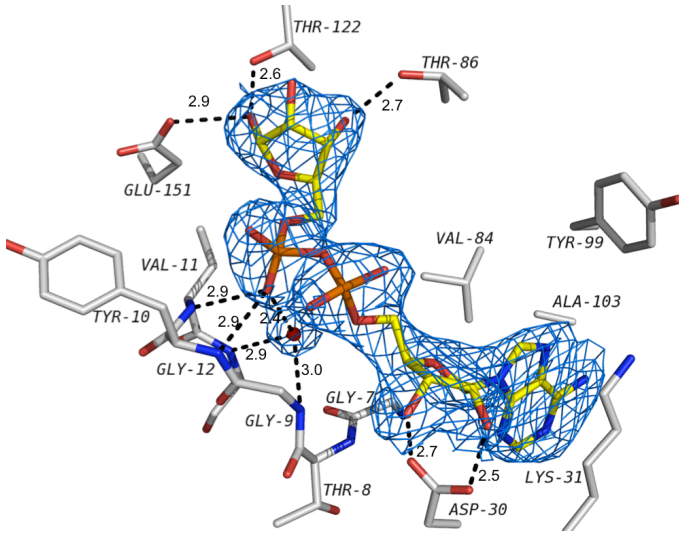


Figure 5.5. *The cofactor NAD$^+$/NADH binding pocket of UgdG.*
The electron density at the cofactor binding pocket shows a docked NAD(H) fragment, where the nicotinamide ring is absent in the 2mFo-DFc map. The cofactor fragment was therefore modelled as an incomplete NAD$^+$, resembling an adenosine-diphospho-ribose (ADP-ribose), represented as sticks coloured yellow for carbon, blue for nitrogen, red for oxygen and orange for phosphorus, and representing interacting UgdG residues with carbon atoms in grey, and a conserved water molecule (see text) as a red sphere. 2mFo-DFc electron density map around the cofactor fragment at 1.5 σ is represented as a blue mesh. Hydrogen bond interactions are shown in black dashed lines with distances in Å.

One conserved residue of the Rossmann fold, Asp30 forms hydrogen bonds with the hydroxyls of the adenine ribose ring. The pyrophosphate group is bound to the protein by the GXGXXG (glycine-rich) motif, known as one of the *fingerprint* regions of NAD(H) binding proteins with a classical Rossmann fold (Bellamancina 1996). The first glycine residue of the loop, Gly7, allows a tight turn of the peptide chain whilst the other two, Gly9 and Gly12, interact with the pyrophosphate group, allowing the correct positioning of the cofactor molecule.

Table 5.4. *Hydrogen bonding interactions within UGDs cofactor binding pockets.* UgdG *(this structure),* BceC with NAD(H) *in silico* model (see *section 3.*), *P. gingivalis* (Bonanno *et al*. 2009, not published, PDB entry 3GG2) with NAD(H) *in silico* model, *H. sapiens* (Kavanaugh *et al*. 2007, not published, PDB entry 2Q3E) *and S. pyogenes* (Campbell *et al*. 2000, PDB entry 1DLJ) UGDs are represented. The cofactor is divided into adenosine, diphosphate, ribose and nicotinamide sections (first column). Equivalent atom labels of reaction substrate/product (second column) appear in the residue rows (columns 3-5) to which they form hydrogen bonds (defined up to 3.2 Å initial cut-off distance, but homologous interactions up to 4 Å were included). Conserved residues are highlighted in bold; residues bridged to the cofactor *via* a solvent water molecule are presented in italic within parenthesis. Direct hydrogen-bonding distance values (Å) are within parentheses.

		UgdG	**BceC *(section 3.)* with NAD(H) *in silico* model**	***P. gingivalis* UGD (3GG2) with NAD(H) *in silico* model**	***H. sapiens* UGD (2Q3E)**	***S. pyogenes* UGD (1DLJ)**
Adenosine	N1A					*(Ser67_OG) (2.7)* *(Tyr71_OH) (2.8)* *(Glu103_OE1) (2.8)*
	N6A	*(Tyr99_O) (3.0)*	*(Tyr98_O) (2.7)*	Tyr176_O (2.7)		
	N7A	*(Tyr99_O) (3.0)*				
	O2B	**Asp**30_OD1 (2.5)	**Asp**30_OD1 (3.9)	**Asp**108_OD1 (2.8)	**Asp**36_OD1 (2.6) *(Asn38_ND2) (3.1)*	**Asp**29_OD1 (2.8)
	O3B	**Asp**30_OD2 (2.7)	**Asp**30_OD2 (4.2) Lys35_NZ (2.7)	**Asp**108_OD2 (3.5)	**Asp**36_OD2 (2.8) Arg41_NH2 (2.9) *(Asn38_ND2) (3.1)*	**Asp**29_OD2 (2.7) Lys34_NZ (3.0)
Diphosphate	O1A		*(Arg338_NH1) (2.9)*		*(Arg346_NH1) (2.8)*	Arg327_NH1 (3.1)
	O2A				Arg41_NH1 (3.1) *(Arg41_NH2) (3.0)*	*(Lys34_NZ) (2.9)*
	O1N			Arg412_NH2 (4.3)	Arg346_NH1 (3.0) Arg346_NH2 (2.9) *(Ser275_OG) (2.8)*	Arg327_NH2 (2.9)
	O2N	**Tyr**10_N (2.9) **Val**11_N (2.9) ***(Gly**9_N) (3.0)* ***(Gly**12_N) (2.9)* *(Ala83_O) (2.8)*	**Tyr**10_N (3.5) **Val**11_N (3.0) ***(Gly**9_N) (3.0)* ***(Gly**12_N) (2.9)* *(Ala83_O) (2.6)*	**Tyr**88_N (2.5) **Val**89_N (3.0) ***(Gly87**_N) (3.0)* ***(Gly90**_N) (2.9)* *(Ala161_O) (2.7)*	**Tyr**14_N (3.1) **Val**15_N (3.0) ***(Gly**13_N) (2.9)* ***(Gly**16_N) (2.9)* *(Ser88_O) (2.9)*	**Tyr**10_N (3.2) **Val**11_N (3.1) ***(Gly**9_N) (3.0)* ***(Gly**12_N) (2.8)* *(Ala80_O) (2.8)*
	O5D				*(Ser275_OG) (2.8)*	
Ribose	O2D	**Lys**267_NZ (3.4) ***(Thr**122_OG1) (3.2)* ***(Asp**268_OD2) (3.0)*	**Lys**273_NZ (2.7) ***(Thr**121_OG1) (2.8)* ***(Asp**274_OD2) (2.8)*	**Lys**347_NZ (2.7) ***(Thr**199_OG1) (2.9)* ***(Asp**348_OD2) (2.5)*	**Lys**279_NZ (2.8) ***(Thr**131_OG1) (2.6)* ***(Asp**280_OD2) (2.6)*	**Lys**263_NZ (2.9) ***(Thr**118_OG1) (2.5)* ***(Asp**264_OD2) (2.6)*
	O3D	**Thr**86_OG1 (2.7) **Thr**122_N (3.2)	**Thr**86_OG1 (2.9) **Thr**121_N (2.7)		**Thr**91_OG1 (2.7) **Thr**131_N (3.0)	**Thr**83_OG1 (2.5) **Thr**118_N (3.2)
Nicotinamide	N1N	**Thr**122_OG1 (2.6) **Glu**151_OE2 (2.9)				
	N7N	N/A	Clash with **Tyr**10 Clash with **Glu**154	Clash with **Tyr**88	**Glu**165_OE2 (2.8)	**Glu**145_OE2 (3.1)

However, whilst the main-chain amines of Tyr10 and Val11, two conserved residues of the phosphate-binding loop, hydrogen bond directly with the pyrophosphate, the main-chain amines of these two glycines contiguous to them interact with the group via a water molecule which has been frequently described as characteristic of the classical dinucleotide-binding Rossmann fold proteins (Bottoms *et al*. 2002; see above Figure 5.5.).

5.4.7. The substrate binding pocket of UgdG

Despite all the attempts, it was not possible to obtain UgdG as a ternary complex with cofactor and substrate or analogues. Having obtained the homologous structure of *Burkholderia cepacia* UGD, BceC, in complex with its reaction product, UDP-GlcA (see *section 3*.), its coordinates were used to map the UgdG substrate-binding pocket (Figure 5.6.).

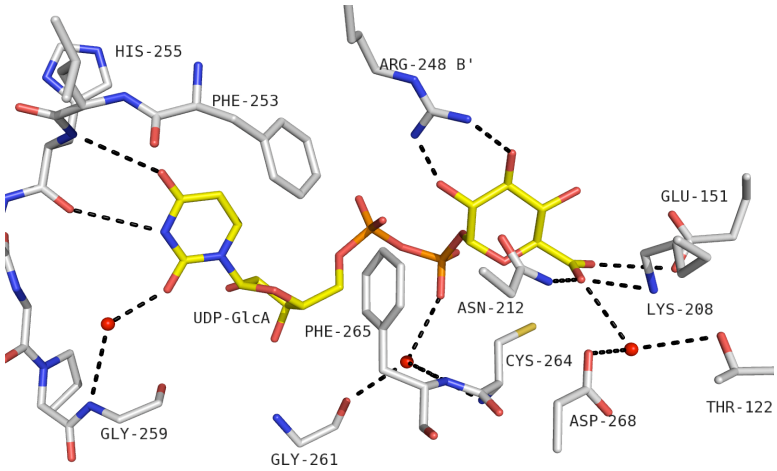


Figure 5.6. *The model of UgdG substrate binding pocket.*

The 3D structure of BceC with a docked UDP-GlcA was superimposed onto the homologue UgdG, generating a hydrogen bonding network, illustrated by dotted lines, between UgdG and a UDP-GlcA model in its substrate binding pocket. Bonds represented as sticks, in yellow for ligand carbon atoms, grey for protein carbon atoms, blue for nitrogen, red for oxygen, orange for phosphorus and dark yellow for sulphur. Solvent water molecules represented as red spheres.

Table 5.5. *Hydrogen bonding interactions within UGDs substrate binding pockets.* UgdG with UDP-GlcA *in silico* model, BceC (see *section 3.*), *P. gingivalis* (Bonanno *et al*. 2009, not published, PDB entry 3GG2), *H. sapiens* (Kavanaugh *et al*. 2007, not published, PDB entry 2Q3E) *and S. pyogenes* (Campbell *et al*. 2000, PDB entry 1DLJ) UGDs are represented. The substrate/product is divided into uridine, diphosphate, and glucose/glucuronate moieties (first column). Equivalent atom labels of reaction substrate/product (second column) appear in the residue's rows (columns 3-7) to which they form hydrogen bonds (defined up to 3.2 Å initial cut-off distance, but homologous interactions up to 4 Å were included). Conserved residues are highlighted in bold; residues bridged to the substrate *via* a water molecule are presented in italic within parenthesis. Hydrogen-bonding distance values (Å) are within parentheses. * Represents the catalytic cysteine (or serine in *S. pyogenes* Cys260Ser mutant UGD).

		UgdG with UDP-GlcA *in silico* model	**BceC (See *section 3.*)**	***P. gingivalis* (3GG2)**	***H. sapiens* (2Q3E)**	***S. pyogenes* (1DLJ)**
Uridine	O4	His255_N (3.3)	Tyr261_N (3.0)	Tyr335_N (2.9)	Lys267_N (3.1)	Asn250_N (3.1) Asn251_N (2.9)
	N3	His255_O (3.4)	Tyr261_O (3.0)	Tyr335_O (2.9)	Lys267_O (2.8)	Asn251_O (2.9)
	O2	Arg414_NH2 (4.3)	*(Asn229_ND2)(3.1)* *(Gly265_N) (3.0)* Arg431_NH1 (3.2)	*(Asn303_ND2)(3.1)* *(Gly339_N)(3.1)* Arg495_NH1 (3.1)	Ser269_OG (2.7)	Ser253_OG (2.8)
	O2* O2C O2D	Arg414_NH2 (3.7)	*(Glu406_OE1)(2.6)* *(Glu406_OE2)(2.8)*	*(Glu470_OE1)(3.0)*	Arg442_NH1 (2.9) Arg442_NH2 (3.1) *(Glu416_OE2)(2.8)*	*(Gly255_O) (2.9)* *(Asn380_ND2)(2.9)* Asp402_OXT (2.7)
	O3* O3C O3D	**Gly**267_N (3.6) Phe324_O (3.8)	**Gly**267_N (3.0) Phe330_O (3.0)	**Gly**341_N (3.0) Phe404_O (2.8)	**Gly**273_N (2.9) Phe338_O (2.7)	**Gly**257_N (2.9) Met319_O (2.8)
Diphosphate	O1A	**Lys**326_NZ (3.7)	***(Lys**331_NZ)(2.7)*	**Lys**405_NZ (2.7)	**Lys**339_NZ (2.8) *(Arg260_NH1)(3.0)*	**Lys**320_NZ (3.2) *(Asp402_OD1)(2.5)*
	O2A	*(**Arg**248_NH1)(3.4)*				Tyr249_OH (2.7)
	O3A	**Lys**326_NZ (2.6)	**Lys**331_NZ (3.3)		**Lys**339_NZ (3.2)	**Lys**320_NZ (3.4)
	O1B	*(**Gly261**_O)(2.7)* *(**Cys**264_N)(3.1)* *(Phe265_N)(2.9)*	*(**Glu158**_OE1)(3.2)* *(**Gly267**_O)(2.8)* *(**Cys**270_N)(3.2)* *(Phe271_N)(2.9)*	*(**Gly341**_O)(2.9)* *(**Cys**344_N)(3.2)* *(Phe345_N)(2.9)*	*(**Gly**273_O)(2.7)* *(Cys276_N) 3.0)* *(Phe277_N)(3.1)*	*(**Glu**145_OE2)(2.7)* *(**Gly**257_O) (3.3)* *(*Ser**260_N) (3.1)* *(Leu261_N) (2.6)*
	O2B	**Lys**326_NZ) (2.4)	**Glu**158_N (3.4) **Lys**331_NZ (2.7)	**Glu**236_N (2.8)	**Glu**165_N (3.0) *(**Glu**165_OE2)(3.2)*	**Glu**145_N (2.9) **Lys**320_NZ (2.8)
Glucose / Glucuronic moiety	O2'	*(**Arg**248_NH1)(2.1)*	**Arg**254_NH1 (2.9)	**Arg**328_NH1 (2.8)	**Arg**260_NH1 (2.8)	**Arg**244_NH1 (2.9)
	O3'	**Phe**152_O (4.2) *(**Arg**248_NH2)(2.5)*	*(**Phe**155_O) (2.8)* **Arg**254_NH2 (2.9) ***(Lys**214_NZ)(3.2)* *(**Asn**218_OD1)(2.8)*	**Phe**233_O (2.7) **Arg**328_NH2 (3.0) ***(Lys**288_NZ)(3.7)* *(**Asn**292_OD1)(2.7)*	**Phe**162_O (2.6) **Arg**260_NH2 (3.0) ***(Lys**220_NZ)(3.5)* *(**Asn**224_OD1)(2.7)*	**Phe**142_O (2.7) **Arg**244_NH2 (3.1) *(**Lys**204_NZ) (3.2)* *(**Asn**208_OD1)(2.8)*
	O4'	**Leu**153_O (3.7) **Lys**208_NZ (2.6)	**Leu**156_O (2.7) **Lys**214_NZ (3.0)	**Leu**234_O (2.7) **Lys**288_NZ (3.0)	**Leu**163_O (2.7) **Lys**220_NZ (2.9)	**Leu**143_O (2.6) **Lys**204_NZ (3.0)
	O61 O6' O'P	**Lys**208_NZ (2.8) **Asn**212_ND2 (2.9) *(**Thr**122_OG1)(3.2)* *(**Asp**268_OD2)(3.0)* ***Cys**264_SG (3.4)	**Lys**214_NZ (2.8) **Asn**218_ND2 (3.0) *(**Thr**121_OG1)(2.8)* *(**Asp**274_OD2)(2.8)* ***Cys**270_SG (3.4)	**Lys**288_NZ (2.8) **Asn**292_ND2 (2.9) *(**Thr**199_OG1)(2.9)* *(**Asp**348_OD2)(2.5)* ***Cys**344_SG (3.7)	**Lys**220_NZ (2.8) **Asn**224_ND2 (2.7) *(**Thr**131_OG1)(2.6)* *(**Asp**280_OD2)(2.6)* ***Cys**276_SG (3.6)	**Lys**204_NZ (2.8) **Asn**208_ND2 (2.9) *(**Thr**118_OG1)(2.5)* *(**Asp**264_OD2)(2.6)* ***Ser**260_OG (3.4)
	O62 O'Q	**Glu**151_OE2 (3.1) ***Cys**264_SG (3.3)	**Tyr**10_OH (2.5) **Glu**154_OE1 (2.6) ***Cys**270_SG (3.2)	**Tyr88**_OH (2.5) ***Cys**344_SG (3.5)	N/A	***Ser**260_OG (2.3)

The majority of the protein contacts with the modeled UDP-GlcA reside within the helical subdomain that connects the N- and C-terminal nucleotide binding Rossmann-like domains. A detailed description of all available UGDs hydrogen-bonding network is given on Table 5.5. (see above). A 15-residue stretch of the polypeptide chain (residues 253-269) makes an extensive series of contacts with the sugar nucleotide, from the uracyl moiety and running lengthwise to the glucuronate moiety. This includes contacts from several residues in the active site signature sequence GGXCXXXD (residues 261-268) characteristic of the nucleotide-sugar dehydrogenases (Campbell *et al.* 2000). In contrast to the absence of direct contacts between the enzyme and the adenine moiety of NAD(H) cofactor, a number of direct and water mediated hydrogen bonds were modelled between UgdG and the uracyl moiety of the reaction product. Interestingly, many of these contacts are made by backbone atoms of the protein, which may limit the accommodation of different orientations and types of bases. It might be one of the reasons why UGDs are specific to UDP-sugars. Phe253 stacks against the uracyl ring through van der Waals interactions. Three residues within the signature sequence, Gly261, Cys264, and Phe265, also coordinate to make a contact via the same water molecule to the O2B atom of pyrophosphate group. A fourth signature residue, Asp268, hydrogen bonds the carboxylate group of UDP-GlcA via a water molecule. 3D analysis shows that this is the catalytic water (Ge *et al*. 2004; Easley *et al*. 2007), held in place at the active site and rendering the enzyme ready for catalysis initiation, even if the substrate is not yet present, as in the case of UgdG. A single residue from the N-terminal dinucleotide binding fold, Glu151, may interact through its side-chain with the carboxylate group of the reaction product. This interaction is only possible when no cofactor is bound or as in this case, if the cofactor lacks its nicotimide ring (see above).The strictly conserved Thr122 on the other hand, establishes an indirect contact using the catalytic water molecule, kept in place by Asp268. Arg248 from the other dimer subunit

establishes two hydrogen bonds with the sugar hydroxyls. Others have pointed this residue as being involved in substrate specificity (Campbell *et al.* 2000).

5.4.8. 3D phyletic analysis of the UGD/GMD family

A previous phylogenetic analysis of sequences from UgdG and other 46 members of the UDP-glucose/GDP-mannose dehydrogenase family (31 and 12 eukaryotic and prokaryotic UGDs, respectively, and five prokaryotic GMDs) led to a systematisation of this family into four groups: eukaryotic UGDs group, prokaryotic UGDs groups I and II, and prokaryotic GMDs group (Granja *et al.* 2007). The actual availability of 3D structures from these four groups now allows a 3D structural based phyletic analysis of this NDP-sugar-6-dehydrogenase family (Figure 5.7.).

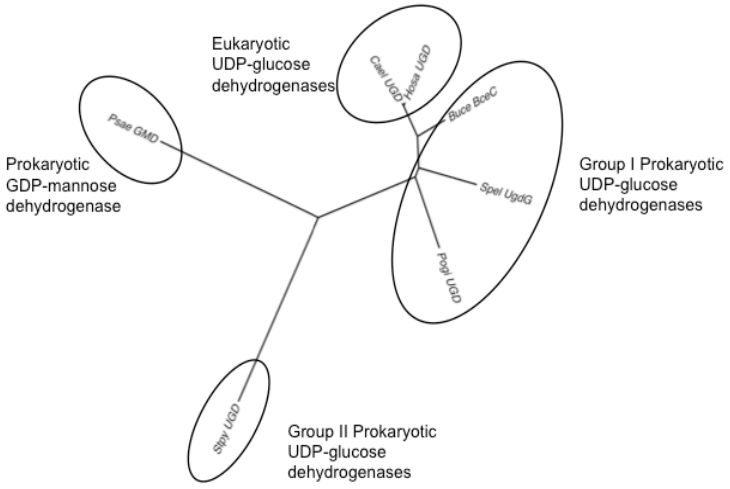


Figure 5.7. *Three-dimensional phyletic dendogram of the UGD/GMD family.*
There is at least one 3D representative for each of the four UGD/GMD sub-families, highlighted in a circle. Structures are identified by the first two letters of the genus followed by the first two letters of the species: Stpy_UGD, PDB entry 1DLJ (Campbell *et al.* 2000); Hosa_UGD, PDB entry 2Q3E (Kavanagh *et al*. 2007, not published); Buce_BceC (see *section 3*.); Spel_UgdG, this structure; Pogi_UGD, PDB entry 3GG2 (Bonnano *et al.* 2009, not published); Psae_GMD, PDB entry 1MV8 (Snook *et al.* 2003).

Similarly to UGDs, the available GMD structure shows two dinucleotide binding Rossmann-like domains, at its N- and C-terminal sequence regions (Snook *et al.* 2003). However, the domains spatial arrangement does not follow the compact monomer fold of UGDs, but they rather adopt the positioning of the complementary domains from the dimer mates in an equivalent UGD, thus enabling an effective overlap of UGDs and GMDs dimers. Such a domain swap in *P. aeruginosa* GMD relative to UGDs results from the backbone angles in the Asp248 to Leu253 loop, the hinge loop between the first two helices of GMD helical region (Snook *et al.* 2003). The GMD dimer has therefore hybrid active sites, with similar contributions from residues of the N- and C-terminal domains of both monomers. A chimeric GMD "monomer" with N- and C-terminal domains from each dimer mate was used in the 3D-structural comparison with UGDs, in order to avoid bias in the phyletic distance calculations due to its different quaternary structure. The resulting dendogram corroborates the 1D structural results of Granja *et al.* 2007, where GMD constitutes a branch of the UGD/GMD family. UgdG belongs to the prokaryotic UGDs group I, closer to the eukaryotic UGDs group than to the other prokaryotic UGDs group II. The prokaryotic group II UGD structure from *S. pyogenes* shows the same overall fold, but is *ca.* 30 residues shorter at the C-terminus. This bends into the protein interior, into the substrate/product binding pocket, and thus limits its exposure to the solvent (Campbell, 2000). It is noteworthy that the other sequences in this group show a high conservation degree, probably indicating similar substrate accessibility. In group I prokaryotic UGDs, although with a longer C-terminus, the substrate/product binding pocket is less occluded. The polypeptide bends to the exterior, with a further helix and adds up to two strands to the central β-sheet. These additional secondary structure elements certainly contribute to the divergence within prokaryotic UGDs and their division into two groups. Only with more structures may these differences be further clarified.

5.5. Acknowledgements

The authors gratefully acknowledge ESRF, Grenoble, for provision of synchrotron radiation. JR is a recipient of a PhD fellowship from Fundação para a Ciência e Tecnologia (FCT), Portugal, SFRH/BD/24216/2005.
This work was partially supported by FCT grants POCTI/BME/38859/2001, POCI/BIO/58401/2004 and PTDC/QUI/67925/2006.

Nitroreductase from *Rhodobacter capsulatus* B10 (NprA)

6. Crystal Structure of NprA, the major nitroreductase from *Rhodobacter capsulatus* B10 bound to acetate

The work here described corresponds to the X-ray crystallographic part of a paper in preparation in collaboration with C. Moreno-Vivián and M. D. Roldán, Uni. Córdoba, Spain, E. Mitchell, ESRF, France and C. Frazão, B. Victor and C. Soares, ITQB-UNL, Portugal.

6.1. Abstract

Rhodobacter capsulatus B10 is a purple nonsulphur photosynthetic bacterium that exhibits a wide variety of metabolic pathways and can be found in a multitude of natural environments. Its major nitroreductase (NR) enzyme can catalyse the reduction of several nitroaromatic compounds, such as 2,4-dinitrophenol, a powerful energetic uncoupler often found as a soil contaminant, making this enzyme a suitable candidate for bioremediation processes.

The *R. capsulatus* B10 NR, termed NprA, belongs to the Type I oxygen-insensitive NRs, and uses both NADH and NADPH as electron donors. Several members of this family have already been characterised, both biochemically and structurally. However, the physiological role(s) of these enzymes, as well as the residues involved in catalysis are still unknown. In the present work, native NprA was expressed, purified and crystallised. Several attempts to also crystallise NprA complexed with different ligands were unsuccessful, thus mapping its catalytic residues using crystallographic techniques alone proved to be an impossible task. Here we report the crystal structure of the *R. capsulatus* B10 NprA enzyme bound to FMN and acetate, at 2.0 Å resolution, solved by the molecular replacement method. Crystals belong to the monoclinic space group C2 and contain two molecules in the asymmetric unit.

6.2. Introduction

The technological development human society has experienced since the 18th century with the Industrial Revolution, has contributed to an improved quality of life. However, this fast industrial rising was accompanied with a massive mobilisation of natural resources and the production of chemical compounds that were inadvertently released and accumulated in the environment, becoming a major environmental concern (Kulkarni & Chaudhari 2007). Most of these xenobiotic compounds persist in the environment because they exhibit structural elements or substituents that are not common in nature, thus limiting their biodegradability and potentially causing a huge impact on living organisms and ecosystems (Esteve-Núñez *et al*. 2001). The most common amongst such compounds are the nitroaromatic or nitroheterocyclic compounds used in the synthesis of explosives, pesticides, herbicides, dyes, pharmaceuticals, etc. They are characterised by the presence of one or more nitro (NO_2) groups on a heterocyclic or aromatic nucleus that confer additional stability, rendering their degradation difficult (Blasco & Castillo 1997; Kulkarni & Chaudhari 2007). Nevertheless, some microorganisms, mostly bacteria, have developed the capacity to metabolise some nitroaromatic compounds, either by aerobic or anaerobic processes. Aerobic degradation occurs mainly through the action of mono- and di-oxygenases and it is the principal degradation mechanism for mononitroaromatic and even for some dinitroaromatic compounds (Roldán *et al*. 1998; Caballero *et al*. 2005a). In the case of polynitroaromatic compounds, the deficiency of electrons in the ring caused by the nitro groups impairs the action of oxygenases, and thus, reduction is favoured, both in aerobic or anaerobic conditions (Johnson & Spain 2003; Kim and Song 2005). This process can occur either by the reduction of the aromatic ring through hydride-Meisenheimer complex formation or by the action of bacterial nitroreductases.

Nitroreductases belong to the family of flavin-coupled dehydrogenases and reductases that are able to reduce a broad range of nitroaromatic compounds as well as quinones and flavins, using both reduced nicotinamide adenine dinucleotide (NADH) or reduced NADPH for catalysis (Haynes *et al*. 2002; Chaignon *et al*. 2006; Choi *et al*. 2008). Enzymatic nitroreduction can be performed either through a one- or two-electron mechanism, generating the nitroso, hydroxylamine and amino derivatives. In the first case, the catalysts can only mediate the reduction of nitroaromatics under anaerobic conditions (Watanabe *et al*. 1998; Lee *at al*. 2008). They are termed oxygen-sensitive or type II nitroreductases, because the formed radicals are easily reoxidised to the parent compounds by O_2 generating superoxide, in the so called "futile cycle". In the latter case, the enzymes catalyse the reduction of nitro groups through the addition of sequential electron pairs from NAD(P)H, and are denominated oxygen-insensitive or type I nitroreductases, because this process does not produce superoxide. Nitroreductases are widely distributed in the bacterial kingdom but although several nitroreductases have been characterised both structurally and biochemically, particularly the minor nitroreductase from *E. coli*, NfnB (Zenno *et al*. 1996b, c; Parkinson *et al*. 2000; Race *et al*. 2002; Johanson *et al*. 2003), the physiological role of these proteins is still an unanswered question. Nevertheless, due to their capacity to reduce many types of polynitrated compounds, their role in detoxification has been well accepted but the potential application of emergent nitroreductases from different bacterial sources remains unexplored (Roldán *et al*. 2008). *Rhodobacter capsulatus* B10 is a purple nonsulphur gram-negative bacterium that exhibits an extensive range of metabolic capacities (Pérez-Reinado *et al*. 2005), including photosynthesis, lithotrophy, and aerobic and anaerobic respiration. This biochemical versatility together with its genetic characteristics allows the bacterium to survive in a variety of habitats, including nitroaromatic contaminated environments (Blasco & Castillo 1997). Genomic studies on this bacterium have shown the presence

of several nitroreductases (Blasco & Castillo 1993) one of which, classified as its major nitroreductase and termed NprA, is highly induced by several aromatic, nitroaromatic and heterocyclic compounds, including benzopyrene, nitrofurazone, the prodrug CB1954, paraquat and 2,4-dinitrophenol, and inhibited by dicoumarol or copper (Pérez-Reinado *et al*. 2008). Our interest in *R. capsulatus* NprA arises from its potential in different biotechnological applications. Here we report the three-dimensional structure of NprA with the bound flavin mononucleotide (FMN) and acetate, a known inhibitor of the type I oxygen-insensitive nitroreductases (Haynes *et al*. 2002).

6.3. Materials and Methods

6.3.1. Expression and purification of *npra* gene from *R. capsulatus* B10

The expression and purification of the major nitroreductase from *R. capsulatus* have been described elsewhere (Pérez-Reinado *et al*. 2005). Briefly, cells of *Escherichia coli* JM109 harbouring the His6-tagged *npra* gene were grown in 2*L* of LB medium supplemented with 100 $\mu g \cdot ml^{-1}$ of ampicillin at 310 K, until an OD_{600nm} of approximately 0.6 was reached. The cells were then induced with 0.3 m*M* isopropyl-*β*-D-thiogalactoside (IPTG) for 3h at 301 K, harvested by centrifugation (10 000*g*, 30 min, 277 K) and the obtained pellets were stored at 253 K, until further use. Cells were resuspended in TIDC Buffer (50 m*M* Tris-HCl pH 8.0, 3 m*M* imidazole, 0.2 m*M* 2,4-dinitrophenol, supplemented with a cocktail of protease inhibitors) and disrupted by sonication (90W for 15s). Unless stated differently, all of the following purification protocol was manually executed at 4°C. Crude cell extract was obtained by centrifugation at 27 000*g* for 40 min, and 1.5 *M* of NaCl and 10% (v/v) of glycerol were added to the collected supernatant. The clarified extract was then applied onto a 1 ml Ni-

NTA affinity column (Qiagen), pre-equilibrated with buffer A (50 m*M* Tris-HCl pH 8.0, 3 m*M* imidazole, 0.2 m*M* 2,4-dinitrophenol, 1.5 *M* NaCl, and 10 % (v/v) glycerol) at a flow rate of approximately 0.5 ml·min^{-1}. The column was washed thoroughly with buffer A, followed by buffer B (50 m*M* Tris-HCl pH 8.0, 3 m*M* imidazole, 0.2 m*M*, 2,4-dinitrophenol, 0.15 *M* NaCl, and 5 % (v/v) of glycerol) to remove any unbound protein, before a gradient of buffer B and buffer C (50 m*M* Tris-HCl pH 8.0, 250 m*M* imidazole, 0.2 m*M* 2,4-dinitrophenol, 0.15 *M* NaCl, 5 % (v/v) of glycerol) was applied. The eluted fractions of NprA were immediately pooled, buffer-exchanged with 50 m*M* Tris-HCl pH 8.0, 75 m*M* NaCl, and 0.2 m*M* 2,4-dinitrophenol, and concentrated using Vivaspin 500 concentrators (Vivascience Ltd, UK) with a 10 kDa cut-off. The composition and purity of the eluted fractions were confirmed by SDS-Page analysis and only homogenous fractions were used subsequently. The protein concentration was determined by absorbance at 280 nm in a NanoDrop ND-1000 Spectrophotometer using an extinction coefficient of 26930 M^{-1}·cm^{-1}. The protein was concentrated in a Vivapore 10/20 concentrator (Vivascience Ltd, UK) to approximately 10 mg·ml^{-1} before storage at 193 K.

