Structure and Function Relationships of Urease and Cytochrome *c*-553 from *Bacillus pasteurii*

Stefano Benini

A Thesis Submitted to the University of York for The Degree of Doctor of Philosophy

(Faculty of Science)

University of York

Department of Chemistry

July, 2000

to my father

Abstract

The study of structure-function relationships of two metalloproteins from *B. pasteurii*, urease and cytochrome *c*-553, by characterisation at the atomic level using X-ray crystallography is the subject of this thesis. I purified, crystallised and solved the structure of both of these proteins.

Urease is a nickel enzyme produced by plants, fungi, algae, and bacteria. It is involved in nitrogen turnover and in crop fertilisation as well as in human and animal pathologies. It catalyses the hydrolysis of urea giving rise to ammonia and carbon dioxide. Beside its medical, ecological and economical importance, the charm of urease resides in the fact that it was the first enzyme to be crystallised in 1926 by Sumner [1]. In 1975 Dixon discovered that nickel is required for the catalysis, first showing the biological role of nickel [2, 3]. The structures of *B. pasteurii* urease in complex with inhibitors belonging to different chemical classes (thiols, phosphoramides, hydroxamates, phosphates) solved to high resolution (1.5-2.0 Å), provide information about binding site and inhibition mechanisms. Based on the comparison between the native and the diamidophosphate inhibited structures a reaction mechanism is proposed that reconciles previous biochemical data.

Cytochrome *c*-553 is a small electron transfer protein. The characterisation at atomic resolution (0.97 Å) provides the most accurate haem structure to date giving hints about the function and the reason for its low reduction potential. The high resolution of the data allowed the solution of the structure by an unconventional method.

Acknowledgements

"And money wasn't what I had in mind.
Oh God, no, what I wanted was to do good.
I was dying to do something good." [1]

Acknowledgements are always a difficult task. Written words can never express what is in someone heart to the extent he really wants. First of all I would like to thank my University supervisor Keith Wilson and my EMBL supervisor Wojciech Rypniewski. They trusted me and believed I would have seen through to the end. They made of me an independent scientist and for this I will always be grateful. Thanks to Stefano Mangani who first introduced me to the *art* of crystallisation and to the synchrotron world. Thanks to Stefano Ciurli who shared with me the initial enthusiasms and frustrations in the fascinating world of science. I would like to thank Guy Dodson who during my *viva voce* using a metaphor borrowed from rugby told me: "*you had the ball in your hands, you had it and you dropped it!*". Many thanks are due to all the friends at the EMBL for human, "philosophical" and scientific discussions. I will never forget them for the friendly support in the dark days of loneliness. Thanks to Mauro for the wonderful grilled eel and the thousands of e-mails. Special thanks are due to my mother and my family for their strong and passionate support and their trust. I apologise for having "excaped", listening to the call of science, and leaving my mother alone in times of need. A huge acknowledgement goes to my girlfriend Marilena, who despite the distance, the troubles, the arguments, the telephone bills, was always close to me whenever I needed some words of consolation.

I would like to thank my father for his love, his example of humanity, honesty, dignity and for his smile whose memory will accompany me for ever.

[1] Saul Bellow: "*Humboldt's Gift*".

Author's declaration

I declare that the work presented in this thesis, except where otherwise stated, is based on my own research and has not been submitted previously for a degree in this or any other University. Parts of the work reported in this thesis have been published in the papers listed in the Appendix: Publications

Contents

List of Figures

List of Tables

Part I

The Reaction Mechanism of *Bacillus pasteurii* Urease from Structural Studies of Native and Inhibited Enzyme

Chapter 1

Urease a Nickel Metalloenzyme

1.1 Nickel and Life

Nickel is a metal used in the manufacture of steel, and as an alloy to make coins and kitchen utensils. It is coupled with cadmium in batteries and present, as nickel carbonate, in electronic components such as vacuum tubes and transistors. The concentration of nickel in seawater ranges from 0.1 to 0.5 $\mu g L^{-1}$, in natural soil from 5 to 500 ppm with typical values of 10-90 ppm while in agricultural soil the concentration may vary between 3 and 1000 ppm. Despite the toxicity observed for some nickel compounds (nickel carbonyl has a LC_{50} (Lethal Concentration for the death of 50% of the population of mice) of 35 ppm in 30 minutes [4]), states of nickel deficiency have been observed with decrease of weight gain and liver damage in calves and sheep and in rats with retarded growth and low haemoglobin levels. The requirement for nickel in human diet has not been demonstrated and its role in animal metabolism is unknown [4].

The discovery by Dixon in 1975 [2] of the presence of Ni^{2+} ions in urease from

Canavalia ensiformis (jack bean) showed for the first time its importance in an enzyme. Since then other nickel containing enzymes have been discovered in microorganisms: nickel superoxide dismutase, Ni-Fe hydrogenases, the bifunctional two-subunit protein carbon monoxide dehydrogenase/acetyl-CoA synthase (CODH/ACS), and methyl CoM reductase [5]. Their function requires the presence of a set of accessory proteins to transport the nickel ions through the cellular membrane, and shuttle and incorporate them into the active site of the target enzyme [6].

To study the mechanism of nickel uptake by microorganisms the ^{63}Ni isotope has been used to detect the increase of concentration in the cell. A high affinity nickel transport system was discovered in *Alcaligenes eutrophus* which was able to accumulate nickel 280 fold over the growth medium concentration. Nickel transport was found to be energy dependent and highly selective [7]. Under anaerobic growth conditions *Escherichia coli* expresses an ATP-dependent nickel-specific transport system composed of five different proteins whose genes are grouped on the DNA as the *nik*ABCDE operon. NikA is a periplasmic nickel binding protein, NikB and NikC are protein subunits embedded in the membrane and NikD and NikE are nucleotide binding proteins, since energy is required for the functioning of the process [7].

In contrast to the energy dependent active transport, permeases do not require energy. In *Alcaligenes eutrophus* a single membrane protein HoxN has been identified as nickel permease [8]. A similar protein, NixA has been identified in *Helicobacter pylori* [9].

Upon entering the cell nickel must be sequestered by ligand proteins, because of

its toxicity, and carried to the specific targets. The metal must then be incorporated in a selective way into the active site often requiring additional ligands or modifications. This selectivity is achieved by means of accessory proteins inserting the correct metal into the correctly folded or folding protein [10].

1.2 Urease

Large quantities of urea are released into the environment as a product of the catabolism of nitrogen containing compounds such as purine, arginine, agmantine and allantoin [11] by living organisms. Urea is excreted by mammals as a detoxification product; human urine for example contains a concentration up to 0.4-0.5 M, equivalent annually to about 10 kg per adult. Urea was the first organic molecule to be synthethised. The presence of urease in crops is exploited to use urea as a fertiliser. Nowadays urea is the most used fertiliser in agriculture because of its high nitrogen content (46%) and relatively low cost.

The spontaneous hydrolysis of urea is a slow process. Urea is a quite stable molecule with a half life of 3.6 years at 38 °C in water. The products of the uncatalysed reaction are ammonia and cyanic acid [12] (reaction 1).

1) $$(NH_2)_2CO + H_2O \rightarrow H{-}N{=}C{=}O + NH_3 + H_2O$$

Urease, a cytoplasmatic nickel containing enzyme, catalyses another hydrolytic reaction at a rate 10^{14} times faster than the uncatalysed one [12] possibly because going through a different intermediate (reaction 2). The catalytic reaction products are one molecule of ammonia and one of carbamic acid, the

latter spontaneously decomposes to another molecule of ammonia and one of carbon dioxide.

2) $$(NH_2)_2CO + H_2O \xrightarrow{urease} H_2N{-}COO^- + NH_4^+$$

The presence of urease has been detected in bacteria, algae, fungi and higher plants [11]. In the soil, urease (released upon cellular lysis) is bound to organic and inorganic soil components (clays and humic substances) and protected from degradation by extracellular proteolytic enzymes and microorganisms [13]. Urease is used by bacteria to transform urea to ammonia as a nitrogen source. In plants urease is used to recycle nitrogen often available only in small amounts in the soil.

Mutagenic studies showed that urease activity is necessary to prevent accumulation of urea produced by degradation of arginine. Arginine is a major nitrogenous transport and storage compound in plants. In the absence of urease the accumulation of urea may not be lethal for large protein-rich seeds like soyabean, but for small seeds like *Arabidopsis thaliana* it may retard or block germination. It has been proposed that urease might be used in the seeds as a chemical defence. In seeds of jack bean, microbial or insect attacks or wounds or infections in the germinating embryo will lead to release of arginase from ruptured mitochondria. Arginase activity will generate urea which is then hydrolysed to ammonia by urease [14].

1.2.1 Urease and Accessory Proteins

Genetic studies revealed the presence of a so called urease cluster in which genes encoding structural and accessory proteins are grouped together. In bacteria

the structural genes are *ure*A, *ure*B and *ure*C (encoding for the γ, β, α subunit respectively) with the exception of *Helicobacter pylori* in which only the *ure*A and *ure*B genes are present because a gene fusion has occurred (*ure*A encodes γ and β subunit in a single polypeptide chain). The high molecular weight urease monomer in plants, such as *Canavalia ensiformis*, is the result of the fusion of the three α, β, γ bacterial subunits into one chain.

In *Klebsiella aerogenes* the accessory proteins for *in vivo* assembly of functional urease are UreD, UreE, UreF, UreG [15]. Studies on mutants in which part or all of the DNA sequence encoding for the accessory proteins was absent showed that they are essential for obtaining a completely functional enzyme [16].

Urease apoprotein (apoU) is proposed to form a complex with the accessory proteins, whose specific role and mode of binding are still unknown. After insertion of the nickel ions in the protein active site the complex is disassembled and the individual components released for a new incorporation. A UreD-apoU complex has been obtained *in vitro* and proposed to be the first step toward the insertion of the metal [16]. *In vitro*, partial activation and nickel incorporation of apoU can be achieved in the presence of CO_2 (as bicarbonate buffer), required to carbamylate a lysine essential in its carbamylated form to build the active site by bridging two nickel ions *via* its terminal oxygens [17]. When the complex of ureD-apoU *in vitro* is incubated in the presence of nickel ions and bicarbonate buffer a higher level of incorporation is reached, further increased for the complex UreD-UreF-apoU and the whole complex UreD-UreF-UreG-apoU [18, 19]. *In vivo* formation of the whole UreD-UreF-UreG-apoU complex could be either by sequential assembly of the single components onto the apoprotein or after UreD-UreF-UreG complex formation.

A putative nucleotide binding site has been found in UreG [19] and the requirement for energy during the process of nickel insertion into apoU has been shown recently to require GTP hydrolysis [20]. Models of the urease metallocentre assembly have been proposed [10, 21] and are shown in Figure 1.1. The

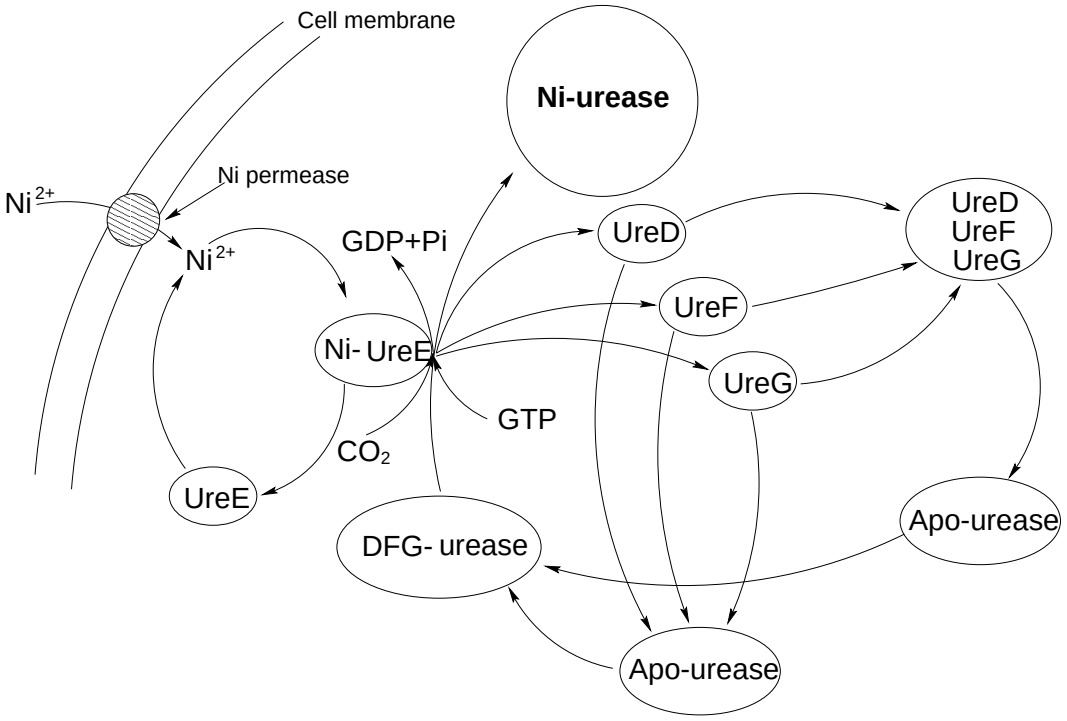


Figure 1.1: Model for urease nickel incorporation *in vivo*. Two different ways of formation of the DFG-urease complex are presented: either sequential with Apo-urease binding UreD, UreF, UreG or with Apo-urease binding to a preformed UreD-UreF-UreG (DFG) complex [21].

role of the nickel binding protein UreE, proposed to act as a metallochaperone is still unclear. It is supposed to facilitate the incorporation of nickel into apoU but a role of storage or protection from free nickel ions is also possible. The identification of metal-binding residues in *Klebsiella aerogenes* UreE and

their localisation at the interface of the homodimer has been determined by site directed mutagenesis, NMR and EXAFS spectroscopy [22]. In *B. pasteurii* and other bacteria a *ure* gene cluster has been sequenced [23] whose homology with *Klebsiella aerogenes* suggests that the nickel incorporation mechanism is the same for all bacterial ureases.

1.2.2 Regulation of Urease Expression

The expression of urease is regulated in different ways in different organisms. In some bacteria urease production is repressed when ammonia or nitrogen compounds are present and is de-repressed under conditions of nitrogen starvation. Other bacterial ureases are expressed in the presence of urea while in *B. pasteurii* urease is constantly produced [11].

The requirement in *B. pasteurii* for a constitutive urease is related to the need for the increase of pH of the growth medium and because of the lack of an active ammonium transport system. *B. pasteurii* therefore relies on a passive diffusion of ammonia into the cell. Moreover the lack of a glutamine synthase in this bacterium may also require high ammonium concentration [24].

1.2.3 The Role of Urease in Pathogenesis

Urease produced by some microorganisms is implicated in the pathogenesis of a number of diseases [11, 15, 25]. **Infection stones** accounting for 15 to 20% of all the urinary stones are generated by supersaturation of soluble polyvalent ions such as Ca^{2+} and Mg^{2+} when the pH increases from 6.5 to 9.0 due to urease activity. In humans, *Proteus mirabilis* is the most common organism implicated in stone formation. Urease inhibitors have been tested in therapy. The most commonly used is acetohydroxamic acid which has been shown to

lower the pH of patients's urine and to decrease the size of urinary stones but has very negative side effects. An increase in pH in *in vitro* studies, conducted by inoculation of human or synthetic urine with *Proteus mirabilis* or *Canavalia ensiformis* enzymes, caused crystallisation and encrustation of salts not observed in the presence of urease inhibitors.

Pyelonephritis is defined as an acute or chronic inflammation of the kidneys. *Proteus mirabilis* is the primary source of this disease in humans. The ammonia liberated from ureolysis causes the alkalinisation of urine and may account, in part, for the necrosis of kidney tissue associated with pyelonephritis.

Urinary catheter obstruction because of encrustation of calcium, phosphor and magnesium compounds has been found in patients with long term indwelling catheters. Encrustation was found in several cases to be due to the presence of *Proteus* strains. Catheter encrustation has been demostrated to be reduced by the administration of urease inhibitors [11].