6.3.2. Protein crystallisation

Protein crystallisation screens were performed in a Cartesian PixSys 4200 robot (Genomic Solutions, U.K.) using the vapour-diffusion method. 384 sitting drops consisting of 100 nl of protein solution at 10 mg·ml^{-1} plus 100 nl of precipitant solution were equilibrated against 100 µl of precipitant solutions from crystallisation screens Index and Crystal Screens I & II from Hampton Research (Aliso Viejo, USA), JCSG and PACT screens from Qiagen Canada Inc. (Montreal, Canada). A crystalline precipitate was found in various drops from the different screens, but only solution B7 (100 m*M* sodium acetate pH 4.6 and 8% (w/v) polyethylene glycol 4000) from the JCSG screen gave visible

single crystals. Manual reproduction and further optimisation of these conditions to promote crystal growth proved to be unsuccessful, and X-ray diffraction analysis was performed with the micro-crystals obtained in the above condition directly from the robot. Cryoprotection of crystals was achieved by transferring them into a solution of mother liquor with higher precipitant concentration (12% (w/v) of PEG 4K instead of the initial 8%) supplemented with 25% (v/v) glycerol. The same crystallisation screens were also tried with NprA at 10 $mg \cdot ml^{-1}$ incubated (for 30 minutes) with different ligands namely (1) 200 μ*M* of CB1954 prodrug, (2) 200 μ*M* of 6,7-dimethyl-5,6,7,8-tetrahydropterine ($DMPH_4$), (3) 160 μ*M* furazolidone and (4) 0.2 m*M* of reduced nicotinamide adenine dinucleotide (NADH) (Pérez-Reinado *et al*. 2008). Due to their reduced solubility, the first three ligands were prepared in 20% DMSO. All the different ligands produced crystalline material only with solution G2 (100 m*M* HEPES pH 7.5, 22% (w/v) polyacrylic acid 5100 sodium salt and 20 m*M* magnesium chloride) of the JCSG screen, but once again the original conditions were found not to be reproducible by hand. In the end, these tries led only to NprA-NADH plate-like crystals, which were suitable for manipulation and X-ray analysis. Cryoprotection was obtained by quick soaking into a solution of mother liquor supplemented with 200 μ*M* of NADH (to avoid back soaking and release of the ligand) and 25% (v/v) glycerol. All ligands were purchased from Sigma (Sigma-Aldrich, France).

6.3.3. X-ray diffraction analysis and data processing

Cryoprotected native and NADH-bound NprA crystals were flash frozen in liquid nitrogen and diffraction data collected at the European Synchrotron Radiation Facility (ESRF), Grenoble, France, at station ID23-2. Reflection intensities were processed with *MOSFLM* (Leslie 1992), scaled and merged

together with *SCALA,* and reduced to structure factor amplitudes using *TRUNCATE* from the CCP4 Suite (Collaborative Computational Project, Number 4, 1994).

6.3.4. Structural solution and refinement

NprA phases were obtained by molecular replacement using *PHASER* (McCoy *et al*. 2005) and the model of *Enterobacter cloacae* nitroreductase (NTR; PDB code 1KQB; Haynes *et al*. 2002) as the search structure that shares, according to the *CLUSTALW* alignment program (Higgins *et al*. 1994), a sequence identity of 30% with the *R. capsulatus* NprA. *ARPWARP* was used to correct the sequence and extend the docked model (Perrakis *et al*. 2001), which was iteratively completed and corrected on a graphics workstation using *COOT* (Emsley & Cowtan 2004) and structurally refined with *PHENIX.REFINE* (Afonine *et al*. 2005). Whenever persistent positive (mFo-DFc) (Read 1986) density blobs were found, solvent molecules were added as water, acetate or glycerol molecules, based on the particular electron density shape and chemical environment. The progress of the refinement was monitored by the use of R_{free}, where 5% of the reflections were consistently set aside for cross-validation.

6.3.5. Structural analysis of NprA

Molecular stereochemistry was evaluated with *PROCHECK* (Laskowsky *et al*. 1993), intermolecular contacts and crystallographic packing was checked with *PISA* (Krissinel & Henrick 2007), structural comparisons among produced models were performed with *SSM* (Krissinel & Henrick 2004). The CCP4 (CCP4 1994) program *BAVERAGE* was used to determine average atomic displacement parameters (*a.d.p.*). Images were prepared using Pymol (Delano 2008).

6.4. Results and discussion

6.4.1. NprA crystal characterisation

The eluted protein fractions obtained using the described procedures (*section 6.3.1.)* were analysed on an SDS-Page gel. The protein migrates as a single polypeptide with an estimated molecular mass of 25.3 kDa (23.0 kDa of the peptide plus 2.3 kDa of the tag). Only the fractions containing pure NprA (Figure 6.1.) were used in the crystallisation trials.

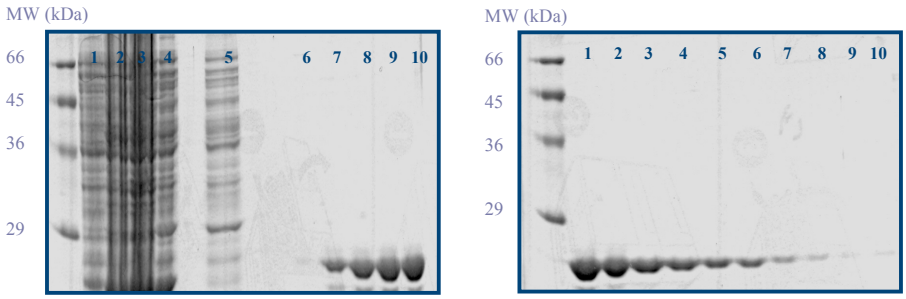


Figure 6.1. *Purification of the NprA native protein*.

SDS-Page analysis of purified NprA using Coomassie blue staining. Fractions containing pure NprA were pooled together and concentrated up 10 $mg \cdot ml^{-1}$. *Legend Gel1*: 1. Before induction; 2. After induction; 3. Insoluble fraction; 4. Soluble fraction; 5. Column flow-through; 6-10. Eluted NprA fractions; *Legend Gel2*: 1-10. Eluted fractions of NprA.

A data set to 2.0 Å resolution (Figure 6.2.) from the best native NprA crystal was collected. The crystal belongs to space group C2, with unit cell dimensions of 103.19, 53.95 and 79.11 Å, and β=101.31. An estimation of the number of molecules in the asymmetric unit (*a.u.*) based on the resolution dependent distribution of the Matthews coefficient (Matthews 1968; Kantardjieff & Rupp 2003) indicates the presence of two molecules in the *a.u.*, with a probability of

99 %, which corresponds to a Matthews Coefficient V_M= 2.35 Å^3/Dalton and a solvent content of 47.6 % (see Table 6.1. for diffraction data statistics).

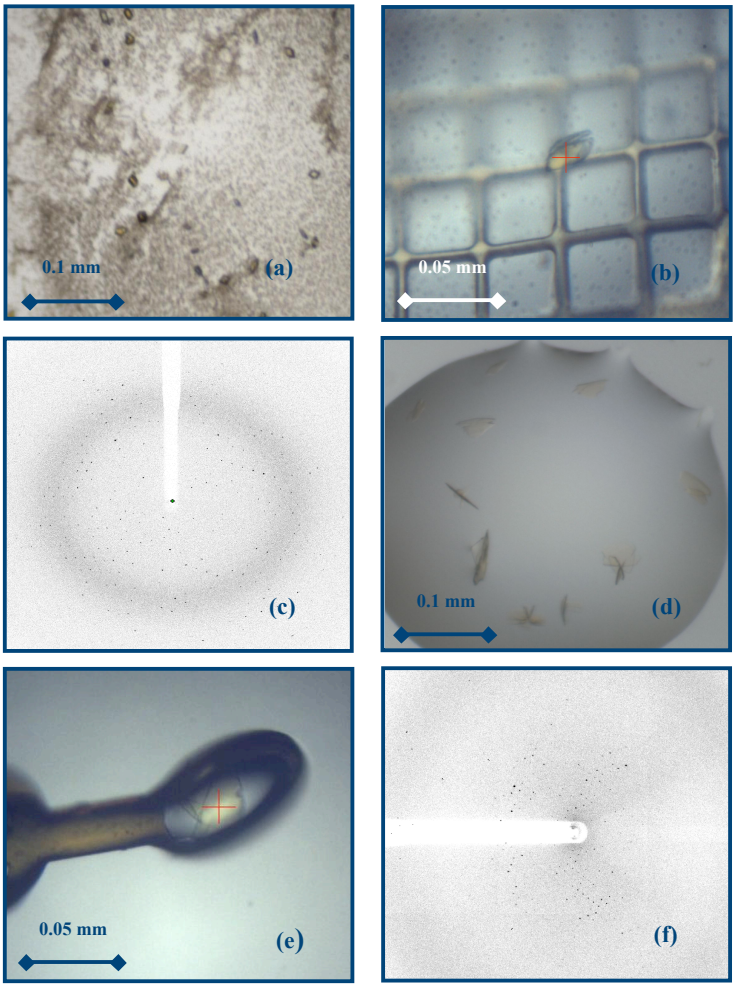


Figure 6.2. *Crystals of NprA and their diffraction pattern.*

(a) NprA forms yellow polyhedral crystals in the nano droplets in the crystallisation robot, after 6h of growing time. (b) NprA crystal mounted on a mesh-loop at the beamline ID23-2 at the ESRF (France) ready for data collection. (c) Diffraction image showing spots up to 2 Å resolution of the NprA crystal taken at the micro focus beamline ID23-2 at the ESRF (France) (d) NprA-NADH crystals in the crystallisation robot droplets after 3 days of growth (e) NprA-NADH crystal mounted on a litho-loop (f) Diffraction pattern showing spots up to 8 Å resolution for the NprA-NADH crystal taken at ID23-2 beamline at the ESRF (France), prior to data collection.

Table 6.1. *Crystal and diffraction data statistics for native NprA.*

Crystal	***NprA_native***
ESRF beamline	ID23-2
Wavelength (Å)	0.873
Resolution (Å)	39.01-2.0 (2.11-2.00)
Space group	C2
Unit cell dimensions a, b, c (Å)	103.19, 53.95, 79.11
β (°)	101.31
No. of measured reflections	115101 (16425)
No. of unique reflections	29030 (4165)
Reflections redundancy	4.0 (3.9)
<I>/<σ (I)>	5.0 (1.5)
R_{pim} [a]	0.079 (0.29)
R_{rim} [b]	0.160 (0.58)
R_{sym} [c]	0.138 (0.50)
Completeness (%)	100.0 (100.0)
Mosaicity (º)	0.7
V_M (Å^{3}Da^{-1})	2.35
Solvent content (%)	47.6
Wilson B (Å^2)	13
Estimated no. of molecules in ASU	2

Data within parenthesis refer to the outer resolution shell; [a] $R_{pim} = \Sigma_h [1/(N-1)]^{1/2} \Sigma_i |I_i(h) - \langle I(h) \rangle| / \Sigma_h \Sigma_i I_i(h)$, where N is the data redundancy, I is the observed intensity and $\langle I \rangle$ is the average intensity of multiple observations from symmetry-related reflections. It is an indicator of the precision of the final merged and averaged data set; [b] $R_{rim} = R_{meas} = \Sigma_\eta [N/(N-1)]^{1/2} \Sigma_i |I_i(h) - \langle I(h) \rangle| / \Sigma_h \Sigma_i I_i(h)$, where N is the data redundancy, I is the observed intensity and $\langle I \rangle$ is the average intensity of multiple observations of symmetry-related reflections. It is an indicator of the average spread of the individual measurements; [c] $R_{sym} = \Sigma_h \Sigma_i |I_i(h) - \langle I(h) \rangle| / |\Sigma_h \Sigma_i I_i(h)$, where I is the observed intensity and $\langle I \rangle$ is the average intensity of multiple observations from symmetry-related reflections.

Regarding the fragile NADH-bound plate-like crystals (see above Figure 6.2.), diffraction data was very poor (≈ 8 Å resolution), and no further X-ray analysis was performed.

6.4.2. Phasing and structural refinement of NprA

A molecular replacement search by *PHASER* (McCoy *et al*. 2005) using the 30% sequence identity model of NTR from *Enterobacter cloacae* (PDB code 1KQB; Haynes *et al*. 2002) as target structure, led to final translational z-scores of 5.0 and 18.3, for each of the two molecules.

Readily interpretable electron density maps were produced, however due to the initial treatment of the *E. cloacae* NTR model with *CHAINSAW*, truncating all non-conserved residues to alanines, the molecular replacement solution lacked most of the side-chains. The MR solution was visually inspected using *COOT* and some regions of the structure were missing. The initial NprA model was then fed into the automated building program *ARPWARP* (Perrakis *et al*. 2001), extending the protein chain to 374 (out of 420) residues, all correctly docked in the sequence. The protein and solvent structure were examined at a graphics workstation against their σ_A-weighted maps (Read 1986), and the model completed and improved to fit their features. Consecutive cycles of structure refinement and manual editing were repeated until convergence of R_{work}/R_{free} to 18.3/20.6 %. Using all data the final R factor is 18.0 %. The stereochemical analysis of the refined model of native NprA, using *PROCHECK* (Laskowsky *et al*. 1993) revealed 99.3%, 0.7%, 0.0% and 0.0% of residues in the most favourable, additional allowed, generously allowed and disallowed regions of the Ramachandran plot, respectively.

Table 6.2. *Refinement and model quality statistics for native NprA*.

Parameter	***NprA (Native)***
R_{work} (%)	18.3
R_{free} (%)	20.6
R (%)	18.0
Average <*a.d.p.*> protein (Å^2)	14.3
Flavin mononucleotide (FMN) <*a.d.p.*> (Å^2)	8.6
Number of residues	425
Solvent water molecules	234
Ligand molecules (FMN, acetate, glycerol)	5
Number of atoms (monomer A)	1663
Number of atoms (monomer B)	1618
Results for monomers (A and B)	
Average <*a.d.p.*> protein (Å^2)	14.1, 14.4
Average <*a.d.p.*> main-chain (Å^2)	12.1, 12.6
Average <*a.d.p.*> side-chain (Å^2)	16.3, 16.4
Average <*a.d.p.*> Flavin mononucleotide (Å^2)	8.1, 9.0
Average <*a.d.p.*> acetate (Å^2)	17.8, 17.6
Deviations from the ideal	
Bond distances (Å)	1.549
Bond angles (º)	0.018

6.4.3. NprA crystal contents

Even though a methionine residue is reported as the first residue in the NprA sequence, the DNA manipulations and the cloning of the *npra* gene into the pQE32 vector (Pérez-Reinado *et al*. 2005) have created a linker of about 10 residues between the 6x histidine tag and the first protein residue, modifying it to a glycine. This linker has the sequence "GIFENGLYFQ", however, only the last five residues (GLYFQ) of this long linker are visible in the electron density map, and only in one of the monomers. All remaining 210 residues could be found in the electron density for each monomer, although some side-chains, not entirely visible in the electron density maps needed to be modelled stereochemically. They belonged to hydrophilic residues exposed to the solvent, and their refined *a.d.p.* values reflect their disordered behaviour. Each monomer of NprA binds a flavin mononucleotide (FMN) molecule, and an acetate ion was found near the prosthetic group binding site. The final model includes a total of 425 amino acid residues, 2 FMN cofactors and 2 acetate ions as ligands, 234 solvent molecules modelled as waters and 1 molecule as glycerol (see Table 6.2. above for further details).

6.4.4. NprA overall structure

Although only low to moderate sequence identities exist between *Rhodobacter capsulatus* NprA and the other members of the family with available tertiary structure, being 30% for the closest related structure, the *E. cloacae* NTR (Haynes *et al*. 2002), the three-dimensional superposition of NprA with other nitroreductase (NR) structures shows a conserved overall fold. Both the *E. cloacae* NTR and the major nitroreductase from *R. capsulatus* belong to group B of the Type I oxygen-insensitive NRs, which reduce a wide variety of nitroaromatic compounds, and use NADH and NADPH as electron donors.

In effect, the ability to use both molecules distinguishes these proteins from their homologues of group A, represented by the *E. coli* NfsA (PDB entry 1F5V; Kobori *et al*. 2001), which use only NADPH as the electron donor, even though their overall fold is essentially the same (but see below).

NprA is a homodimer of about 46 kDa. Each monomer exhibits an α+β fold with a 3-layer αβα-sandwich architecture common to the group B NRs, such as the representative member NfnB from *E. coli*, the NTR from *E. cloacae* or the FRaseI from *V. fischeri* (PDB entries 1DS7, 1KQB and 1VFR, respectively; Koike *et al*. 1998; Parkinson *et al*. 2000; Haynes *et al*. 2002). However, the domain organisation of NR family members has been differently characterised. Parkinson and co-workers have considered the *E. coli* NfnB as a three-domain protein, with a major central domain, a smaller flexible α-helical domain, and a C-terminal region tightly involved with the second monomer (Parkinson *et al*. 2002). NR dimerisation involves residues from different regions of the peptide chains (discussed below) other than the C-terminal residues as referred for NfnB. In the classification used for the *E. cloacae* NTR (Haynes *et al*. 2002) and the *V. fischeri* FRaseI (Koike *et al*. 1998), both authors consider a two-domain protein with a major central domain and a minor flexible domain. In the NprA structure described here we will follow the latter classification where each of the 210 residue monomers is composed of two domains (Figure 6.3.): a major central core formed by a twisted β-sheet surrounded on one side by two large α-helices, α2 and α6, and on the other side by two smaller α-helices, α3 and α7; and a minor and flexible domain, comprising two α-helices, α4-α5 located at one edge of the major domain β-sheet. This minor domain is thought to be important for substrate specificity (Koike *et al*. 1998; Parkinson *et al*. 2000).

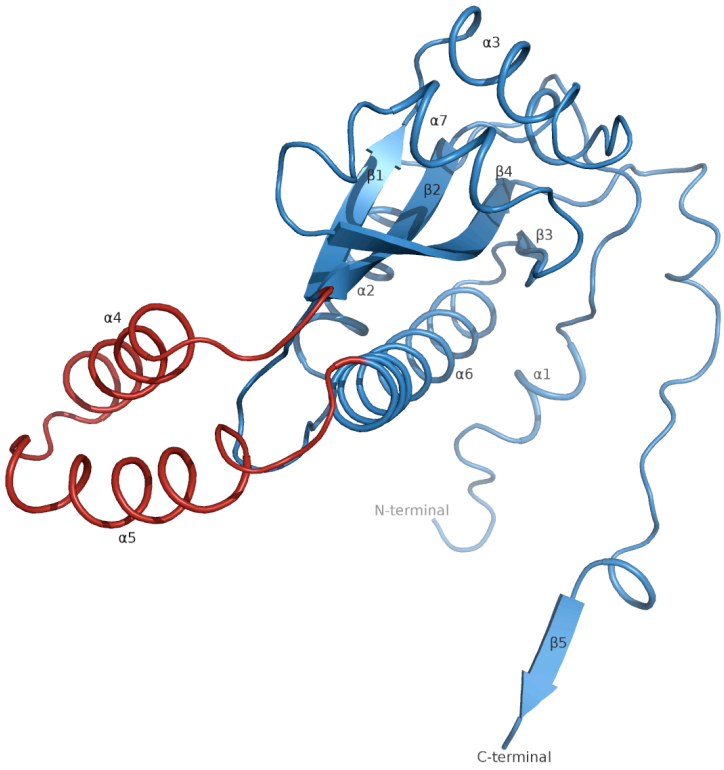


Figure 6.3. *The monomeric structure of R. capsulatus B10 nitroreductase (NprA).* NprA exhibits a three-layer (αβα) sandwich architecture. It comprises a major core (blue) formed by a central four-stranded anti-parallel twisted β-sheet surrounded by helical elements, two large α-helices on one side and two smaller ones on the other side; and a minor flexible domain (red) composed by two α-helices, thought to be involved in substrate specificity. Secondary structure elements are labelled numerically for both α-helical and β-strand components.

When the protein dimerises, each minor domain blocks the access to the FMN of the dimer partner (see Figure 6.4. a, b). The minor helical domain is therefore responsible for the size of the substrate binding pocket, which explains in part, why NRs exhibit different substrate specificities and functions.

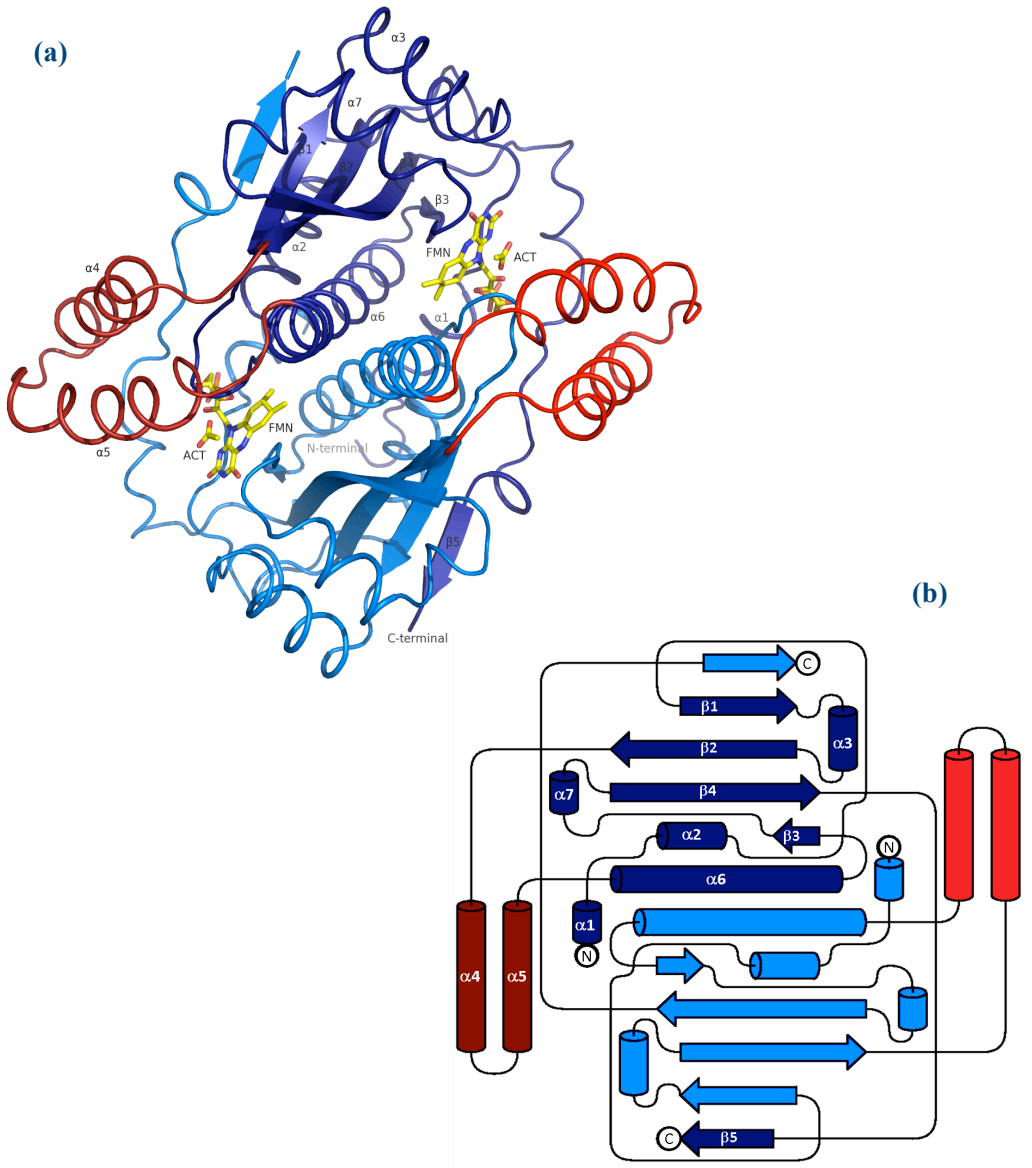


Figure 6.4. *Three-dimensional structure of the dimeric NprA from R. capsulatus B10.*
(a) Cartoon representation of the physiological dimer of NprA with bound flavin mononucleotide and acetate ion (in stick representation and colour coded yellow for carbon, blue for nitrogen, red for oxygen and orange for phosphorus). The major domain is represented in blue and the smaller domain in red (dark and light colours are used for each monomer, respectively). Both polypeptide termini of the dark colours monomer are labelled. (b) Topology diagram of the NprA dimer, with α-helices represented as cylinders and β-strands as arrows (coloured as in a). Image prepared with TopDraw (Bond 2003).

Structural analysis of NprA shows an extensive interaction between the two monomers, involving their major and minor domains and both cofactors, pointing the dimer as the physiological unit, as reported for the other NRs.

These interactions involve residues from the large central helix (α6), the α2 helix, and interactions within the central mixed β-sheet. Similar to the other group B structures, the central β-sheet contains four anti-parallel strands, β1 to β4, joined to a fifth strand, β5, from the C-terminal part of the other dimer-mate, which runs parallel to strand β1, thus forming a five-stranded sheet (see Figure 6.4. above). When compared to the other NRs, NprA shows a short β3-strand (residues 151-153) whilst other structures have β3 with five residues. An analysis of the dimer structure indicates an interface area of 3180Å^2 (26% of the total area of the monomer) corroborating the dimeric nature of NprA. It involves 81 residues (39%) of each monomer in a total of 66 polar interactions of which 53 hydrogen bonds and 13 salt bridges. The interface spans throughout the molecule, maintaining the enzyme as a compact dimeric unit, where N- and C-terminal residues are in close contact with the other monomer.

The overall fold of NprA is similar to that of NfsA from *E. coli* (PDB entry 1F5V; Kobori *et al*. 2001), representative of Group A NRs.

The most obvious difference between them is in the small domain (Figure 6.5. a, b). In NprA it consists of helices α4 and α5 involving 30 residues, located between β2 and α6 of the major core domain, while in NfsA it consists of 16 residues that include helix (αE), located between β2 and αF, of its major domain. However, 60 extra residues at the C-terminus of NfsA create an equivalent region (helices I and J), which superposes with the smaller domain of NprA second monomer (Figure 6.5. b). These differences make the substrate-binding pockets of the two NRs different from each other.

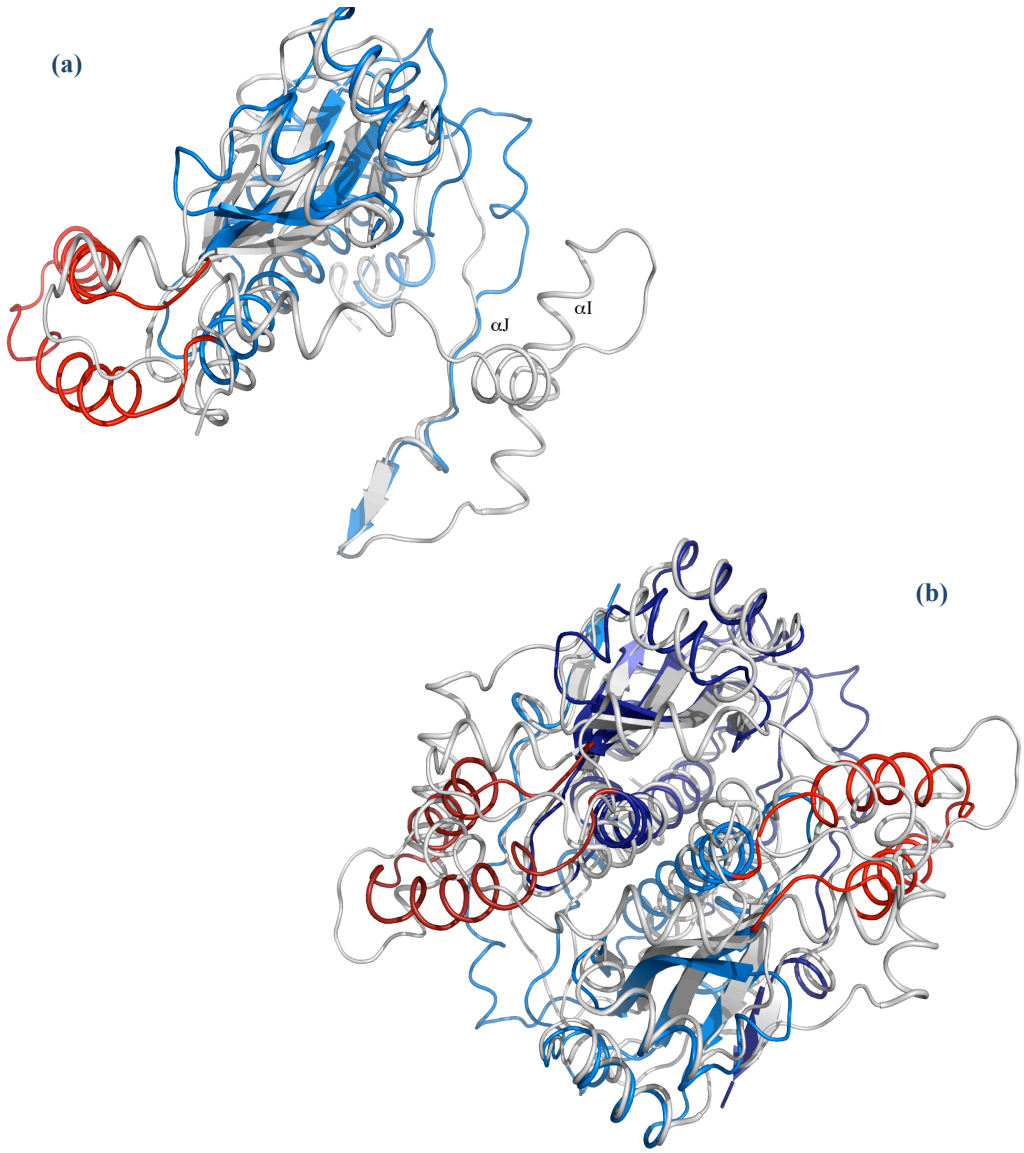


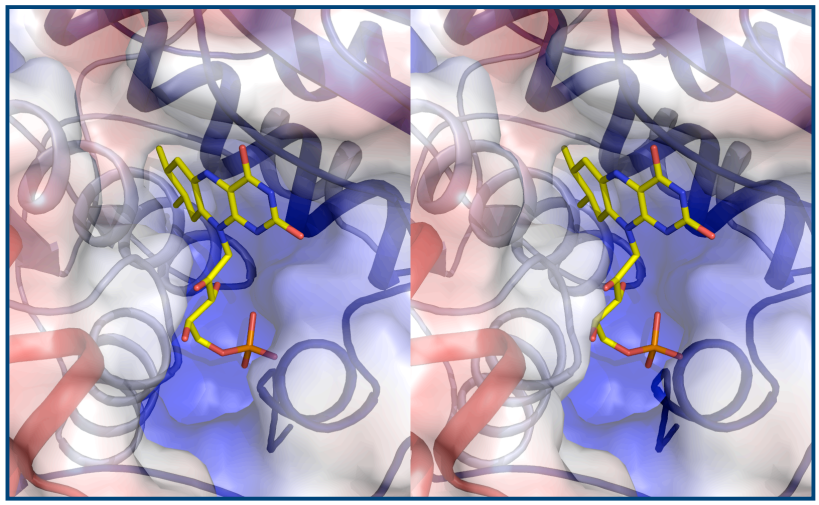
Figure 6.5. *The 3D superposition of NprA and NfsA Type I nitroreductases*.
The two enzymes belong to different families of the Type I NRs. NfsA (PDB entry 1F5V; Kobori *et al*. 2001) uses only NADP(H) as electron donor and is a member of Group A, whilst NprA uses both NADPH or NADH and belongs to the Group B (see *Thesis Background* for more information). (a) 3D superposition of NprA and NfsA monomers (coloured as in panel. (b) 3D superposition of the NprA and NfsA dimers. Macromolecules in cartoon representation coloured blue and red for the major and minor domains of NprA (lighter colours for the dimer partner) and grey for NfsA.

While in the NprA, the smaller domain of one monomer hinders the access to the cofactor of the other subunit and models the substrate binding cavity (see Figure 6.4. and below), in NfsA, the cofactor and substrate binding pockets are modelled by a single peptide chain. Moreover, NprA can use either NADH or NADPH as an electron donor (Perèz-Reinado *et al*. 2005), while NfsA only uses NADPH (Kobori *et al*. 1999). Interestingly, another type of nitroreductase has been identified in bacteria, namely the hypothetical nitroreductase from *Bacillus cereus* (PDB entry 1YWQ, Zhang *et al*. to be published) or the minimal ydjA from *E. coli* (PDB code 3BM1; Choi *et al*. 2008; see Figure 1.16 in *section 1.2.* for further details). These, as yet biochemically uncharacterised type I oxygen-insensitive nitroreductases, have the same fold as the Group B NRs, except for the absence of the minor helical domain. Because these different secondary structure elements are located near the FMN cofactor and substrate sites, in both Group A and Group B, it is probably the explanation for the different substrate specificity, but also to the different biological roles attributed to the several members of the NR family. Nevertheless, the similar overall fold reinforces the evolutionary relationship between the different group members and their origin in a common ancestral flavoprotein.

6.4.5. The flavin mononucleotide (FMN) binding pocket

Each prosthetic group is located in a deep pocket at the interface of the NprA dimer, where they are firmly held in place by interactions with residues from both monomers (Figure 6.6). The flavin stacks against β3, at one edge of the central β-sheet and the phosphate interacts with the residues from the loop following α1. Residues from the other monomer, between helix α2 and strand β1, cap the other side of the flavin and create an interface crevice where the latter is nestled.

(a)



(b)

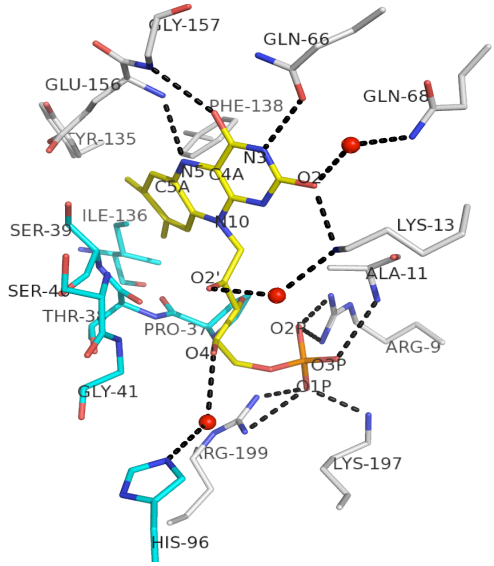


Figure 6.6. *The flavin mononucleotide binding pocket of NprA.*

(a) Cross-eyes stereo representation of the FMN (with bonds as sticks coloured in yellow for carbon, blue for nitrogen, red for oxygen and orange for phosphorous) bound to the NprA dimer represented as a cartoon with a transparent representation of its solvent accessible surface mapped with a calculated electrostatic potential. NprA major and minor domains are coloured in dark blue and red respectively, and the other subunit is represented in the correspondent light colours. (b) FMN hydrogen-bonding interactions with peptide and solvent molecules (represented in black dashed lines). Solvent water molecules represented as red spheres. Residues are represented as sticks with carbon atoms in grey for one monomer and cyan for the dimer mate, nitrogen in blue and oxygen in red.

The cofactor exhibits a bent conformation characteristic of the NR family (Haynes *et al*. 2002). This distortion of the isoalloxazine ring has been termed as the "butterfly conformation" to indicate a bending about the N5–N10 axis (Johansson *et al.* 2003). Only a few residues amongst those involved in FMN interactions are conserved, although the protein-cofactor interactions are similar to the NR Type I family. In the NprA structure there are a total of four hydrogen bonds between the isoalloxazine ring and the protein although only Gly157 is conserved. Its main-chain amine forms a hydrogen bond with O4 from the ring system (see above Figure 6.6.). The Glu156 main-chain amine interacts with N5, which is thought to be the site for the hydride transfer between the flavin and NAD(P)H (Koike *et al*. 1998; Johansson *et al*. 2003). In the FMN reduced state N5 is protonated and therefore the interaction with the main-chain amine of Glu156 cannot occur. Studies on different oxidation states of NfnB and NTR show an increase of the "butterfly bending" effect upon reduction, with an increase of the distance between the amine and N5, which moves out of the C4, N5 and C5 plane (see Figure 6.6. b; Johansson *et al*. 2003). The Gln66 OE1 establishes a hydrogen bond with the N3 atom. In the *V. fischeri* FRaseI or the *E. cloacae* NTR structures, this residue is substituted by an asparagine (Asn73 or Asn71, respectively) while in the *Streptococcus pyogenes* structure (PDB code 2HAY; Kim *et al*. 2006, to be published) the residue at this position is a glycine (Gly71). A 3D superposition of these structures shows that the regions where these residues are located exhibit secondary structure variability, although the prosthetic groups of each superpose closely (see Figure 6.7.). Whilst in the *S. pyogenes* structure the main-chain interacts directly with FMN, the corresponding interaction in FRaseI and NTR is obtained via the side-chain of an Asn residue in an α-helix, and in the case of NprA the interaction comes from a loop located further away from the FMN, but using a residue with a longer side-chain, Gln66.

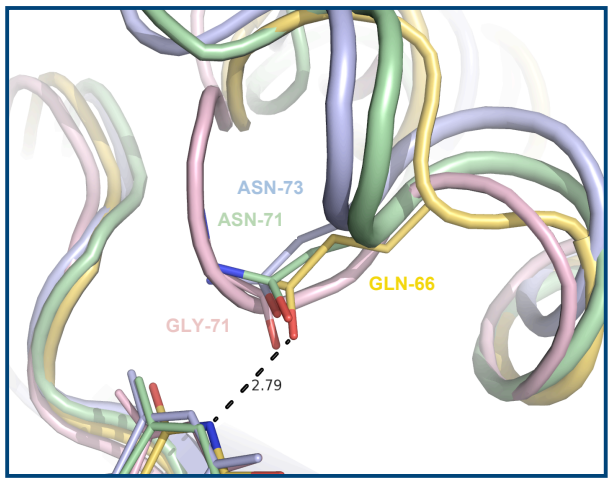


Figure 6.7. *Variability of the peptide region harbouring the interaction with FMN N3 atom*. The 3D-structural superposition of *S. pyogenes* NR (pink) (PDB code 2HAY; Kim *et al*. 2006, to be published), *V. fisherii* FRaseI (light blue) (PDB code 1VFR; Koike *et al*. 1998), *E. cloacae* NTR (green) (PDB code 1KQB; Haynes *et al*. 2002) and *R. capsulatus* NprA (in yellow) show the secondary structure variability of the region harbouring the conserved interaction between the peptide chain and the N3 atom from the cofactor. The NprA cofactor and Glu66 are represented in sticks coloured in yellow for carbons, blue for nitrogens and red for oxygen. The remaining cofactors are coloured according to the respective peptide structure.

The O2 atom from the flavin interacts with NZ from Lys13. The positive charge at this position seems to be conserved throughout almost all the family members, with either an Arg or a Lys residue at this position. The exceptions are the putative nitroreductase ydfN from *B. subtilis* (PDB code 3BEM; Joint Centre for Structural Genomics, not published) and the uncharacterised nitroreductase ydjA from *E. coli* (PDB code 3BM1; Choi *et al*. 2008) where a serine can be found at this position, and the putative nitroreductase from *Streptococcus mutans* UA159 (PDB code 2IFA; Forouhar *et al*., not published) with a tyrosine. The O2 atom is also involved in two water-mediated hydrogen bonds with OE1 of Gln66 and NE2 of Gln68, respectively. Whilst the polar

region of the isoalloxazine ring is accessible to the solvent, the benzenoid portion of the ring system is buried in a hydrophobic patch composed of residues from both monomers (Figure 6.5. b), namely Tyr135, Phe138 and Pro154 from the monomer where the FMN docks, and Ser39 and Ile136 from the monomer that caps the other side of the flavin. The Ser39 side-chain lies above the central ring of the flavin, while the other residues establish van der Waals interactions with the apolar region of the FMN. Although the ribityl moiety of FMN shows no interactions with protein residues, but solely with solvent water molecules, it is well defined in the electron density as both its ends are firmly docked in the structure. The O2' and O4' hydroxyl groups interact indirectly with Lys13 NZ and His96 NE2 from the dimer mate through two ordered water molecules. The same is not true for the phosphate group, which interacts extensively with the protein residues side-chains and main-chain amines. O1P atom establishes hydrogen bonds with Lys197 NZ and NH1 and NH2 from Arg199. O2P forms hydrogen bonds with NH1 and NH2 from the strictly conserved Arg9, while Ala11 donates its main-chain amine to hydrogen bond with O3P from the pyrophosphate moiety. In fact, in Group B NRs one observes a conservation of the positively charged network, although the involved residues are not always sequentially homologous.

6.4.6. The acetate binding pocket

During the structure refinement and completion steps, triangular shaped features of positive density were detected above the isoalloxazine ring of each FMN (Figure 6.8). They were modelled as acetate ions in accordance with the crystallisation media (with an acetate concentration *ca*. 100 m*M*) and to the shape and size of the electron density. Upon refinement, their *a.d.p.* values converged to about 18 $Å^2$ (see Table 6.2. above), in accordance with the average value for other solvent molecules. The acetate ions are held in place by

hydrophobic interactions with the isoalloxazine rings (at *ca.* 3.3 Å distance) and by hydrogen interactions with the ribityl O2' hydroxyl and with the main-chain amine of residue Ser40 (from the dimer mate).

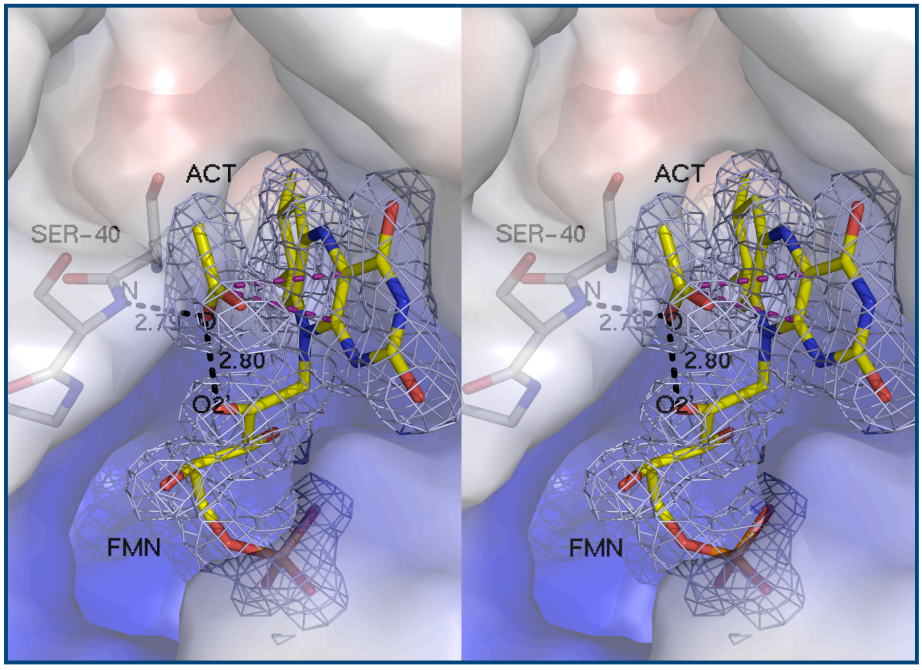


Figure 6.8. *Cross-eye stereo view of NprA substrate binding pocket with bound acetate.* The small inhibitor molecule binds directly above C4 and C10 atoms of the cofactor's central ring at approximately 3.3 Å establishing van der Waals interactions (magenta dashes). It interacts with both the cofactor and Ser40 main-chain amine from the dimer mate. NprA represented by its solvent accessible surface mapped with its electrostatic potential. Residues and ligands are represented as sticks, in red for oxygen, blue for nitrogen, orange for phosphate and grey or yellow for carbons in the peptide or ligand, respectively. The electron density map (2mFo-DFc) represented as light blue mesh contoured at 1.5σ.

Acetate is a known inhibitor of Type I nitroreductases as it is capable of acting competitively with the substrate NAD(P)H for the first half reaction, the flavin reduction, and uncompetitively for the second part of the reaction, the reduction

of the nitroaromatic (Koder *et al*. 1998; Haynes *et al*. 2002). Both the NTR from *E. cloacae* and *E. coli* NfnB structures have been determined in the presence of this small ligand (PDB code Haynes *et al*. 2002; PDB code 1YLR; Race *et al*. 2005). Other NR structures, determined in the presence of different ligands, helped to understand the inhibition phenomenon, especially the case of the NfnB from *E. coli*, bound to nicotinic acid (PDB code 1ICR, Lovering *et al*. 2001), a substrate analogue for the first half of the reaction, with its carboxylic group at exactly the same position as the carboxylate of the small organic inhibitor. According to Haynes and co-workers the binding of carboxylate groups near the positive charges of NfnB Lys14 and Lys74, causes electrostatic interactions with these residues and may be the principle reason for the binding of these inhibitors (Haynes *et al*. 2002). In the NprA structure Lys14 and Gln68 occupy these equivalent positions.

A 3D structural comparison of *E. coli* and *E. cloacae* NRs against NprA shows a bigger substrate cavity for the latter, with wider putative access routes (Figure 6.9. a). It is noteworthy, that in NprA one does not find a residue homologous to Phe124 of *E. coli* and *E. cloacae* NRs, which was found stacking with aromatic rings of different substrates (Figure 6.9. b; Parkinson *et al*. 2000; Lovering *et al*. 2001; Haynes *et al*. 2002). Due to the broad specificity of substrates and the low amino acid sequence conservation of Type I nitroreductases, it is still not clear what residues are relevant in the electron transfer with FMN, and what is the physiological role(s) of these proteins in their biological systems. Several attempts were carried out in order to elucidate the residues involved both in substrate binding and electron transfer. NprA was incubated with different ligands, including the 6,7-dimethyl-5,6,7,8-tetrahydropterine ($DMPH_4$), suggested as *in-vivo* product analogue, as NprA seems to have dihydropterine reductase activity (Péréz-Reinado *et al*. 2008).

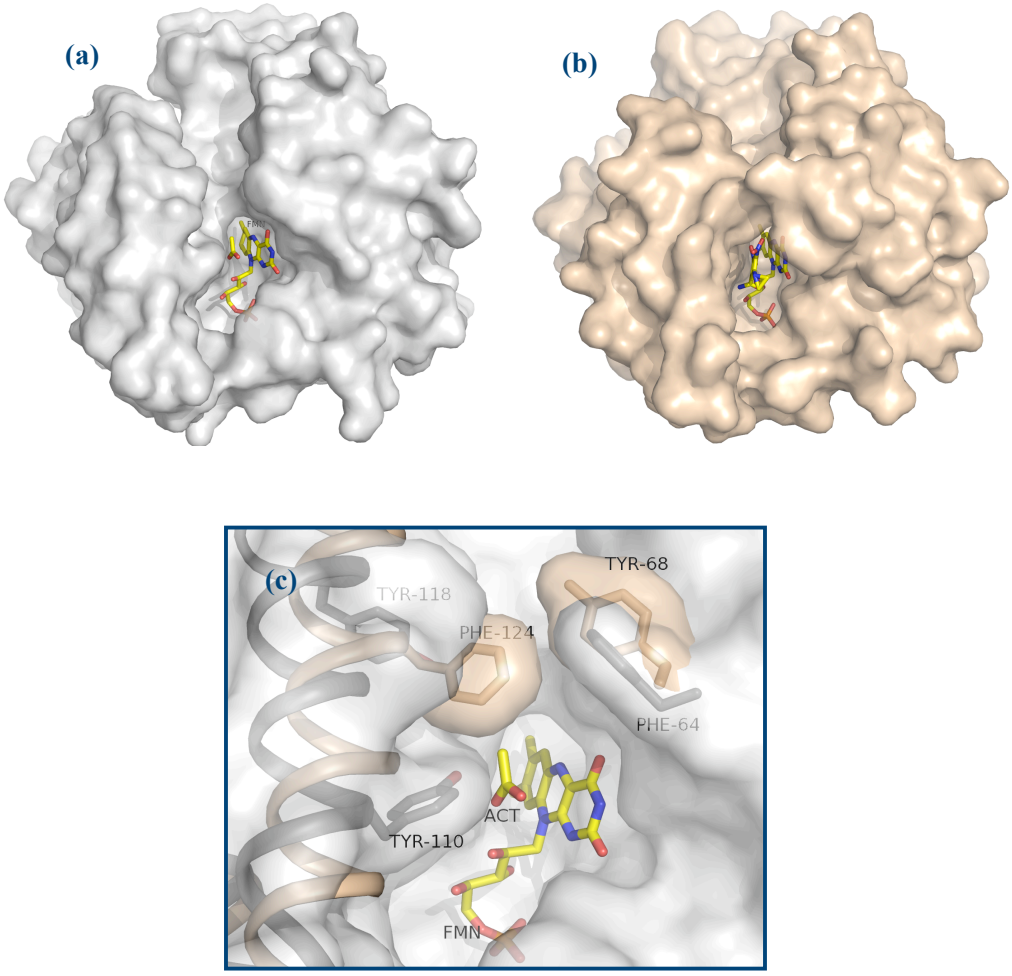


Figure 6.9. *The substrate cavities of NprA and E. coli NfnB.*

(a) NprA solvent accessible surface with docked FMN and acetate represented as sticks coloured in yellow for carbon, blue for nitrogen, red for oxygen and orange for phosphorus. (b) NfnB with docked FMN and CB1954 molecules, represented as in A. (c) Zoomed view of NprA and NfnB superposition, represented as in A and B, showing only NprA ligands. In NprA, the small domain (grey cartoon) shows an open conformation relative to NfnB (orange cartoon). In NfnB, Tyr68 and Phe124 define the substrate binding pocket. While Tyr68 of NfnB seems homologous to NprA Phe64, Phe124, considered an important residue in NfnB selectivity, has no apparent homologue in NprA. However, other aromatic residues, like Tyr110 and Tyr118, may play its role.

Crystallisation trials were assayed with the Nanodrop Crystallisation Robot in order to obtain crystals of NprA complexes with different ligands, using several crystallisation screens (see *section 6.3.2.*). Only plate-like crystals of NprA-NADH could be obtained. They were very fragile, diffracted poorly up to 8 Å resolution (Figure 6.2. d, e, f) and were rather sensitive to radiation damage. Several attempts were tried to reproduce and optimise these crystals at micro litre scale, but they remained unsuccessful, as found with the native NprA crystals.

6.5. Acknowledgements

The authors would like to thank the staff at the High Throughput Crystallisation Laboratory at EMBL and at the European Synchrotron Radiation Facility (ESRF) in Grenoble, for the use of facilities and synchrotron radiation, respectively. Joana Rocha (JR) acknowledges the PhD fellowship SFRH/BD/24216/2005, from Fundação para a Ciência e Tecnologia (FCT). This work was funded by PTDC/QUI/67925/2006.

Eva Pérez-Reinado (EPR) at the Biochemistry and Molecular Biology Department, University of Córdoba did the cloning of the *npra* gene. The protein expression and purification was done by JR at Instituto de Tecnologia Química e Biológica (ITQB) following EPR experience and protocols. JR at the ESRF did the crystallisation, data collection, phase solution and structural determination and analysis.

General Conclusions and Future Perspectives

7. General Conclusions and future perspectives

The work described in this thesis brings new information to the research community in both structural determination of two uridine-5'-diphospho-glucose dehydrogenases (UGDs) from *Burkholderia cepacia* IST 408 and *Sphingomonas elodea* ATCC 31461, as well as new insights in their enzymatic mechanism, revealing the role of a conserved residue. Additionally, a Type I nitroreductase from *Rhodobacter capsulatus* B10, an enzyme with potential biotechnological applications, was also structurally characterised. The key results from four years of scientific research are summarised below.

7.1. UDP-Glucose Dehydrogenase from *Burkholderia cepacia* IST 408 (BceC)

BceC was produced and crystallised, and its native crystal structure was determined. Phases were obtained by molecular replacement (MR) using the homologous structure of *Sphingomonas elodea* UGD as the search model. The BceC crystal structure was refined to 1.75 Å resolution. BceC crystals belong to orthorhombic space group $P2_12_12_1$ and exhibit 4 molecules of the binary complex BceC:UDP-GlcA in the asymmetric unit. BceC monomer has dimensions of approximately 75, 45 and 35 Å, and it is composed of two domains with α/β topology, connected by a four-helical bundle. Both N- and C-terminal domains exhibit dinucleotide-binding Rossmann-like folds in a 3-layer αβα sandwich architecture, suggesting possible gene duplication in its origin. Curiously, whilst the N-terminal domain is used for cofactor (NAD^+) binding, the C-terminal domain is not conventionally used for nucleotide-sugar binding. Instead, the substrate (UDP-Glc) binds in a cavity of the connecting α-helical

subdomain. This contains both the catalytic site and the homodimerisation responsible region. These characteristics are common to all known structures of the UGD/GMD family. In fact, all its members are known as active dimers, evidenced also in BceC by *in vitro* studies and the crystal packing analysis.

A detailed 3D structural comparison of known UGDs shows that protein-ligand(s) interactions are highly conserved in substrate and cofactor binding pockets. Similarly to *P. gingivalis* UGD, BceC shows an extra hydrogen bond between UDP-GlcA and its Tyr10 hydroxyl group. Tyr10 belongs to the N-terminal Rossmann fold glycine rich signature motif, which in UGD/GMD family includes a further conserved tyrosine, GXG**Y**XG. Except for *P. gingivalis* UGD, all other known UGD structures were obtained with a docked cofactor, which localisation necessarily obviated such an interaction due to a stereochemical overlap. A set of Tyr10 BceC mutants were produced, Y10F, Y10S and Y10K, of which the two latter could be crystallised, and their structures determined. Enzymatic studies of the mutants showed them to be almost inactive, thus highlighting the Tyr10 catalytic importance.

In spite of the several literature available studies, with mutagenic and structural work on the enzymatic mechanism of the UGD/GMD family, the residue(s) involved in the final catalytic step, the hydrolysis of a thioester-intermediate, have as yet remained elusive. The BceC structural and mutagenic results presented in this Thesis show that Tyr10 functions as the catalytic, final proton conveyer from the aqueous system to a forming thiolate group, upon hydrolysis of the intermediate thioester, which results in UDP-GlcA.To further confirm this thesis, the kinetic analysis of BceC Y10E mutant, as well as mass spectrometry assays of the existing mutants are desirable. It is expected that Y10E will still enhance hydrolysis but, due to its negative charge, with efficiency lower than that of native BceC. Additionally, mass spectrometry studies of the almost inactive mutants may lead to identification of the thioester

intermediate covalently bound to BceC, if incubated with the reaction substrates.

Another purpose of the present work was the comparison of BceC against human UGD, in order to explore potential medicinal applications. However, the high similarity of both structures may render difficult the design of a potentially selective inhibitor. Nevertheless, the insight into BceC structure and the explanation of its complete catalytic mechanism (and consequently, of the whole UGD/GMD family) should at least help and facilitate efforts towards the design of potential antibiotics based on structure-based inhibitors, to specifically combat this pathogen by hindering its capacity for cepacian production and biofilm formation.

7.2. UDP-Glucose Dehydrogenase from *Sphingomonas elodea* ATCC 31461 (UgdG)

Both native and seleno-methionine derivatised UgdGs were produced and crystallised, and diffraction data of their crystals were collected. The selenium derivative data enabled the crystal phasing at 3.4 Å resolution by single wavelength anomalous dispersion (SAD). The resulting structure was used in molecular replacement (MR) of the native crystals, which led to the crystallographic structure of UgdG at 2.37 Å resolution.

UgdG is a member of the UGD/GMD family, which includes NAD^+ dependent nucleotide sugar 6-dehydrogenases that convert alcohols into carboxylic acids. It is a macromolecule of 438 amino acids, where N- and C-terminal halves show analogous tertiary structures but irrelevant sequence homologies, suggesting probable gene duplication of the ancestral protein. UgdG comprises three-domains: a canonical Rossmann domain at the N-terminal part, which functions as core for the NAD^+ cofactor binding, a central α-helical sub-

domain, responsible for both the catalysis and homodimerisation, and a second Rossmann-like domain at the C-terminal part. A bound cofactor molecule was found present in the structure, even though the nicotinamide ring of NAD(H) was not visible in the electron density. Therefore it was modelled as an adenosine-diphosphoribose. Attempts to co-crystallise or soaking the complete cofactor remained unsuccessful, suggesting that the cofactor suffers degradation under the crystalisation conditions. A 3D structural comparison with other UGD-NAD(H) complexes revealed that the protein-cofactor (visible region) interactions were almost totally conserved, involving mainly residues of the Rossmann fold *fingerprint* region, the GXGXXG motif, the carboxylic acid (Asp30 in UgdG) and a conserved structural water molecule, characteristic in NAD-binding proteins. Similar to its bacterial homologues, UgdG is a dimeric protein, with the two subunits related by a non-crystallographic 2-fold rotation axis, where the N-terminal of one is in close proximity with the C-terminal of the other. A 3D phyletic comparison of the UGD/GMD family members has confirmed what the previous 1D analysis had suggested. The family is divided into four groups, with the prokaryotic members separated into two subfamilies. UgdG belongs to group I of prokaryotic UGDs closer to the eukaryotic members than the group II prokaryotic UGDs. Future work on this project may include an analysis of UGD/GMD family oligomerisation, in an attempt to correlate the different oligomeric states observed in the four subfamilies.

7.3. Nitroreductase from *Rhodobacter capsulatus* B10 (NprA)

The major Type I oxygen-insensitive NprA was produced and crystallised. Its crystal structure was solved by molecular replacement, using *Enterobacter cloacae* NTR as search model, and refined at 2.0 Å resolution.

Each NprA monomer is an α+β folded 210 residues peptide, exhibiting a 3-layer αβα sandwich architecture, common of Type I group B nitroreductases. It is composed of two domains, a major central core formed by a twisted β-sheet surrounded by α-helices, and a minor flexible domain, comprising two helices that pack against an edge of the core β-sheet. NprA dimerisation occludes 26% of each monomer solvent accessible area, involving helices α6 and α2 of the major domain and residues from the C-terminus that surround the dimer mate and originate a five-stranded twisted β-sheet. In the dimer, the minor domain of one subunit caps the FMN cofactor of the other subunit, and helps form the substrate-binding pocket. The dimer contains two docked flavin mononucleotide (FMN) molecules, one for each chain, bound at the interface and establishing contacts with both subunits. Only a few residues are conserved in protein-cofactor hydrogen bonding interactions, as these involve mainly main-chain atoms and/or water molecules.

Detailed 3D comparison of available Type I oxygen-insensitive NR structures from groups A and B shows a similar overall fold, with the major difference lying on the presence or absence of the minor domain, and its location in the structure. These differences may explain the diversity of substrates and the different biological roles attributed to the several members of the Type I NRs. Moreover, they reinforce the evolutionary relationship between the different NR members and the possibility of a common ancestral flavoprotein.

Several attempts to obtain NprA-ligand complexes were carried out aiming to determine the residues involved in substrate binding and catalysis, in order to clarify its physiological role within *R. capsulatus*. Among the ligands tried in NprA complexes formation, we included NAD(H), one of the electron donors used by NRs, the prodrug CB1954, used in cancer therapy, as well as some nitroheterocyclic substrates, and dimethyltetrahydropterine ($DMPH_4$), as NprA was pointed as a possible dihydropteridine reductase (DHPR). Unfortunately,

all complexes crystallisation attempts failed. One should however bear in mind that even the native crystals were obtained only twice, and their crystallisation conditions could never be reproduced or optimised, in spite of numerous attempts.

Future work on this project should include an optimisation of the purification protocol by minimising manual procedures and by using mostly automated chromatographic systems, in order to reduce the actual level of crystallisation irreproducibility; further attempts of protein-ligands complex crystallisation, as well as *in silico* ligand docking studies. Such an approach would allow the use of actual 3D data to map the residues that most probably are involved in interactions with *e.g.* the CB1954 prodrug. Together, the results may not only elucidate the important residues involved in NprA activity but also give hints on the physiological role of NprA, and possibly extend it to the Type I oxygen-insensitive family of NRs.

Appendices

Appendix I. Signature Amidases

8. The Amidase from *Achromobacter xylosoxidans* (Ana)

This appendix describes the expression, purification, crystallisation and preliminary crystallographic studies of an amidase from *Achromobacter xylosoxidans* (Ana). The *Escherichia coli* clone harbouring the constructed plasmid pAna, was kindly given by **Dr. Gang Cai and co-workers** from the *Laboratory of Molecular Biology, Institute of Plant Physiology and Ecology, of the Chinese Academy of Sciences* (Shanghai, China: see reference Cai *et al*. 2005).

8.1. Amidase signature (AS) family

Peptide bonds are the most commonly occurring class of carbon-nitrogen bonds found in Nature. However, there is a wide range of other C-N bonds whose metabolism is not so well understood. These comprise the nitriles (R-C≡N), acid amides (R-C(=O)-NH_2), secondary amides (R-C(=O)-NH-R') and others, which are the substrates for the nitrilase superfamily. This superfamily groups 13 different branches based on their structure and sequence similarities. However, the substrate specificity is only known for some of these branches and the physiological role of these enzymes is still not clearly understood (Brenner 2002; Sharma *et al*. 2009). Historical reasons led to the classification of these enzymes as nitrilase-related, but only one of these branches exhibits true nitrilase activity (hydrolysis of nitrile generating the corresponding acid plus ammonia), whilst the remaining eight have apparent amidase or amidase-condensation activities (see Figure 8.1.; Pace & Brenner 2001).