Peptic ulceration and cancer may be caused by the ureolytic bacterium *Helicobacter pylori* a spiral Gram-negative microaerophilic bacterium that infects humans. It is recognized as the major aetiological factor in chronic gastritis, and gastric and duodenal ulceration and as a risk factor for gastric cancer; approximately 30 to 50% of the patients in Western Europe have been shown to be infected. *H. pylori* normally inhabits the mucus-lined surface of the antrum of the human stomach where it induces inflammation, but its presence is otherwise often asymptomatic. A variety of virulence factors appear to play a role in pathogenesis, including the vacuolating cytotoxin VacA, urease

and motility. Urease is critical for *H. pylori* colonisation of the human gastric mucosa, providing protection against the acidic environment. *In vitro*, the bacterium is quite sensitive to the effect of low pH, but the presence of urea allows the creation, *via* urease activity, of a neutral environment compatible with bacterial life. The recent development of a mouse model for the direct study of *H. pylorii* infections will probably help in finding vaccines to be used against this disease [26].

1.2.4 Urease in the Rumen

Microbial ureases play an important role in the nitrogen metabolism of ruminants such as cattle, sheep, and others. Substantial amounts of animal-derived urea, diffusing from the bloodstream or entering with saliva are recycled into the rumen, where microbial ureolytic activity releases ammonia, utilised by the ruminal bacteria for their needs. Microbial biomass is subsequently degraded by the ruminant digestive system and supplies nitrogen-containing compounds to the animal (Fig. 1.2).

Urea feeding is used to compensate nitrogen-poor feedstocks and to enhance the quality of ruminant diet at a lower price than with proteic supplement. 35% of bacteria detected in the rumen belongs to ureolytic species, amongst them *Staphylococcus spp.*, *Lactobacillus casei* and *Klebsiella aerogenes* [11].

1.2.5 Urease in the Environment

Urease activity is widely spread in soil and aquatic environments where it takes part in the nitrogen turnover pathway. In agricultural settings urea is the most used fertiliser because of its low cost, ease in handling, and high nitrogen content. High urease activity in the soil is however often a problem because it may

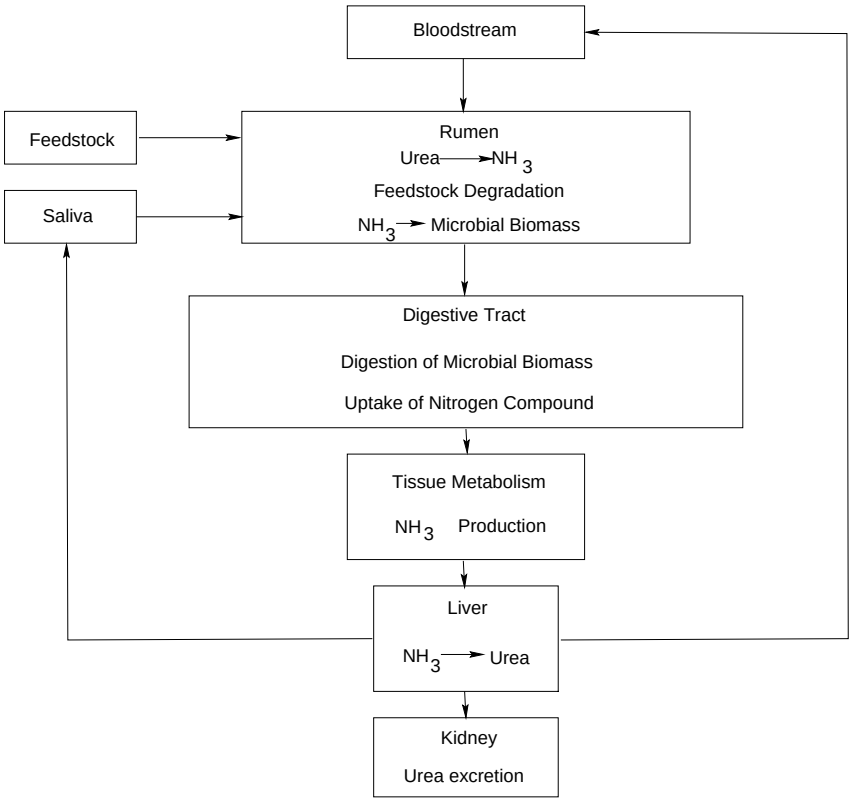


Figure 1.2: Urea recycling in ruminants

lead to plant damage from ammonia toxicity or elevated pH (mainly regarding seed germination and seedling growth). In basic soils ammonia is volatilised and lost to the atmosphere.

Studies have been carried out to examine the possibility of adding urease inhibitors to the fertiliser urea to reduce the rate of hydrolysis, to minimize crop damage, and enhance nitrogen uptake by plants. Hydroxamic acid was shown to reduce the rate of ammonia volatilisation in field trials and phosphoramides were shown to retard urea hydrolysis. The application of phenylphosphorodiamidate in field evaluation did not however result in an increase in the yield of corn [11]. At present the discovery of new active molecules is still based solely

on screening tests. Recently ^{13}C-NMR has been used to develop a screening system based on measurement of the time of decrease in the signal of ^{13}C-urea in the presence of urease. The effect on urease activity of known inhibitors, hydroxamic acids, L-ascorbic acid, 2,2-dipyridyl disulfide and ninhydrin, was tested and measured by this method [27].

1.3 Urease Inhibitors

Hydroxamic acids. In 1969 Blakeley published the first evidence of inhibition by acetohydroxamic acid (AHA) in jack bean urease [28]. Using a pH-stat method [29], the pH during urea hydrolysis is kept constant by injection of HCl whose total amount is measured to quantify the activity. AHA was found to be a slow, tight-binding competitive inhibitor. No mechanism of interaction with the enzyme could be proposed because of the lack of information about the active site and of the Ni^{2+} ions.

The first evidence for the presence of two nickel ions in each active site and their direct involvement in the reaction and inhibition mechanism was due to Dixon's studies [2, 3]. The effect of AHA addition to the enzyme solution was a pronounced shift in the curve in the UV-Vis spectrum corresponding to the metal centre.

Competitive inhibitors have been studied extensively by Todd *et al.* in *K. aerogenes* using a colorimetric assay [30]. The slow binding competitive inhibition by AHA is proposed to follow a mechanism in which the inhibitor competes with the substrate for enzyme to form an initial enzyme-inhibitor complex which is then slowly transformed to a more stable one [31]. The mode of bind-

ing of this inhibitor has been proposed to be initialy monodentate with only one nickel coordinated to the AHA carbonyl oxygen. The more stable complex involves both nickel atoms in a bidentate mode binding the carbonyl oxygen to one nickel atom while the other AHA oxygen bridges between them.

The structure of the *K. aerogenes* Cα319A mutant in complex with acetohydroxamate was recently published [32]. In this structure AHA is bound through its carbonyl oxygen to Ni-1 and makes a hydrogen bond with Hisα219, the other AHA oxygen bridges the two nickels. An additional hydrogen bond is made between the AHA nitrogen and the carbonyl oxygen of Alaα363. However the accuracy of the structure was not good enough to allow precise determination of the metal ligand distances. The bridging AHA oxygen was reported to be 2.60 and 1.83 Å from Ni-1 and Ni-2 respectively - quite unusual for coordination bonds.

Despite its negative side effects acetohydroxamic acid is still the most widely used in human therapy. Inhibition of DNA synthesis, depression of bone marrow synthesis and, when used in high dose, teratogeny are negative side effects during the treatment [11].

Phosphoramides have often been shown to be stronger inhibitors than hydroxamates. They range from simple phosphoramidate and diamidophosphate to more complex substituted groups like phenylphosphorodiamidates. Inhibition by phosphoramidate was already studied by Dixon in 1979 [33] by spectrophotometric experiments showing the slow process of inhibition and reactivation of jack bean urease, and its direct binding to the nickel ions in the active site. The tight binding species were proposed to mimic a hypothet-

ical tetrahedral intermediate of urea hydrolysis [31]. It has been suggested that urease hydrolyses larger compounds (e.g., phenylphosphorodiamidate) to form diamidophosphate and the identical dissociation rates for different compounds actually represent the dissociation constant of diamidophosphate. Diamidophosphate is alternatively defined either as a substrate or an inhibitor of urease [34]. Phenylphosphorodiamidate (PPD) was used for *K. aerogenes* to quantitate the number of active sites per native enzyme molecule by the graphical method of Ackerman-Potter plotting the concentration of the enzyme against the velocity of reaction at different PPD concentrations [31]. The same method has been used for *B. pasteurii* urease [35]. The effects of different phosphoramides on urease from jack bean and *B. pasteurii* have been compared by McCarty *et al.* [36]. The result of this study, finding different degrees of interaction for different compounds, was the discovery that plant ureases have higher affinity for the inhibitors than bacterial urease. This was supposed to be due to the different optimal pH, close to neutrality for the plant enzyme and close to pH 8.0 for the bacterial one [13, 37].

Thiols have been useful for the characterisation of the nickel metallocentre but they are poor inhibitors since they show only a competitive mechanism. The interaction of β-mercaptoethanol (BME) with urease produces a shift in the region of the UV-Vis absorption spectrum, associated with the nickel ions. The dissociation constant for BME determined spectrophotometrically (0.95 ± 0.05 mM at pH 7.12 25 °C) is in agreement with its role of competitive inhibitor in the hydrolysis of urea [33]. In *K. aerogenes* urease, kinetic analysis for a range of thiol compounds revealed that the inhibition constant is influenced by the presence of other charged groups in the molecule. Cysteamine (β-mercaptoethanolamine), containing a β-amino group, has the highest affinity

($K_i \approx 0.01$ mM) followed by β-mercaptoethanol which has an alcoholic functional group ($K_i \approx 0.55$ mM) while other thiolates containing anionic carboxyl groups are in general poor inhibitors (e.g., $K_i \approx 21$ mM for ethanethiol) [31]

Phosphate competitively inhibits urease between pH 5 and 7. Plotting the value of pH against $\log(K_i)$ the pH dependence of inhibition shows two different slopes from pH 5 to pH 6.3 and from pH 6.3 to pH 7. This behaviour requires the presence of two ionizable groups one with pK_a of 6.3 and the other with a pK_a less than 5 (not determined because of the loss of nickel below pH 5), due to changes in the protonation state, in the inhibitor, in the enzyme or in both [31]. Fully protonated phosphoric acid is a competitive inhibitor of urease with $K_i \approx 0.12$ μM. At pH 6.5 $H_2PO_4^-$ has a $K_i \approx 6$ mM but little inhibition is observed at neutral pH because partially protonated phosphate binds poorly to the active site [11].

Boric acid and boronic acids are competitive inhibitors of *P. mirabilis* urease, with K_i values ranging from 0.1 to 1.2 mM at pH 7.5 [11] while for *K. aerogenes* K_i is 0.33 mM [31]. From pH dependence studies with boric acid it has been shown that the inhibitory molecule is the trigonal $B(OH)_3$ rather than the tetrahedral $B(OH)_4^-$. Boric acid is used in hospitals as a preservative of urine for microbiological tests. The disadvantages due to its inhibitory action have been discussed by Jones [38].

1.4 Urease from *Klebsiella aerogenes*

The enzyme from the enteric bacterium *K. aerogenes* is the best characterised urease. The importance of histidines in urease was proposed on the basis of pH dependence of enzyme activity, consistent with the presence of a group acting as a general base (with pK_a = 6.55) and another as a general acid (with pK_a = 8.85) [39]. The reaction of diethylpyrocarbonate with urease's histidines was in agreement with the presence of one of them acting as a general base with pK_a = 6.55 [40]. In 1993, two years before the structure was determined, Park and Hausinger, starting from a sequence alignement of six ureases, identified 10 conserved histidines and began an extensive site-directed mutagenesis study [40]. The enhanced reactivity of DEP in the presence of apo-urease confirmed the involvement of histidines in nickel coordination. Histidines His γ96, His β39, His β41, His α134, His α136, His α219, His α246, His α312, His α320 and His α321 were mutated one by one to alanine. His α134, His α136 and His α246 were shown to be involved in nickel ligation since the metal content was reduced to 53, 6 and 21% respectively compared to the wild type enzyme [41]. The enormously increased K_m for His α219A (1100 $\pm$0.2 mM compared to the wild type K_m=2.3$\pm$0.2 mM) and its low specific activity (1.9% of the wild type) suggested a direct role for this group in binding the substrate. The almost total loss of specific activity for His α320A (0.003 % of the wild type) and its increase in K_m (8.3$\pm$0.2 mM) confirmed its participation in urea hydrolysis [41].

In 1995 the solution of the crystal structure of *K.aerogenes* urease by Jabri *et al.* [42] (PDB code 1kau) rationalised all the previous information but did not provide sufficient detail about metal coordination in the active site because of the limited resolution of the data. The enzyme was composed of a

trimer of three different subunits α, β, γ forming a so-called trimer of trimers $(\alpha\beta\gamma)_3$. The *ure* structural genes *ure*A, *ure*B, *ure*C, express respectively the γ (M_r = 11.1 kDa), β (M_r = 11.7 kDa) and α (M_r = 60.3 kDa) chains. The α, containing the binuclear nickel active site, was composed of two structural domains. The first and catalytic domain represented a typical $(\alpha\beta)_8$ barrel domain. Very important for the catalytic activity was the flap covering the active site groove, composed of three helices and very mobile. In the second domain eight and four stranded β sheets formed a U shaped channel. The γ and β subunits revealed novel folds. The γ subunit was composed of four helices and a two-stranded antiparallel β-sheet. The β subunit featured a unique six stranded antiparallel β jellyroll. The three $(\alpha\beta\gamma)$ subunits made very extensive contacts in order to pack tightly and stabilise the enzyme. Many conserved residues held the trimers together by hydrogen bonds between the different subunits [43]. The active site was characterised by the presence of two Ni^{2+} ions 3.5 Å apart. The coordination for Ni-1 was proposed to be full-filled by three protein ligands: His α246, His α272 and the carbamylated lysine Lys α217*. Its final coordination was proposed to be between three and four because of a partially occupied disordered water. Ni-2 was pentacoordinated with His α134, His α136, Asp α360 and the carbamylated Lys α217* as protein ligands. A tightly bound water proposed to act as nucleophile during urea hydrolysis completed the coordination sphere. The geometry for Ni-1 and Ni-2 was described as pseudotetrahedral (with the weakly occupied water as fourth ligand) and distorted trigonal bipyramidal or distorted square pyramidal with His α136 as apical ligand respectively. The active site of native *K. aerogenes* urease is depicted in Figure 1.3.

The proposed coordination of both metal ions was far from the optimal octahe-

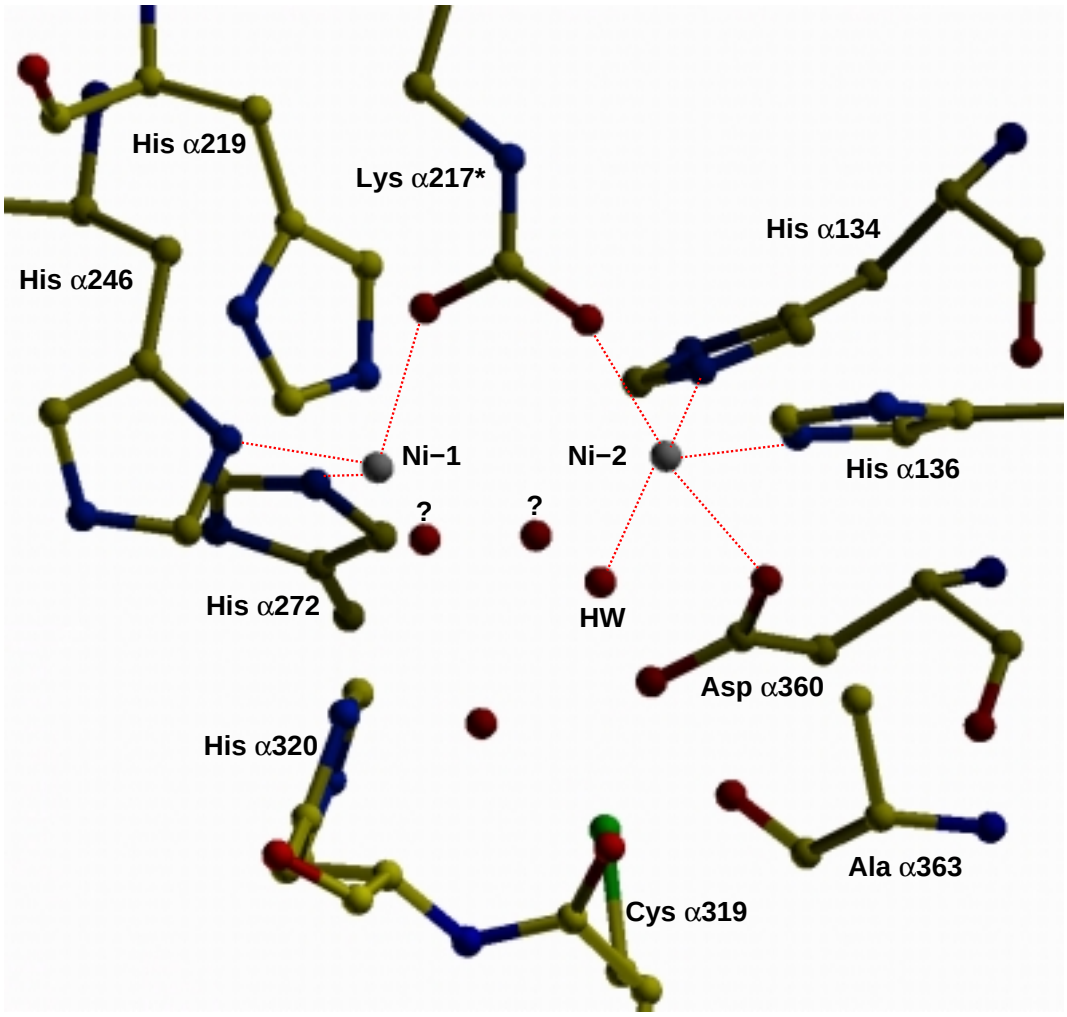


Figure 1.3: Model of native *K. aerogenes* urease active site. Waters marked with ? are the waters missing from the first model published.

dral with six ligands, and in conflict with all spectroscopic evidence. Extended X-ray Absorption Spectroscopy (EXAFS) had demonstrated the presence of two nickels with an average coordination between 5 and 6 [44]. Structural studies of the apo-enzyme and two mutants (Hα219A and Hα320A) revealed no differences in the overall structure [43]. Unfortunately because of the lack of

electron density no information about the missing partner for the protonated Nϵ of His α219 could be provided. Furthermore, the putative catalytic base His α320 appeared to be too far from the putative nucleophilic water coordinated to Ni-2, in order to activate it [43].

Cys α319 was thought to be essential for the catalysis on the basis of enzyme inactivation by thiol reagents [45]. Biochemical studies on Cys α319 mutants demonstrated its influence on the enzymatic activity but excluded its essentiality. Cα319A, Cα319S, Cα319D and Cα319Y were found to possess 48, 4.5, 0.03 and 0.00% activity respect to the wild type urease under normal assay conditions [46]. Structural studies on the Cys α319 mutants showed differences in the mobility and position of the flap while constantly having three water molecules bound to the two nickels, one bridging and two monodentate. The variation in activity was related to the shift of His α320, important in catalysis.

The new water structure in the active site was also adopted for the native enzyme correcting the very anomalous nickel coordination. Ni-1 changed from tricoordinated to pentacoordinated and Ni-2 from penta to hexacoordinated (PDB code 1FWJ). The presence of a water bridging between Ni-1 and Ni-2 has been confirmed based on molecular mechanics calculations and model compounds [47].

The enhanced stability of the mutated *K. aerogenes* Cα319A enzyme in the presence of β-mercaptoethanol relied on the lack of the cysteine thiol group otherwise involved in a mixed disulfide. According to the authors the binding of AHA into the Cα319A active site *via* its carbonyl oxygen to Ni-1, further stabilised by a hydrogen bond with His α219, was the same for urea. The re-

placement of the bridging water, by the second AHA oxygen, and its possible role in catalysis were not taken into account.

The results of the structural studies on the Cα319A mutant and AHA inhibited enzyme led the authors to propose a reaction mechanism [32, 48] which relied on overturning the role of His α320 previously thought to be a general base [40] and now proposed to act as a general acid [48]. In the absence of conformational changes in the active site, urea was proposed to bind to Ni-1 *via* its carbonyl oxygen and the water bound to Ni-2 was proposed to be the hydrolytic water (HW). His α320 was too far away to act as a general base in activating HW, while in its role of general acid it could provide the proton to the urea amide. The activated HW would attack the urea carbon giving rise to the tetrahedral intermediate which is immediately dissociated into one molecule of ammonia and one of highly unstable carbamic acid.

Weak points in this proposal were: the missing base needed to activate HW and the fact that an enzyme with optimal pH of 8 would require a protonated His α320 whose pK_a $\approx$6.55. The introduction of the mechanism of reverse protonation though answering the question about the protonated histidine would imply that only 0.3% of urease molecules are in the correct protonation state for activity [48]. This hypothesis was not compatible with the velocity of the reaction.

All these questions waited to be answered.

1.5 Urease from *Helicobacter pylori*

Gastritis and gastroduodenal ulcers can be caused by infection with *H. pylori*, known to be the main risk factor for the development of stomach cancer and lymphomas. The bacterium was isolated and for the first time grown on agar plates by Marshall and Warren [49]. Based on its constant presence in a high percentage of biopsies from patients affected by ulcers they proposed it to be the cause of the disease. To prove this Marshall exposed himself to contamination with a bacterial culture and got infected [49].

H. pylori has a peculiar biology being confined in such a very unhospitable environment as the surface of stomach epithelial cells. It is protected from acidity by the mucus layer. It is a Gram negative bacterium provided with flagella for motility. Urease production is the main factor allowing the bacterium to survive in the very acidic environment of the stomach lumen before reaching the mucus layer, a natural protection against the low pH. VacA and urease are the best characterised virulence factors. The vacuolating toxin VacA causes vacuoles formation in the stomach cells cytoplasm leading eventually to the complete filling of the cell. The process is aided by the presence of ammonium ions and starts from small vacuoles in the perinuclear zone of the stomach epithelial cell. An increased permeability to metal ions like Fe^{3+} and Ni^{2+} is also induced by VacA in order to supply the bacterium with the high amount of nickel required for the production of urease. The main role of urease in *H. pylori* infection is the creation of a "cloud" of ammonia used not only for bacterial cell protection but also cooperatively to help VacA. Damaging the epithelial cells allows vacuolation to proceed faster [50].

It has recently been demonstrated that *H. pylori* urease although cytoplasmic

is associated with the cell surface in late log-phase bacteria. In the absence of surface-bound urease, *H. pylori* is not able to survive in acidic media even in the presence of urea. *In vitro* experiments showed that *H. pylori* cultures in early and mid-log phases do not survive the acidic environment even in the presence of urea and possess only cytoplasmic urease. Late log-phase bacteria contain surface bound cytoplasmic enzymes and do survive acidic condition when cultured in the presence of urea. The mechanism allowing the surface association of urease is proposed to be an "altruistic autolysis". Genetically progammed autolysis of a fraction of the bacterial population would occur in order to release urease and other cytoplasmic enzymes then adsorbed onto the cellular surface of neighbouring bacteria. The survival of *H. pylori* in the highly acidic stomach is even more surprising if we consider that 99.9 % of *Salmonella* and *Vibrio sp.* are killed by gastric acid [51].

1.6 Urease from *Canavalia ensiformis*

In seeds of legumes urease is involved in the mobilisation of nitrogen stored as arginine or canavanine and is mobilised upon germination when the demand for nitrogen compounds is higher and cannot be satisfied by root absorption [52]. In the metabolism of adult plants urease is used to convert urea to ammonia which is then assimilated as a source of nitrogen [53, 14].

Urease from jack-bean (*Canavalia ensiformis*) was the first enzyme to be crystallised by Sumner [1] in 1926. The importance of the discovery goes far beyond the simple achievement of crystals from a protein, indeed it represents a fundamental step in protein science. Before Sumner's discovery no enzyme had been purified and their chemistry was completely unknown. Although

jack bean urease cystals have been available since then the structure has not yet been solved. Crystals have been recently obtained in a cubic space group $F4_132$ symmetry with $a = 364$ Å. They diffracted to 3.5 Å when using synchrotron radiation [54]. The molecular weight of this enzyme is about 545.34 kDa. It is composed of six subunit of 90.77 kDa. The individual subunits can be seen as derived from the fusion of the α, β, γ bacterial subunits forming a single polypeptide chain [55]. Studies with electron microscopy clearly showed the arrangement of the subunit to form hexagonally shaped molecules.

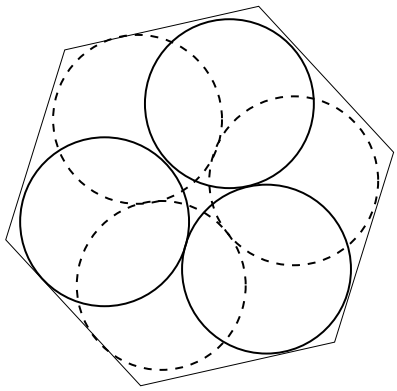


Figure 1.4: Model of the subunits arrangement in jack bean urease. The two layers are drawn with solid and dashed bold lines.

This hexagonal shape can be explained by an arrangement of the six subunits with D_3 symmetry obtained by stacking two trimers to form a trigonal antiprism (Fig. 1.4) [12]. The resulting octahedron can be imagined as formed by the pairing of two bacterial urease molecules. Jack bean urease was the first protein shown to contain nickel in the active site. In 1975 Dixon *et al.* [2] in the paper reporting the discovery say: "*In 1926, Sumner isolated the first enzyme, urease, and defined the proposition that enzymes could be proteins devoid of organic coenzymes and metal ions. It is therefore with some sadness that in this communication we adduce evidence which strongly indicates that urease is a nickel metalloenzyme*". The presence of six-coordinate octahedral Ni^{2+} in jack bean urease was first proposed on the basis of optical absorption spectroscopic studies [2, 3, 33, 56], which almost completely excluded the presence of four or five-coordinated nickel ions. The metal content was quantified by atomic absorption spectrocopy, but a better procedure involves dry ashing

followed by spectrophotometric analysis using dimethyl glyoxime [57]. The nickel atoms were 2.0 ±0.12 per 96.6 g of protein indicating the presence of a binuclear active site per subunit [12].

X-ray absorption spectroscopy (XAS) data were interpreted with a model involving distorted octahedral Ni^{2+} ions bound to five or six nitrogen or oxygen donors at an average distance of 2.06 Å. Extended X-ray Absorption Fine Structure (EXAFS) studies revealed the appearance of a new peak in the Fourier transform upon addition of β-mercaptoethanol to jack bean urease. This peak was fitted using a model with two nickel ions 3.26 Å apart. The presence of histidines as metal ligands was also proposed [44].

1.7 Urease from *Bacillus pasteurii*

The Gram positive aerobic alkalophilic bacterium *B. pasteurii* is our hero. Found in soil, water, sewage and incrustation on urinals, *B. pasteurii* has the highest ureaclastic capacity from ureolytic bacteria. The isolation of a pure culture from soil only requires the presence of urea in the culture medium and temperatures between 26 and 30 °C. Ammonia production from urease activity selects in few days, among all bacteria developed, only the most viable: *B. pasteurii*. Its importance is mainly related to agricultural and environmental concerns, because of the release of ammonia upon urea degradation.

Urease from *B. pasteurii* was the first purified from a bacterial source since it is constitutively highly produced [37]. Previous SDS-PAGE electrophoresis studies indicating the subunit composition to be α_4 (with α=65.5 kDa) [58], or $\alpha_2\beta_2\gamma_2$ [35], were based on Coomassie stained bands intensities. The wrong

interpretations were probably due, in the first case, to bands running out from the gel, in the second to the different affinity of polypeptide chains for the dye (different ratio between aromatic residues and polypeptide mass). Genetic studies indicated the presence of three chains $\alpha\beta\gamma$ of molecular weight 61.4, 14.0 and 11.1 kDa [59].

The first structural studies on the *B. pasteurii* urease active site were carried out using Ni-edge X-ray absorption spectroscopy (XAS) suggesting an average coordination of the Ni^{2+} ions of five or six nitrogen or oxygen ligands, with two of the ligands being histidines [60]. The distances found for the first coordination sphere of nickel in *B. pasteurii* urease are the same as those found in the EXAFS spectra of jack bean urease and *K. aerogenes* [61]. The edge and EXAFS analysis of the phenylphosphorodiamidate (PPD) inhibited *B. pasteurii* urease indicate that the first coordination sphere was not significantly affected by the inhibitor binding, and that phosphorous was present in the outer shells [60].

The structural studies by X-ray crystallography on *B. pasteurii* urease are one of the two subjects of this thesis.

Chapter 2

Material and Methods

2.1 Bacterial Growth and Protein Purification

Bacillus pasteurii (DSM, type strain 33), was grown[1] under aerobic conditions, by blowing compressed air into a liquid medium containing yeast extract (20 g/L), urea (20 g/L), and 1 mM $NiCl_2$ at pH 9 and a temperature of 28 °C. The cells were allowed to grow for 16 hours giving a yield of 5 g per liter of culture medium. The cells were harvested by centrifugation at 30,000 g for 15 minutes and washed three times with 50 mM phosphate buffer pH 7.5 containing 1 mM EDTA (ethylenediaminotetraacetic acid) to chelate free metal ions and inhibit metalloproteinases, and 50 mM Na_2SO_3 to protect the enzyme from oxidation, (buffer A). About 100 g (wet weight) of cells were resuspended in buffer A and disrupted using a French press operating at 20,000 psi. The extract was subjected to differential centrifugation first at 30,000 g for 30 minutes and then ultracentrifuged at 150,000 g for 2 hours 30 minutes, in order to remove cell wall debris and aggregates.

[1] Bacterial growth and protein purification were carried out at the Institute of Agricultural Chemistry, University of Bologna (Italy).

The crude extract, about 400 mL, was dialysed overnight using dialysis membrane with a molecular weight cut-off of 2,000 Dalton against buffer A. After centrifugation at 30,000 *g* for 15 minutes, to remove precipitate, the dialysed protein solution was loaded onto a Q Sepharose XK 50/20 Pharmacia anionic exchange column, previously equilibrated with buffer A, and washed with buffer A, at a flow rate of 6 mL/min, until the elution profile approached 0. A gradient procedure was used to elute the protein with steps of increasing ionic strength. NaCl was used up to concentrations of 150, 250, 450 mM (in buffer A) to elute with 1 L of solution for each step. Active fractions detected in the last step were pooled and the ionic strength raised to 1 M $(NH_4)_2SO_4$ using a 3 M ammonium sulphate solution in buffer A (50 mL concentrated solution added to 100 mL protein solution).

After centrifugation at 30,000 *g* for 15 minutes, to remove precipitate, the solution was loaded onto a Phenyl Sepharose XK 26/20 Pharmacia hydrophobic interaction column equilibrated with 1 M $(NH_4)_2SO_4$ and eluted with a linear gradient from 1 M to 0 M at a flow rate of 3 mL/min. Urease, eluted with 400 mM $(NH_4)_2SO_4$, was concentrated using an Amicon ultrafiltration cell with membrane of 100,000 Dalton molecular weight cut-off and loaded onto a Sephacryl S300 XK 26/60 Pharmacia gel filtration column equilibrated with buffer A and eluted at a flow rate of 1 mL/min.

The peak shown to contain urease was loaded onto a Mono Q 10/10 FPLC Pharmacia anionic exchange column, equilibrated with buffer A, and urease was eluted using a linear gradient from 0 to 500 mM. The active fractions were concentrated and further purified to homogeneity using a Superose 12 HR 10/30 FPLC Pharmacia gel filtration column equilibrated with buffer A,

plus 150 mM NaCl in order to prevent aspecific interaction between the protein and the resin, at a flow rate of 0.5 mL/min. Urease active fractions from the last step were concentrated to 11 mg/mL and stored in buffer A at 4°C. Figure 2.1 shows a flowchart of the purification.