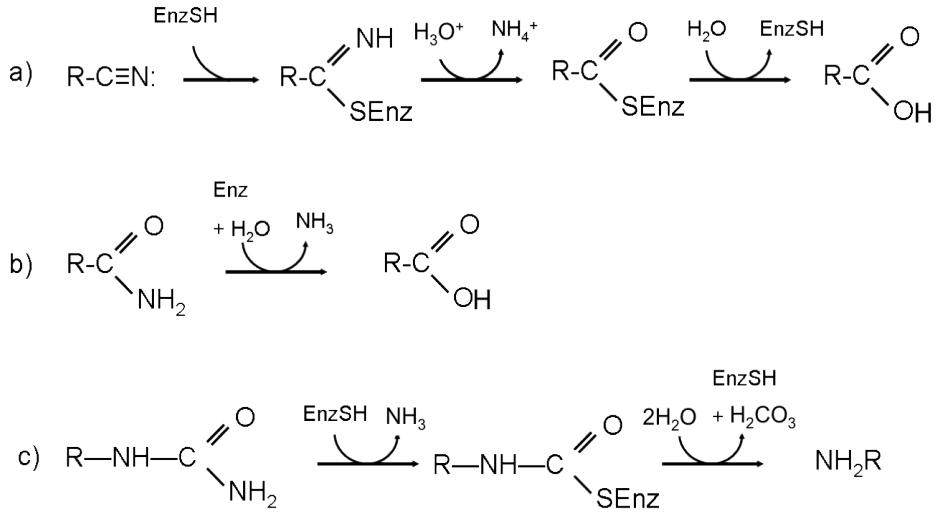


Figure 8.1. *Types of reactions catalysed by the nitrilase superfamily members.*
(a) The nitrilase reaction is performed by branch one of the superfamily, where nitrile is hydrolysed to the corresponding acid and ammonia; (b) The amidase reaction is the most common throughout all of the superfamily, and involves enzymes from branches two, three, four, seven and eight, including the acylamide amidohydrolases; (c) the carbamylase reaction is a special case of the amidase reaction and it is catalysed by enzymes of branches five and six. Branch nine comprises the N-acyltransferases that catalyse the reverse amidase reaction (not shown), transferring a fatty acid to the N-terminal of polypeptides. Figure adapted from Pace and Brenner 2001.

Acylamide amidohydrolases (EC 3.5.1.4) included in this superfamily cleave C-N bonds and share the same α/β hydrolase fold, but can be divided into two different groups on the basis of their catalytic site and preferred substrate. The first group of amido hydrolases has a catalytic site composed of a Glu-Lys-Cys triad and can rapidly hydrolyse short-chain aliphatic amides like acetamide and acrylamide (Cilia *et al*. 2005).

Amino acid sequence comparisons have shown that this group has high similarity and motif conservation with the nitrilase branch of the superfamily. Members of this group possess the invariant cysteine and glutamate residues involved in the nitrilase catalysis, display a similar catalytic mechanism and are therefore also able to act on several nitrile compounds (Bork & Koonin 1994; Sharma *et al*. 2009). *Pseudomonas aeruginosa* (Andrade *et al*. 2007) and *Helicobacter pylori* (Skouloubris *et al*. 1997) amidases belong to this group.

The second group of amido hydrolases includes the amidase signature (AS) members acting on mid-chain amides, some arylamides, α-aminoamides and α-hydroxyamides, such as isobutyramide, valeramide, benzamide, phenylpropionamide and nicotinamide amongst others (Fournand & Arnaud 2001). Enzymes included in this group are, for example, the enantioselective amidases from *Bacillus* sp. BR449 (Kim & Oriel 2000) and *Rhodococcus* sp. (Mayaux *et al*. 1991). Their biochemical role varies widely but they all possess an active centre composed of a Ser-*cis*Ser-Lys catalytic triad. Besides this classical triad, some members of the family possess a putative additional Cys-*cis*Ser-Lys catalytic centre, which confers their bifunctionality, acting on both amides and nitriles (Cilia *et al*. 2005). This is the case of the well characterised signature amidases from *Rhodococcus rhodochrous* J1 (RhorhJ1; Kobayashi *et al*. 1993; Kobayashi *et al*. 1998) and *Sulfolobus solfataricus* (SsAH; Scotto d'Abusco *et al*. 2001). All of the AS members exhibit acyl transfer activity in the presence of hydroxylamine, leading to the formation of the corresponding hydroxamic acids (Fournand & Arnaud 2001; Andrade *et al*. 2007). The AS sequence was first identified by Mayaux (Mayaux *et al*. 1991) in a group of enzymes containing a highly conserved stretch of approximately 50-130 serine-rich and glycine-rich amino acids, located in the centre of the protein structure (Chebrou et al. 1996; Shin *et al*. 2002). Moreover, this segment contains the consensus motif GGSS, which is strictly conserved in this class of amidases.

The AS family is a large class of enzymes spread throughout all of the three domains of living organisms. They are ubiquitous enzymes that catalyse the hydrolysis of C-N bonds (other than peptide bonds), and act on various endogenous and foreign aliphatic and aromatic amides (Scotto d'Abusco *et al*. 2005; Andrade *et al*. 2007). A search in the protein databases shows more than 200 members (most of them are putative, being derived from DNA sequences) distributed amongst 90 different organisms ranging across Bacteria, Archaea and Eukarya (Labahn *et al*. 2002; Shin *et al*. 2003). Whilst the biological function (the cleavage of the RCO-NH_2 bond) of the AS family is common to all of the members, the biochemical function of these enzymes varies widely. It is important to point out that one microorganism can contain several amidases, like *Rhodococcus* sp. R312 (bacteria), from which a L-α-aminoamidase, an enantioselective amidase, an aliphatic amidase and specific formamide, nicotinamide and urea amidases have been annotated (Fournand & Arnaud 2001). Several amidases have also been described for *Aspergillus nidulans* (a fungus), namely a formamidase (with formamide and glycinamide as substrates), an acetamidase (cleaving butyramide and valeramide amongst others) and a wide-spectrum amidase (hydrolysing aliphatic amides, but also arylamides like benzamide and phenylacetamide) (Hynes & Pateman 1970a; 1970b; Frazer *et al*. 2001). *Candida utilis* (yeast) also possesses a wide-spectrum amidase which efficiently breaks down acetamide, proprionamide, acrylamide, butyramide amongst others (Brady 1969). Kammermeier-Steinke and co-workers purified a peptide amidase from oranges (plants), a highly selective enzyme that hydrolyses C-terminal amide groups in peptides or N-protected amino acids (Kammermeier-Steinke *et al*. 1993). However, most of the currently known and described amidases belong to bacteria, and span many different genera, such as *Rhodococcus*, *Pseudomonas*, *Bacillus*, *Streptomyces*, *Alcaligenes* (also known as *Achromobacter*), *Helicobacter* and others (Fournand & Arnaud 2001).

Three-dimensional structures have been determined for several members of the acylamide amidohydrolases (EC 3.5.1.4), both for the short-chain aliphatic amidases and the signature amidases (mid-chain aliphatic amidases) (groups I and II, respectively), either in their native form, or with bound substrates (Table 8.1.). Bacterial amidases are widespread and their ability to hydrolyse several amides into the corresponding carboxylic acids and ammonia is of growing interest in biotechnology with potential applications in chemical and pharmaceutical industries, as well as in bioremediation. Moreover, their additional acyl transfer activity in the presence of hydroxylamine leads to the production of hydroxamic acids known for their chelating properties (Fournand *et al*. 1998). These compounds have a great variety of biological functions and low toxicities leading to potential use in different therapeutic applications (Vanjari & Pande 2003). As potent inhibitors of metalloproteases, they have also been investigated for treatment of anti-human immunodeficiency virus (Gao *et al*. 1995) and even applied to the treatment of ureaplasma infections or anemia (Holmes 1996). They have also been described for use as anti-malarial (Tsafack *et al*. 1995), antitumoural and anticancer agents (Kikushi *et al*. 2002), or as antibiotics (Komatsu *et al*. 2001). Apart from these medical applications, hydroxamic acids (particularly mid- and long-chain hydroxamic acids) have also been explored for their potential in wastewater treatments, as a way to eliminate contaminating metal ions (Koide *et al*. 1987).

8.2. Ana from *Achromobacter xylosoxidans*

Achromobacter xylosoxidans (*A. xylosoxidans*), also known as *Alcaligenes denitrificans* subsp. *xylosoxydans*, is an aerobic gram-negative motile rod with flagella.

Table 8.1. Structures of several acylamide amidohydrolases (EC 3.5.1.4) belonging either to group I, the short-chain aliphatic amidases, or group II, the amidase signature enzymes.

	Organism	Protein	PDB code	Reference
Short-chain aliphatic amidases (Group I)	*Pseudomonas aeruginosa*	Amidase	2UXY	Andrade *et al*. 2007
	Geobacillus pallidus	Amidase	2PLK	Agarkar *et al*. 2006
	Helicobacter pylori	Formamidase (AmiF)	2DYU, 2DYV, 2E2K, 2E2L	Hung *et al*. 2007
Amidase Signature (AS) enzymes (Group II)	*Thermatoga maritima*	Amido transferase	2GI3	Joint Center for Structural Genomics (JCSG)
	Thermus thermophilus	Probable amidase	2DC0	RIKEN Structural Genomics/ Proteomics Initiative (RSGI)
	Stenotrophomonas maltophilia	Peptide amidase (Pam)	1M21, 1M22	Labahm *et al*. 2002
	Rattus norvegicus	Fatty acid amide hydrolase	1MT5, 2VYA, 2WAP	Bracey *et al*. 2002; Mileni *et al*. 2008; Ahn *et al*. 2009
	Staphylococcus aureus	tRNA-Dependent Amidotransferase GatCAB	2F2A, 2DF4, 2G5H, 2G5I	Nakamura *et al*. 2006
	Bradyrhizobium japonicus	Malonamidase E2	1OBL, 1O9O, 1OBK, 1OCK, 1OCH, 1OCL, 1O9P, 1OBJ, 1O9Q, 1O9N, 1OBI	Shin *et al*. 2003

In recent years, this microorganism has become an infrequent, but potentially serious nosocomial pathogen and an important cause of bacteremia, with a predilection for immunocompromised individuals, including cystic fibrosis patients. However, since it is often confused in clinical specimens with other non-fermentative, gram-negative rods, especially *Pseudomonas* species, its role as a significant pathogen may be underestimated. The strains involved in the infection are usually multiply resistant to antimicrobial therapy. This is probably due to the fact that this bacterium harbours three kinds of β-lactam acylases or β-lactamases (EC 3.5.2.6), capable of degrading β-lactam antibiotics such as penicillins, cephalosporins (which are relatively resistant to β-lactamase activity), cephamycins, and carbapenems. These antibiotics contain the four-atom β-lactam ring as a common element (Figure 8.2.).

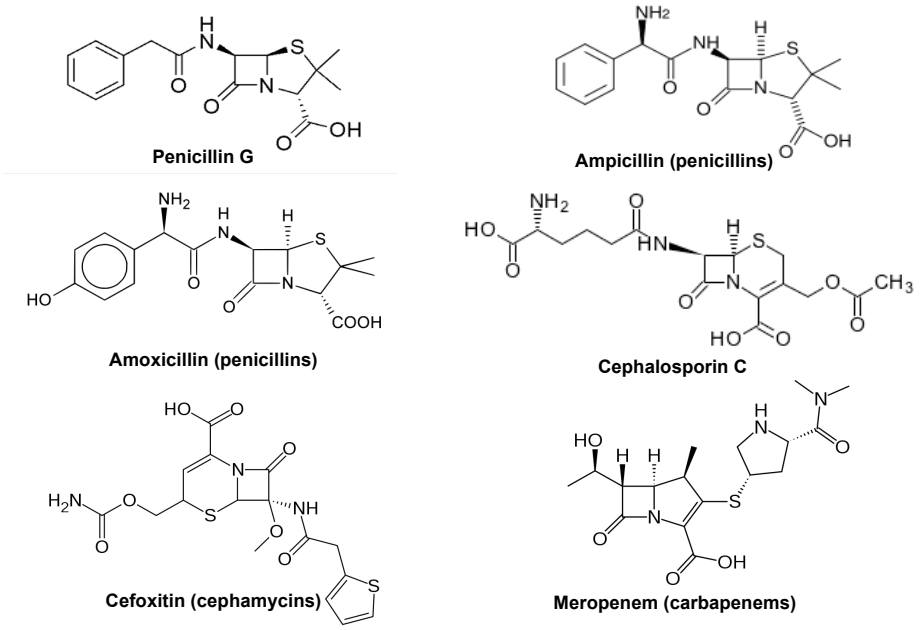


Figure 8.2. *Structure of several antibiotics, belonging to the penicillin, cephalosporin, cephamycin and carbapenems classes*.

The β-lactamase enzymes are capable of degrading one or several of these compounds. Amidase (ana) from *A. xylosoxidans* has acylase activity towards some synthetic derivatives of cephalosporin. Figure adapted from www.wikipedia.org.

The lactamase enzyme breaks the β-lactam ring open, deactivating the molecule's antibacterial properties. *A. xylosoxidans* genome harbours a thermostable penicillin G acylase (Cai *et al*. 2004), a putative cephalosporin acylase (Zhu *et al*. 2003) and a β-lactam acylase capable of degrading ampicillin, cephalexin and amoxicillin (Plhackova *et al*. 2003). It is probably one of the few bacterial strains yet characterised with three kinds of β-lactam acylases. Whilst screening for a cephalosporin acylase in *A. xylosoxidans*, Cai and co-workers found a novel amidase (Ana) with deacylation activity for several cephalosporin structural analogues (Cai *et al*. 2005). This amidase is a 509 residue peptide that belongs to the signature amidase family. In this work the overexpression, purification, crystallisation trials as well as preliminary analysis of the X-ray data for this novel amidase from *A. xyloxidans* are described, with the aim of structural characterisation and comparison with other members of the acyl amidohydrolase family.

8.3. Expression and purification of Ana

Overexpression of the *ana* gene was carried out by cultivation of *Escherichia coli* BL21 (DE3) transformants harbouring the plasmid *pAna*, in 1000 ml of LB medium supplemented with 50 μg.ml^{-1} of kanamycin at 310 K, until an OD_{600nm} of 0.8 was reached. The cells were then induced with 0.5 m*M* isopropyl *β*-D-thiogalactoside (IPTG) for approximately 4h at 299 K, harvested by centrifugation in a Beckman Instrument (10 000*g*, 30 min, 277 K) and the pellets obtained were stored at 253 K. Cells were resuspended in 50 m*M* sodium phosphate buffer at pH 8.0, supplemented with 10 m*M* imidazole, 500 m*M* NaCl, and 10 μl of benzonase (Novagen), incubated for 10 min at 4°C and disrupted in a French press. The crude cell extract was obtained by centrifugation at 27 000*g* for 40 min at 277 K (Beckman Instrument; JA-20 rotor) and the supernatant was applied onto a 5 ml HisTrap FF column (GE

Healthcare), previously equilibrated with 50 m*M* phosphate buffer pH 8.0, supplemented with 10 m*M* imidazole and 0.5 *M* NaCl (buffer A), and connected to an ÄKTA Explorer Instrument (GE Healthcare) according to the manufacturer's recommendation. The column was washed with buffer A to remove any unbound protein, and a gradient of concentration of imidazole (10-500 m*M*) was applied. The *A. xylosoxidans* amidase eluted at approximately 200 m*M* imidazole in a symmetrical chromatography peak (Figure 8.3.). The composition and purity of the eluted fractions were confirmed by SDS-Page analysis and only homogenous fractions were used (Figure 8.4.). The protein samples were immediately pooled and buffer-exchanged with 25 m*M* Tris-HCl pH 8.0, 100 m*M* NaCl and 1 m*M* dithiothreitol (DTT), using a PD10 desalting column (GE Healthcare).

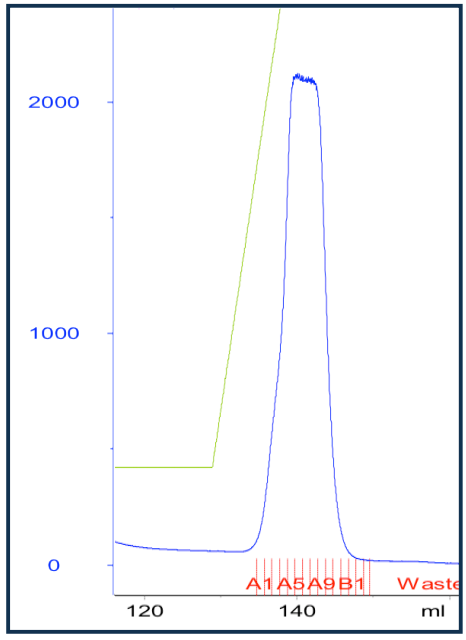


Figure 8.3. *Affinity purification chromatogram of Ana from A. xylosoxidans*.
The protein absorbance (mAU) was measured at 280 nm (blue line). The protein eluted in a symmetrical peak at approximately 200 *mM* imidazole concentration (green line), on the HisTrap FF (5 ml) column (GE Healthcare).

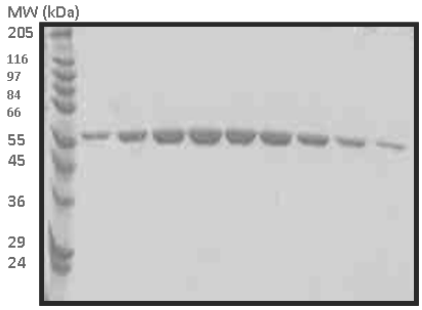


Figure 8.4. *SDS-page gel from the purification of Ana from A. xylosoxidans*.

The eluted fractions were loaded on a 10% acrylamide SDS-Page gel and stained with Coomasie Blue dye. Ana migrates as a single peptide band with molecular weight of ≈ 53 kDa. The Sigma Marker Wide Range (range from 6,500-200,000 Da; Sigma Aldrich) was used for calibration. All of the fractions showing pure protein were pooled and the protein concentrated up to 10 mg·ml^{-1} and stored at 193 K, until further use.

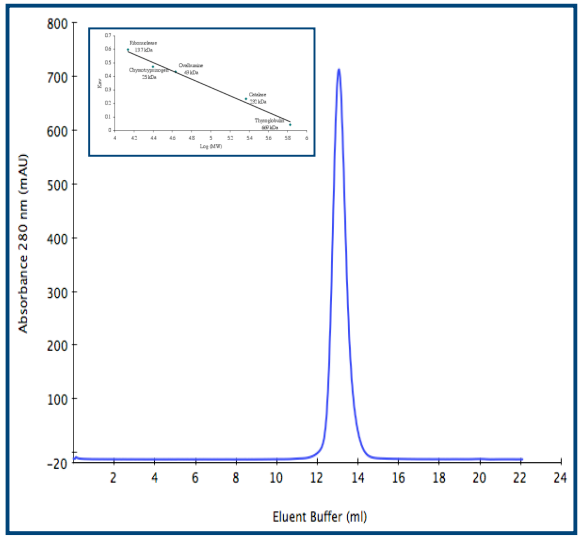


Figure 8.5. *Molecular weight estimation of Ana using size exclusion chromatography*.

Size exclusion chromatogram for Ana (10 mg·ml^{-1}) using a 24 ml Superdex 200 GL 10/300 (GE Healthcare), shows a single peak eluting with an approximate volume of 13.1 ml. Eluent buffer composition as follows: 25 m*M* Tris-HCl buffer pH 8.0 + 150 m*M* of NaCl + 1 m*M* DTT. The calibration curve of the molecular weight standards (Biorad) used (Thyroglobulin 669 kDa (9 ml), Catalase 232 kDa (12 ml), Ovalbumin 43 kDa (15.2 ml), Chymotrypsinogen 25 kDa (15.7 ml) and Ribonuclease 13.7 kDa (17.7 ml)) is shown (inset).

The protein was concentrated in a Vivapore 10/20 concentrator (Vivascience Ltd, UK) to approximately 10 $mg \cdot ml^{-1}$ before storage at 193K. The protein concentration was determined by direct measurement of absorbance at 280 nm in a NanoDrop ND-1000 Spectrophotometer using an extinction coefficient of 47 245 $M^{-1} \cdot cm^{-1}$. The homogeneity and oligomerisation state of the protein in the referred solution was analysed by size exclusion chromatography using a 24 ml Superdex 200 GL 10/300 (GE Healthcare). The chromatographic results show a unique symmetrical peak, indicating a homogeneous species with a unique oligomerisation state (see Figure 8.5. above). The protein elutes from the column at 13.1 ml of buffer volume. This corresponds to an estimated molecular weight of about 110 kDa, suggesting that Ana displays a dimeric arrangement (under these experimental conditions), as the monomer has a molecular weight of 52.8 kDa.

8.4. Crystallisation of Ana

Protein crystallisation screens were performed using a Cartesian PixSys 4200 crystallisation robot (Genomic Solutions, U.K.) using the vapour-diffusion method. 576 sitting drops consisting of 100 nl of protein solution at 10 $mg \cdot ml^{-1}$ (or 5 $mg \cdot ml^{-1}$) plus 100 nl of precipitant solution were equilibrated against 100 µl of precipitant solutions from crystallisation screens PACT, JCSG, Index, Crystal Screen I and II from Hampton Research (Aliso Viejo, USA) and QuickScreen, Grid screens Ammonium Sulphate, Sodium Malonate/Sodium Formate, MembFac and Natrix from Qiagen Canada Inc. (Montreal, Canada). Crystalline precipitates were found in various drops from the different screens, but only the Index and the JCSG screens gave three-dimensional crystals (Figure 8.6. a, b). Manual reproduction and optimisation of these conditions were tried in order to decrease nucleation and promote further crystal growth. Finally, only condition E7 from the JCSG screen (200 m*M* $Zn(CH_3COO)_2$,

100 m*M* $Na(CH_3)_2AsO_2$ pH 6.5 and 10% (v/v) 2-propanol) showed reproducible results at the µl scale. Crystals were further improved (size and nucleation) by changing several parameters such as the ratio of protein *vs.* precipitant, the use of additives and detergents (Hampton Research, Aliso Viejo, USA), or even by the change of the initial precipitants, for example, use of ethylene glycol and glycerol instead of 2-propanol.

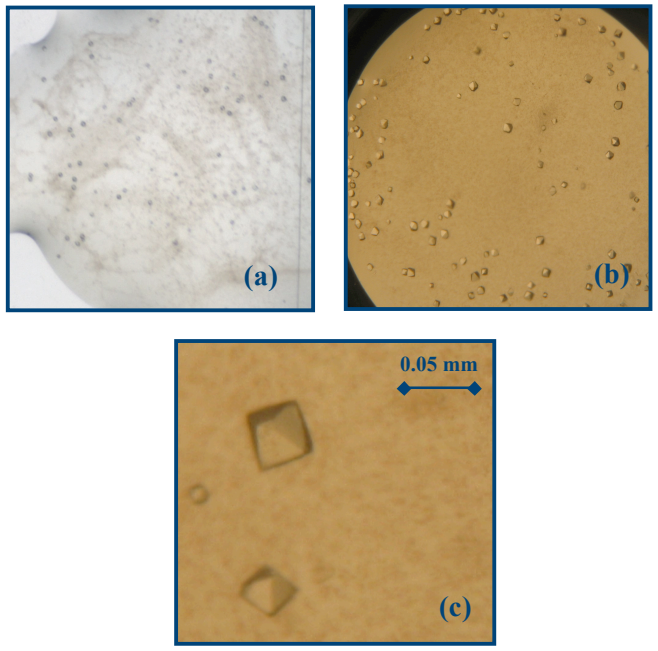


Figure 8.6. *Crystals of amidase from Achromobacter xylosoxidans*.
(a) Crystals obtained with the crystallisation robot in the precipitant solution E7 of the JCSG Screen from Hampton (200 m*M* $Zn(CH_3COO)_2$, 100 m*M* $Na(CH_3)_2AsO_2$ pH 6.5 and 10% (v/v) 2-propanol); (b) Manually reproduced crystals in the same conditions; (c) Optimised crystals (microlitre scale) obtained with the precipitant solution 200 m*M* $Zn(CH_3COO)_2$, 100 m*M* $Na(CH_3)_2AsO_2$ pH 5.5, 8% (v/v) 2-propanol and 3% (w/v) 6-aminohexanoic acid. Crystals are quadrangular pyramidal shaped with dimensions of approximately 0.05, 0.05, 0.08 mm.

The best shaped crystals (see Figure 8.6. c) were obtained at 293 K from sitting drops of 1 µl of protein solution at 5 $mg \cdot ml^{-1}$ and 1 µl of precipitant solution (200 m*M* $Zn(CH_3COO)_2$, 100 m*M* $Na(CH_3)_2AsO_2$ pH 5.5, 8% (v/v) 2-propanol

and 3% (w/v) of 6-aminohexanoic acid (aminocaproic acid) as additive, against 500 µl of reservoir solution. Cryoprotection was achieved by passing the crystals through a cryostabilising solution of mother liquor supplemented with 25% glycerol. The crystals were then cryocooled by plunging into liquid nitrogen and stored in the same conditions, until X-ray diffraction analysis was performed.

8.5. X-ray diffraction analysis of the Ana crystals

Crystals were tested and measured at the ESRF (Grenoble, France), on the beamline ID14-2, and the majority diffracted up to approximately 2.5 Å resolution (Figure 8.7.). Diffraction data were indexed and reduced, belonging to space group $P2_12_12_1$ with unit cell dimensions of a=85.0 Å, b=86.3 Å, and c=264.4 Å. Data processing statistics for the best data set are given in Table 8.2. The Matthews coefficient (Matthews 1968) calculation, points to the existence of four molecules of amidase in the asymmetric unit of the crystals.

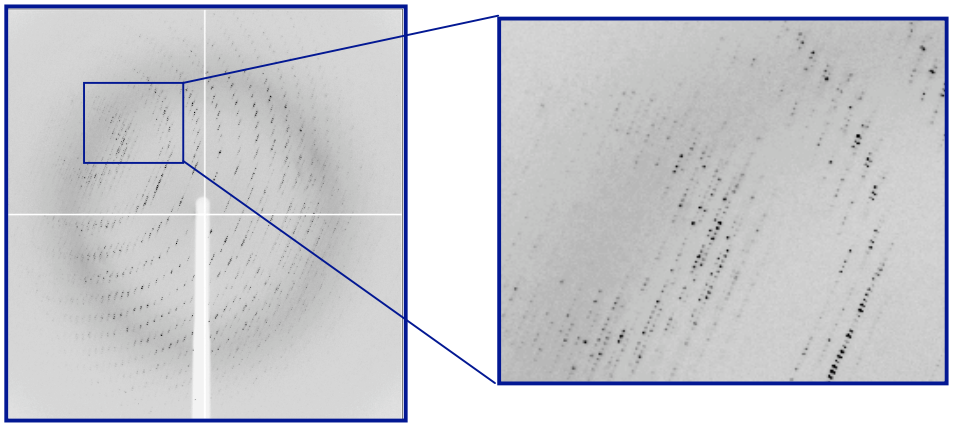


Figure 8.7. *Diffraction images for Ana crystals*.
The crystals were tested at the ESRF (Grenoble, France) at ID14-2 beamline. Crystals diffracted up to 2.45 Å and belong to space group $P2_12_12_1$.

Table 8.2. Data processing statistics for native Ana's crystal. Reflections were integrated with *MOSFLM* (Leslie *et al*. 1992), intensities scaled, merged with *SCALA*, and further reduced to structure factors magnitudes with *TRUNCATE*, programs from the CCP4 suite (CCP4 1994).

Crystal	Native Ana
ESRF beamline	ID14-2
Wavelength (Å)	0.933
Resolution (Å)	44.68-2.45 (2.58-2.45)
Spacegroup	$P2_12_12_1$
a, *b*, *c* (Å)	85.0, 86.3, 264.4
No. measured refs.	523575 (73248)
No. unique refs.	72582 (10466)
Redundancy	7.2 (7.0)
I/σ(I)	3.4 (1.9)
R_{pim} [a]	0.059 (0.154)
R_{rim} [b]	0.160 (0.411)
R_{sym} [c]	0.148 (0.380)
Completeness (%)	100.00 (100.00)
Mosaicity (º)	0.63
V_M ($Å^3Da^{-1}$)	2.30
Solvent content (%)	46.5
No. molecules in *a.u.*	4
Wilson B ($Å^2$)	38.0

Data within parentheses refer to the outer resolution shell; [a]Rpim = $\Sigma_h [1/(N-1)]^{1/2} \Sigma_i |I_i(h) - \langle I(h) \rangle| / \Sigma_h \Sigma_i I_i(h)$, where N is the data redundancy, *I* is the observed intensity and <*I*> is the average intensity of multiple observations from symmetry-related reflections. It is an indicator of the precision of the final merged and averaged data-set; [b]Rrim = Rmeas = $\Sigma_\eta [N/(N-1)]^{1/2} \Sigma_i |I_i(h) - \langle I(h) \rangle| / \Sigma_h \Sigma_i I_i(h)$, where N is the data redundancy, *I* is the observed intensity and <*I*> is the average intensity of multiple observations of symmetry-related reflections. It is an indicator of the average spread of the individual measurements; [c]Rsym = $\Sigma_h \Sigma_i |I_i(h) - \langle I(h) \rangle / | \Sigma_h \Sigma_i I_i(h)$, where *I* is the observed intensity and <*I*> is the average intensity of multiple observations from symmetry-related reflections.

8.6. Preliminary results and discussion

According to the sequence alignment performed with *CLUSTALW* (Thompson *et al*. 1997), the closest homologues of Ana having an available 3D structure, are the Glutamyl-tRNA amidotransferase subunit A (tm1272) from *Thermatoga maritima* (pdb code 2GI3, JCSG *not published*), the peptide amidase (Pam) from *Stenotrophomonas maltophilia* (pdb code 1M21, Labahn *et al*. 2002), the malonamidase E2 from *Bradyrhizobim japonicum* (pdb code 1OBK, Shin *et al*. 2003), the fatty acid amide hydrolase from *Rattus norvegicus* (pdb code 1mt5, Bracey *et al*. 2002) and the tRNA-dependent amidotransferase (GatCAB) from *Streptococcus aureus* (pdb code 2F2A, Nakamura *et al*. 2006). Molecular replacement was attempted using these structures as search models, either singularly or combined as an ensemble and also using the programs *PHASER* (McCoy *et al*. 2005) or *BALBES* (Long *et al*. 2008). Unfortunately, all of the programs used failed to produce a solution. This is probably due to the low homology between Ana and the structures available, which is around 25-30%.