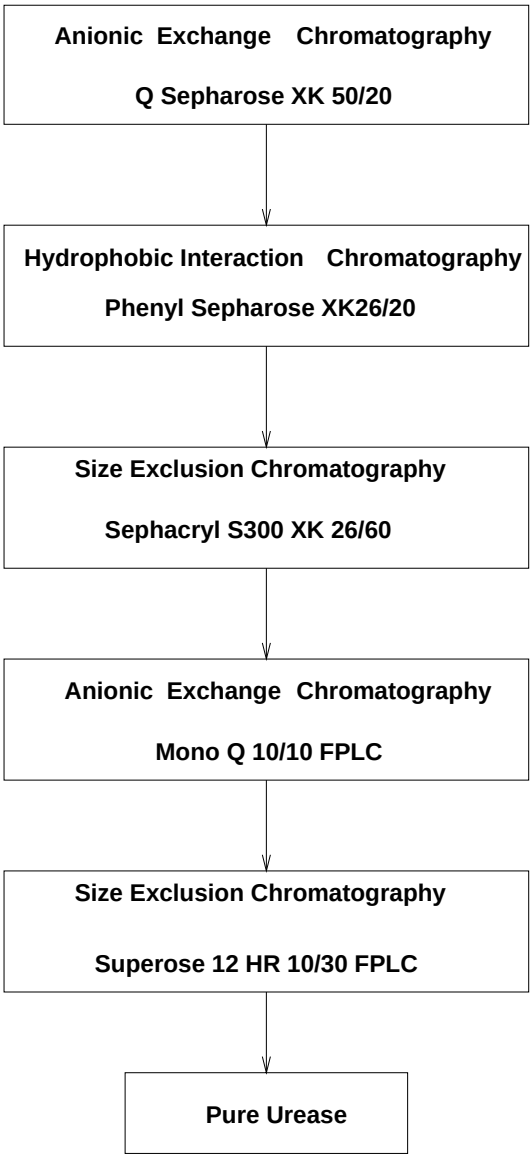


Figure 2.1: Flowchart of urease purification.

Native and SDS gels were performed using 4-20% polyacrylamide gradient gels. The gels were stained with Commassie R250 Blue. A single band was detected in the native gel and three well resolved bands were detected in the SDS gel given by the three subunit α, β, γ running at molecular weights of about 60, 14, 11 kDa corresponding to 61.397, 13.958, 11.146 kDa as determined by gene sequencing [59].

2.2 Activity Assay

2.2.1 Qualitative Assay

Qualitative assays were performed at each step of the purification to identify the active fractions in the eluates from the columns. 5 μL of enzyme containing solution were added to 50 μL of a solution containing 1 M urea, 2 mM phosphate buffer pH 7.0 and 0.01 w/v cresol red as pH indicator. The assay solution is yellow/red and turns to purple as the urease activity raises the pH of the solution to 9.

2.2.2 Quantitative Assay

Specific activity was quantified at 30°C, using a pH-stat method [29] and a Crison Micrott 2022 equipment. Activity was measured for 3 minutes after the addition of 100 μL from a solution of 4 M urea to 10 mL 20 mM phosphate buffer pH 7.0 containing previously added urease, under continuous stirring. The pH was mantained constant by the addition of small quantities of 10 mM HCl. The quantity of enzyme was measured according to the consumption of HCl needed to neutralise the pH increase. One unit of urease activity is defined as the amount of enzyme needed to hydrolyse 1 μmol of urea at 30 °C $\cdot$ min^{-1}.

B. pasteurii urease has a specific activity of 2500, which means that 1 mg of enzyme is able to hydrolyse 2500 ± 100 μmol of urea at 30 °C · min^{-1} [35]. The old colorimetric assay based on the indophenol reaction [30] was not used because the use of ammonium sulphate during the purification interfered with the measurement and because it was time consuming.

2.3 Protein Crystallisation

The hanging drop vapour diffusion method was used for all the crystallisation trials and cocrystallisation was chosen to obtain the urease-inhibitor complexes. The protein buffer was exchanged in a Centricon ultrafiltration unit, in order to preserve the native enzyme or to achieve inhibition before crystallisation. The native enzyme was exchanged with 20 mM Tris-HCl pH 7.5 containing 50 mM of Na_2SO_3. The inhibitor complexes were prepared in 20 mM Tris-HCl pH 8.0 by adding either 4 mM of β-mercaptoethanol (BME), or 4 mM of phenylphosphorodiamidate (PPD), or 4 mM mercaptoethanolamine (MEA), or 4 mM acetohydroxamic acid (AHA) respectively. The phosphate (PO_4) inhibited protein was stored in 50 mM phosphate buffer pH 7.5 containing 50 mM of Na_2SO_3.

The protein concentration in all the trials was 11 mg/mL. Aliquots of 200 μL of urease were stored in liquid nitrogen for later use. A volume of 2 μL of protein solution mixed with 2 μL of precipitant solution was equilibrated against 1 mL of precipitant in 24-well Linbro plates.

2.3.1 Initial Crystallisation Screening

An extended screening was performed using the native enzyme at two different temperatures 4 and 20°C by a sparse matrix method [62] and different concentrations of ammonium sulphate (AMS), polyethylene glycol (PEG) 6000 and PEG 6000 plus LiCl at various pH. The buffers used to obtain the different pHs were 100 mM sodium acetate at pH 5.0, 100 mM sodium citrate at pH 6.3, 100 mM sodium citrate at pH 7.0, 100 mM Tris-HCl at pH 8.0, 100 mM Tris-HCl at pH 9.0. The concentration of precipitants ranged between 1.6 M and 3.2 M for AMS, from 15 to 30% for PEG 6000 and from 15 to 30% of PEG 6000 and from 0.5 to 1 M of LiCl for PEG 6000 plus LiCl. Numerous small crystals grew within 7 days at 4°C from a solution containing 30% (v/v) 2-methyl-2.4-pentanediol (MPD) and 200 mM $CaCl_2$ in 100 mM sodium acetate pH 4.6, or using 28% (v/v) PEG 400 and 200 mM $CaCl_2$ in 100 mM HEPES pH 7.5, or again using 55% saturated ammonium sulphate in 20 mM Tris-HCl pH 7.0 very thin plates were formed. Very small prismatic crystals were obtained using 60% of a solution containing saturated AMS and 1.2 M LiCl, as co-precipitant, in 20 mM sodium citrate pH 6.3 in presence of amorphous precipitate. The same conditions as above but at 20°C gave only amorphous precipitate except for one containing 60% of saturated AMS and 1.2 M LiCl that gave single rice-shaped crystals. Optimising this crystallisation condition by varying slightly the concentration of precipitant gave bigger crystals, suitable for data collection.

2.3.2 Optimisation of Crystallisation Conditions

A fine-tuning of the conditions was carried out in order to improve the dimensions and therefore the diffraction quality of the crystals. A range of concentra-

tions was set up from 50 to 60% of the stock solution (saturated AMS + 1.2 M LiCl in 20 mM sodium citrate buffer pH 6.3). The best conditions were found around 52-56% with little variation in size and number of crystals obtained in the drops. Crystals grew on average within one week to a maximum size of about $0.2\times0.2\times0.5$ mm^3. A screening of different concentrations of precipitant for all the inhibitors was carried out based on the previous results obtained on the native enzyme by adding the appropriate inhibitor to the buffers.

- For β-mercaptoethanol (BME) inhibited urease the best conditions were found between 50 and 55% of a stock solution of saturated AMS plus 1.2 M LiCl in 20 mM sodium citrate buffer pH 6.3 + 4 mM BME.

- For the PPD inhibited urease new crystallisation conditions were found between 1.8 and 2.0 M AMS + 4mM PPD in 100 mM sodium citrate buffer pH 6.3. A second set of conditions based on the native urease crystallisation was found between 51 and 54% (saturated AMS + 1.2 M LiCl) in 100 mM sodium citrate pH 6.3 containing 4 mM PPD.

- For AHA inhibited urease the crystallisation conditions were 42-46% (saturated AMS + 1.2 M LiCl) in 100 mM sodium citrate pH 6.3 + 4mM AHA.

- For MEA inhibited urease the crystallisation conditions were 62-70% AMS 3.2M + 4mM mercaptoethanolamine.

- For the phosphate inhibited enzyme the protein, previously stored in 50 mM sodium phosphate pH 7.5 containing 50 mM Na_2SO_3 to prevent the active site from oxidation, was equilibrated against 52-58% (AMS 3.2 M) 100 mM sodium phosphate buffer pH 6.3.

2.4 Collecting Diffraction Data

Preliminary tests with urease crystals mounted in a capillary demonstrated that their life-time in the intense X-ray beam was too short for data collection at room temperature. After a few frames the resolution dropped from 1.6 to 3.5 Å. Vitrifying the crystals at 100 K prolonged their life-time making it possible to collect a complete dataset from a single crystal. All data were collected using the synchrotron radiation from the BW7B wiggler line [63, 64], apart for the DAP inibited which was collected from the BW7A wiggler line at the EMBL outstation Hamburg (Germany). The radiation for both beam lines was provided by the DORIS storage ring of the Deutsches Elektronen Synchrotron (DESY).

For cryoprotection, the crystals were soaked for a short time in a solution containing 20% ethylene glycol in the mother liquor. The crystals were fished out with a nylon loop (Hampton Research) and rapidly exposed to a cold nitrogen stream (Oxford Cryosystems Cryostream). The dimensions of the crystals ranged from about $0.1 \times 0.1 \times 0.2$ mm^3 to $0.3 \times 0.3 \times 0.6$ mm^3.

A special procedure was used for DAP inhibited urease in order to prevent crystal cracking. A crystal, of dimensions $0.1 \times 0.1 \times 0.3$ mm^3, was first transferred from the crystallisation drop into 500 mL of a soaking solution (2.4 M AMS, 4 mM PPD, 100 mM sodium citrate, pH 6.3). After equilibration the concentration of the cryoprotectant was gradually increased from 0 to 15% by adding 100 mL portions of 20% ethylene glycol, at 1 min. intervals. The crystal was transferred to a full concentration soaking solution containing 20% ethylene glycol and, after 1 min, it was scooped up in a cryo-loop and rapidly exposed to the cold nitrogen stream.

Diffraction images were recorded using a 30-cm Mar Research imaging plate

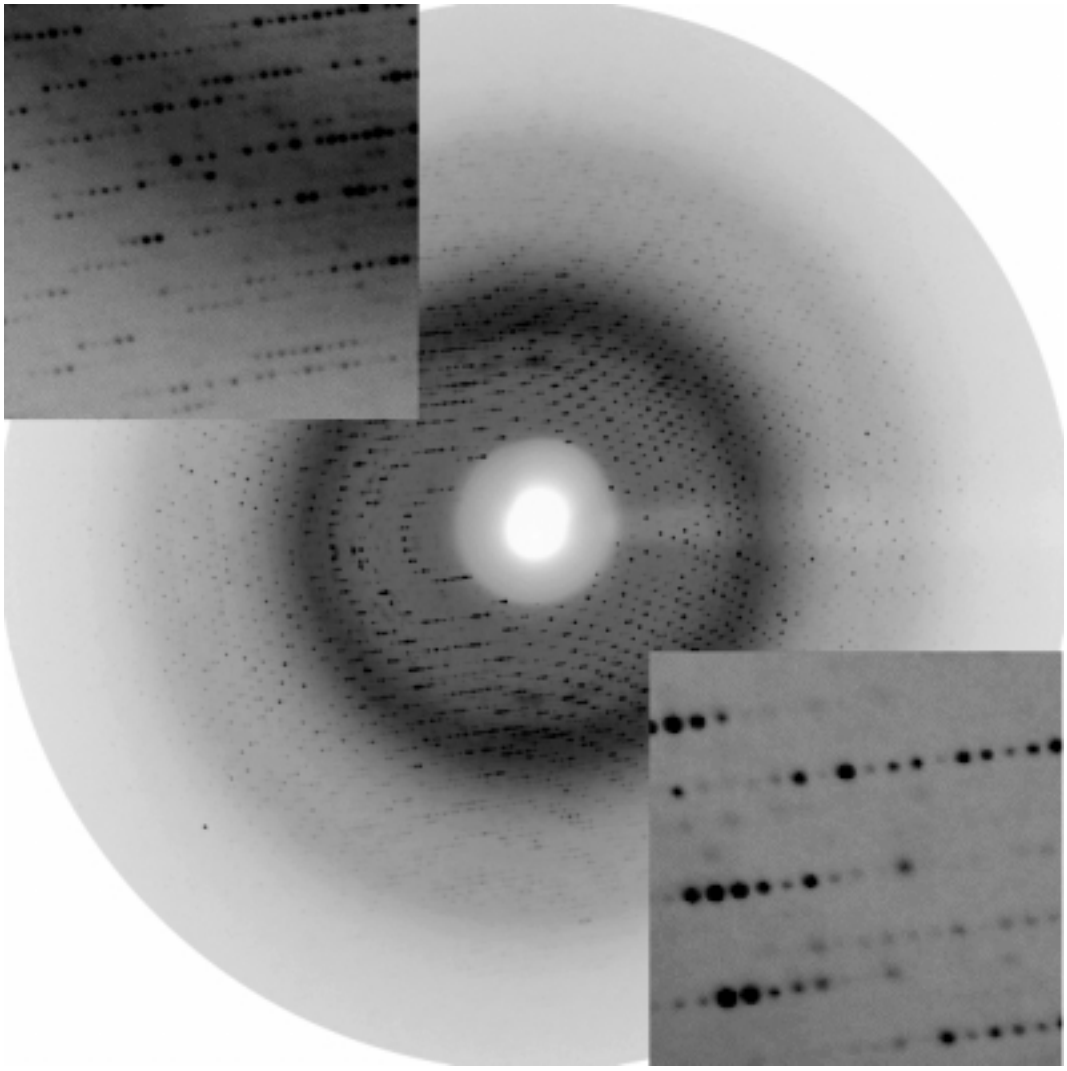


Figure 2.2: Diffraction image from a BME inhibited *B. pasteurii* urease crystal between 1.65 Å and 16 Å. The insets are magnification of the pattern.

scanner in two or three sweeps at different exposure time, crystal to detector distance and plate size in order to measure accurately both high and low

resolution diffraction spots and obtain higher redundacy and good I/σ ratio. Figure 2.2 shows a typical *B. pasteurii* urease diffraction pattern.

The Bravais lattice was identified using the autoindexing procedure implemented in the program DENZO [65, 66] as primitive hexagonal $P622$ with cell dimensions of about $a = b$ =131 Å, c =190 Å, with the six-fold axis coincident with the long crystal axis. The correct space group was identified as $P6_322$ by looking at the systematic absences. Crystals were mounted in the cryo-loop with the c axis nearly parallel to the spindle axis which minimised overlaps and allowed data collection to be completed in only 30° of rotation. To avoid overlaps due to the crystal mosaicity, (0.3-0.6°), and the big cell dimensions, data were collected in ϕ slices of 0.3° oscillation, for the high resolution passes, and of 0.6-1° for the low resolution passes.

Data were processed using the program DENZO, scaled, merged and postrefined with SCALEPACK [65, 66]. Table 2.4 reports a summary of data collection statistics and reduction for the native and inhibited *B. pasteurii* urease. The quality of the data can be judged by looking at Figure 2.3 showing a Wilson plot and a I/σ *vs.* resolution plot of BME inhibited *B. pasteurii* urease. Assuming one $\alpha\beta\gamma$ fragment (86.5 kDa) per asymmetric unit, the volume-to-mass ratio, V_m is 2.7 Å^3 Da^{-1} in the normal range for proteins (1.7-3.5 Å^3 Da^{-1} [67]) giving a solvent content of 54%. Four urease molecules ($\alpha_3\beta_3\gamma_3$) lie on special positions of point symmetry 3.

	NAT	BME	DAP
Wavelength (Å)	0.885	0.885	0.999
High Res. (Å)	2.0	1.65	2.0
Low Res. (Å)	20.0	14.0	18.0
R_{merge} %	9.7	7.6	15.0
Raw measurements	219,698	477,866	353,642
Unique reflections	63,765	114,679	65,301
Redundancy	3.44	4,17	5.41
Completeness (%)	96.7	98.7	99.9
Greater than 3σ (%)	70.2	71.1	67.4
I/σ in high res. bin	2.4	2.2	2.8
Space group	$P6_322$	$P6_322$	$P6_322$
Cell parameters			
$a = b$ (Å)	131.36	131.34	131.95
c (Å)	189.76	190.01	189.00
	AHA	MEA	PO_4
Wavelength (Å)	0.999	0.855	0.834
High Res. (Å)	1.55	1.9	1.85
Low Res. (Å)	32.5	24.0	30.0
R_{merge} %	7.1	5.0	9.7
Raw measurements	791,805	224,007	381,291
Unique reflections	138,830	82,433	82,718
Redundancy	5.70	2.70	4.61
Completeness (%)	99.5	84.5	99.3
Greater than 3σ (%)	65	76.1	63.4
I/σ in high res. bin	2.8	2.2	1.85
Space group	$P6_322$	$P6_322$	$P6_322$
Cell parameters			
$a = b$ (Å)	131.15	131.09	131.49
c (Å)	189.36	189.75	189.489

Table 2.1: Data collection summary for the *B. pasteurii* urease crystals

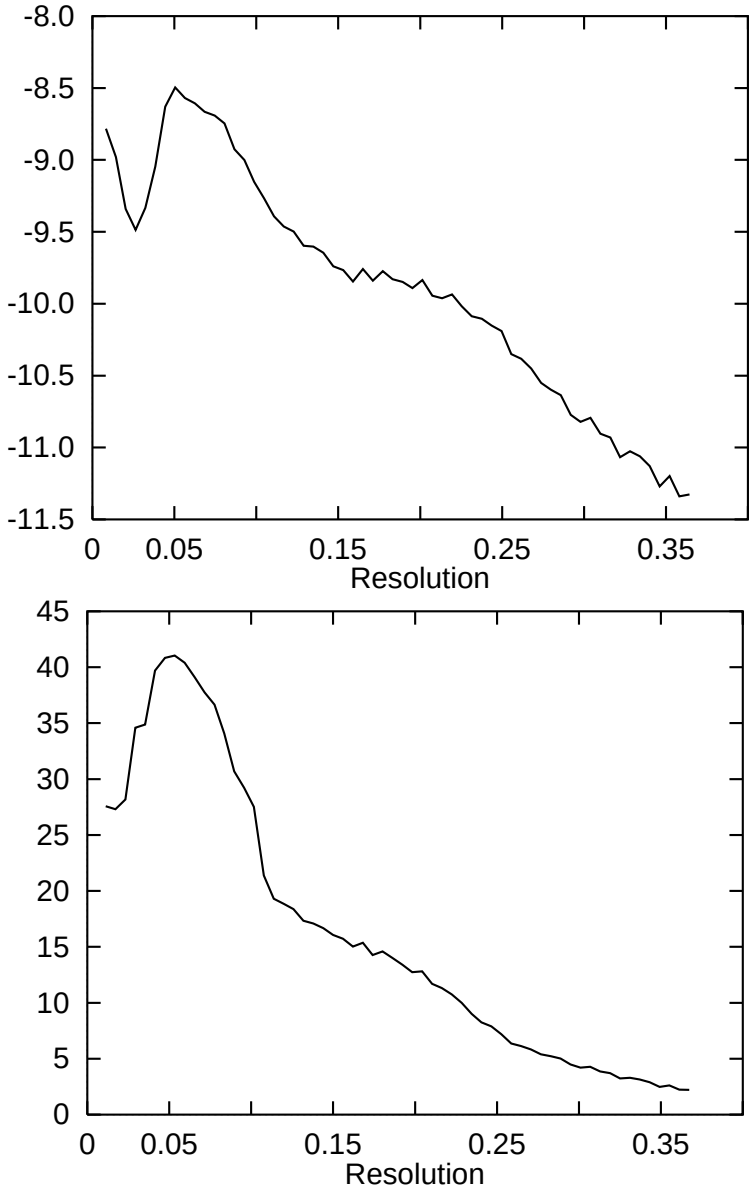


Figure 2.3: Top: Wilson plot and bottom: I/σ *vs.* resolution plot of BME inhibited *B. pasteurii* urease. The resolution is reported as $1/d^2$. In the Wilson plot the 'Y' axis is: $ln\frac{\overline{I_{(hkl)}}}{\sum_i (f_i^0)^2}$

2.5 Solving the Structure with Molecular Replacement

The structure was determined very easily by molecular replacement using the program AMoRe [68] and the data from BME-inhibited *B. pasteurii* urease in the range of resolution between 8 and 3.5 Å. The $\alpha\beta\gamma$ chains of urease from *K. aerogenes* (PDB code 1kau), solved in 1995 by Jabri *et al.* [42], were used as search model. The high sequence identity between *B. pasteurii* and *K. aerogenes* urease (63% for the α, 46% for the β and 61% for the γ subunits respectively), the highly conserved fold and the good quality of the data, allowed the structure to be solved at the first attempt.

For the rotation function the search model was placed in a $P1$ cell with dimensions $90 \times 90 \times 90$ Å. The integration radius for the Patterson synthesis was 35 Å. The solutions, sorted by correlation coefficients (CC) clearly indicated the correct one. Table 2.5 shows the progress of molecular replacement, with the second highest peak included for comparison.

2.6 Model Building, Refinement and Validation

The appropriate rotation and translation was applied to the model, and maps with coefficients $3F_{obs} - 2F_{calc}$ and $F_{obs} - F_{calc}$, when using least-squares refinement (SFALL, PROTIN, PROLSQ [69]), or $2mF_{obs} - DF_{calc}$ and $mF_{obs} - DF_{calc}$) when using maximum likelihood (REFMAC [70]), were calculated and used for inspection and model building with the graphics program "O" [71].

Rotation	α	β	γ	X	Y	Z	CC	R
Solution 1	38.34	67.14	94.36	0.00	0.00	0.00	12.2	0.0
Solution 2	10.37	121.04	344.10	0.00	0.00	0.00	5.4	0.0
Translation	α	β	γ	X	Y	Z	CC	R
Solution 1	38.34	67.14	94.36	0.538	0.823	0.394	47.4	43.8
Solution 2	10.37	121.04	344.10	0.119	0.144	0.028	5.5	56.8
Rigid Body:	α	β	γ	X	Y	Z	CC	R
Solution 1	38.97	67.40	94.49	0.537	0.822	0.394	52.1	41.9
Solution 2	9.67	120.84	343.22	0.119	0.145	0.028	7.8	56.0

Table 2.2: Molecular replacement solutions.

When the least-squares procedure was used for the BME inhibited and the native enzyme, several cycles of refinement and manual intervention were necessary to eliminate the initial model bias and build the final *B. pasteurii* urease model based on the sequence [59]. A significant improvement of refinement and map quality was observed when maximum likelihood was used, as implemented in REFMAC [70], resulting in faster refinement and improved model quality for all *B. pasteurii* urease structures.

For all datasets 2% of the reflections were flagged as R_{free} and used as a cross-validation test [72, 73] to monitor the improvement of the model and to assess quality of refinement. R_{free} is also used internally by REFMAC to optimise the refinement. Geometric targets such as bond lengths and angles were those of Engh and Huber [74] and restraints were calculated using the program PROTIN [69] before each cycle of REFMAC. Automatic insertion and updating of water molecules were carried out using the program ARP [75, 76].

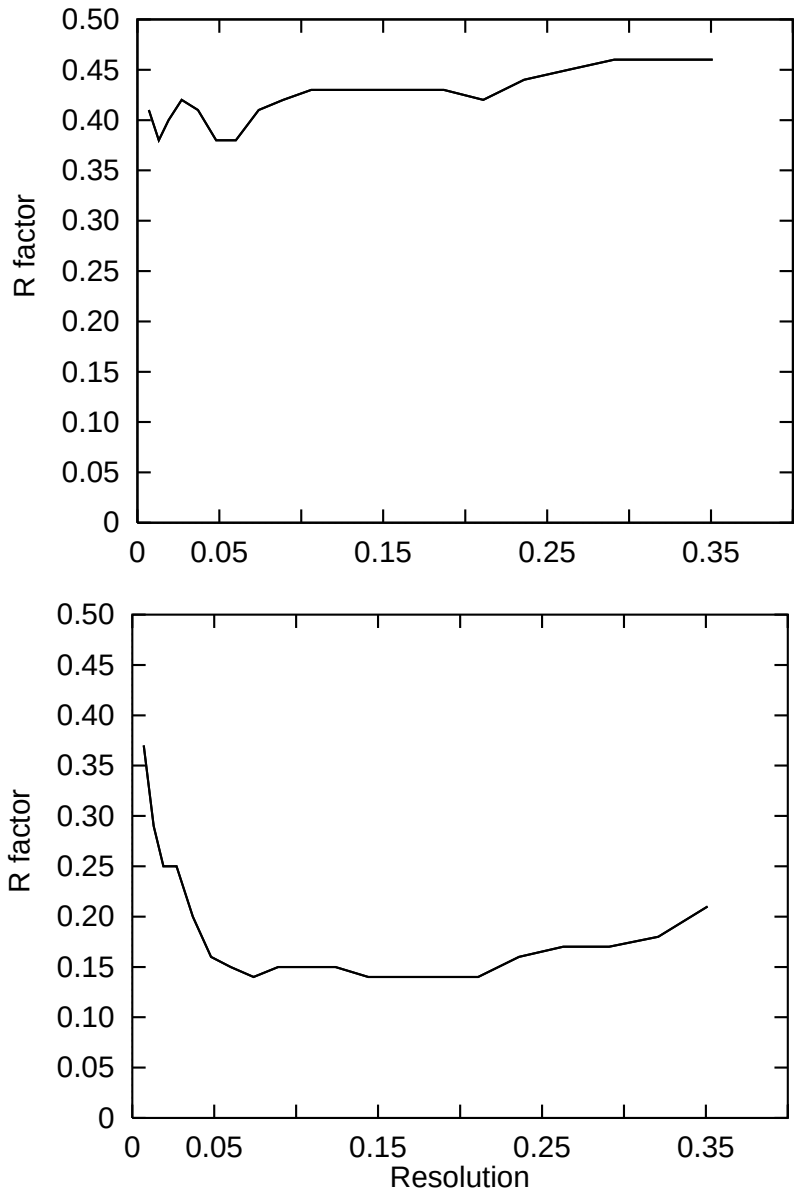


Figure 2.4: R factor *vs.* resolution for BME inhibited *B. pasteurii* urease data between 1.65 Å and 14.0 Å. Top: the initial model, bottom: the final model. The resolution is reported as $1/d^2$.

Water molecules were added if the $F_{obs} - F_{calc}$ or $mF_{obs} - DF_{calc}$ map had density higher than 3 σ and kept only if their density was higher than 1.3 σ in the $3F_{obs} - 2F_{calc}$, or $2mF_{obs} - DF_{calc}$ and if located within 2.3 and 3.3 Å from

neighbouring oxygen or nitrogen atoms. Figure 2.4 shows the R factor *versus* resolution of the initial and the final BME inhibited urease model.

The RMSD between the initial model after AMoRe and the final BME inhibited model was 6.9 Å. The BME structure was used as the starting model for native urease which itself was subsequently used for all the other complexes. Tables from 2.3 to 2.8 report a summary of structure quality for native and inhibited urease.

The stereochemical quality of the final models was checked using validation programs. PROCHECK [77] was used to check the correctness of nomenclature in accordance with the IUPAC naming conventions and to validate the secondary structure against structures in the PDB. WHATIF [78, 79] was used to check the orientation of residues involved in hydrogen bonding, the correctness of cell parameters (inaccurate determination of λ at the synchrotron results in systematic errors in bond lengths) and to list and check bad contacts. The following tables show the summary of the final models.

High Res. (Å)	2.00
Low Res. (Å)	20.00
Protein atoms	6054
Metal ions	2
Solvent atoms	881
Temperature factors (Å ^{2}):	
overall protein	21.1
main chain	17.5
side chain and solvent	24.3
solvent	33.9
Ni-1, Ni-2	19.1, 17.9
RMSD bond length	0.013
RMSD bond angle	0.028
RMSD planes	0.033
R factor, R_{free} (%)	16.2, 20.4

Table 2.3: Summary of native *B. pasteurii* urease

High Res. (Å)	1.65
Low Res. (Å)	14.00
Protein atoms	6054
Metal ions	2
Solvent atoms	1011
Temperature factors (Å ^{2}):	
overall protein	17.7
main chain	13.8
side chain and solvent	20.9
solvent	29.1
ligand (average)	17.6
Ni-1, Ni-2	14.3,12.3
RMSD bond length	0.009
RMSD bond angle	0.024
RMSD planes	0.041
R factor, R_{free} (%)	17.1, 19.4

Table 2.4: Summary of BME inhibited *B. pasteurii* urease

High Res. (Å)	2.00
Low Res. (Å)	18.00
Protein atoms	6054
Metal ions	2
Solvent atoms	841
Temperature factors (Å ^{2}):	
overall protein	15.8
main chain	12.7
side chain and solvent	18.5
solvent	27.7
ligand (average)	16.7
Ni-1, Ni-2	21.2, 16.8
RMSD bond length	0.009
RMSD bond angle	0.028
RMSD planes	0.033
R factor, R_{free} (%)	16.2, 20.4

Table 2.5: Summary of DAP inhibited *B. pasteurii* urease

High Res. (Å)	1.55
Low Res. (Å)	32.50
Protein atoms	6054
Metal ions	2
Solvent atoms	746
Temperature factors (Å ^{2}):	
overall protein	26.6
main chain	26.5
side chain and solvent	26.7
solvent	41.0
ligand (average)	21.6
Ni-1, Ni-2	22.1, 20.5
RMSD bond length	0.013
RMSD bond angle	0.030
RMSD planes	0.038
R factor, R_{free} (%)	15.1, 19.1

Table 2.6: Summary of AHA inhibited *B. pasteurii* urease

High Res. (Å)	1.90
Low Res. (Å)	24.00
Protein atoms	6054
Metal ions	2
Solvent atoms	801
Temperature factors (Å ^{2}):	
overall protein	22.7
main chain	21.2
side chain and solvent	24.1
solvent	28.0
ligand (average)	19.8
Ni-1, Ni-2	20.0, 20.4
RMSD bond length	0.006
RMSD bond angle	0.025
RMSD planes	0.032
R factor, R_{free} (%)	17.6, 21.5

Table 2.7: Summary of MEA inhibited *B. pasteurii* urease

High Res. (Å)	1.85
Low Res. (Å)	30.00
Protein atoms	6054
Metal ions	2
Solvent atoms	896
Temperature factors (Å ^{2}):	
overall protein	30.7
main chain	26.8
side chain and solvent	34.3
solvent	47.1
ligand (average)	40.9
Ni-1, Ni-2	35.9, 34.16
RMSD bond length	0.012
RMSD bond angle	0.033
RMSD planes	0.044
R factor, R_{free} (%)	17.3, 21.1

Table 2.8: Summary of phosphate inhibited *B. pasteurii* urease

2.7 The "*Not Allowed*" Ile β99

The stereochemistry of the main chain dihedral angles shows tight clustering with 89.7% of the residues in the most favoured region in the Ramachandran plot and 9.2% in the additional allowed regions (Fig. 2.5). One residue (Ile