While preparing this manuscript, the structure of the 6-aminohexanoate cyclic dimer hydrolase from *Arthrobacter* sp. KI72 was deposited (PDB entries 3A2P and 3A2Q; Yasuhira *et al*. 2010). By amino acid sequence alignment (Figure 8.8.), this protein proved to be the closest homologue to Ana ever reported (32 % and 51 % of sequence identity and similarity, respectively). For this reason, a molecular replacement was performed using the program *MOLREP* (Vagin & Teplyakov 1997) and the *Arthrobacter* hydrolase as a search model. The program was able to locate the four monomers present in the asymmetric unit of the crystals, with score values (contrast) of 3.07, 3.15, 5.19 and 5.93 for each of the four molecules.

```
Ana          MGTADRVDSSSFDAMALGAEIAQGATTAQAAMQQAVERVAARNPSINAAC   50
3A2P         MSKVDLWQ----DATAQAELVRSGEISRTELLEATIAHVQAVNPEINAVI   46
             *...*  :    ** * .  : .*  :    :: :: :* * **.***.

Ana          GVQAELGLDLARALDDELATLTAEQRLALLRERPFLGVPTLLKDLGTAAL  100
3A2P         -----------------IPLFEKARRESELASGPFAGVPYLLKDLTVVSQ   79
                           :. :   :* : * . ** *** ***** ..:

Ana          G--LPSAMGSVLYGQVEWNVDAEIVKRYRRAGLIPFGRSTSAELGLSPTS  148
3A2P         GDINTSSIKGMKESGYRADHDAYFVQRMRAAGFVLLGKTNTPEMGNQVTT  129
             *   .*:: .:  .  . : ** :*:* * **:: :*::.:.*:* . *:

Ana          ESPVYGAPTQNPWKAGHSAGGSSGGAGAALASGMVRIAHGSDGGGSIRIP  198
3A2P         EPEAWGA-TRNPWNLGRSVGGSSGGSGAAVAAALSPVAHGNDAAGSVRIP  178
             *. .:** *:***: *:*.******:***:*:.:  :***.*..**:***

Ana          ASCCGVLGLKPSRGMMPLGPLKGEG--WGGLATEHMMTLSVRDCAAALDI  246
3A2P         ASVCGVVGLKPTRGRISPGPLVTDSDNVAGAAHEGLFARSVRDIAALLDV  228
             ** ***:****:** :. ***  :.   .* * * ::: **** ** **:

Ana          SAGADVGAPYAAPARPGESYRDIVARVSADPRSSPRRRIAYINTTYEGEA  296
3A2P         VSGHRPGDTFCAPTAS----RPYAQGISENPGS--LRVGVLTHNPVGDFA  272
              :*   * .:.**: .    *  .  :* :* *   *  .  :..  . *

Ana          IHPEVAAAVDEAARLFAGLGHELVVAAPP-VGSEEVLSPMLPLIASAAAN  345
3A2P         LDPECAAAARGAAAALAALGHDVNDAYPEALGDRSFLKDYSTICDVAIAR  322
             :.** ***.  **  :*.***::  * *  :*....*.   .:   * *.

Ana          AIDSFVGARNRRLAADELQPTTLGAREYARAISGAQYVACVDTCHEITRR  395
3A2P         EIERNGELIGRPLTEDDVEWTSWEMVKRADQVTGRAFAACVDELRYYAGK  372
              *:      .* *: *::: *:    : *  ::*  :.****  :  : :

Ana          IGRFLHREGTQGHDLFLSPVLAQLPAPIGRYAMDNADYLAYR--LGKKGV  443
3A2P         VERWWEAG----WDLLILPTVTRQTPEIGELMLAKGTDLEGRQSAFISGS  418
             : *: .       **:: *.::: .. **.  : :.  *  *     .*

Ana          IGYSPFAPLANLTGMPAISIPFGLSADGLPIGIQVMGPLGSEAWLLELAA  493
3A2P         LQMLAFTVPFNVSGQPAISLPIGMSSDGMPIGVQIVAAYGREDLLLQVAA  468
             :   .*:   *::* ****:*:*:*:**:***:*::.. * *  **::**

Ana          QVEALRPWQRVAPIAR---------  509
3A2P         QLEGALPWVARRPQLLNPSRKIPAA  493
             *:*.  **    *
```

Figure 8.8. *Amino acid sequence alignment of Ana and its homologue from Arthrobacter sp. (PDB entry 3A2P).*

The sequence alignment performed with the program ClustalW (Thompson *et al*. 1997) shows the high sequence homology (marked with "*" or ":" for identical or homologue residues, respectively) between the two members of the amidase signature (AS) family. The GGSS motif that characterises these proteins is shown in blue (see above *section 8.1.*).

RESOLVE, with automatically implemented four-fold NCS averaging and the "prime and switch" procedure (Terwilliger 2004b) was used to improve the phases of the molecular replacement solution. Readily interpretable electron density maps were produced with a *figure of merit* for phasing of 0.68, and a model with 1285 out of 2036 possible residues (of which 308 docked in sequence) was produced (Figure 8.9. a, b).

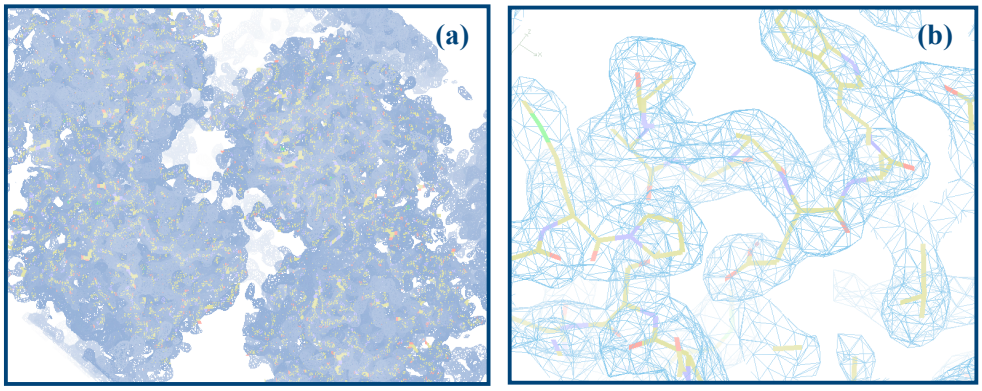


Figure 8.9. *Electron density maps and amidase model obtained after density modification with the program RESOLVE (Terwilliger 2004b).*

(a) 2mFo-DFc electron density map (blue mesh) contoured at 1.0σ covering the whole Ana crystal asymmetric unit with its four monomers. (b) Section of the 2mFo-DFc electron density map (light blue mesh) contoured at 1.0σ showing a partial model with the sequence docked. The model is represented in sticks, coloured yellow for carbon, blue for nitrogen, red for oxygen and green for sulphur atoms. Images were prepared with *COOT* (Emsley & Cowtan 2004).

The protein and solvent structure were examined at a graphics workstation against their σ_A maps (Read 1986), and the model was extended and improved accordingly, using the program *COOT* (Emsley & Cowtan 2004). Steps of consecutive model editing and refinement with the program *PHENIX* (Afonine *et*

al. 2005) are now being carried out in order to obtain a complete three-dimensional structure for the amidase of *Achromobacter xylosoxidans*.

Note: Due to the fact that this manuscript was at its final stage of preparation when the homologue structure of *Arthrobacter* sp. hydrolase became available, no further work regarding the amidase from *Achromobacter xylosoxidans* was performed. This lead to very preliminary and yet incomplete results on the 3D structure determination of Ana, hence this section was included as an appendix, instead of being a complete chapter of the PhD thesis. Carrying out the structural refinement and complete characterisation of Ana is one of the future objectives and perspectives.

Appendix II. Macromolecular Crystallography

9. General Concepts

X-ray crystallography is an efficient way of obtaining three-dimensional models of molecules that precipitate as single crystals revealing the exact position of most of its atoms. For enzymes these models can give detailed information about activity, recognition mechanisms towards substrate and/or cofactor binding and the conformational changes that those proteins might undergo. In this sense, 3D structures of macromolecules allow us to understand biological processes at the most basic level: how molecules interact, how enzymes catalyse reactions, how drugs act and, in some cases, it may help to understand the basics of diseases and help to develop new drugs. The determination of a three-dimensional protein structure is not a trivial process, although the dramatic development of X-ray crystallography in the last twenty years, linked partly with the availability of third generation synchrotron sources, has pushed forwards the understanding of some of the issues related with the technique, such as resolving the phase problem (see below).

9.1. The need for X-rays

The human eye has difficulty in distinguishing details less than 0.1 mm in size, and for an enhanced visualisation of such small objects, we currently make use of optical microscopes. These work with visible light, making use of lenses that allow the formation of a magnified image of the object. However, even the most advanced optical microscopes fail to separate details less than about 0.5 μm. Visible light is composed of electromagnetic waves spanning from wavelengths of about 350 nm (far violet) to 700 nm (deep red).

Considering that it is not possible to resolve features smaller than the wavelength of the light used in their analysis, to use visible light with wavelengths in the order of the hundreds of nanometres makes, for example, the distinction of objects separated by only 1 Å (0.1 nm) impossible. This is the case of covalently bound atoms at about 1-2 Å apart, or the case of strong polar interactions and hydrogen bonds, where distances vary between 2.5-3.5 Å. Thus, using light of equal wavelength or even smaller, to the size of the object under study, is imperative. X-rays are therefore essential in protein structure determination, as they can provide the correct wavelength, typically between 1.9 and 0.6Å (with corresponding energies of 6 keV and 20 keV, respectively) to resolve atoms in a biological molecule (Blow 2002).

In a conventional optical microscope, light strikes the object and is scattered in several directions and a lens collects the scattered rays and overlaps them, forming an image of the object. Unfortunately, microscopes that can collect X-rays scattered from an object and use a lens or device to form an image of it are not available.

9.2. The need for protein crystals

In protein crystals the scattering results from a large 3D array of protein molecules, arranged in an ordered manner and in the same orientation (Figure 9.1.). Protein crystals can thus act as signal amplifiers; working with single molecules would give rise to weak signals extremely difficult to detect. Of course, radiation damage (see below) will always occur, even for a protein crystal, because just as for any organic or living material, proteins are sensitive to ionising radiation (in this case, X-ray).

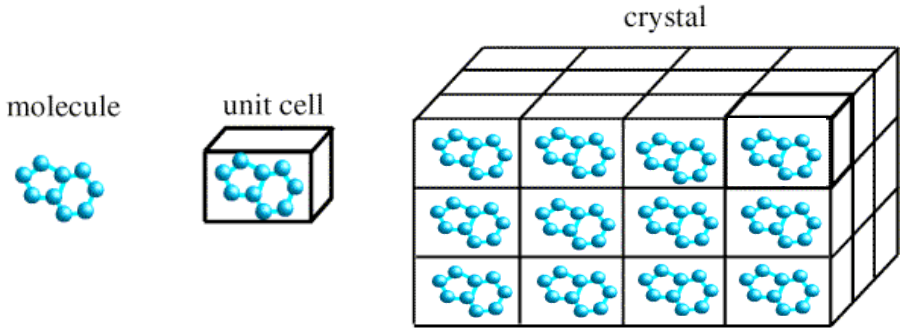


Figure 9.1. *Arrangement of molecules in a crystal.*
The crystal structure is a periodic arrangement of a motif in a lattice. The motif or *asymmetric unit* can be a single atom or a protein molecule or a combination of related molecules. The *unit cell* describes the basic building block for the crystal, and is characterised by the lengths of its three edges (a,b,c) and the angles between them (α,β,γ). The crystal is built from unit cells arranged into a three-dimensional lattice, also known as the *crystal lattice* (adapted from http://www-structmed.cimr.cam.ac.uk/Course/Overview/Overview.html).

The ionisation events created by X-rays interacting with the crystal result in the production of free radicals, which can destroy the protein crystal, particularly if one uses the very intense beams from synchrotron sources. However, even if the X-rays destroy some of the protein molecules, the overall effect on the scattering will not be as serious as for just one molecule. X-ray sources used in home laboratories are less intense, but less deleterious for the crystal sample.

9.3. Protein crystallisation

There are a number of potential bottlenecks in determining a three-dimensional structure of a protein, but growing a useful crystal is one of the most serious and time consuming.

The first stages of any protein crystallographic study are the purification and crystallisation of the protein sample, that is, the production of well-ordered single crystals of sufficient quality to diffract an X-ray beam, and to record the corresponding diffraction. However, to succeed in crystallisation one needs to control many different experimental variables both of thermodynamic and kinetic nature. Their refinement is still achieved today, only via empirical procedures. Crystallisation is the process in which a protein is transferred from a soluble state to a highly organised solid state (see Figure 9.1. above). Firstly, the formation of a periodically ordered initial aggregate of critical size has to occur (the nucleation step; Figure 9.2.). These *nuclei* have to be thermodynamically stable to allow more molecules to orderly aggregate, promoting the growth of a crystal (the growth step; Figure 9.2.).

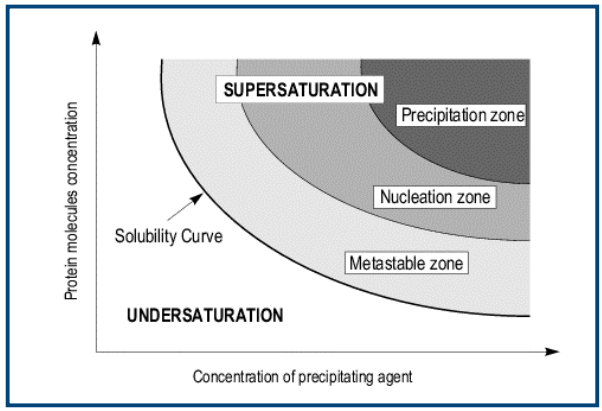


Figure 9.2. *The phase diagram for protein crystallisation.*
The two-dimensional phase diagram illustrates the change of protein concentration against the precipitant concentration. The concentration space is divided by the protein solubility curve into the undersaturated and supersaturated states. The supersaturated region comprises the metastable, nucleation and precipitation zones (adapted from Ducruix and Griege 1992).

In both cases, a non-equilibrium super-saturated state is required that will cause the protein molecules to pass from the soluble to a solid crystalline state (Riès-

Kautt 1999). A number of techniques have been developed for bringing a protein solution into a supersaturation state. Amongst them, the most frequently used methods in protein crystallisation are vapour diffusion, micro-batch and dialysis, and less frequently the use of matrix gels, like agarose (Unge 1999). Supersaturation of the protein solution can be achieved by all of these techniques, however the underlying principles of these methods vary. The vapour diffusion technique is the most common method in protein crystallisation, as it requires relatively small amounts of protein, thereby allowing many variables to be screened with the protein available (Figure 9.3.).

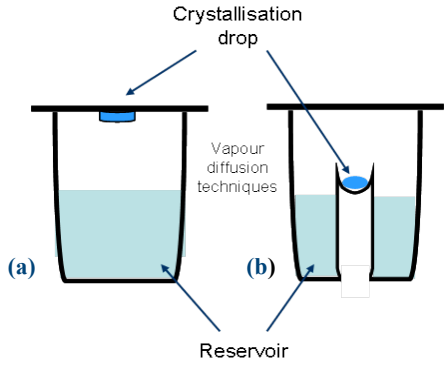


Figure 9.3. *Vapour diffusion experimental set-up in a) hanging drop and b) sitting drop techniques.*

In a sealed compartment, a drop mixture of protein solution and precipitant is allowed to equilibrate against a reservoir containing the precipitant solution at higher concentrations. The drop can be set on the compartment cover (hanging drop technique) or in a well inside the reservoir (sitting drop technique).

This technique uses the evaporation and diffusion of water between solutions of different concentrations as a means of approaching and achieving supersaturation of protein macromolecules. In the experimental set-up, a protein solution (in concentrations ranging from 2-50 $mg \cdot ml^{-1}$) is mixed in a drop with the reservoir solution containing precipitant(s).

When the precipitant concentration in the protein/precipitant drop is lower than in the reservoir solution, water evaporates from the drop into the reservoir, increasing both protein and precipitant concentrations in the drop, thus promoting supersaturation. Crystals may form when the concentration of protein in solution passes its limit of solubility, reaching the nucleation zone of the supersaturated state (see above Figure 9.2.). When the first crystals appear, concentration of protein in solution decreases. Ultimately, these crystals will grow until the concentration of the protein in the drop reaches the solubility curve. Finding the best precipitant solution is the result of an empirical search.

A common starting point is the so called sparse matrix approach, in which the above procedure is repeated hundreds or thousands of times, sampling different combinations of buffers and precipitating agents to find at least one condition that may lead to crystal formation. Optimisation of parameters such as protein concentration, protein/precipitant ratio, temperature, pH, ionic strength and additives can also be assayed once a first crystallisation condition has been found. In recent years, automation has played an increasing role in protein crystallisation since robots are able to produce, routinely and reliably, protein drops of about 100 nl. They are able to set thousands of conditions in a few hours, using less protein than previously necessary (Bergfors 1999). The visualisation of the crystallisation drops is also becoming increasingly automated, overcoming the problems of tedious manual inspection of such small and numerous of drops. Finally, once a suitably sized crystal has been found it has to be tested for X-ray diffraction and reproducibility. In this field, automation has also evolved tremendously over the last few years and the auto-mounting of crystals is now routine at many synchrotron beamlines, allowing rapid crystal evaluation.

As referred to earlier, X-ray radiation is ionising, interacting with the molecules within the crystal and causing what is known as radiation damage. The extent of this damage is related to the absorbed radiation dose. The chemical damage in organic materials by X-rays or other forms of ionising radiation is believed to consist of two kinds of damage: primary and secondary. Primary damage is caused by direct interaction between the radiation beam and electrons, which may lead to the breaking of chemical bonds by the radiation, thus causing the destruction of molecules. Secondary damage arises from the resulting radiolytic products, e.g. free radicals that diffuse through the crystal, causing further chemical reactions that alter the structure of the molecules in the crystal lattice and damage the intermolecular contacts that stabilise the crystal (Teng & Moffat 2000; Holton 2009). Whilst primary radiation damage depends upon the energy and number of photons absorbed, secondary damage varies with the nature of the solvent and with factors such as temperature or the presence or absence of free-radical scavengers that affect the mobility and reactivity of the radiolytic products (Garman & Owen 2006). One way of decreasing the secondary radiation damage caused by X-rays is to cool the crystals using a nitrogen gas stream at about 100 K, thus slowing down the kinetics of radical diffusion (Garman 2003). However, one has to be careful not to produced ice during the cooling process, due to the solvent (water) content of protein crystals. If ice is formed during cryocooling, its expansion in the crystal solvent channels will create heterogeneities with concomitant crystal breakdown. Cryoprotectants, are compounds that increase the viscosity of the mother liquor, hindering aqueous hydrogen bond formation and prevent ice during cooling. Glycerol or polyethylene glycol (PEG), already present or added to the mother liquor, allow effective cryoprotection of protein crystals (Garman & Owen 2006).

9.4. Interaction of X-rays and matter: the diffraction experiment

When a protein crystal is exposed to X-ray radiation, the electron magnetic waves that constitute the X-ray beam will make the atomic electrons vibrate at the same frequency. This oscillation of the electrons will modify the electromagnetic field in the space around the crystal. Each unit cell in the crystal will behave as an emitter of electromagnetic waves that interact and interfere with the waves from the other unit cells. If the difference in the lengths of the paths taken by each wave, from different scattering elemental volumes is a multiple of the radiation wavelength, then the waves will scatter *in phase*, their amplitudes will add up and this will give rise to a diffraction spot, recorded at the detector device. Diffraction spots are often called reflections, because one can compare crystals to sets of thousands of "mirrors" that reflect the X-rays. These "mirrors" are called Bragg planes (Figure 9.4.).

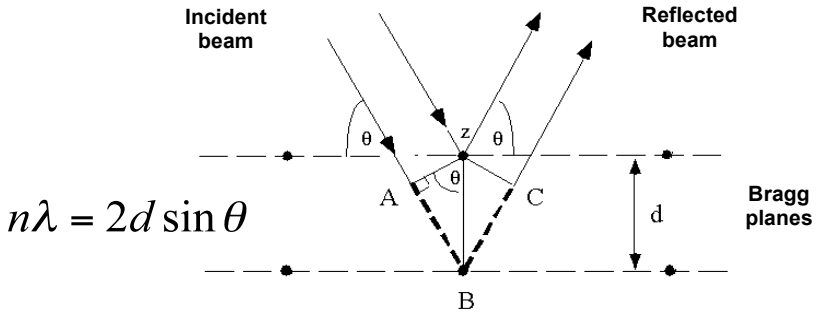


Figure 9.4. *Diagram of Bragg's law for the X-ray diffraction.*

When an X-ray beam (wavelength λ) strikes the crystal surface, the rays of the incident beam are always in phase and parallel up to the point at which the top beam strikes the top layer at atom z. The second beam continues to the next layer where it is scattered by atom B. The second beam must travel the extra distance AB + BC if the two beams are to continue travelling adjacent and parallel. This extra distance must be an integral (n) multiple of the wavelength (λ) for the phases of the two reflected beams to be the same (adapted from www.eserc.stonybrook.edu/ProjectJava/Bragg).

In order for the unit cells to diffract in phase, the Bragg planes must pass through the same points in all the unit cells in the crystal, this is to say, if the sets of planes are separated by one unit or an integral fraction of the unit cell edge, then they will pass through equivalent atoms in the different unit cells. On the other hand, if the planes divide the unit cell edge by a non-integral number, the different unit cells will diffract out of phase and the waves will cancel out. This is described by *Bragg's Law,* described mathematically by the equation, $n\lambda=2d\sin\theta$. The record of the intensity and position of all these reflections is called X-ray diffraction (see Figure 9.5.). The diffraction pattern consists of reflections of different intensity. Typically many hundreds or thousands of reflections need to be collected to record a complete data set.

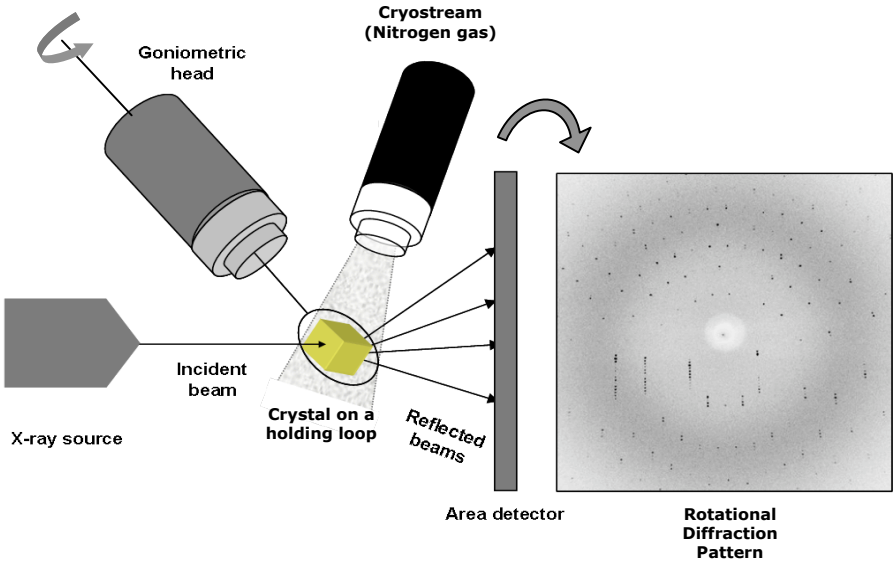


Figure 9.5. *The X-ray diffraction experiment.*

The protein crystal is exposed to an incident X-ray beam from which diffraction results, upon interaction with the electrons present in the crystal. The intensity and position of the diffracted beams are recorded as images, from small incremental crystal rotations, the so-called "rotation method". Data collection proceeds until a sufficient amount of diffraction data has been collected for structure determination.

9.5. The final stage: 3D-structure determination

9.5.1. Crystal symmetry and space groups

Protein crystals are aggregates of ordered and regular three-dimensional arrays of a polypeptide, in a repetitive and periodic geometrical arrangement. This geometrical arrangement is described by the crystal symmetry. Several symmetry operations can occur in a three-dimensional body, such as rotation, inversion, reflection, screw axis, glide planes or translation. A crystallographic *point group* is a combination of rotation, reflection and inversion operations that when applied to a crystal, leave a point fixed whilst moving every other atom to an equivalent position, making the crystal look the same before and after these symmetry operations are applied. A *Bravais lattice* corresponds to an infinite set of points that are generated by translational symmetry applied to all the atoms in the crystal. The crystal looks the same when viewed from any of these lattice points. In the case of proteins, some of the symmetry operations are not allowed. Proteins lack a centre of symmetry and are therefore called *chiral.* They possess "handedness", because their C-alpha carbons are linked to four different moieties. An inversion of the naturally occurring L-amino acids (that compose proteins) would produce D-amino acids. This is also valid for mirror symmetry, as the reflection in a mirror of a left-handed object (e.g. L-amino acid) will give rise to a right-handed object (e.g. D-amino acid). In effect, only rotational and translational symmetry are allowed in protein crystals. However, crystal space and unit cell shape impose restrictions on these two symmetry operations. The stacking of unit cells in the crystal has to fill the complete space while obeying the rotational symmetry. Therefore, rotational symmetry follows *n-fold* symmetry (where *n* is an integer and an integral fraction of 360°), but only the 2-fold, 3-fold, 4-fold and 6-fold rotational axes may exist. It is not

possible to fill the complete crystal space with regular pentagons or heptagons, therefore no crystal lattice exists with 5, 7- or higher-fold rotational symmetry. Molecular assemblies, on the other hand, can have such symmetries inside the unit cells.

Based on the possible combinations of the different symmetry operations and restrictions, 230 space groups exist (International Tables of Crystallography, Vol. I). Due to chirality of protein molecules, only 65 space groups are allowed for protein crystals, which are described in Table 9.1. Diffraction data collection involves several steps, such as testing the crystal to radiation (whether it diffracts or not), how much it diffracts (resolution), the dimensions of the unit cell (the repeating unit of the crystal), the internal arrangement of the molecules inside the crystal (space group), how it is ordered (mosaicity), etc. This is often achieved by analysis of a few diffraction images, at different crystal orientations. There are several software programs that allow the *indexing* (space group and unit cell determination) and *reduction* of the data (overall corrections of the experiment, accounting for errors and crystal damage, by scaling the data). In this thesis, different software packages were used for indexing and data reduction: *MOSFLM / SCALA* (Leslie 1992; Evans 2006) and *DENZO / SCALEPACK* (Otwinowski and Minor 1997).

9.5.2. The phase problem

The process of data collection of a biological macromolecule crystal in a diffraction experiment produces an enormous amount of data. Each diffraction spot (see Figure 9.5. above) represents a set of diffracted waves, defined by an amplitude and a phase (related to the intensity of the spot and to the relative arrival time of diffracted photons, respectively). Mathematically, each diffracted wave can be treated as vector, with amplitude and a phase or, in crystallographic language, a structure factor.

Table 9.1. *The non-centrosymmetric space groups for protein crystals* (Adapted from Blundell & Johnson 1976)

Crystal System	Axes of symmetry (rotational)	Geometrical lattice constrains	Possible Bravais Lattices	Point groups	Space groups
Triclinic	No axes of symmetry	$a\neq b\neq c$ $\alpha\neq\beta\neq\gamma$	P	1	P1
Monoclinic	2-fold axis parallel to b	$a\neq b\neq c$ $\alpha=\gamma=90°\neq\beta$	P, C	2	P2, $P2_1$, C2
Orthorhombic	3 orthogonal 2-fold axes	$a\neq b\neq c$ $\alpha=\beta=\gamma=90°$	P, C, I, F	222	P222, $P2_12_12_1$, $P2_12_12$, $P222_1$, C222, $C222_1$, I222, $I2_12_12_1$, F222
Tetragonal	4-fold axis parallel to c	$a=b\neq c$ $\alpha=\beta=\gamma=90°$	P, I	4 422	P4, $P4_1$, $P4_2$, $P4_3$, I4, $I4_1$ P422, $P42_12$, $P4_122$, $P4_12_12$, $P4_222$, $P4_22_12$, $P4_322$, $P4_32_12$, I422, $I4_122$
Trigonal	3-fold axis parallel to c	$a=b\neq c$ $\alpha=\beta=90°$ $\gamma=120°$ $a=b=c$ $\alpha=\beta=\gamma<120°, \neq 90°$	P (or R)	3 32	P3, $P3_1$, $P3_2$, R3 P312, P321, $P3_121$, $P3_112$, $P3_212$, $P3_221$, R32
Hexagonal	6-fold axis parallel to c	$a=b\neq c$ $\alpha=\beta=90°$ $\gamma=120°$	P	6 622	P6, $P6_1$, $P6_2$, $P6_3$, $P6_4$, $P6_5$ P622, $P6_122$, $P6_222$, $P6_322$, $P6_422$, $P6_522$
Cubic	four 3-fold axis along the diagonals of the cube	$a=b=c$ $\alpha=\beta=\gamma=90°$	P, I, F	23 432	P23, F23, I23, $P2_13$, $I2_13$ P432, $P4_132$, $P4_232$, $P4_332$, F432, $F4_132$, I432, $I4_132$

The inverse Fourier transform of the structure factors is the electron density (Figure 9.6.), which maps out all the crystal contents, in particular the electron density of the protein molecule. The amplitude of the structure factor is the square root of the observed intensity in each measured spot. However, the relative phase of the waves is not directly measurable during the diffraction experiment. This is known as the phase problem and, together with the crystallisation stage, is often one of the major bottlenecks in protein crystallography.

$$\rho(x,y,z) = \frac{1}{V}\sum_{h}\sum_{k}\sum_{l} F_{(h,k,l)} \exp\left[-2\pi \cdot i(hx + ky + lz)\right]$$

Figure 9.6. *The electron density equation.*

The x, y and z variables are the atomic coordinates that is, the position of each protein atom in the considered crystal volume, V. The Miller indices (h, k, l) define each set of Bragg planes, whereas $F_{(h,k,l)}$ is the corresponding structure factor, originated by the plane diffracted wave.

Phases are crucial for the reconstruction of structural information and a large part of crystallography is devoted to solve the phase problem. Furthermore, the contribution of the phase information is dominant with respect to structure factor amplitudes in the calculation of the electron density map. It is therefore important to obtain accurate phases to determine a macromolecular structure.

A miscalculated set of phases can lead to wrong or "fake" electron density map (www.ysbl.york.ac.uk/~cowtan/fourier/fourier.html). There are several methods by which the phases can be determined, requiring different diffraction experiments with different experimental conditions, or even the chemical modification of the protein crystal under study. These methods are briefly outlined below; two of which were used in the present studies: the molecular replacement (MR) and the single anomalous dispersion (SAD) methods.

1. Molecular Replacement. This method is used when a model from a similar protein exists, as in the case of homologous proteins. As the availability of newly determined protein structures increases every day, the possibility that an unknown structure shares common features with one or more characterised proteins (even if not necessarily participating in the same biological processes) becomes greater. The molecular replacement method consists of positioning the available 3D model in the unit cell of the target crystal, in such a way that the calculated diffraction of the model matches the experimentally measured data. In other words, making the model to fit the experimental data, and retrieving an electron density map using the calculated phases and the measured amplitudes. The positioning of the model (probe) is given by rotational and translational parameters, which fully describe the orientation and position of the probe in the unit cell (Figure 9.7.). There are several software packages available to perform molecular replacement. For the work described in this thesis, the programs *PHASER* (McCoy *et al*. 2005) and *MOLREP* (Vagin & Teplyakov 1997) were used.

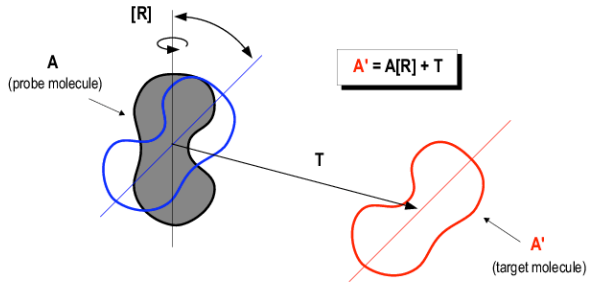


Figure 9.7. *Visual representation of the principles underlying the Molecular Replacement method.*

The model or probe is represented by **A** and the target molecule by **A'**. Rotating (orientating) and then translating (positioning) the probe, superimposes it onto the target molecule. **[R]** is the rotation matrix, and **T** is the translation vector. (Adapted from http://www-cryst.bioc.cam.ac.uk/~dima/whitepapers/mr-in-action).