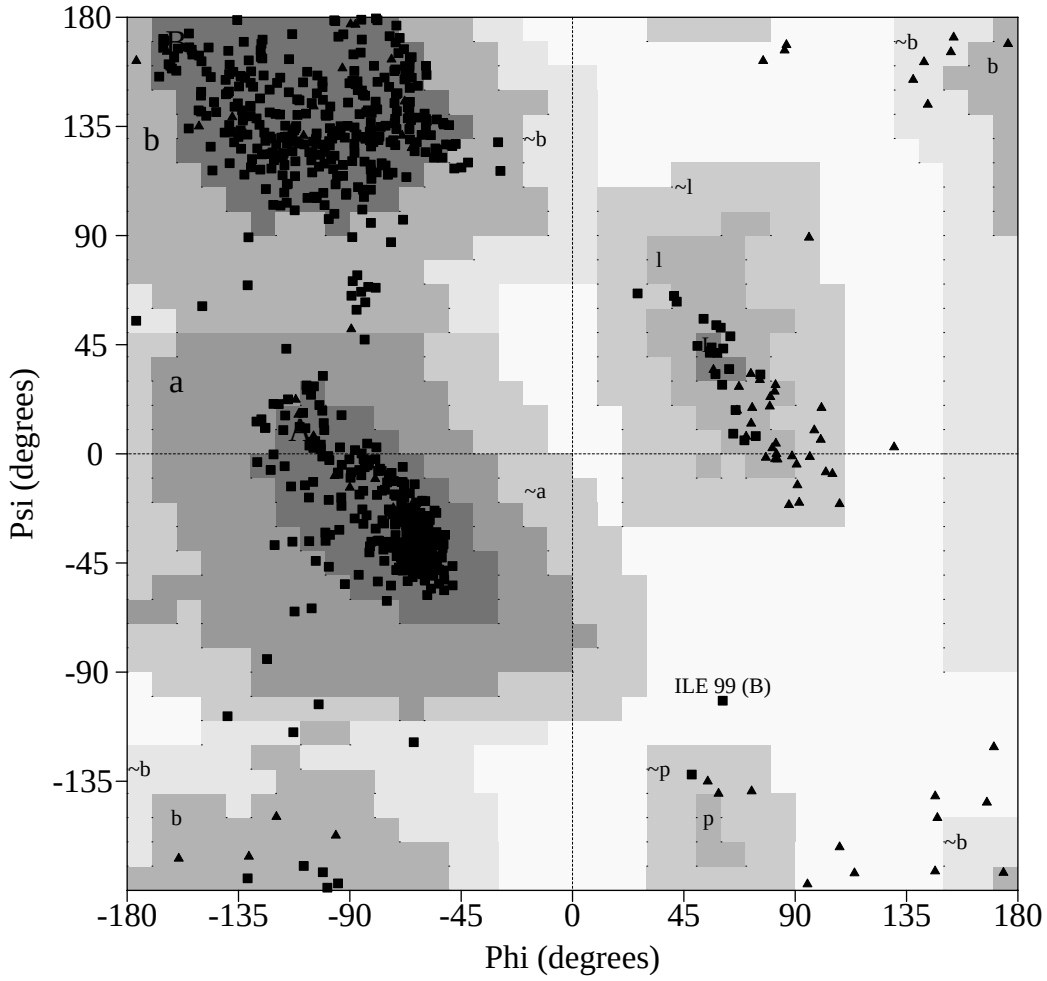


Figure 2.5: Ramachandran plot of the 1.55 Å resolution AHA inhibited *B. pasteurii* urease and the "not allowed" Ile β99.

β99) is located in a disallowed region with unusual values for the ϕ and ψ an-

gles ($\phi = 58°$, $\psi =$ -103°). Ile β99 is situated in clear electron density (Fig.2.6) on a turn between two strands (residues β95-β96 and β105-β106) forming an antiparallel sheet. In this region the β subunit makes extensive contacts with the α and α' subunits. The interactions stabilising Ile β99 in its conformation are hydrogen bonds between the main chain nitrogens and carbonyl oxygens and the hydrophobic interactions between the side chain and the core of a hydrophobic pocket formed between the subunits β, α and α'. In *K. aerogenes* urease the corresponding Phe β93has similar ϕ and ψ angles ($\phi = 59°$, $\psi =$ -120°) although the *K. aerogenes* amino-acid sequence is 21 residues shorter than *B. pasteurii* and does not feature any definite secondary structure after the antiparallel β sheet formed by residues β21-β28 and β76-β83. Figure 2.6 compares *B. pasteurii* urease β98-β101 and *K. aerogenes* urease β92-β95.

B. pasteurii Ile β99 N is 3.0 Å from Glu α52 Oϵ2, and Ile β99 O is 3.05 Å from Ala α'230 N. Glu α52 Oϵ1 further binds Thr α'228 Oγ1 at 2.52 Å and Ser α'231 Oγ at 2.77 Å. The tight turn is stabilised by the hydrogen bonds between Gly β98 O with Ser β100 N at 2.84 Å and Asp β101 N again at 2.84 Å. Gly β98 N is at 2.89 Å from Asn β104 Oδ1. Ser β100 N forms a bond with Glu β50 Oϵ1 at 2.87 Å.

The coordinates and the structure factors of *B. pasteurii* urease have been deposited at the Protein Data Bank (PDB) with accession codes 2ubp for the native, 1ubp for the β-mercaptoethanol, 3ubp for the diamidophosphate inhibited, 4ubp for the acetohydroxamic acid inhibited, 5ubp for cysteamine inhibited and 6ubp for the phosphate inhibited.

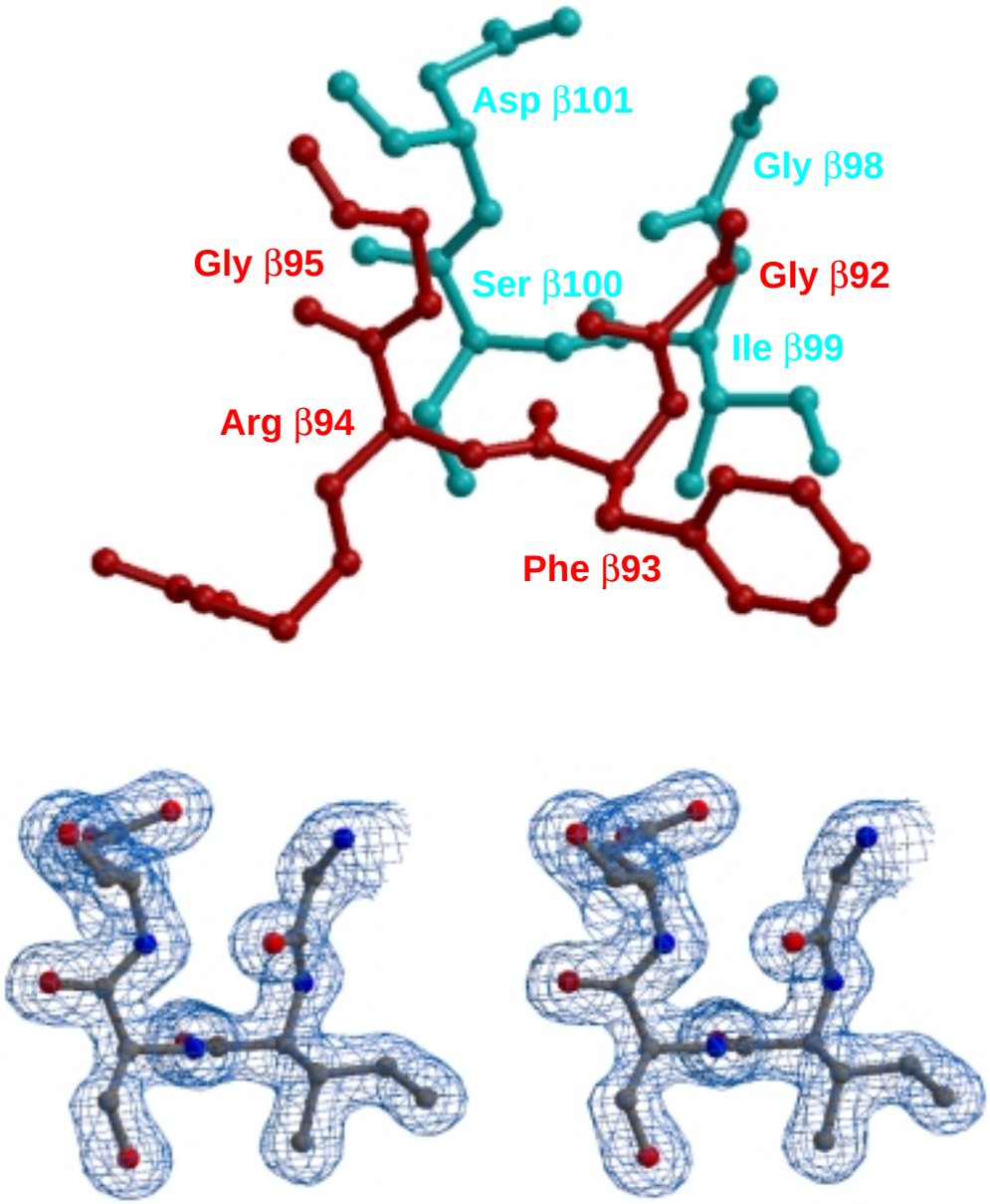


Figure 2.6: Top: Structural comparison between *B. pasteurii* urease β98-β101 (cyan) and *K. aerogenes* urease β92-β95 (red). Bottom: Stereoview of the $2mF_{obs} - DF_{calc}$ electron density map showing the tight turn and the very well defined electron density in the region between Gly β98 and Asp β101.

.

2.8 The Overall Structure

The urease quaternary structure is characterised as a trimer of trimers $(\alpha\beta\gamma)_3$ forming a big triangular molecule with dimensions approximately 120 Å on each side (Fig. 2.7). The T-shaped basic unit $\alpha\beta\gamma$ (Fig. 2.8), has dimensions about $75 \times 80 \times 80$ Å, arranged around the three-fold molecular axis (coincident with the crystallographic axis of the $P6_322$ cell) generates trimers denoted as $\alpha\beta\gamma$, $\alpha'\beta'\gamma'$ and $\alpha''\beta''\gamma''$. Each subunit makes extensive contacts with the others, stabilising the structure by means of numerous hydrogen bonds and hydrophobic interactions. The sides of the triangle are represented by $\alpha\ \alpha'\ \alpha''$ subunits (570 residues each) packed together in a head to tail arrangement. The $\beta\ \beta'\ \beta''$ subunits (121 residues each) are located at the vertices of the triangle formed by the $\alpha\ \alpha'\ \alpha''$ subunits. β interacts with $\alpha\ \alpha'$, β' with $\alpha'\ \alpha''$ and β'' with $\alpha''\ \alpha$. The small $\gamma\ \gamma'\ \gamma''$ subunits (100 residues each) are located on one side of the $\alpha\ \alpha'\ \alpha''$ subunits and pack closely around the three fold symmetry axis. The γ subunit makes extensive interactions with $\alpha\ \alpha'$ and with both γ' and γ''. An interesting interaction takes place between $\gamma\ \gamma'\ \gamma''$ involving a network of hydrogen bonds between Gln γ12, Gln γ'12 and Gln γ''12 around the three-fold axis (Fig. 2.9). The active sites are located in each of the $\alpha\ \alpha'\ \alpha''$ subunits about 50 Å from each other and characterised by the presence of two Ni^{2+} ions in each.

The $(\alpha\beta\gamma)$ unit is composed of four structural domains. The α subunit consists of two structural domains while β and γ are single domains. The α subunit catalytic domain consists of an $(\alpha\beta)_8$ barrel in which the active site is located and an additional mobile flap, involved in the reaction mechanism, formed by a helix-loop-helix motif (residues 311-339). About 100 residues at the C-terminus of the α subunits are wrapped around the $(\alpha\beta)_8$ domain like a hoop

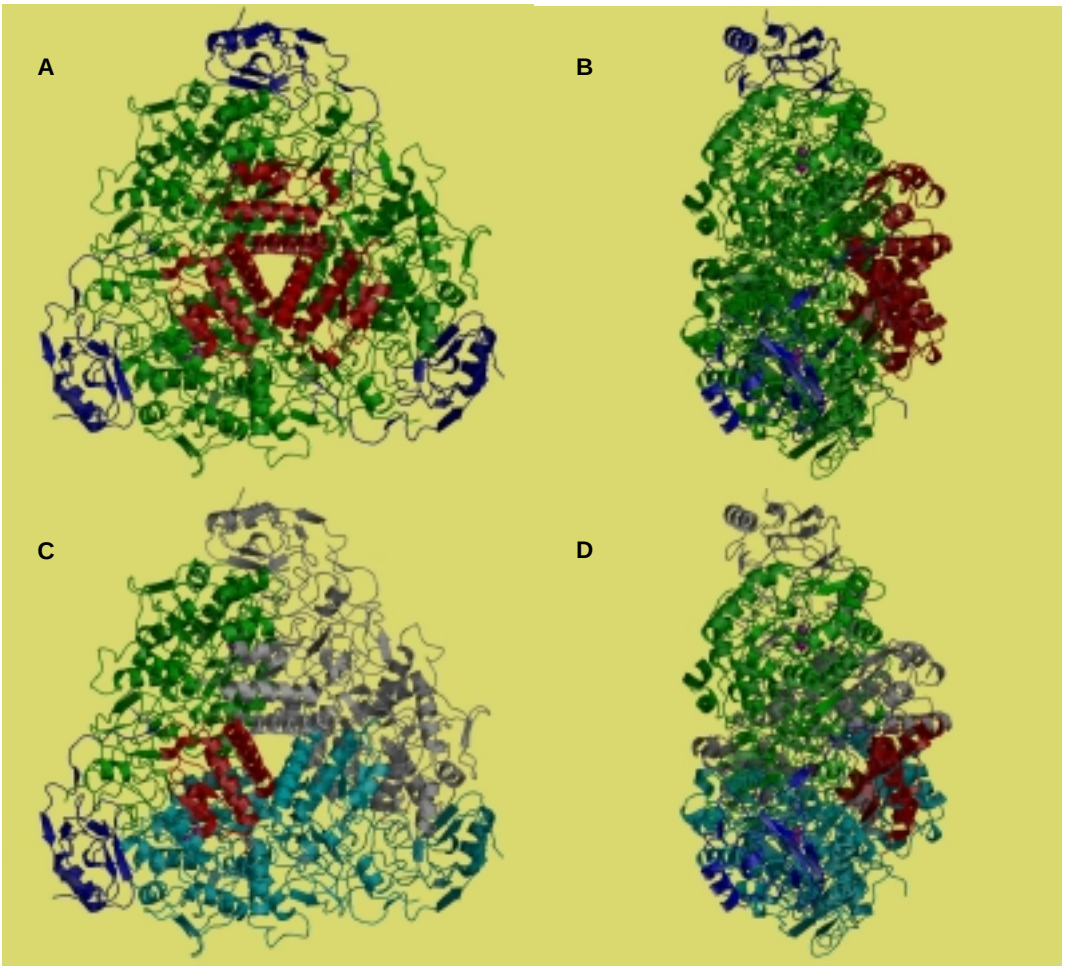


Figure 2.7: Quaternary structure representation of *B. pasteurii* urease. In figures A and B α, β and γ are coloured in green, blue and red. In figures C and D α, β and γ are still coloured in green, blue and red while $\alpha'\beta'\gamma'$ is coloured in cyan and $\alpha''\beta''\gamma''$ in grey. Nickel ions are depicted as magenta spheres. The figure was made with "O" [71].

of a wine barrel making it more stable. The second structural domain in the α subunit consists of an arrangement of β-sheets to form the walls of a U-shaped valley. The β subunit structural domain is mainly a β-structure with a six-stranded antiparallel β-jellyroll, two short helices, two antiparallel β-strands

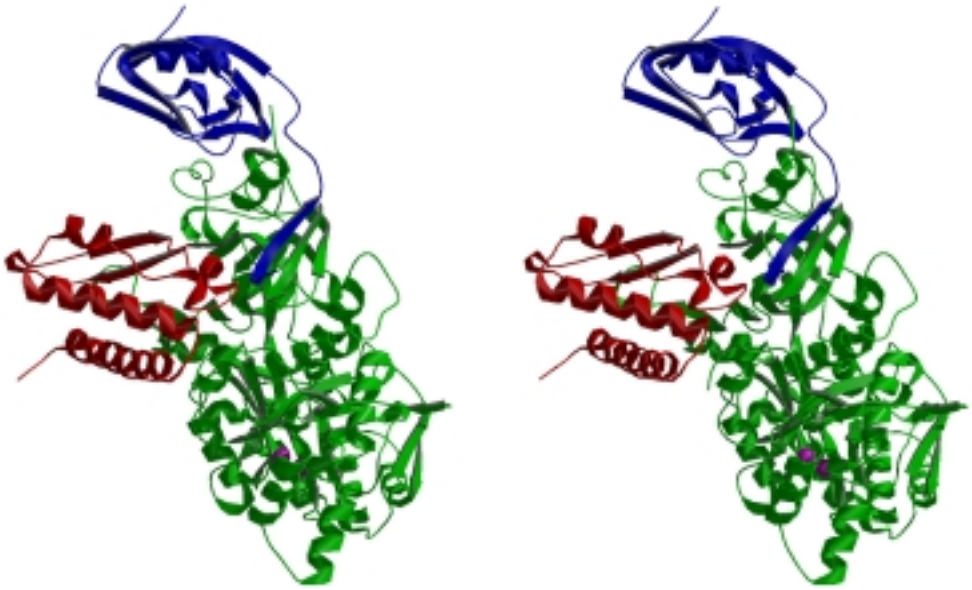


Figure 2.8: Stereo representation of the *B. pasteurii urease* $\alpha\beta\gamma$ trimer. α, β and γ are coloured in green, blue and red. Nickel ions are depicted as magenta spheres. The figure was made with BOBSCRIPT [80] and Raster3D [81].

and a long helix (residues 109-120) at the C-terminus. The N-terminal residues of the β subunits form two antiparallel β sheets with the N-terminal residues of the α subunits. The γ subunit features five helices and two antiparallel β strands. Two of the helices and the strands pack tightly together like the helices of a four-helix bundle with the same right handed up and down topology seen for that family. A complete description of the secondary structure according to DSSP [82] is shown in Figure 2.10 in a topology diagram.