2. Single or Multiple Isomorphous Replacement (SIR/MIR). This method is based on the comparison of at least two protein crystals diffraction data sets. They should differ in one (or more) strong scatterer(s) present in one of the crystals but absent in the other, which will cause comparable differences between the measured intensities. This is achieved by crystallising the protein with (*derivative*) and without (*native*) the strong scatterer, which is normally a heavy atom, such as Hg or Pt. An alternative method is to soak native crystals in a solution containing heavy atoms, hoping that the heavy atom will bind tightly to a specific site(s) of the protein. This method requires that the native and derivative crystals have the same unit cell dimensions and symmetry (isomorphism). Of course, introducing these atoms in a protein will always disturb the protein structure (in the worst cases, completely destroying the crystal), so isomorphism is never perfect. The determination of the heavy atom substructure and combination of the native and derivative data, allows the calculation of preliminary experimental phases for the native protein, which in turn will be used to build an experimental protein model. The difference between the SIR and MIR methods is the number of derivatives used for the phase determination. With the SIR method one cannot determine a unique set of phases, still having to resolve the true hand of the structure. This can be performed using more derivatives (the MIR method). Several programs to calculate the experimental phases from a *MIR/SIR* experimental data are available, such as *SOLVE/RESOLVE* (Terwilliger 2004a), *SHARP* (Vonrheim *et al*. 2005) and *SHELXD/SHELXE* (Sheldrick 2008).

3. Multiple and Single-wavelength Anomalous Dispersion (MAD/SAD). These techniques are dependent upon the presence of (mostly heavy) atoms in a protein structure, which cause small changes in the diffraction pattern

of a protein crystal. If the incident X-ray beam has the correct energy to excite the inner electrons in these atoms (absorption edge) part of the energy will be absorbed and electronic transitions occur. When the electrons return to the resting state, fluorescence will be emitted, and waves will be scattered. These waves will have a different phase but also a different intensity, as part of the energy is absorbed during the transition. This phenomenon is called *anomalous scattering*. The use of these methods has become very popular in phase determination firstly, due to the development of techniques to modify natural proteins by substitution of the sulphur (in methionine or cysteine amino acids) into selenium. Secondly, with the increase on the number of tuneable beam lines at synchrotrons, which allow the X-ray wavelength to be set to values close to an absorption edge of an anomalous scatterer. A fluorescence edge scan allows the measurement of the absorption edge of the anomalous scatterer(s) in the crystal, and should always be performed prior to a *MAD/SAD* diffraction experiment. The programs used on a *MAD/SAD* experiment are the same as for isomorphous replacement (see above). In this work only the programs *SOLVE/RESOLVE* (Terwilliger 2004a) and *SHELXD/SHELXE* (Sheldrick 2008) were used.

4. Multiple Isomorphous Replacement with Anomalous Scattering (MIRAS). This method uses a combination of the MIR and MAD methods (as outlined above), and is effective when the heavy atom has enough measurable anomalous signal at the working wavelength. The differences in intensities from both the isomorphous signal and the anomalous scattering due to the presence of the heavy metal(s) in the crystal are used to calculate the experimental phases.

5. Ab Initio Phase Determination (Direct Methods). This phasing method relies on the use of arbitrary starting phases and it is based on a series of assumptions, such as that the atomicity condition is satisfied, that is, diffraction data to a resolution equal to or better than 1.2 Å. It has been widely used in small-molecule crystallography, however success of the method is dependent on the size of the molecule, as well as the resolution. Hence, for macromolecular entities, like proteins, the use of this technique is very limited, as proteins have a large number of atoms and in general, they do not diffract to atomic resolution. Several programs are available for direct methods phase determination, such as ACORN (Foadi *et al*. 2000) or *SHELXD/SHELXE* (Schneider & Sheldrick 2002).

9.5.3. Density modification: phase improvement

Once experimental phases have been determined, the Fourier transform of these with the measured amplitudes, results in a calculated electron density map. Depending on several factors, like data quality, resolution and phase errors, the map may be clear or difficult to interpret. However, before attempting to trace a structure into a scarcely interpretable map, one has several possibilities of improving the electron density. These are known as *density modification* methods, and are used for phase improvement.

1. Solvent flattening: Since protein crystals have typically 30-70% of the unit cell volume as disordered solvent, a mean value of electron density can be attributed to the solvent regions, and new "flattened" electron density is obtained. This improved electron-density map is then converted by an inverted Fourier transform into new set of phases, which when combined with the experimental amplitudes originates a new, "improved" density

map. The result is like enhancing the contrast of the protein region of the map, against the solvent regions.

2. Histogram matching: This technique makes use of the known electron density distribution for typical protein structures. The minimum, maximum and intermediate values of electron density are adjusted to those that can be found in a protein electron density map, at the same resolution, again resulting in an improved density map, from which new phases, and ultimately new improved maps are calculated.

3. Non-Crystallographic Symmetry averaging: this map improvement is based on the extra information available when several copies of the molecule exist within the asymmetric unit of a crystal. In fact, protein molecules often crystallise as oligomers, and their internal arrangement is not described by the crystal symmetry. This gives a "local" symmetry within the asymmetric unit, also called *non-crystallographic symmetry* (NCS). In this technique, the electron density of the different monomers is averaged locally, which results in a randomisation of the noise and in increased signal for the molecular features, leading to a more precise representation of the calculated phases that are fed into the calculation of new improved maps.

4. Prime-and-switch as implemented in RESOLVE: Used after molecular replacement. MR relies on the use of the phases obtained from a homologous protein structure, and this poses the problem of *model bias*, as phases are calculated from the positioned probe model. The prime-and-switch method is based upon the definition of a map probability calculated independently from any prior phase information. This allows the intrinsic

model bias from the calculated phases of a molecular replacement solution to be reduced.

The methods of density modification are of great use for the improvement of the initial electron density maps and for eliminating phase errors. Again, several programs are available, using one, or several methods combined. In the present work only *DM* (Cowtan 1994; Cowtan 1998) and *RESOLVE prime-and-switch* (Terwilliger 2004b) were used.

9.5.4. Model building and structure refinement

Once the initial phases are good enough to calculate an interpretable electron density map, a first (usually inaccurate and incomplete) model can be built. Several programs are able to auto-trace an initial model, such as *ARPWARP* (Perrakis *et al*. 1999) or *RESOLVE* (Terwilliger 2004a). Of course, their chance of success is higher when good resolution data are available and accurate phases have been obtained in the previous steps. In less favourable cases, when only poor initial phase information is available, or the resolution is not high enough (current auto-build programs fail at around 3-3.5 Å), human intervention is needed to identify secondary structure features of the protein and to build an initial poly-Ala trace. This can be done manually and interactively using graphics software such as *COOT* (Emsley & Cowtan 2004). Model building will generate a new phase set that can be used to generate improved electron density maps (by combination with initial experimental phases). In general, model building is performed in iterative cycles together with the structure refinement, thus generating improved maps. Structure refinement is the process in which the protein model is altered with the scope of matching, as much as possible, the calculated structure factors F_{calc}, to the experimental ones F_{obs}, aiming to obtain the best fit of the model to the crystallographic

observations. In other words, refinement consists in the determination of the best possible combination of atomic positions, atomic displacement parameters (*B factors*) and atomic occupancies for protein residues, ligands and solvent molecule. The aim of model building and refinement is to construct a model that adequately explains the experimental observations, whilst making physical, chemical and biological sense. The goal is to produce electron density maps that allow detection and correction of the initial model errors. However, one has to be careful when tracing a structure into an electron density map, because one may inadvertently create a model containing errors and artefacts, resulting from errors in the experimental data. In this way, without care on the part of the crystallographer, a refinement program can produce an over-fitted, inaccurate model. Maps from refinement should improve as the model becomes more accurate, hence improving the calculated phases. Due to the limited resolution typically obtained in protein crystallography, the experimental data is complemented with a priori chemical information, for instance geometric information concerning bond lengths and angles. This information is used during the refinement, in the form of geometric parameter dictionaries. Several statistical parameters are used to assess if refinement was successful and if the structural model is being over-fitted to the X-ray data namely, the R_{factor} and R_{free} factors. R_{factor} compares the observed structure factors with the calculated ones (Figure 9.8.). The smaller the differences between the two terms, the better they agree.

$$R_{factor} = \frac{\sum_h \left| \left|F_{obs}\right| - \left|F_{calc}\right| \right|}{\sum_h \left|F_{obs}\right|}$$

Figure 9.8. *The R_{factor} equation.*

The R_{factor} is calculated over a group of reflections (h), which can be all of the reflections, or a particular group, like the case of the "test set" used in the R_{free} calculation.

The progress of the refinement can be monitored by the fall on the R_{factor}. This parameter is dependent upon the data resolution and crystal quality. Moreover, it can be reduced by erroneous adjustments in the structure, like refining individual atomic displacement parameters or modelling alternative conformations at resolutions where this is not warranted; by removal of data (low-resolution cut-offs) or adding spurious entities, such as solvent molecules. Structural models with low $R_{factors}$ do not necessarily represent accurate and correct three-dimensional structures. For this reason, a cross-validation tool was introduced in macromolecular crystallography, the R_{free} factor (Brunger 1992). The idea is to set aside a small fraction of the data, the "test set", not used in the refinement, but for which an *R* factor is calculated. This parameter does not have the problem of bias (it is not used in refinement) and decreases of the R_{free} value represent real improvements in the model. Because the R_{free} is calculated with reflections that are not used in the refinement process, the value is higher than the conventional R_{factor}. Comparison of both conventional R_{factor} and the "unbiased" R_{free} factor values, their behaviour and the gap between values during the building/refinement cycles reveals the extent to which the model has been overfitted to the experimental data as well as about the quality of model and data. In this thesis, the programs *REFMAC5* (Murshudov *et al*. 1997) and *PHENIX.REFINE* (Afonine *et al*. 2005) were used for the structural refinement of the proteins under study.

9.5.5. Structure quality and validation

As referred to above, the overall quality of a final structure can be assessed by the crystallographic *R*-values, and the difference between them. However, reasonable *R* factors can be obtained for an overall correct molecule, even if small parts of it may be wrong in terms of geometry and chemical environments. In this sense, parameters like the stereochemistry, deviations

from ideality of bond lengths, bond angles and violations of dihedral, flatness and chirality restraints, for both side chains and main chain (Ramachandran plot; Figure 9.9.; Ramachandran *et al*. 1963), side-chain rotamer analysis, average *B* values for ordered and disordered protein residues, ligand, substrate and solvent molecules (waters or others), intermolecular interactions and contacts, have to be analysed and validated before a structure can be submitted to the Protein Data Bank (PDB; Berman *et al*. 2000). However, values deviated from ideality can sometimes be correct, if supported by the electron density maps, and may highlight particularly interesting (and unusual) parts of the structure under study.

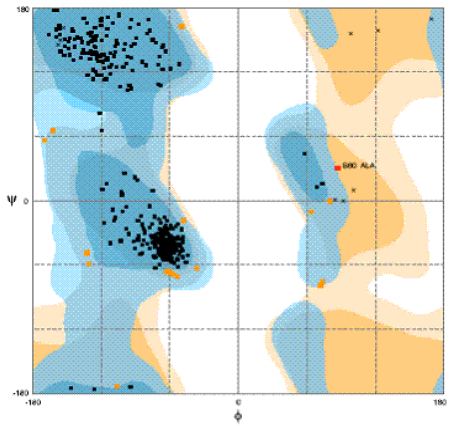


Figure 9.9. *The Ramachandran plot.*

This graphical representation is based on the analysis of the ϕ (C(O)-N-Cα-C(O)) and ψ (N-Cα-C(O)-N) torsion angles (which are affected by steric hindrance of residues side-chains), of several protein structures with resolutions equal or better to 2.0 Å and R_{factor} not greater than 20%. The plot is divided in four different colour-coded areas: the generally favoured (blue); generally allowed (cyan); glycine favoured (orange); and glycine allowed (light orange). A good model is expected to have more than 98% in the favoured regions. Residues in "disallowed" areas may highlight poor areas of the structure, refinement/modelling errors, or interesting residues relevant to the protein folding or catalytic activity (adapted from http://www.ccp4.ac.uk).

Several crystallographic validation programs are available either in software packages or on Internet servers. In the present studies, the programs *WHATCHECK* (Hooft *et al*. 1996), *PROCHECK* (Laskowski *et al*. 1993) and *MOLPROBITY* (Davis *et al*. 2007) were used.

Bibliographic References

Afonine, P. V.; Grosse-Kunstleve, R. W. and Adams, P. D. (2005) "A robust bulk-solvent correction and anisotropic scaling procedure." *Acta Crystallogr D Biol Crystallogr*, 61: 850-855.

Agarkar, V. B.; Kimani, S. W.; Cowan, D. A., Sayed, M. F. and Sewell, B. T. (2006) "The quaternary structure of the amidase from *Geobacillus pallidus* RAPc8 is revealed by its crystal packing". *Acta Crystallogr* Sect.F, 62: 1174-1178.

Ahmad, F. and Hughes, J.B. (2002) "Reactivity of partially reduced arylhydroxylamine and nitrosoarene metabolites of 2,4,6-trinitrotoluene (TNT) toward biomass and humic acids". *Environ Sci Technol*, 36: 4370–4381.

Ahn, K.; Johnson, D. S.; Mileni, M.; Beidler, D.; Long, J. Z.; Mckinney, M. K.; Weerapana, E.; Sadagopan, N.; Liimatta, M.; Smith, S. E.; Lazerwith, S.; Stiff, C.; Kamtekar, S.; Bhattacharya, K.; Zhang, Y.; Swaney, S.; Vanbecelaere, K.; Stevens, R. C. and Cravatt, B. F. (2009) "Discovery and Characterisation of a Highly Selective Faah Inhibitor that Reduces Inflammatory Pain". *Chem Biol*, 16: 411-420

Andrade, J.; Karmalik, A.; Carrondo, M. A. and Frazão, C. (2007) "Structure of Amidase from *Pseudomonas aeruginosa* showing a trapped acyl transfer reaction intermediate state". *J Biol Chem*, 282: 19598-19605.

Andreeva, A.; Howorth, D.; Chandonia, J. M.; Brenner, S. E.; Hubbard, T. J.; Chothia, C. and Murzin, A. G. (2008) "Data growth and its impact on the SCOP database: new developments". *Nucleic Acids Res*, 36: 419-425.

Anlezark, G. M.; Melton, R.G.; Sherwood, R.F.; Coles, B.; Friedlos, F. and Knox, R. J. (1992) "The Bioactivation of 5-(aziridin-1-yl)-2, 4-dinitrobenzamide (CB1954) - I. Purification and properties of a nitroreductase enzyme from *Escherichia coli* – a potential enzyme for antibody directed enzyme prodrug therapy (ADEPT)". *Biochem Pharmacol*, 44: 2289–95.

Anlezark, G. M.; Vaughan, T.; Fashola-Stone, E.; Michael, N. P.; Murdoch, H.; Sims, M. A.; Stubbs, S.; Wigley, S. and Minton, N. P. (2002) "*Bacillus amyloliquefaciens* orthologue of *Bacillus subtilis* ywrO encodes a nitroreductase enzyme which activates the prodrug CB1954". *Microbiology*, 148: 297-306.

Aragão, D., Fialho, A. M.; Marques, A. R.; Mitchell, E. P.; Sá-Correia, I. and Frazão, C. (2007) "The complex of *Sphingomonas elodea* ATCC 31461 glucose-1-phosphate uridylyltransferase with glucose-1-phosphate reveals a novel quaternary structure, unique among nucleoside diphosphate-sugar pyrophosphorylase members". *J Bacteriol*, 189(12): 4520-8.

Arrecubieta, C., Garcia, E. and López, R. (1996) "Demonstration of UDP-glucose dehydrogenase activity in cell extracts of *Escherichia coli* expressing the pneumococcal cap3A gene required for the synthesis of type 3 capsular polysaccharide". *J Bacteriol*, 178(10): 2971-4.

Bajaj, I. B.; Survase, S. A.; Saudagar, P. S. and Singhal, R. S. (2007) "Gellan Gum: Fermentative Production, Downstream Processing and Applications". *Food Technol Biotechnol*, 45 (4): 341-354.

Bar-Peled, M., Griffith, C. L.; Ory, J. J. and Doering, T. L. (2004) "Biosynthesis of UDP-GlcA, a key metabolite for capsular polysaccharide synthesis in the pathogenic fungus *Cryptococcus neoformans*". *Biochem J*, 381(Pt 1): 131-6.

Bayman, P. and Radkar, G.V. (1997) "Transformation and tolerance of TNT (2,4,6-trinitrotoluene) by fungi". *International Biodeterioration and Biodegradation*, 39: 45–53.

Bellamacina, C. R. (1996) "The nicotinamide dinucleotide binding motif: a comparison of nucleotide binding proteins". *FASEB Journal*, 10: 1257-1269.

Bergfors, T. M. (1999) "Crystallisation Strategies." In *Protein crystallisation: techniques, strategies, and tips: a laboratory manual*. Terese. M. Bergfors (ed), Internat'l University Line, 69-76.

Berman, H. M.; Westbrook, J.; Feng, Z.; Gilliland, G.; Bhat, T. N.; Weissig, H.; Shindyalov, I. N. and Bourne, P. E. (2000) "The Protein Data Bank". *Nucleic Acids Res*, 28: 235-242.

Blanch, M., Legaz, M. E. and Vicente, C. (2006) "Purification and properties of an unusual UDP-glucose dehydrogenase, NADPH-dependent, from *Xanthomonas albilineans*". *Microbiol Res*, Article in Press.

Blasco, R. and Castillo, F. (1992) "Light-dependent degradation of nitrophenols by the phototrophic bacterium *Rhodobacter capsulatus* E1F1". *Appl Environ Microbiol*, 58: 690-695.

Blasco, R. and Castillo, F. (1993) "Characterisation of a nitrophenol reductase from the phototrophic bacterium *Rhodobacter capsulatus* E1F1". *Appl Environ Microbiol*, 59: 1774-1778.

Blasco, R. and Castillo, F. (1997) "Characterisation of 2,4-Dinitrophenol Uptake by *Rhodobacter capsulatus*". Pesticide Biochemistry and Physiology, 58: 1–6.

Blow, D. (2002) "Outline of Crystallography for biologists." *Oxford University Press*.

Blundell, T. L. and Johnson, L. N. (1976) "Protein Crystallography". Academic Press.

Bonanno, J. B.; Freeman, J.; Bain, K. T.; Chang, S.; Sampathkumar, P.; Wasserman, S.; Sauder, J. M.; Burley, S. K. & Almo, S. C. (2009). "Crystal Structure of UDP-glucose 6-dehydrogenase from *Porphyromonas gingivalis* bound to product UDP-glucuronate": New York SGX Research Centre for Structural Genomics (NYSGXRG).

Bond, C. S. (2003) "TopDraw: a sketchpad for protein structure topology cartoons". *Bioinformatics*, 19: 311-312.

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

Bottoms, C. A., Smith, P. E. and Tanner, J. J. (2002) "A structurally conserved water molecule in Rossmann dinucleotide-binding proteins." *Protein Sci*, 11: 2125-2137.

Bracey, M. H.; Hanson, M. A.; Masuda, K. R.; Stevens, R. C. and Cravatt, B. F. (2002) "Structural Adaptations in a Membrane Enzyme That Terminates Endocannabinoid Signaling". *Science*, 298: 1793-1796.

Bradford, M. M. (1976) "A rapid and sensitive method for the quantification of microgram quantities of protein utilizing the 521 principle of protein-dye binding." *Anal Biochem*, 72: 248-254.

Brady, B. L. (1969) "The acylamide amidohydrolase of *Candida utilis*: purification and properties". *Journal of General Microbiology*, 59: 47-55.

Brenner, C. (2002) "Catalysis in the nitrilase superfamily". *Curr Opin Struct Biol*, 12: 775-782.

Broadbent, J. R.; McMahon, D. J.; Welker, D. L.; Oberg, C. J. and Moineau, S. (2003) "Biochemistry, Genetics and Applications of Exopolysaccharide Production in *Streptococcus thermophilus*: A review". *J Dairy Sci*, 86: 407-423.

Brunger, A. T. (1992) "Free R value – a novel statiscal quantity for assessing the accuracy of crystal structures". *Nature*, 355: 472-475.

Brunger, A. T. and Rice, L. M. (1997) "Crystallographic refinement by simulated annealing". *Methods Enzymol*, 277: 243-269.

Bryant, C. & DeLuca, M. (1991) "Purification and characterisation of an oxygen-insensitive NAD(P)H nitroreductase from *Enterobacter cloacae*". *J Biol Chem*, 266: 4119–4125.

Bryant, C.; Hubbard, L. and McElroy, W. D. (1991) "Cloning, nucleotide sequence, and expression of the nitroreductase gene from *Enterobacter cloacae.*" *J Biol Chem*, 266: 4126–4130.

Bryant, C.; McCalla, D. R.; Leeksma, M. and Laneuville, P. (1981) "Type I nitroreductases of *Escherichia coli.*" *Can J Microbiol*, 27: 853–859.

Bylund, J.; Burgess, L. A.; Cescutti, P.; Ernst, R. K. and Speert, D. P. (2006) "Exopolysaccharides from *Burkholderia cenocepacia* inhibit neutrophil chemotaxis and scavenge reactive oxygen species". *J Biol Chem*, 281: 2526–2532.

Caballero, A.; Esteve-Núñez, A.; Zystra, G. J. and Ramos, J. L. (2005a) "Assimilation of nitrogen from nitrite and trinitrotoluene in *Pseudomonas putida* JLR11." *J Bacteriol*, 187: 396-399.

Caballero, A.; Lázaro, A.; Ramos, J. L. and Esteve-Núñez, A. (2005b) "PnrA, a new nitroreductase-family enzyme in the TNT degrading strain *Pseudomonas putida* JLR11." *Environ Microbiol*, 7: 1211–1219.

Cai, G.; Zhu, S. C.; Yang, S.; Zhao, G. P. and Jiang, W. H. (2004) "Cloning, overexpression, and characterisation of a novel thermostable penicillin G acylase from *Achromobacter xylosoxidans*: Probing the molecular basis for its high thermostability". *Appl Env Microbio*, 70 (5): 2764-2770.

Cai, G.; Zhu, S.; Wang, X. and Jiang, W. (2005) "Cloning, sequence analysis and expression of the gene encoding a novel wide-spectrum amidase belonging to the amidase signature superfamily from *Achromobacter xylosoxidans*". *FEMS Microbiol Lett*, 249:15-21.

Campbell, R. E. and Tanner, M. E. (1999) "UDP-Glucose Analogues as Inhibitors and Mechanistic Probes of UDP-Glucose Dehydrogenase", *J Org Chem*, 64: 9487-9492.

Campbell, R. E.; Mosimann, S. C.; van De Rijn, I.; Tanner, M. E. and Strynadka, N. C. (2000) "The first structure of UDP-glucose dehydrogenase reveals the catalytic residues necessary for the two-fold oxidation". *Biochemistry*, 39: 7012-7023.

Campbell, R. E.; Sala, R. F.; van de Rijn, I. and Tanner, M. E. (1997) "Properties and kinetic analysis of UDP-glucose dehydrogenase from group A streptococci. Irreversible inhibition by UDP-chloroacetol." *J Biol Chem*, 272 (6): 3416-3422.

CCP4 (1994) "Collaborative Computational Project, N.4" *Acta Crystallogr Sect D Biol Crystallogr*, 50: 760-763.

Cerantola, S.; Bounery, J.-D.; Segonds, C.; Marty, N. and Montrozier, H. (2000) "Exopolysaccharide production by mucoid and non-mucoid strains of *Burkholderia cepacia*". *FEMS Microbio Lett*, 185(2): 243-6.

Cerantola, S.; Lemassu-Jacquier, A. and Montrozier, H. (1999) "Structural elucidation of a novel exopolysaccharide produced by a mucoid clinical isolate of *Burkholderia cepacia*. Characterisation of a trisubstituted glucuronic acid residue in a heptasaccharide repeating unit." *Eur J Biochem*, 260: 373-383.

Cescutti, P.; Bosco, M.; Picotti, F.; Impallomeni, G.; Leitão, J. H.; Richau, J. A. and Sá-Correia, I. (2000) "Structural study of the exopolysaccharide produced by a clinical isolate of *Burkholderia cepacia"*. *Biochem Biophys Res Comm*, 273: 1088-1094.

Chaignon, P.; Cortial, S.; Ventura, A. P.; Lopes, P.; Halgand, F.; Laprevote, O. and Ouazzani, J. (2006) "Purification and identification of a *Bacillus* nitroreductase: Potential use in 3,5-DNBTF biosensoring system". *Enzyme and Microbial Technology* 39: 1499-1506.

Chakuad, A.; Egger, S.; Yue, W. W.; Guo, K.; Sethi, R.; Filippakopoulos, P.; Muniz, J. R. C.; von Delft, F.; Bountra, C.; Arrowsmith, C. H.; Weigelt, J.; Edwards, A. M.; Kavanagh, K. L.; Nidetzky, B. and Oppermann, U. (2009) " Structure of Human UDP-glucose dehydrogenase Glu161Gln, in complex with thiohemiacetal intermediate", Structural Genomics Consortium (SGC). (To be published)

Chandrasekaran, R. and Radha, A. (1995) "Molecular architectures and functional properties of gellan gum and related polysaccharides", *Trends in Food Science & Technology*, 6: 143-148.

Chang, K. W.; Weng, S. F. and Tseng, Y. H. (2001) "UDP-glucose dehydrogenase gene of *Xanthomonas campestris* is required for virulence." *Biochem Biophys Res Commun*, 287: 550-555.

Chebrou, H.; Bigey, F.; Arnaud, A. and Galzy, P. (1996) "Study of the amidase signature group." *Biochimica and Biophysica Acta*, 1298: 285-293.

Chen, J.; Dai, R. J.; Tong, B.; Xiao, S. Y. and Meng, W. (2007) "Reduction of 4-nitrophenol catalyzed by nitroreductase." *Chinese Chemical Letters*, 18: 10-12.

Chi, Z. and Fang, Y. (2005) "Exopolysaccharides from marine bacteria." *J Ocean Univer China*, 4(1): 67-74.

Choi, J.-W.; Lee, J.; Nishi, K.; Kim, Y.-S.; Jung, C.-H. and Kim, J.-S. (2008) "Crystal Structure of a Minimal Nitroreductase, ydjA, from *Escherichia coli* K12 with and without FMN Cofactor". *J Mol Biol*, 377: 258-267.

Ciardelli, G.; Chiono, V.; Vozzi, G.; Pracella, M.; Ahluwalia, A.; Barbani, N.; Cristallini, C. and Giusti, P. (2005) "Blends of poly- (epsilon-caprolactone) and polysaccharides in tissue engineering applications". *Biomacromolecules*, 6: 1961-1976.

Cilia, E.; Fabbri, A.; Uriani, M.; Scialdone, G. G. and Ammendola, S. (2005) "The signature amidase from *Sulfolobus solfataricus* belongs to the CX3C subgroup of enzymes cleaving both amides and nitriles. Ser195 and Cys145 are predicted to be the active site nucleophiles". *FEBS J*, 272: 4716-4724.

Conway, B.-A. D.; Chu, K. K.; Bylund, J.; Altman, E. and Speert, D. P. (2004) "Production of Exopolysaccharide by *Burkholderia cenocepacia*: Results in Altered Cell-Surface Interactions and Altered Bacterial Clearance in Mice". *Journal of Infectious Diseases*, 190: 957-66.

Cowtan, K. (1994) "DM: An automated procedure for phase improvement by density modification". *Jnt CCP4 ESF-EACBM Newsl Protein Crystallogr*, 31: 34-38.

Cowtan, K. (1998) "Modified phased translation functions and their application to molecular-fragment location." *Acta Crystallogr D Biol Crystallogr*, 54: 750-756.

Cunha, M. V.; Sousa, S. A.; Leitão, J. H.; Moreira, L. M.; Videira, P. A. and Sá-Correia, I. (2004) "Studies on the involvement of the exopolysaccharide produced by cystic fibrosis-associated isolates of the *Burkholderia cepacia* complex in biofilm formation and in persistence of respiratory infections". *J Clin Microbiol*, 42: 3052–3058.

Davis, I. W.; Leaver-Fay, A.; Chen, V. B.; Block, J. N.; Kapral, G. J.; Wang, X.; Murray, L. W.; Arendall III, W. B.; Snoeyink, J.; Richardson, J. S. and Richardson, D. C. (2007) "MolProbity: all-atom contacts and structure validation for proteins and nucleic acids". *Nucleic Acids Research*, 35: 375-383.

De Luca, C.; Lansing, M.; Crescenzi, F.; Martini, I.; Shen, G.-J.; O'Regan, M. and Wong, C.-H. (1996) "Overexpression, one-step purification and characterisation of UDP-glucose dehydrogenase and UDP-N-acetylglucosamine pyrophosphorylase". *Bioorg Med Chem*, 4(1): 131-41.

De Vuyst, L. and Degeest, B. (1999) "Heteropolysaccharides from lactic acid bacteria". *FEMS Microbiol Rev*, 23:153-177.

DeLano, W.L. (2008) "The PyMOL Molecular Graphics System". DeLano Scientific LLC, Palo Alto, California, USA. World Wide Web http://www.pymol.org.

Dillert, R.; Brandt, M.; Fornefett, I.; Siebers, U. and Bahnemann, D. (1995) "Photocatalytic degradation of trinitrotoluene and other nitroaromatic compounds". *Chemosphere*, 30(12): 2333-2341.

Dougherty, B. A. & van de Rijn, I. (1993) "Molecular characterisation of hasB from an operon required for hyaluronic acid synthesis in group A streptococci. Demonstration of UDP-glucose dehydrogenase activity." *J Biol Chem*, 268(10): 7118-7124.

Ducruix, A. and Giege, R. (1992) "Crystallisation of nucleic acids and proteins. A practical approach." *Oxford University Press*.

Dudman, W. F. (1977) "The role of surface polysaccharides in natural environments." In: Sutherland, I.W. (Editor), *Surface Carbohydrates of the Prokaryotic Cell*, Academic Press, Ed. 3: 357-414.

Easley, K. E.; Sommer, B. J.; Boanca, G., Baricki, J. J. and Simpson, M. A. (2007) "Characterisation of human UDP-glucose dehydrogenase reveals critical catalytic roles for lysine 220 and aspartate 280". *Biochemistry*, 46(2): 369-78.

Emptage, C. D.; Knox, R. J.; Danson, M. J. and Hough, D. W. (2009) "Nitroreductase from *Bacillus licheniformis*: A stable enzyme for prodrug activation". *Biochemical Pharmacology*, 77: 21-29.

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

Esteve-Núñez, A.; Caballero, A. and Ramos, J. L. (2001) "Biological degradation of 2,4,6-trinitrotoluene." *Microbiol Mol Biol Rev*, 65: 335-352.

Evans, P. (2006) "Scaling and assessement of data quality". *Acta Crystallogr D Biol Crystallogr*, 62: 72-82.

Feingold, D. S. and Franzen, J. S. (1981) "Amino acid sequence of the tryptic peptide containing the catalytic-site thiol group of bovine liver uridine diphosphate glucose dehydrogenase". *Trends Biochem Sci*, 6: 103-105.

Felsestein, J. (1993) " Phylip (Phylogeny Inference Package) version 3.5c. Distributed by the author. Department of Genetics, University of Washington, Seatlle.

Fetrow, J. S. (1995). "Omega loops: nonregular secondary structures significant in protein function and stability". *FASEB J,* 9: 708-717.

Fialho, A. M.; Moreira L. M.; Granja, A. T.; Popescu, A. O.; Hoffmann, K. and Sá-Correia, I. (2008) "Occurrence, production, and applications of gellan: current state and perspectives". *Appl Microbiol Biotechnol*, 79: 889-900.

Finn, R. D.; Tate, J.; Mistry, J.; Coggill, P. C.; Sammut, S. J.; Hotz, H. R.; Ceric, G.; Forslund, K.; Eddy, S. R.; Sonnhammer, E. L. and Bateman, A. (2008) "The Pfam protein families database". *Nucleic Acids Res*, 36: 281-288.

Fiorella, P. and Spain, J. (1997) "Transformation of 2,4,6-trinitrotoluene by *Pseudomonas pseudoalcaligenes* JS52." *Appl Environ Microbiol*, 63:2007-2201.

Foadi, J.; Woolfson, M. M.; Dodson, E. J.; Wilson, K. S.; Jia-xing, Y. and Chao-de, Z. (2000) "A flexible and efficient procedure for the solution and phase refinement of protein structures". *Acta Crystallogr D Biol Crystallogr*, 56: 1137-1147.

Fonstein, M. and Haselkorn, R. (1993) "Chromossomal structure of *Rhodobacter capsulatus* SB1003: Cosmid encyclopedia and high-resolution physical and genetic map." *Proc Natl Acad Sci*, 90: 2522-2526.

Forouhar, F.; Chen, Y.; Xiao, R.; Ma, L. C.; Byler, T.; Acton, T. B.; Montelione, G. T.; Tong, L. and Hunt, J. F. "Crystal Structure of the putative nitroreductase (SMU 260) in complex with FMN from *Streptococcus mutans*". Northeast Structural Genomics Consortium (NESG). (Not published)

Fournand, D. and Arnaud, A. (2001) "Aliphatic and enetioselective amidases: from hydrolysis to acyl transfer". *J Appl Microbiol*, 9: 381-393.

Fournand, D.; Bigey, F. and Arnaud, A. (1998) "Acyl transfer activity of an amidase from Rhodococcus sp. R312: Formation of a wide range of hydroxamic acids". *Applied and Environmental Microbiology*, 64: 2844-2852.

Fraser, J. A.; Davis, M. A. and Hynes, M. J. (2001) "The Formamidase Gene of *Aspergillus nidulans*: Regulation by Nitrogen Metabolite Repression and Transcriptional Interference by an Overlapping Upstream Gene". *Genetics*, 157: 119-131.

French, S. and Wilson, K. (1978) "On the treatment of negative intensity observations." *Acta Cryst A*, 34: 517-525.

Funk, S. B.; Crawford, D. L. and Crawford R. L. (1996) "Bioremediation of nitroaromatic compounds". In: *Bioremediation: principles and applications*, Don L. Crawford (Editor), Cambridge University Press, 195-208.

Gainey, P. A. and C. F. Phelps (1975) "Interactions of urdine diphosphate glucose dehydrogenase with the inhibitor urdine diphosphate xylose". *Biochem J*, 145(2): 129-34.

Gao, W. Y.; Mitsuya, H.; Driscoll, J. S. and Johns, D. G. (1995) "Enhancement by hydroxyurea of the anti-human immunodeficiency virus type 1 potency of 29-b-fluoro-29, 39-dideoxyadenosine in peripheral blood mononuclear cells". *Biochem Pharmacol*, 50: 274-276.

Garman, E (2003) "Cool' crystals: macromolecular cryocrystallography and radiation damage." *Curr Opin Struct Biol*, 13:545–551.

Garman, E. F. and Owen, R. L. (2006) "Cryocooling and radiation damage in macromolecular crystallography". *Acta Crystallogr D Biol Crystallogr*, 62: 32-47.

Ge, X.; Campbell, R. E.; van de Rijn, I. and Tanner, M. E. (1998) "Covalent adduct formation with a mutated enzyme: Evidence for a thioester intermediate in the reaction catalyzed by UDP-glucose dehydrogenase." *J Am Chem*, 120: 6613-6614.

Ge, X.; Penney, L. C.; van de Rijn, I. and Tanner, M. E. (2004) "Active site residues and mechanism of UDP-glucose dehydrogenase." *Eur J Biochem*, 271(1): 14-22.

Geremia, R. and Rinaudo M. (2004) "Biosynthesis, structure and physical properties of some bacterial polysaccharides". In: Severian Dimitriu, Editor, 2004. *Polysaccharides: Structural diversity and functional versatility*, CRC Press, Ed.2: 411-430.

Giavasis, I.; Harvey, L. M. and McNeil, B. (2000) "Gellan gum." *Crit Rev Biotechnol*, 20:177–211.

Goldberg, J. (2007) "*Polysaccharides of Burkholderia spp*." In: Coenye T. and Vandamme P., Editors, 2007. *Burkholderia* Molecular Microbiology and Genomics, Horizon Scientific Press, 93-110.

Gonçalves, V. M. F.; Reis, A.; Domingues, M. R. M.; Lopes-da-Silva, J. A.; Fialho, A. M.; Moreira, L. M.; Sá-Correia, I. and Coimbra M. A. (2009) "Structural analysis of gellans produced by *Sphingomonas elodea* strains by electrospray tandem mass spectrometry". *Carbohydrate Polymers*. ARTICLE IN PRESS

González-Pérez, M. M.; van Dillewijn, P.; Wittich, R.-M. and Ramos, J. L. (2007) "*Escherichia coli* has multiple enzymes that attack TNT and release nitrogen for growth". *Environmental Microbiology*, 9(6): 1535-1540.

Govan, J. R. and V. Deretic (1996) "Microbial pathogenesis in cystic fibrosis: mucoid *Pseudomonas aeruginosa* and *Burkholderia cepacia*". *Microbiol Rev*, 60(3): 539-74.

Granja, A. T., Popescu, A.; Marques, A. R.; Sá-Correia, I. and Fialho, A. M. (2007) "Biochemical characterisation and phylogenetic analysis of UDP-glucose dehydrogenase from the gellan gum producer *Sphingomonas elodea* ATCC 31461". *Appl Microbiol Biotechnol*, 76(6): 1319-27.

Green, N. K.; Kerr, D. J.; Mautner, V.; Harris, P. A. and Searle, P. F. (2004) "The nitroreductase/CB1954 enzyme-prodrug system." *Methods Mol Med*, 90: 459-477.

Griffith, C. L.; Klutts, J. S.; Zhang, L.; Levery, S. B. and Doering, T. L. (2004) "UDP-glucose dehydrogenase plays multiple roles in the biology of the pathogenic fungus *Cryptococcus neoformans*". *J Biol* Chem, 279(49): 51669-76.

Harding, N. E.; Patel, Y. N. and Coleman R. J. (2004) "Organisation of genes required for gellan polysaccharide biosynthesis in *Sphingomonas elodea* ATCC 31461". *J Ind Microbiol Biotechnol*, 31(2): 70-82.

Hawari, J.; Beaudet, S.; Halasz, A.; Thiboutot, S. and Ampleman, G. (2000) "Microbial degradation of explosives: biotransformation versus mineralisation". *Appl Microbiol Biotechnol*, 54: 605–618.

Haynes, C. A.; Koder, R. L.; Miller, A.-F. and Rodgers, D. S. (2002) "Structures of nitroreductase in the three states. Effect of inhibitor binding and reduction." *J Biol Chem*, 13: 11513–11520.

Hecht, H. J.; Erdmann, H.; Park, H. J.; Sprinzl, M. and Schmid, R. D. (1995) "Crystal structure of NADH oxidase from *Thermus thermophilus*". *Nat Struct Biol*, 2 (12): 1109-14.

Hempel, J.; Perozich, J.; Romovacek, H.; Hinich, A.; Kuo, I. and Feingold, D. S. (1994) "UDP-glucose dehydrogenase from bovine liver: primary structure and relationship to other dehydrogenases". *Protein Sci*, 3(7): 1074-80.

Herasimenka, Y.; Cescutti, P.; Impallomeni, G.; Campana, S.; Taccetti, G.; Ravenni, N.; Zanetti, F. and Rizzo, R. (2007) "Exopolysaccharides produced by clinical strains belonging to the *Burkholderia cepacia* complex". *Journal of Cystic Fibrosis*, 6: 145-152.

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

Higson, F. K. (1992) "Microbial Degradation of Nitroaromatic compounds." In: *Advances in Applied Microbiology*. Saul L. Neidleman, Allen I Laskin (Editors), Academic Press, 1997, 1-21.

Hirooka, T.; Nagase, H.; Hirata, K. and Miyamoto, K. (2006) "Degradation of 2,4-dinitrophenol by a mixed culture of photoautotrophic microorganisms." *Biochemical Engineering Journal*, 29: 157-162.

Holm, L.; Kääriäinen, S.; Rosenström, P. and Schenkel, A. (2008) "Searching protein structure databases with DaliLite v.3." *Bioinformatics*, 24(23): 2780-2781.

Holmes, L. B. (1996) "Hydroxamic acid: a potential human teratogen that could be recommended to treat ureaplasma". *Teratology*, 53: 227-229.

Holton, J. M. (2009) "A beginner's guide to radiation damage". *J Synchrotron Rad*, 16:133-142.

Honeycutt, M. E.; Jarvis, A S. and McFarland, V. A. (1996) "Cytotoxicity and mutagenicity of 2,4,6-trinitrotoluene and its metabolites". *Ecotoxicol Environ Saf*, 35: 282-287.

Hooft, R. W. W.; Vriend, G.; Sander, C. and Abola E. E. (1996) "Errors in protein structures." *Nature*, 381: 272-272.

Hughes, J. E.; Stewart, J.; Barclay, G. R. and Govan J. R. W. (1997) "Priming of neutrophil respiratory burst activity by lipopolysaccharide from *Burkholderia cepacia.*" *Infect Immun*, 65(10): 4281-7.

Huh, J. W.; Yoon, H. Y.; Lee, H.-J.; Choi, W.-B.; Yang, S.-J. and Cho, S.-W. (2004) "Importance of Gly-13 for the coenzyme binding of human UDP-glucose dehydrogenase". *J Biol Chem*, 279(36): 37491-8.

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

Hung, R.-J.; Chien, H.-S.; Lin, R.-Z.; Lin, C.-T., Vatsyayan, J.; Peng, H.-L. and Chang, H.-Y. (2007b) "Comparative Analysis of Two UDP-glucose Dehydrogenases in *Pseudomonas aeruginosa* PAO1". *The Journal of Biological Chemistry*, 282 (24): 17738-17748.

Hwang, H. Y. and Horvitz, H. R. (2002) "The *Caenorhabditis elegans* vulval morphogenesis gene sqv-4 encodes a UDP-glucose dehydrogenase that is temporally and spatially regulated". *Proc Natl Acad Sci*, 99: 14224-14229.

Hynes, M. J. and Pateman, J. A. (1970a) "The use of amides as nitrogen sources by *Aspergillus nidulans*". *Journal of General Microbiology* 63: 317-324.

Hynes, M. J. and Pateman, J. A. (1970b) "The genetic analysis of regulation of amidase synthesis in *Aspergillus nidulans*. II. Mutants resistant to fluoroacetamide." *Molecular and General Genetics*, 108: 107-116.

Iyer, A.; Mody, K. H. and Jha, B. (2004) "Accumulation of hexavalent chromium by an exopolysaccharide producing marine *Enterobacter cloaceae*." *Mar Poll Bull*, 49: 974-977.

Jansson, B.; Hagerstrom, H.; Fransen, N.; Edsman, K. and Björk, E. (2005) "The influence of gellan gum on the transfer of fluorescein dextran across rat nasal epithelium in vivo." *Eur J Pharm Biopharm*, 59(3): 557-64.

Jay, A. J.; Colquhoun, I. J.; Ridout, M. J.; Brownsey, G. J.; Morris, V. J.; Fialho, A. M.; Leitão, J. H. and Sá-Correia, I. (1998) "Analysis of structure and function of gellans with different substitution patterns". *Carbohydr Polym*, 35:179–188.

Johansson, E.; Parkinson, G. N.; Denny, W. A. and Neidle, S. (2003) " Studies on the nitroreductase prodrug-activating system. Crystal Structure of complexes with the inhibitor dicoumarol and dinitrobenzamide prodrugs and of the enzyme active form." *J Med Chem*, 46 (19): 4009-4020.

Johnson, G. R. and Spain, J. C. (2003) "Evolution of catabolic pathways for synthetic compounds: bacterial pathways for degradation of 2,4-dinitrotoluene and nitrobenzene". *Appl Microbiol Biotechnol*, 62: 110-123.

Joint Center for Structural Genomics (JCSG) "Crystal structure of putative nitroreductase ydfN (2632848) from *Bacillus subtilis* at 1.65 A resolution." (To be published)

Joint Center for Structural Genomics (JCSG) "Crystal structure of Glutamyl-tRNA (Gln) amidotransferase subunit A (tm1272) from *Thermatoga maritima* at 1.80 A resolution." (To be published)

Jolly, L.; Vincent, S. J.; Duboc, P. and Neeser, J. R. (2002) "Exploiting expolysaccharides from lactic acid bacteria". *Antonie Van Leeuwenhoek*, 82(1-4): 367-74.

Jones, A. M.; Dodd, M. E. and Webb, A. K. (2001) "*Burkholderia cepacia*: current clinical issues, environmental controversies and ethical dilemmas". *Eur Respir J*, 17(2): 295-301.

Kahng, H.-Y.; Lee, B.-U.; Cho, Y.-S. and Oh, K.-H. (2007) "Purification and Characterisation of the NAD(P)H-nitroreductase for the Catabolism of 2,4,6-trinitrotoluene (TNT) in *Pseudomonas* sp. HK-6". *Biotechnology and Bioprocess Engineering*, 12: 433-440.

Kammermeier-Steinke, D.; Schwarz, A.; Wandrey, C. and Kula, M.-R. (1993) "Studies on the substrate specificity of a peptide amidase partially purified from orange flavedo". *Enzyme and Microbial Technology*, 15: 764-769.

Kanekar, P.; Doudpure, P. and Sarnaik, S. (2003) "Biodegradation of nitroexplosives." *Indian Journal of Experimental Biology*, 41: 991-1001.

Kaneko, T. and Kang, K. S. (1979) "Agar-like polysaccharide produced by a *Pseudomonas* species: Taxonomical studies." *Abstracts of the 79th Annual Meeting of the American Society for Microbiology*, Washington DC, USA.

Kaneko, T.; Katoh, T.; Sato, S.; Nakamura, A.; Asamizu, E. and Tabata, S. (2000) "Structural analysis of *Arabidopsis thaliana* chromosome 3.II. Sequence features of the 4,251,695 bp regions covered by 90 P1, TAC and BAC clones." *DNA Res*, 7(3): 217-221.

Kantardjieff, K. and Rupp, B. (2003) "Matthews coefficient probabilities: Improved estimates for unit cell contents of proteins, DNA, and protein-nucleic acid complex crystals". *Protein Sci*, 12: 1865-1871.

Kärkönen, A. (2005) "Biosynthesis of UDP-GlcA: Via UDPGDH or the myo-inositol oxidation pathway?" *Plant Biosystems*, 139 (1): 46-49.

Kärkönen, A. and Fry, S. C. (2006) "Novel characteristics of UDP-glucose dehydrogenase activities in maize: non-involvement of alcohol dehydrogenases in cell wall polysaccharide biosynthesis". *Planta*, 223(4): 858-70.

Kärkönen, A.; Murigneux, A.; Martinant, J.-P.; Pepey, E.; Tatout, C.; dudley, B. J. and Fry, S. C. (2005) "UDP-glucose dehydrogenases of maize: a role in cell wall pentose biosynthesis". *Biochem J*, 391(Pt 2): 409-15.

Kavanagh, K. L.; Guo, K.; Bunkoczi, P.; Savitsky, E.; Pilka, E.; Bhatia, C.; Smee, C.; Berridge, G.; Von Delft, F.; Wiegelt, J.; Arrowsmith, C.; Sundstrom, M.; Edwards, A. and Oppermann, U. (2007a) "Structure of human UDP-glucose dehydrogenase complexed with NADH and UDP-glucose". Structural Genomics Consortium (SGC). (To be published)

Kavanagh, K. L.; Guo, K.; Bunkoczl, P.; Savitsky, E.; Pilka, E.; Bhatia, C.; Smee, C.; Berridge, G.; Von Delft, F.; Wiegelt, J.; Arrowsmith, C.; Sundstrom, M.; Edwards, A. and Oppermann, U. (2007b) "Structure of human UDP-glucose dehydrogenase complexed with NADH and UDP-glucuronate". Structural Genomics Consortium (SGC). (To be published)

Khan, T. A.; Bhadraand, R. and Hughes, J. (1997) "Anaerobic transformation of 2,4,6-TNT and related nitroaromatic compounds by *Clostridium acetobutylicum*". *J Ind Microbiol Biotechnol*, 18: 198–203.

Kikuchi, T.; Itoh, F.; Toyota, M.; Suzuki, H.; Yamamoto, H.; Fujita, M.; Hosokawa, M. and Imai, K. (2002) "Aberrant methylation and histone deacetylation of cyclooxygenase 2 in gastric cancer". *Int J Cancer*, 97: 272-277.

Kim, H.-Y. and Song, H.-G. (2005) "Purification and characterisation of NAD(P)H-dependent nitroreductase I from *Klebsiella* sp. C1 and enzymatic transformation of 2,4,6-trinitrotoluene". *Appl Microbiol Biotechnol*, 68: 766-773.

Kim, H.-Y.; Bennett, G. N. and Song, H.-G. (2003) "Degradation of 2,4,6-trinitrotoluene by *Klebsiella* sp. isolated from activated sludge". *Biotechnol Lett*, 24: 2023–2028.

Kim, S.-H. and Oriel, P. (2000) "Cloning and expression of the nitrile hydratase and amidase genes from *Bacillus* sp. BR449 into *Escherichia coli*". *Enzyme and Microbial Technology*, 27: 492-501.

Kim, Y.; Duggan, E.; Clancy, S. and Joachimiak, A. (2006) "The Crystal Structure of the Putative NAD(P)H-Flavin Oxidoreductase from *Streptococcus pyogenes* M1 GAS". (To be published)

Kimura, N.; Shinozaki, Y.; Suwa, Y. and Urushigawa, Y. (2000) "Phylogenetic and phenotypic relationships of microorganisms that degrade uncoupler compound, 2,4-dinitrophenol". *J Gen Appl Microbiol*, 46: 317–322.

Kirner, S.; Hammer, P. E.; Hill, D. S.; Altman, A.; Fischer, I.; Weislo, L. J.; Lanahan, M.; van Pee, K. H. and Ligon, J. M. (1998) "Functions encoded by pyrrolnitrin biosynthetic genes from *Pseudomonas fluorescens*". *J Bacteriol*, 180: 1939-1943.

Klinghammer, M. and Tenhaken, R. (2007) "Genome-wide analysis of the UDP-glucose dehydrogenase gene family in *Arabidopsis*, a key enzyme for matrix polysaccharides in cell walls". *Journal of Experimental Botany*, 58 (13): 3609-3621.

Kobayashi, M.; Goda, M. and Shimizu, S. (1998) "The catalytic mechanism of amidase also involves nitrile hydrolysis." *FEBS Letters*, 439: 325-328.

Kobayashi, M.; Komeda, H.; Nagasawa, T.; Nishiyama, M.; Horinouchi, S.; Beppu, T.; Yamada, H. and Shimizu, S. (1993) "Amidase coupled with low molecular-mass nitrile hydratase from Rh. rhodochrous J1". *European Journal of Biochemistry*, 217:327-336.

Kobori, T.; Sasaki, H.; Lee, W. C.; Zenno, S.; Saigo, K.; Murphy, M. E. P. and Tanokura, M. (2001) "Structure and site-directed mutagenesis of a flavoprotein from *Escherichia coli* that reduces nitrocompounds: alteration of pyridine nucleotide binding by a single amino acid substitution". *J Biol Chem*, 276: 2816-2823.

Koder, R. L. and Miller, A. F. (1998) "Overexpression, isotopic labelling and spectral characterisation of *Enterobacter cloacae* nitroreductase". *Protein Expr Purif* , 13: 53–60.

Koder, R. L.; Oyedele, O. and Miller, A. F. (2001) "Retro-nitroreductase, a putative evolutionary precursor to *Enterobacter cloacae* strain 96-3 nitrorreductase." *Antioxid Redox Signal*, 3: 747-755.

Koide, Y.; Uchino, M. and Yamada, K. (1987) "Studies of collectors. IX. The flotation of a trace amount of uranium by using 2-(alkylamino)propionohydroxamic acid and cotelomer-type surfactants bearing hydroxyaminocarbonyl and pyridyl groups". *Bull Chem Soc Jpn*, 60: 3477-3483.

Koike, H.; Sasaki, H.; Kobori, T.; Zenno, S.; Saigo, K.; Murphy, M. E. P.; Adman, E. T. and Tanokura, M. (1998) "1.8Å Crystal structure of the major NAD(P)H:FMN oxidoreductase of a bioluminescent bacterium, *Vibrio fisheri*: overal structure, cofactor and substrate-binding, and comparison with related flavoproteins." *J Mol Biol*, 280: 259-273.

Komatsu, Y.; Tomizaki, K. Y.; Tsukamoto, M.; Kato, T.; Nishino, N.; Sato, S.; Yamori, T.; Tsuruo, T.; Furumai, R.; Yoshida, M.; Horinouchi, S. and Hayashi, H. (2001) "Cyclic hydroxamic-acid-containing peptide 31, a potent synthetic histone deacetylase inhibitor with antitumor activity". *Cancer Res*, 61: 4459-4466.

Kong, J.; Lee, H.; Hong, J.; Kang, Y.; Kim, J.; Chang, M. and Bae, S. (1998) "Utilisation of a cell-bound polysaccharide produced by the marine bacterium *Zooglea* sp.: New biomaterial for metal adsorption and enzyme immobilisation". *J Mar Biotechnol*, 6: 99-103.

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

Krissinel, E. and Henrick, K. (2007) "Inference of macromolecular assemblies from crystalline state". *J Mol Biol*, 372: 774-797.

Kulkarni, M. A. and Chaudhari, A. B. (2007) "Microbial remediation of nitro-aromatic compounds: an overview". *Journal of Environmental Management*, 85(2): 496-512.

Kulkarni, M.A. and Chaudhari, A. B. (2006) "Biodegradation of p-nitrophenol by *P. putida*". *Bioresource Biotechnology*, 97: 982-988.

Kumar, A. S.; Mody, K. and Jha, B. (2007) "Bacterial exopolysaccharides – a perception." *Journal of Basic Microbiology*, *47*: 103-117.

Kutty, R. and Bennett, G. N. (2005) "Biochemical characterisation of trinitrotoluene transforming oxygen-insensitive nitroreductases from *Clostridium acetobutylicum* ATCC 824." *Arch Microbiol*, 184: 158-167.

Kutzenko, A. S.; Lamzin, V. S. and Popov, V. O. (1998) "Conserved supersecondary structural motif in NAD-dependent dehydrogenases." *FEBS Lett*, 423 (1): 105-109.

Labahn, J.; Neumann, S.; Buldt, G.; Kula, M. R. and Granzin, J. (2002) "An alternative mechanism for amidase signature enzymes." *J Mol Biol*, 322:1053-1064.

Landstein, D.; Graves, M. V.; Burbank, D. E.; deAngelis, P. and Van Etten, J. L. (1998) "*Chlorella* virus PBCV-1 encodes functional glutamine: fructose-6-phosphate amidotransferase and UDP-glucose dehydrogenase enzymes". *Virology*, 250(2): 388-96.

Lang, D.; Thoma, R.; Henn-Sax, M.; Sterner, R. and Wilmanns, M. (2000) "Structural Evidence for Evolution of the β/α Barrel Scaffold by Gene Duplication and Fusion." *Science*, 289: 1546-1550.

Laskowski, R. A.; MacArthur, M. W.; Moss, D. S. and Thornton, J. M. (1993) "PROCHECK: a program to check the stereochemical quality of protein structures". *J App Cryst*, 26: 283-291.

Lee, B.-U.; Park, S.-C.; Cho, Y.-S.; Kahng, H.-Y. and Oh, K.-H. (2008) "Expression and Characterisation of the TNT Nitroreductase of *Pseudomonas* sp. HK-6 in *Escherichia coli*". *Curr Microbiol*, 56: 386–390.

Lei, B.; Liu, M.; Huang, S and Tu, S.-C. (1994) "*Vibrio harveyi* NADPHflavin oxidoreductase: cloning, sequencing and overexpression of the gene and purification and characterisation of the cloned enzyme". *J Bacteriol*, 176: 3552-3558.

Leid, J. G.; Willson, C. J.; Shirtliff, M. E.; Hassett, D. J.; Parsek, M. R. and Jeffers, A. K. (2005) "The Exopolysaccharide Alginate Protects *Pseudomonas aeruginosa* Biofilm Bacteria from IFN-γ-Mediated Macrophage Killing". *Journal of Immunology*, 7512-7518.

Leslie, A. G. W. (1992) "Recent changes to the MOSFLM package for processing film and image plate data." *Jnt CCP4/ESF-EACBM News Protein Crystallogr*, 26.

Leszczymski, J. F. & Rose, G. D. (1986) "Loops in globular proteins: a novel category of secondary structure." *Science*, 234: 849-55.

Levander, F. and Rådström, P. (2001) "Requirement for phosphoglucomutase in exopolysaccharide biosynthesis in glucose- and lactose-utilizing *Streptococcus thermophilus*". *Appl Environ Microbiol*, 67: 2734-2738.

Li, J.; Kamath, K. and Dwivedi, C. (2001) "Gellan film as an implant for insulin delivery." *J Biomater Appl*, 15:321–343.

Lindhorst, T. K. (2007) "Essentials of Carbohydrate Chemistry and Biochemistry." Ed.3, Wiley-VHC Publications.

Linker, A.; Evans, L. R. and Impallomeni, G. (2001) "The structure of a polysaccharide from infectious strains of *Burkholderia cepacia*." *Carbohydr Res*, 335 (1): 45-54.

Liochev, S. I.; Hausladen, a. and Fridovich, I. (1999) "Nitroreductase A is regulated as a member of the *sox*RS regulon of *Escherichia coli.*" *Proc Natl Acad Sci*, 96: 3537-3539.

Long, F.; Vagin, A.; Young, P.and Murshudov, G. N. (2008) "BALBES: a Molecular Replacement Pipeline". *Acta Crystallogr D Biol Crystallogr*, 64: 125-132.

Lupyan, D.; Leo-Macias, A. and Ortiz, A. R. (2005) "A new progressive-iterative algorithm for multiple structure aligment." *Bioinformatics*, 21 (15): 3255-3263.

Maeda, T.; Nakamura, R.; Kadokami, K. and Ogawa, H. I. (2007) "Relationship between mutagenicity and reactivity or biodegradability for nitroaromatic compounds." *Environmental toxicology and chemistry*, 26(2): 237-41.

Mahenthiralingam, E.; Baldwin, A. and Vandamme, P. (2002) "Burkholderia cepacia complex infection in patients with cystic fibrosis." *J Med Microbiol*, 51: 533-538.

Malashkevich, V. N.; Sauder, J. M.; Burley, S. K.; Almo, S. C. (2009) "Crystal structure of NDP-N-acetyl-D-galactosaminuronic acid dehydrogenase from *Methanosarcina mazei* Go1". (To be published)

Margolls, S. A.; Howell, B. F. and Schaffer, R. (1976) "Purification and Analysis of the Purity of NADH." *Clin Chem*, 22 (8): 1322-1329.

Marques de Oliveira, I.; Pêgas Henriques, J. A. and Bonatto, D. (2007) "In silico identification of a new group of specific bacterial and fungal nitroreductase-like proteins." *Biochem Biophys Res Commu*, 355: 919-925.

Martins, L. O. and Sá-Correia, I. (1993) "Temperature profile of gellan gum synthesis and activities of biosynthetic enzymes, *Biotechnol Appl Biochem*, 20: 385–395.

Marvin-Sikkema, F. D. and Bont, J. A. (1994) "Degradation of nitroaromatic compounds by microorganisms." *Applied microbiology and biotechnology*, 42(4): 499-507.

Matthews, B. W. (1968) "The Solvent Content of Protein Crystals". *Mol Biol*, 33: 491-497.

Mayaux, J.-F.; Cerbelaux, E.; Soubrier, F.; Yeh, P.; Blanche, F. and Pétré, D. (1991) "Purification, cloning, and primary structure of a new enantiomer-selective amidase from a *Rhodococcus* strain: structural evidence for a conserved genetic coupling with nitrile hydratase". *Journal of Bacteriology*, 173: 6694-6704.

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

McFee, R. B.; Caraccio, T. R.; McGuigan, M. A.; reynolds, S. A. and Bollomeyer, P. (2002) "Dying to be thin—hyperpyrexia and weight loss: a case report of a dinitrophenol (DNP) related fatality". *Journal of Toxicology Clinical Toxicology*, 57: 622.

McRee, D. E. (1999) "XtalView/Xfit – A Versatile Program for Manipulating Atomic Coordinates and Electron Density." *J Structural Biology*, 125*:* 156-165.

Miksic, J. R. and Brown, P. R. (1978) "Reactions of Reduced Nicotinamide Adenine Dinucleotide in Acid: Studies by Reversed-Phase High-pressure Liquid Chromatography." *Biochem*, 17 (11): 2234-2238.

Mileni, M.; Johnson, D. S.; Wang, Z.; Everdeen, D. S.; Liimatta, M.; Pabst, B.; Bhattacharya, K.; Nugent, R. A.; Kamtekar, S.; Cravatt, B. F.; Ahn, K. and Stevens, R. C. (2008) "Structure-Guided Inhibitor Design for Human Faah by Interspecies Active Site Conversion". *Proc Natl Acad Sci* USA, 105: 12820-12824.

Mijakovic, I.; Petranovic, D. & Deutscher, J. (2004) "How Tyrosine Phosphorylation Affects the UDP-Glucose Dehydrogenase Activity of Bacillus subtilis YwqF." *J Mol Microbiol Biotechnol*, 8:19–25

Moreira, L. M.; Videira, P. A.; Sousa, S. A.; Leitão, J. H.; Cunha, M. V. and Sá-Correia, I. (2003) "Identification and physical organisation of the gene cluster involved in the

biosynthesis of *Burkholderia cepacia* complex exopolysaccharide." *Biochem Biophys Res Commun*, 312(2): 323-33.

Morin, A. (1998). "Screening of polysaccharide-producing microorganisms, factors influencing the production and recovery of microbial polysaccharides". In: Severian Dumitriu, Editor, 1998, *Polysaccharides - Structural Diversity and Functional Versatility*, Marcel Dekker Inc. Publication, 275-296.