The structural alignment carried out with LSQKAB [69] between *B. pasteurii* and the previously published *K. aerogenes* native urease [42] shows the almost perfect match of the secondary structural elements (Fig. 2.11). The alignment made using the Cαs gives an overall RMSD of the distances between the two

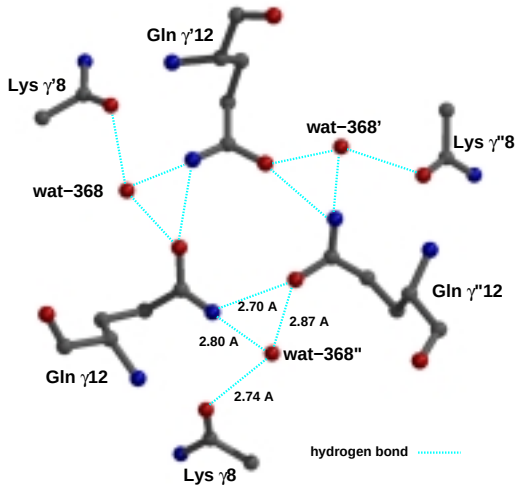


Figure 2.9: One of the interactions holding together the $\gamma\gamma'\gamma''$ subunits, looking down the symmetry three fold axis in native *B. pasteurii* urease. The figure was made with BOBSCRIPT [80] and Raster3D [81].

structures of 6.28 Å because of the large deviation of the β subunit. The α and γ subunits almost fit perfectly apart from the stretch of residues α320-α334 where the average deviation is 3.9 Å (His α324 has the maximum deviation with 5.4 Å). If the fit between the two structures is made only using the α and γ subunits the RMSD drops to 1.02 Å and the average RMSD of the conserved active site residues is only 0.48 Å. The biggest difference is in the β subunit with a twelve amino acid long helix (β109- β120) at the C-terminus in *B. pasteurii* urease whose β subunit sequence is twenty residues longer than that of *K. aerogenes*. Furthermore the helix-turn-helix motif, proposed to act as a mobile flap important for regulating the access into the active site groove [32], is in the open conformation in native *B. pasteurii* urease, while the equivalent region in *K. aerogenes* native urease is in the closed conformation (Fig. 2.11). This helix-turn-helix motif in native *B. pasteurii* urease is characterised by high mobility as judged by the temperature factors and the high B-factors of

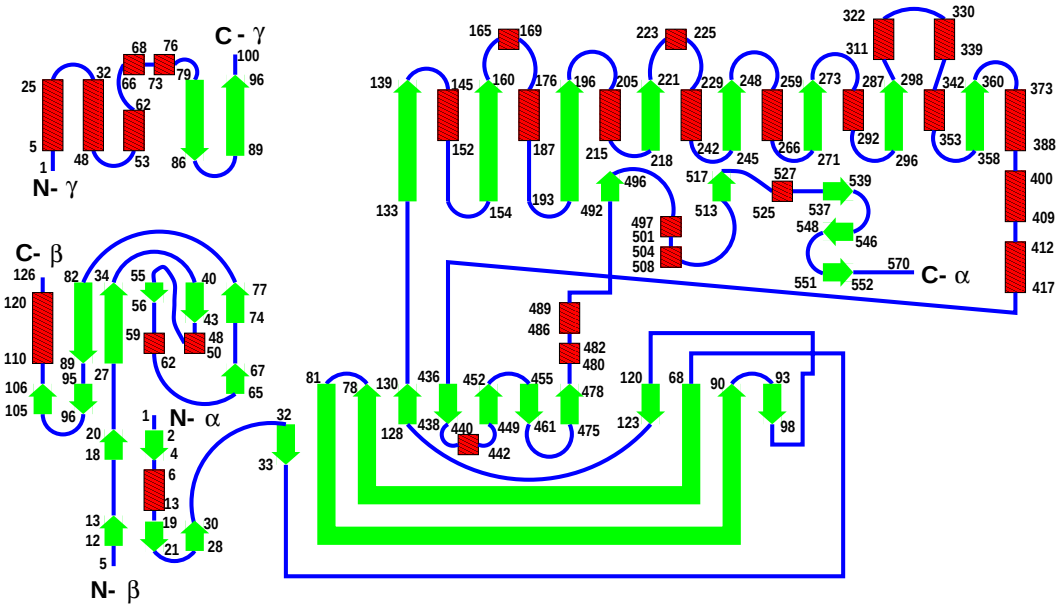


Figure 2.10: Topology diagram of *B. pasteurii* urease subunits showing the four structural domains in the $\alpha\beta\gamma$ assemble. The numbers represent the first and the last residue in the stretch. Inspired by Jabri *et al.* [42]

the flap residues (α311-α340) also indicate a high degree of flexibility/mobility in native *K. aerogenes* urease [32]. All the urease structures presented in this study are well ordered, in the native structure, for instance, the average temperature factor for the main chain is 17.5 and for the side chains 21.2 Å^2.

Two regions in the α subunit, α315-α336 and α388-α398, have high mobility as indicated by their high temperature factors. Figure 2.12 shows the main chain (B-factor) plots for the native structure taken as representative for all *B. pasteurii* urease structures. The first stretch of disordered residues (α315-α336) is the mobile flap regulating the access to the active site while the second one (α388-α398) is a loop region located right after the last helix of the $(\alpha\beta)_8$ barrel domain but far from the active site, thus not participating in the reac-

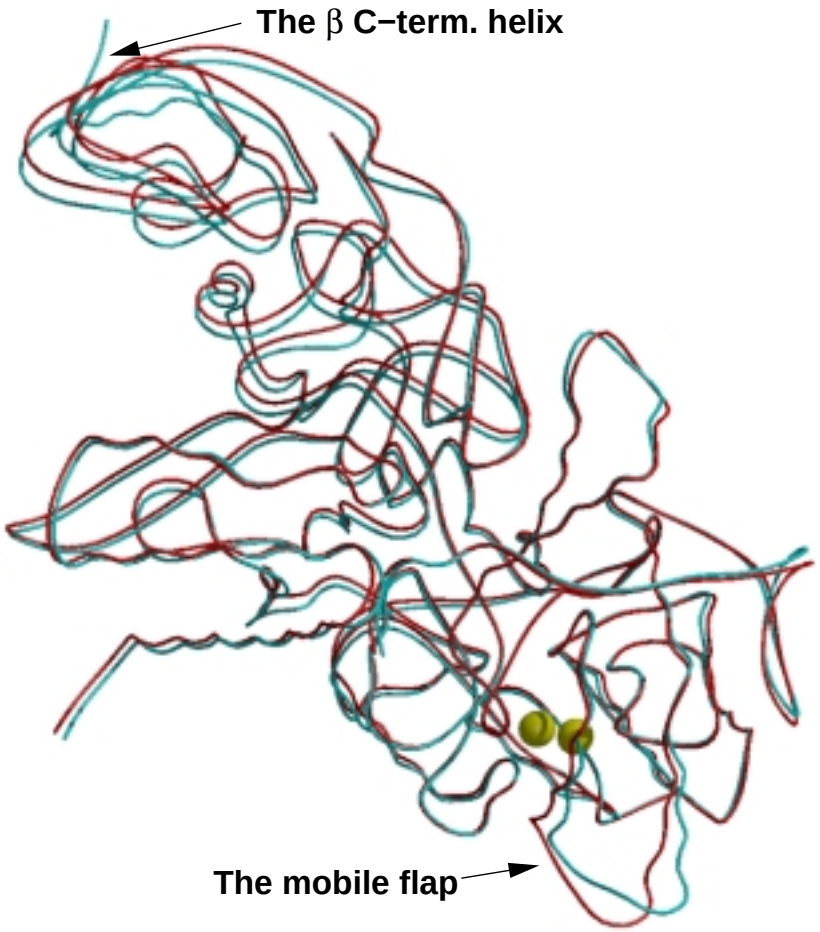


Figure 2.11: Structural superposition of *B. pasteurii* and *K. aerogenes* urease. The figure was made with MOLSCRIPT [83] and Raster3D [81]

tion. This second flexible region might be involved in the interaction between apo-urease and the accessory proteins.

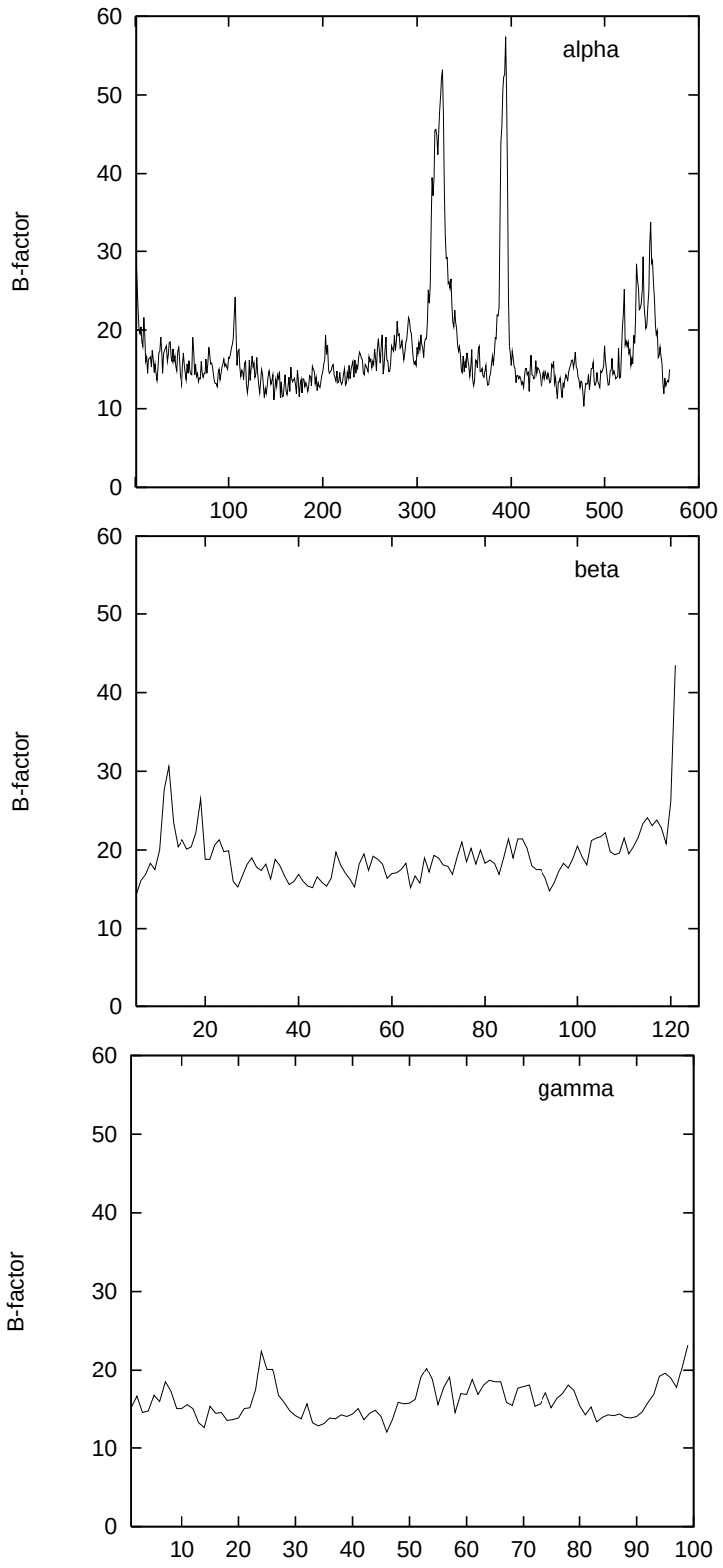


Figure 2.12: Temperature factor plots of main chain atoms in native *B. pasteurii* urease subunits.

2.9 Urease Sequence Alignment

Urease sequences were retrieved from the databases using SRS (Sequence Retrieval System [84]) and the alignment between bacterial enzymes carried out with the program CLUSTALX [85]. The alignment between 22 ureases from different organisms (Figures from 2.13 to 2.18) reveals a very high sequence identity that for the residues making up the active site shows 100% conservation. A structural analysis on *K. aerogenes* based on the conserved residues revealed that those absolutely conserved are mainly involved in trimer building/stabilisation with interaction between residues belonging to the different urease subunits [43]. Hydrogen bonds and hydrophobic interactions confer the great stability of the urease trimer of trimers which is probably the main reason for the conservation. Since structure is usually more conserved than sequence, this very high urease sequence identity indicates a even higher structure similarity. The structures of *B. pasteurii* and *K. aerogenes* urease superimpose very closely and the sequence conservation suggests a completely conserved fold in bacteria and plants with very limited insertion or deletions.

In the Figures from 2.13 to 2.18 is reported the sequence alignment of *B. pasteurii* urease with corresponding residue numbering. The sequence codes correspond to:
ACTPL *Actinobacillus pleuropneumoniae*,
2UBP *Bacillus pasteurii*,
BACSB *Bacillus sp.* (Strain TB-90),
BACSU *Bacillus subtilis*,
BORBR *Bordetella bronchiseptica*,
CLOPE *Clostridium perfringens*,
HAEIN *Haemophilus influenzae*,

HELFE *Helicobacter felis*,

HELHE *Helicobacter heilmannii*,

HELPY *Helicobacter pylori*,

FWJ *Klebsiella aerogenes*,

LACFE *Lactobacillus fermentum*,

MYCTU *Mycobacterium tubercolosis*,

PROMI *Proteus mirabilis*,

PROVU *Proteus vulgaris*,

RHIME *Rhizobium meliloti*,

STAXY *Staphylococcus xylosus*,

STRSL *Streptococcus salivarius*,

SYNY3 *Synechocystis sp.* (Strain PCC 6803),

UREUR *Ureaplasma urealiticum*,

YEREN *Yersinia enterocolitica*,

YERPS *Yersinia pseudotubercolosis*,

B. pasteurii and *K. aerogenes* urease sequences are derived from the structures deposited in the PDB. Residues are colour coded according to CLUSTALX: AVFPMILW are represented in red (small plus hydrophobic, including aromatic but not Y), DE in blue (acidic), RHK in magenta (basic), STYHCNGQ in green (hydroxyl plus amine plus basic but not Q) and others grey. The symbols in the consensus line are:

"*" = identical or conserved residues in all sequences in the alignment

":" = indicates conserved substitutions

"." = indicates semi-conserved substitutions

The line underneath the alignment represents the quality curve, high for conservation and low for substitution.