Moslemy, P.; Guiot, S. R. and Neufeld, R. J. (2004) "Activated sludge encapsulation in gellan gum microbeads for gasoline biodegradation." *Bioprocess Biosyst Eng*, 26:197–204.

Moslemy, P.; Neufeld, R. J.; Millette, D. and Guiot, S. R. (2003) "Transport of gellan gum microbeads through sand: an experimental evaluation for encapsulated cell bioaugmentation". *J Environ Manage*, 69: 249-59.

Moyrand, F. and Janbon, G. (2004) "UGD1, encoding the *Cryptococcus neoformans* UDP-glucose dehydrogenase, is essential for growth at 37 degrees C and for capsule biosynthesis". *Eukaryot Cell*, 3(6): 1601-8.

Murshudov, G. N.; Vagin, A. A. and Dodson, E. J. (1997) "Refinement of Macromolecular Structures by the Maximum-Likelihood method." *Acta Crystallogr D Biol Crystallogr*, 53: 240-255.

Murzin, A. G.; Brenner, S. E.; Hubbard, T. and Chothia, C. (1995) "SCOP: a structural classification of proteins database for the investigation of sequences and structures". *J Mol Biol*, 247: 536-540.

Nakamura, A.; Yao, M.; Chimnaronk, S.; Sakai, N. and Tanaka, I. (2006) "Ammonia channel couples glutaminase with transamidase reactions in GatCAB." *Science*, 312: 1954-1958.

Nelsestuen, G. L. and Kirkwood, S. (1971) "The mechanism of action of uridine diphosphoglucose dehydrogenase. Uridine diphosphohexodialdoses as intermediates." *J Biol Chem*, 246: 3824-3834.

Ohmori, T.; Hagiwara, S.; Ueda, A.; Minoda, Y. and Yamada, K. (1978) "Production of pyoluteorin and its derivatives from n-paraffin by *Pseudomonas aeruginosa* S10B2." *Agric Biol Chem*, 42: 2031–2036.

Ohshima, T.; Sakuraba, H.; Ebihara, A.; Kanagawa, M.; Nakagawa, N.; Kuroishi, C.; Satoh, S.; Kuramitsu, S. and Yokoyama, S. RIKEN Structural Genomics/Proteomics Initiative (RSGI) "Crystal structure of amidase". (To be published)

Omoto, T.; Uno, Y. and Asai, I. (1999) "The latest technologies for the application of gellan gum". *Progr Colloid Polym Sci*, 114: 123-126.

Ordman, A. B. & Kirkwood, S. (1977) "UDP-glucose dehydrogenase. Kinetics and their mechanistic implications." *Biochim Biophys Acta*, 48: 25-32.

Ortega, X. P.; Cardona, S. T.; Brown, a. R.; Loutet, S. A.; Flannagan, R. s.; Campopiano, D. J.; Govan, J. R. W. and Valvano, M. A. (2007) "A putative gene cluster for aminoarabinose biosynthesis is essential for *Burkholderia cenocepacia* viability". *J Bacteriol*, 189(9): 3639-44.

Otwinowski, Z. and Minor, W. (1997) "Processing of X-ray Diffraction Data Collected in Oscillation Mode". *Methods in Enzymology*, 276: 307-326.

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

Pace, S. A. and Pace, A. (2002) "Dinitrophenol oral ingestion resulting in death." *Journal of Toxicology Clinical Toxicology*, 199:683.

Padda, R. S.; Wang, C.; Hughes, J. B.; Kutty, R. and Bennett, G. N. (2003) "Mutagenicity of nitroaromatic degradation compounds". *Environ Toxic and Chem*, 22(10): 2293-7.

Painter, J. and Merritt E. A., (2006) "TLSMD web server for the generation of multi-group TLS models." *J Appl Cryst*, 39: 109-111.

Papagrigoriou, E.; Salah, E.; Turnbull, A. P.; Smee, C.; Burgess, N.; Gileadi, O.; Von Delft, F.; Gorrec, F.; Arrowsmith, C. H.; Weigelt, J.; Sundstrom, M.; Edwards, A. M. and Oppermann, U. (2006) "Crystal structure of human hydroxyisobutyrate dehydrogenase". (To be published)

Parkinson, G. N.; Skelly, J. V. and Neidle, S. (2000) "Crystal structure of FMN-dependent nitroreductase from *Escherichia coli* B: a prodrug-activating enzyme". *J Med Chem*, 43: 3624-3631.

Pearl, F. M. G.; Lee, D.; Bray, J. E.; Sillitoe, I.; Todd, A. E.; Harrison, A. P.; Thornton, J. M. and Orengo, C. A. (2000) "*Assigning genomic sequences to CATH*". *Nucleic Acids Research*, 28(1): 277-282.

Pérez-Reinado, E.; Blasco, R.; Castillo, F.; Moreno-Vivián, C. and Roldán, M. D. (2005) "Regulation and characterisation of two nitroreductase *nprA* and *nprB* genes of *Rhodobacter capsulatus.*" *Appl Environ Microbiol*, 71: 7643-7649.

Pérez-Reinado, E.; Roldán, M. D.; Castillo, F. and Moreno-Vivián, C. (2008) "The NprA nitroreductase required for 2,4-dinitrophenol reduction in *Rhodobacter capsulatus* is a dihydropteridine reductase." *Environ Microbiol*, 10(11): 3174–3183.

Perrakis, A.; Morris, R. and Lamzin, V. S. (1999) "Automated protein model building combined with iterative structure refinement." *Nat Struct Biol*, 6: 458-63.

Plhackova, K.; Becka, S.; Skrob F. and Kyslik, P. (2003) "Isolation and characterisation of a new strain of *Achromobacter* sp. with β-lactam antibiotic acylase activity." *Appl Microbio Biotech*, 62 (5-6): 507-516.

Ploegman, J. H.; Drent, G.; Kalk, K. H.; Hol, W. G.; Heinrikson, R. L.; Keim, P.; Weng, L. and Russell, J. (1978) "The covalent and tertiary structure of bovine liver rhodanese". *Nature*, 273: 124-129.

Pollock, T. J. (1993) "Gellan-related polysaccharides and the genus *Sphingomonas*". *J Gen Microbiol*, 139:1939-1945.

Race, P. R.; Lovering, A. L.; Green, R. M.; Ossor, A.; White, S. A.; Searle, P. F.; Wrighton, C. W. and Hyde, E. I. (2005) "Structural and mechanistic studies of *Escherichia coli* nitroreductase with the antibiotic nitrofurazone". *J Biol Chem*, 280: 13256–13264.

Race, P. R.; Lovering, A. L.; White, S. A.; Grove, J. I.; Searle, P. F.; Wrighton, C. W. and Hyde, E. I. (2007) "Kinetic and structural characterisation of *Escherichia coli* nitroreductase mutants showing improved efficacy for the prodrug substrate CB1954". *J Mol Biol*, 368: 481-492.

Ramachandran, G. N.; Ramakrishnan, C. and Sasisekharan, V. (1963) "Stereochemistry of polypeptide chain conformations." *J Mol Biol*, 7: 95-99.

Read, R. J. (1986) "Improved Fourier coefficients for maps using phases from partial structures with errors". *Acta Crystallogr* A, 42: 140-49.

Reitz, S.; Alhapel, A.; Essen, L. O. and Pierik, A. J. (2008) "Structural and Kinetic Properties of a beta-hydroxyacid Dehydrogenase Involved in Nicotinate Fermentation." *J Mol Biol*, 382: 802-811.

Richau, J. A.; Leitão, J. H. and Sá-Correia, I. (2000a) "Enzymes leading to the nucleotide sugar precursors for exopolysaccharide synthesis in *Burkholderia cepacia*." *Biochem Biophys Res Commun*, 276(1): 71-6.

Richau, J. A.; Leitão, J. H.; Correia, M.; Lito, L.; Salgado, M. J.; Barreto, C.; Cescutti, P. and Sá-Correia, I. (2000b) "Molecular typing and exopolysaccharide biosynthesis of *Burkholderia cepacia* isolates from a Portuguese cystic fibrosis center". *J Clin Microbiol*, 38(4): 1651-5.

Riès-Kautt, M. (1999) "Strategy 2: An Alternative to Sparse Matrix Screens". In *Protein crystallisation: techniques, strategies, and tips: a laboratory manual*. Terese. M. Bergfors (ed), Internat'l University Line, 91-110.

Robb, F. T.; Maeder, D. L.; Brown, J. R.; Di Ruggiero, J.; Stump, M. D.; Yeh, R. K.; Weiss, R. B. & Dunn, D. M. (2001) "Genomic sequence of hyperthermophile, *Pyrococcus furiosus*: implications for physiology and enzymology." *Meth Enzymol*, 330: 134-157.

Roberts, I.S. (1996) "The biochemistry and genetics of capsular polysaccharide production in bacteria". *Annu Rev Microbiol*, 50: 285–315.

Robertson, B. K. and Jemba, P. K. (2005) "Enhanced bioavailability of sorbed 2,4,6-trinitrotoluene (TNT) by a bacterial consortium". *Chemosphere*, 58: 263-270.

Rocha, J.; Granja, A. T.; Sá-Correia, I. Fialho, A. and Frazão, C. (2010a) "Cloning, expression, purification, crystallisation and preliminary crystallographic studies of UgdG, an UDP-glucose dehydrogenase from *Sphingomonas elodea* ATCC 31461". *Acta Crystallographica* Section F, 66: 69-72.

Rocha, J.; Popescu, A.; Sá-Correia, I. Fialho, A.; and Frazão, C. (2010b) "Cloning, expression, purification, crystallisation and preliminary crystallographic studies of BceC, an UDP-glucose dehydrogenase from *Burkholderia cepacia* IST 408." *Acta Crystallographica Section F,* 66: 269-271.

Rocha, J.; Popescu, A.; Sá-Correia, I.; Fialho, A. and Frazão, C. (2010c) "Structure of *Burkholderia cepacia* IST 408 UDP-glucose dehydrogenase (BceC) and the role of Tyr10 in the intermediate thioester hydrolysis." (To be submitted)

Rocha, J.; Fialho, A.; Sá-Correia, I. and Frazão, C. (2010d) "The three dimensional structure of the *Sphingomonas elodea* ATCC 31461 Uridine-5'-diphosphoglucose dehydrogenase (UgdG)." (To be submitted)

Rocha, J.; Roldán, M. D.; Mitchell, E. and Frazão, C. (2010e) "Crystal Structure of NprA, the major nitroreductase from *Rhodobacter capsulatus* B10 bound to the inhibitor acetate." (To be submitted)

Roldán, M. D.; Blasco, R.; Caballero, F. J. and Castillo, F. J. (1998) "Degradation of p-nitrophenol by the phototrophic bacterium *Rhodobacter capsulatus*." *Arch Microbiol*, 169: 36-42.

Roldán, M. D.; Pérez-Reinado, E.; Castillo, F. and Moreno-Vivián, C. (2008) "Reduction of polynitroaromatic compounds: the bacterial nitroreductases". *FEMS Microbiol Rev*, 32: 474–500.

Roman, E.; Roberts, I.; Lidholt, K. and Kusche-Gullberg, M. (2003) "Overexpression of UDP-glucose dehydrogenase in *Escherichia coli* results in decreased biosynthesis of K5 polysaccharide." *Biochem J*, 374: 767-772.

Ruffing, A. and Chen, R. R. (2006) "Metabolic engineering of microbes for oligosaccharide and polysaccharide synthesis." *Microbial Cell Factories*, 5: 25.

Sá-Correia, I.; Fialho, A. M.; Videira, P.; Moreira, L. M.; Marques, A. R. and Albano, H. (2002) "Gellan gum biosynthesis in *Sphingomonas paucimobilis* ATCC 31461: genes, enzymes and exopolysaccharide production engineering". *J Ind Microbiol Biotechnol*, 29(4): 170-6.

Salamanca-Pinzón, S. G.; Camacho-Carranza, R.; Hernández-Ojeda, S. L. and Espinosa-Aguirre, J. J. (2006) "Nitrocompound activation by cell-free extracts of nitroreductase-proficient *Salmonella typhimurium* strains." *Mutagenesis*, 21: 369-374.

Sali, A. and Blundell, T. L. (1993) "Comparative protein modelling by satisfaction of spatial restraints". *J Mol Biol*, 234: 779-815.

Savadogo, A.; Ouattara, C. A. T.; Savadogo, P. W.; Barro, N.; Ouattara, A. S. and Traoré, A.S. (2004) "Identification of exopolysaccharides-producing lactic acid bacteria from Burkina Faso fermented milk samples". *African Journal of Biotechnology*, 3(3): 189-194.

Schirmer, F.; Ehrt, S. and Hillen, W. (1997) "Expression, inducer spectrum, domain structure, and function of MopR, the regulator of phenol degradation in *Acinetobacter calcoaceticus* NCIB8250." *J Bacteriol*, 179: 1329-1336.

Schneider, T. R. and Sheldrick, G. M. (2002) "Substructure Solution with SHELXD", *Acta Crystallogr D Biol Crystallogr*, 58: 1772-1779.

Scotto d'Abusco, A.; Ammendola, S.; Scandurra, R. and Politi, L. (2001) "Molecular and biochemical characterisation of the recombinant amidase from hyperthermophilic archaeon *Sulfolobus solfataricus*." *Extremophiles*, 5: 183-192.

Scotto d'Abusco, A.; Casadio, R.; Tasco, G.; Giangiacomo, L.; Giartosio, A.; Calamia, V.; Di Marco, S.; Chiaraluce, R.; Consalvi, V.; Scandurra, R and Politi, R. (2005) "Oligomerisation of *Sulfolobus solfataricus* signature amidase is promoted by acidic pH and high temperature". *Archaea*, 1: 411-423.

Sharma, M.; Sharma, N. N. and Bhalla, T. C. (2009) " Amidases: versatile enzymes in nature." *Rev Environ Sci Biotechnol*, 8: 343-366.

Sheldrick, G. M. (2002) "Macromolecular phasing with SHELXE." *Z Kristallogr*, 217: 644-650.

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

Shin, S.; Lee, T.-H.; Ha, N.-C.; Koo, H. M.; Kim, S.-Y.; Lee, H.-S.; Kim, Y. S. and Oh, B.-H. (2002) "Structure of malonamidase E2 reveals a novel Ser-*cis*Ser-Lys catalytic triad in a new serine hydrolase fold that is prevalent in nature". *EMBO J*, 21: 2509-2516.

Shin, S.; Yun, Y. S.; Koo, H. M.; Kim, Y. S.; Choi, K. Y. and Oh, B.-H. (2003) "Characterisation of a novel Ser-*cis*Ser-Lys catalytic triad in comparison with the classical Ser-His-Asp triad". *J Biol Chem*, 278: 24937-24943.

Sieberth, V.; Rigg, G. P.; Roberts, I. S. and Jann, K. (1995) "Expression and characterisation of UDPGlc dehydrogenase (Kfid), which is encoded in the type-specific region 2 of the *Escherichia coli* K5 capsule genes." *J Bacteriol*, 177: 4562-4565.

Silva, E.; Marques, A. R.; Fialho, A. M.; Granja, A. T. and Sá-Correia, I. (2005) "Proteins encoded by *Sphingomonas elodea* ATCC 31461 rmlA and ugpG genes, involved in gellan gum biosynthesis, exhibit both dTDP- and UDP-glucose pyrophosphorylase activities". *Appl Environ Microbiol*, 71(8): 4703-12.

Sist, P.; Cescutti, P.; Skerlajav, S.; Urbani, R.; Leitão, J. H.; Sá-Correia, I. and rizzo, R. (2003) "Macromolecular and solution properties of Cepacian: the exopolysaccharide produced by a strain of *Burkholderia cepacia* isolated from a cystic fibrosis patient". *Carbohydr Res*, 338(18): 1861-7.

Sivaswami, A.; Kelkar, S. M. and Nadkarni, G. B. (1972) "Uridine diphosphoglucose dehydrogenase from rat liver: purification and effect of pH on regulatory properties". *Biochim Biophys Acta*, 276 (1): 43-52.

Skouloubris, S.; Labigne, A. and De Reuse, H. (1997) "Identification and characterisation of an aliphatic amidase in *Helicobacter pylori*". *Molecular Microbiology*, 25: 989-998.

Smith, A. L.; Erwin, A. L.; Kline, T.; Unrath, W. C. T.; Nelson, K.; Weber, A. and Howald, W. N. (2007) "Chloramphenicol Is a Substrate for a Novel Nitroreductase Pathway in *Haemophilus influenzae*". *Antimicrobial Agents and Chemotherapy*, 2820–2829.

Smith, A. M.; Shelton, R. M.; Perrie, Y. and Harris, J. J. (2007) "An initial evaluation of gellan gum as a material for tissue engineering applications." *J Biomater Appl*, 22:241–254.

Snook, C. F.; Tipton, P. A. and Beamer, L. J. (2003) "Crystal structure of GDP-mannose dehydrogenase: a key enzyme of alginate biosynthesis in *P. aeruginosa*". *Biochemistry*, 42: 4658-4668.

Somerville, C. C.; Nishino, S. F. and Spain, J. C. (1995) "Isolation and characterisation of nitrobenzene nitroreductase from *Pseudomonas pseudoalcaligenes* JS45." *J Bacteriol*, 177: 3837-3842.

Sommer, B. J.; Barycki, J. J. and Simpson, M. A. (2004) "Characterisation of human UDP-glucose dehydrogenase. CYS-276 is required for the second of two successive oxidations". *J Biol Chem*, 279: 23590–23596.

Sousa, S. A.; Ulrich, M.; Bragonzi, A.; Burke, M.; Worlitzsch, D.; Leitão, J. H.; Meisner, C.; Eberl, L.; Sá-Correia, I. and Döring, G. (2007) " Virulence of Burkholderia cepacia complex strains in gp91phox-/- mice." *Cell Microbiol,* 9: 2817-2825.

Spain, J. C. (1995) "Biodegradation of nitroaromatic compounds". *Annu Rev Microbiol*, 49: 523–555.

Spain, J. C. (2000) "Biodegradation of nitroaromatic compounds and explosives". In: Jim C. Spain, Joseph B. Hughes, Hans-Joachim Knackmuss (Editors), CRC Press.

Spicer, A. P.; Kaback, L. A.; smith, T. J. and Seldin, M. F. (1998) "Molecular cloning and characterisation of the human and mouse UDP-glucose dehydrogenase genes". *J Biol Chem*, 273(39): 25117-24.

Stewart, D. C. and Copeland, L. (1998) "Uridine 5-Diphosphate-Glucose Dehydrogenase from Soybean Nodules." *Plant Physiol*, 116: 349-355.

Streker, K.; Freiberg, C.; Labischinski, H.; Hacker, J. and Ohlsen, J. (2005) "*Staphylococcus aureus* NfsA (SA0367) is a flavin mononucleotide-dependent NADPH oxidase involved in oxidative stress response". *J Bacteriol*, 187: 2249-2256.

Strominger, J. L.; Kalckar, H. M.; Axelrod, J. and Maxwell, E. S. (1954) "Enzymatic oxidation of uridine diphosphate glucose to uridine diphosphate glucuronic acid." *J Am Chem Soc*, 76: 6411–6412.

Sutherland, I. W. (1994) "Structure-function relationships in microbial exopolysaccharides". *Biotechnol Adv*, 12: 393-448.

Sutherland, I. W. (1997) "Microbial exopolysaccharides – structural subtleties and their consequences ". *Pure and Appl Chem*, 69(9): 1911-1917.

Sutherland, I. W. (1998) "Novel and established applications of microbial polysaccharides." *Trends in Biotechnology*, 16 (1): 41-6.

Sutherland, I. W. (2001) "Biofilm exopolysaccharides: a strong and sticky Framework". *Microbiology*, 147: 3-9.

Sutherland, I. W. (2004) "Microbial Exopolysaccharides". In: Severian Dimitriu (ed.) *Polysaccharides: Structural diversity and functional versatility*, CRC Press, Ed.2: 431-457.

Symons, Z. C. and Bruce, N. C. (2006) "Bacterial pathways for degradation of nitroaromatics." *Natural product reports*, 23(6): 845-50.

Takeda, K.; Iizuka, M.; Watanabe, T.; Nakagawa, J.; Kawasaki, S. and Niimura, Y. (2007) "*Synechocystis* DrgA protein functioning as nitroreductase and ferric reductase is capable of catalyzing the Fenton reaction". *FEBS J*, 274: 1318-1327.

Tanner, J. J.; Lei, B.; Tu, S.-C. and Krause, K. L. (1996) "Flavin reductase P: structure of a dimeric enzyme that reduces flavin". *Biochemistry*, 35: 13531-13539.

Teng, T.-Y. and Moffat, K. (2000) "Primary radiation damage of protein crystals by an intense synchrotron X-ray beam". *J Synchrotron Rad*, 7:313-317.

Terwilliger, T. (2004a) "SOLVE and RESOLVE: automated structure solution, density modification and model building." *J Synchrotron Radiat*, 11: 49-52.

Terwilliger, T. (2004b) "Using prime-and-switch phasing to reduce model bias in molecular replacement". *Acta Crystallogr D Biol Crystallogr*, 60: 2144-2149.

Thompson, J. D.; Higgins, D. G. and Gibson, T. J. (1997) "CLUSTAL W: Improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties, and weight matrix choice". *Nucleic Acids Res*, 22: 4673-4680.

Tichi, M. A. and Tabita, F. R. (2001) " Interactive control of *Rhodobacter capsulatus* redox-balancing systems during phototrophic metabolism." *J Bacteriol*, 183 (21): 6344-6354.

Tickle, J.; Pilka, E. S.; Bunkoczi, G.; Berridge, G.; Smee, C.; Kavanagh, K. L.; Hozjan, V.; Niesen, F. H.; Papagrigoriou, E.; Pike, A. C. W.; Turnbull, A.; Arrowsmith, C. H.; Edwards, A.; Sundstrom, M.; Weigelt, J.; Von Delft, F. and Oppermann, U. (2007) "The Structure of the Cytokine-Like Nuclear Factor N-Pac". (To be published)

Timmis, K. N. and Pieper, D. H. (1999) "Bacteria designed for bioremediation". *Trends Biotechnol*, 17: 201-204.

Tsafack, A.; Golenser, J.; Libman, J.; Shanzer, A. and Cabantchik, Z. I. (1995) "Mode of action of iron (III) chelators as antimalarials. III. Overadditive effects in the combined action of hydroxamate-based agents on *in vitro* growth of *Plasmodium falciparum*." *Mol Pharmacol*, 47: 403-409.

Tukey, R. H and Strassburg, C. P. (2000). "Human UDP-glucuronosyltransferases: metabolism, expression, and disease". *Annu Rev Pharmacol Toxicol*, 40: 581–616.

Unge, T. (1999) "Crystallisation Methods". In *Protein crystallisation: techniques, strategies, and tips: a laboratory manual*. Terese. M. Bergfors (ed), Internat'l University Line, 7-18.

Vagin, A. and Teplyakov, A. (1997) "MOLREP: an automated program for molecular replacement." *J Appl Cryst*, 30: 1022-1025.

Vanhooren, P. and Vandamme, E. J. (1998) "Biosynthesis, physiological role, use and fermentation process characteristics of bacterial exopolysaccharides". *Recent Res Devel Fermen Bioeng*, 1: 253-299.

Vanhooren, P. and Vandamme, E. J. (2000) "Microbial production of clavan, an L-fucose rich exopolysaccharide." In: Bielecki, S., Tramper, J. and Polak, J., Editors, *Food Biotechnology*, Elsevier Science B.V.,109-114.

Vanjari, H. and Pande, R. (2003) "Hydroxamic acids: proton donor and acceptor strength for use in drug design". *J Pharm Biomed Anal*, 33: 783-788.

Vartak, N. B.; Lin, C. C.; Cleary, J. M.; Fagan, M. J. and Saier Jr., M. H. (1995) "Glucose metabolism in *Sphingomonas elodea*: pathway engineering *via* construction of a glucose-6-phosphate dehydrogenase insertion mutant". *Microbiol*, 141: 2339-2350.

Vasudevan, S. G.; Shaw, D. C. and Armarego, W. L. F. (1988) "Dihydropteridine reductase from *Escherichia coli.*" *Biochem J*, 255: 581–588.

Ventura, C. L.; Cartee, R. T.; Forsee, W. T. and Yother, J. (2006) "Control of capsular polysaccharide chain length by UDP-sugar substrate concentrations in *Streptococcus pneumoniae.*" *Mol Microbiol*, 61(3): 723-33.

Vigetti, D.; Ori, M.; Viola, M.; Genasetti, A.; Karousou, E.; Rizzi, M.; Pallotti, F.; Nardi, I.; Hascall, V. C.; De Luca, G. and Passi, A. (2006) "Molecular cloning and characterisation of UDP-glucose dehydrogenase from the amphibian *Xenopus laevis* and its involvement in hyaluronan synthesis". *J Biol Chem*, 281(12): 8254-63.

Vonrhein, C.; Blanc, E.; Roversi, P. and Bricogne, G. (2005) "Automated structure solution with autoSHARP." In *Crystallographic Methods*, S. Doublie, Ed., Humana Press, submitted.

Vuong, C.; Kocianova, S.; Voyich, J. M.; Yao, Y.; Fischer, E. R.; DeLeo, F. R. and Otto, M. (2004) "A Crucial Role for Exopolysaccharide Modification in Bacterial Biofilm Formation, Immune Evasion, and Virulence". *The Journal of Biological Chemistry*, 279(52): 54881–54886.

Wang, H.; Wang, H.-L.; Jiang, W.-F. and Li, Z.-Q. (2009) "Photocatalytic degradation of 2,4-dinitrophenol (DNP) by multi-walled carbon nanotubes (MWCNTs)/TiO2 composite in aqueous solution under solar irradiation". *Water Research*, 43: 204-210.

Wang, L.; Li, S.and Li, Y. (2003) "Isolation and sequencing of glycosyltransferase gene and UDP-glucose dehydrogenase gene that are located on a gene cluster involved in a new exopolysaccharide biosynthesis in *Streptomyces.*" *DNA Seq*, 14: 141-145.

Watanabe, M.; Nishino, T.; Takio, K.; Sofuni, T. and Nohmi, T. (1998) "Purification and characterisation of wild-type and mutant ''classical'' nitroreductases of *Salmonella typhimurium*." *J Biol Chem*, 273: 23922–23928.

Weiss, M. S. (2001) "Global indicators of X-ray data quality." *J Applied Crystallography*, 34: 130-135.

Westhead, D.R.; Slidel, T. W.; Flores, T. P. and Thornton, J. M. (1999) "Protein structural topology: Automated analysis and diagrammatic representation". *Protein Sci*, 8: 897-904.

White, D. C.; Sutton, S. D. and Ringelberg, D. B. (1996) "The genus *Sphingomonas*: physiology and ecology". *Curr Opin Biotechnol*, 7:301-306.

Whiteway, J.; Koziarz, P.; Veall, J.; Sandhu, N.; Kumar, P.; Hoeler, B. and Lambert, I. B. (1998) "Oxygen-insensitive nitroreductases: analysis of the roles of nfsA and nfsB in development of resistance to 5-nitrofuran derivatives in *Escherichia coli*". *J Bacteriol*, 180: 5529-5539.

Xiao, Y.; Wu, J. F.; Liu, H.; Wang, S. J.; Liu, S. J. and Zhou, N.Y. (2006) "Characterisation of genes involved in the initial reactions of 4-chloronitrobenzene degradation in *Pseudomonas putida* ZWL73". *Appl Microbiol Biotechnol*, 73: 166-171.

Yasuhira, K.; Shibata, N.; Mongami, G.; Uedo, Y.; Atsumi, Y.; Kawashima, Y.; Kato, D.; Takeo, M.; Higuchi, Y. & Negoro, S. (2010) "X-ray crystallographic analysis of 6-aminohexanoate cyclic dimer hydrolase: catalytic mechanism and evolution of an enzyme responsible for nylon-6 by-product degradation." *J Biol Chem*, 285 (2): 1239-1248.

Zenno, S.; Kobori, T.; Masaru, T. and Saigo, K. (1998a) "Conversion of NfsA, the major *Escherichia coli* nitroreductase, to a flavin reductase with an activity similar to that of Frp, a flavin reductase in *Vibrio harveyi*, by a single amino acid substitution." *J Bacteriol*, 180: 422-425.

Zenno, S.; Kobori, T.; Tanokura, M. and Saigo, K. (1998b) "Purification and characterisation of NfrA1, a *Bacillus subtilis* nitro/flavin reductase capable of interacting with the bacterial luciferase." *Biosci Biotechnol Biochem*, 62: 1978-1987.

Zenno, S.; Koike, H.; Kumar, A. N.; Jayarman, R.; Tanokura, M. and Saigo, K. (1996a) "Biochemical characterisation of NfsA, the *Escherichia coli* mayor nitroreductase exhibiting a high amino acid sequence homology to Frp, a *Vibrio harveyi* flavin oxidoreductase". *J Bacteriol*, 178: 4508-4514.

Zenno, S.; Koike, H.; Tanokura, M. and Saigo, K. (1996b) "Gene cloning, purification and characterisation of NfsB, a minor oxygeninsensitive nitroreductase from *Escherichia coli*, similar in biochemical properties to FRaseI, the major flavin reductase in *Vibrio fischeri*". *J Biochem*, 120: 736-744.

Zenno, S.; Koike, H.; Tanokura, M. and Saigo, K. (1996c) "Conversion of NfsB, a minor *Escherichia coli* nitroreductase, to a flavin reductase similar in biochemical properties to FRaseI, the major flavin reductase in *Vibrio fischeri*, by a single amino acid substitution". *Appl Environ Microbiol*, 178: 4731-4733.

Zenno, S.; Saigo, K.; Kanoh, H. and Inouye, S. (1994) "Identification of the gene encoding the major NAD(P)H-flavin oxidoreductase of the bioluminescent bacterium *Vibrio fischeri* ATCC 7744". *J Bacteriol*, 176: 3536-3543.

Zeyer, J. and Kocher, H. P. (1988) "Purification and characterisation of a bacterial nitrophenol oxygenase, which converts ortho-nitrophenol to catechol and nitrite". *Journal of Bacteriology*, 170: 1789-1794.

Zhang, R.; Li, H.; Collart, F.; Moy, S. and Joachimiak, A. "Crystal structure of a nitroreductase family protein from *Bacillus cereus* ATCC 14579." (To be published).

Zhang, Y.; Zhan, C.; Patskovsky, Y.; Ramagopal, U.; Shi, W.; Toro, R.; Wengerter, B. C.; Milstein, S.; Vidal, M.; Burley, S. K. and Almo, S. C. (2006) "Crystal Structure of *Caenorhabditis Elegans* Udp-Glucose Dehydrogenase". New York SGX Research Centre for Structural Genomics (NYSGXRG). (To be published)

Zhu, S. C.; Yang, Y. L.; Zhao, G. P. and Jiang, W. H. (2003) "A rapid and specific method to screen environmental microorganisms for cephalosporin acylase activity". *J Microbio Methods*, 54 (1): 131-135.

Zlosnik, J. E. A.; Hird, T. J.; Fraenkel, M. C.; Moreira, L. M.; Henry, D. A. and Speert, D. P. (2008) "Differential Mucoid Exopolysaccharide Production by Members of the Burkholderia cepacia Complex." *J Clinic Microbio*, 46 (4): 1470-1473.