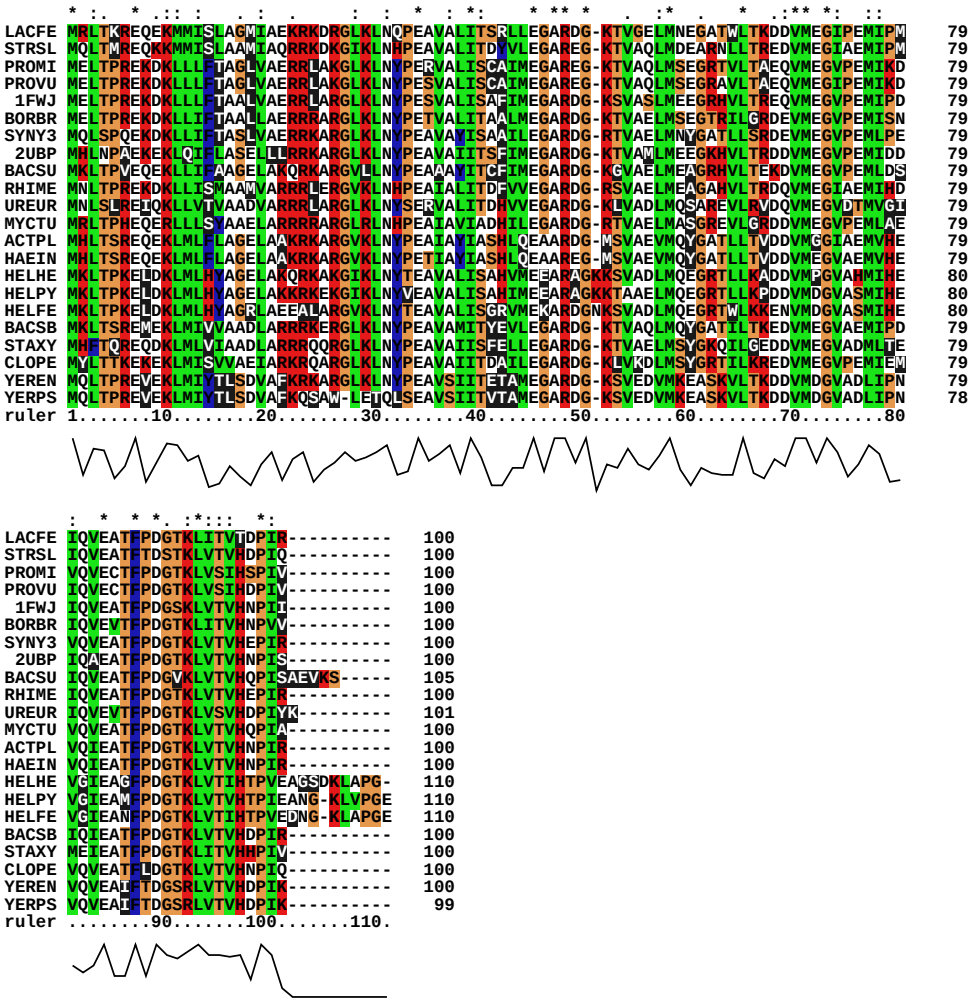


Figure 2.13: Urease sequence alignment (γ subunit).

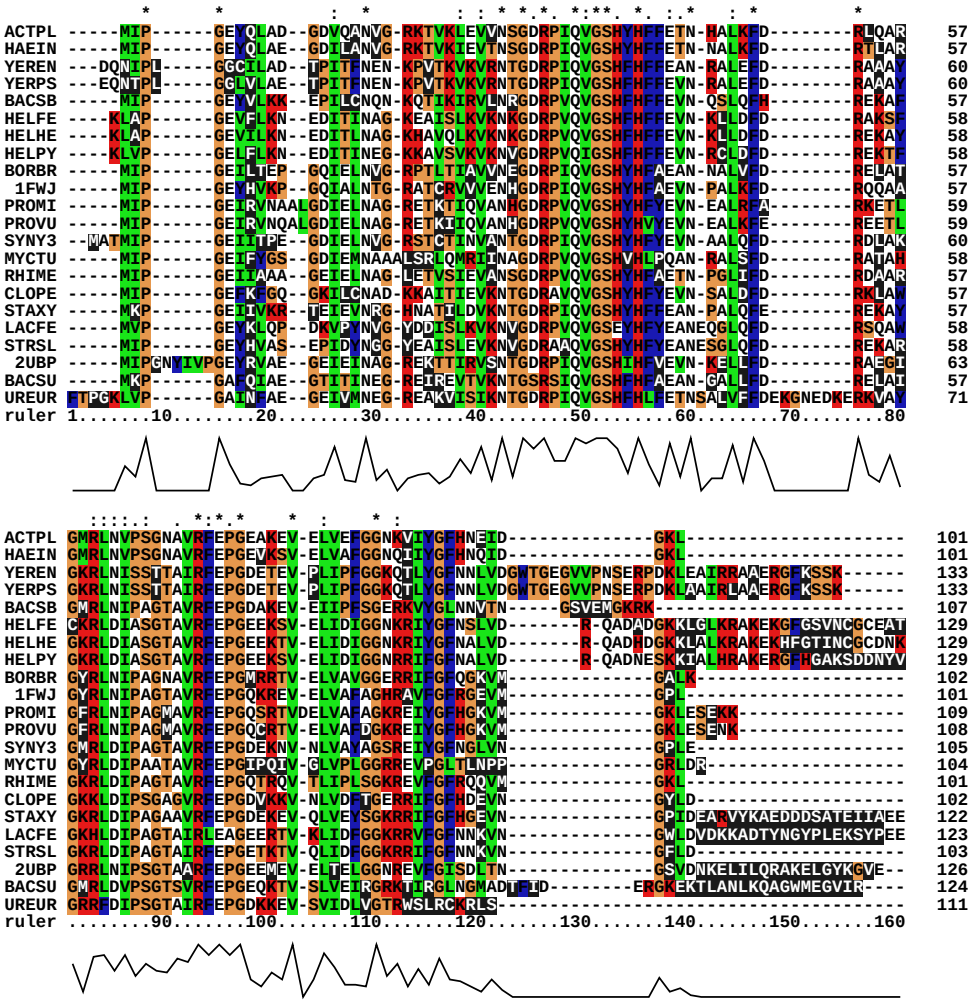


Figure 2.14: Urease sequence alignment (β subunit).

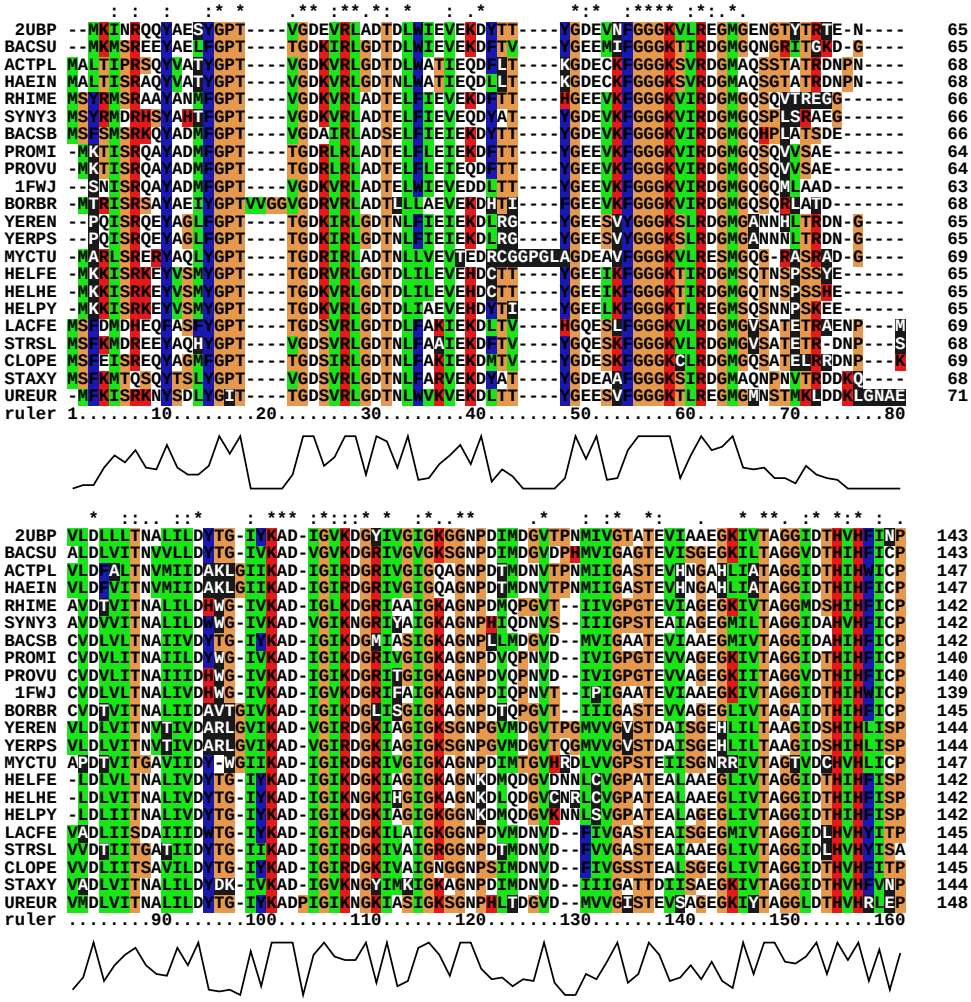


Figure 2.15: Urease sequence alignment (α subunit 1-143).

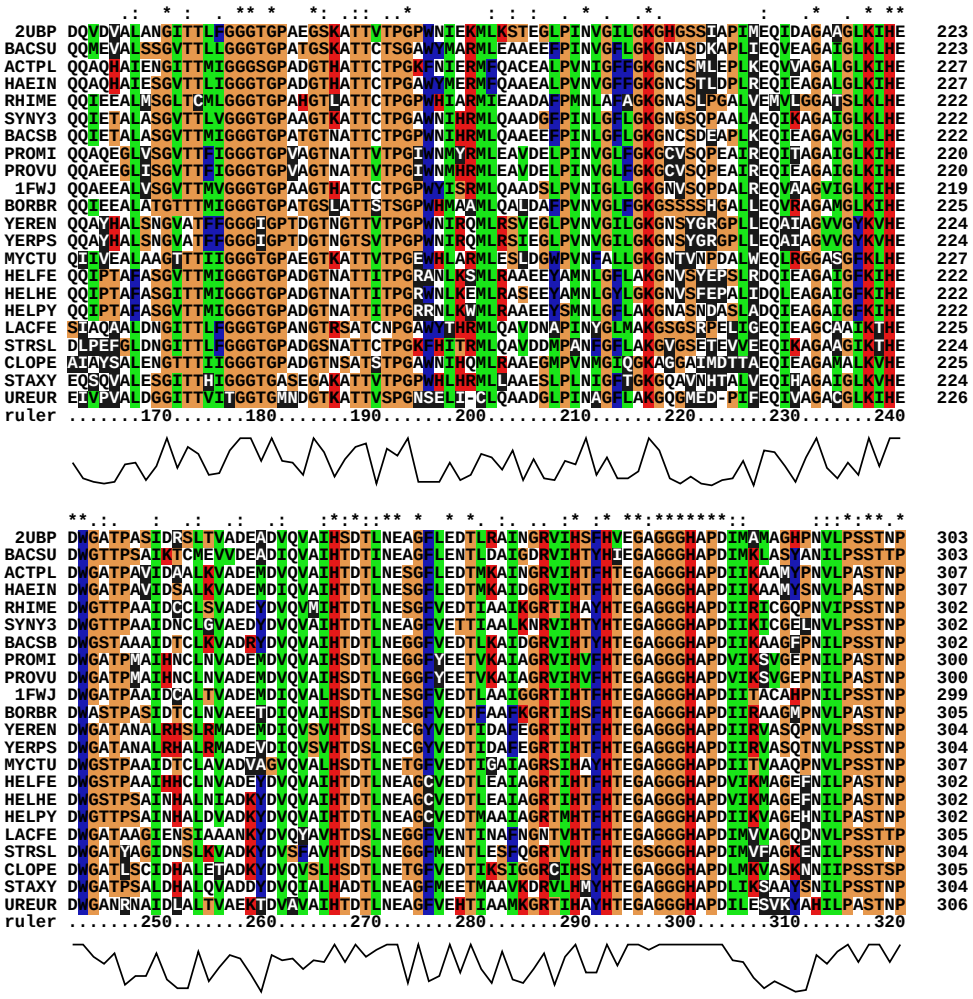


Figure 2.16: Urease sequence alignment (α subunit 144-303).

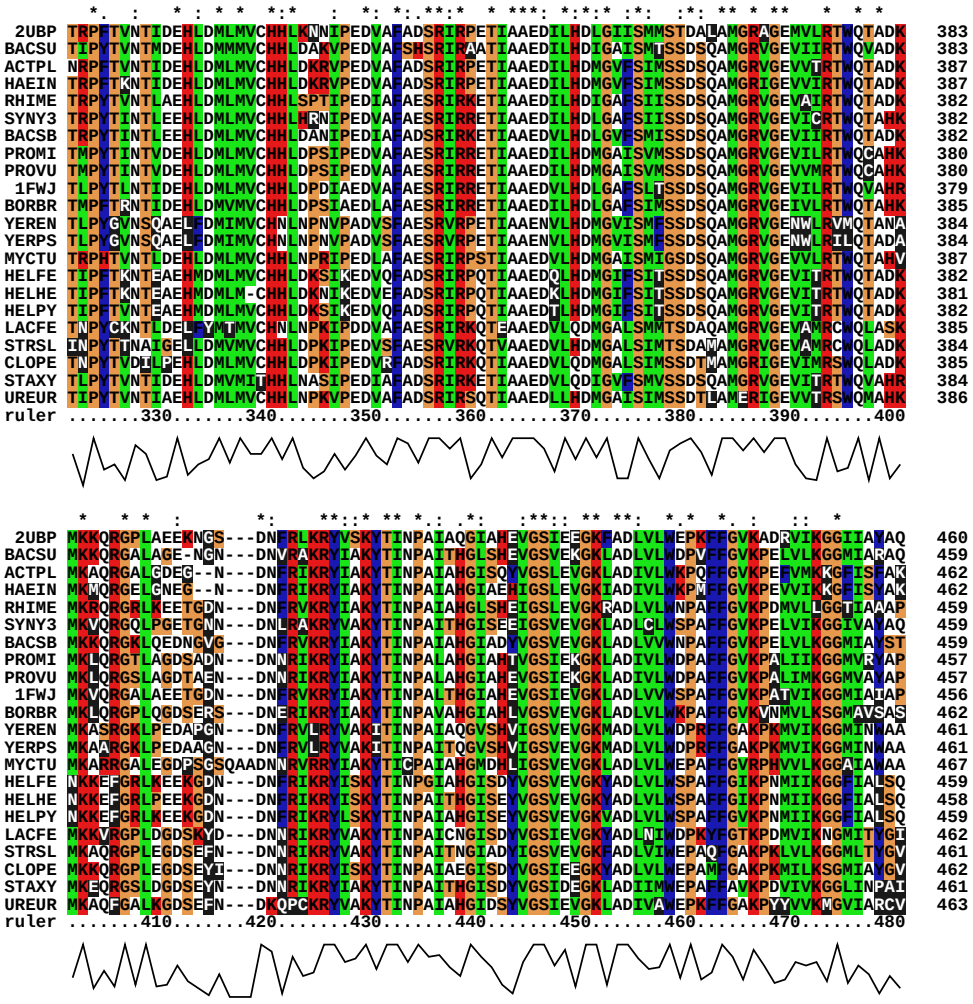


Figure 2.17: Urease sequence alignment (α subunit 304-460).

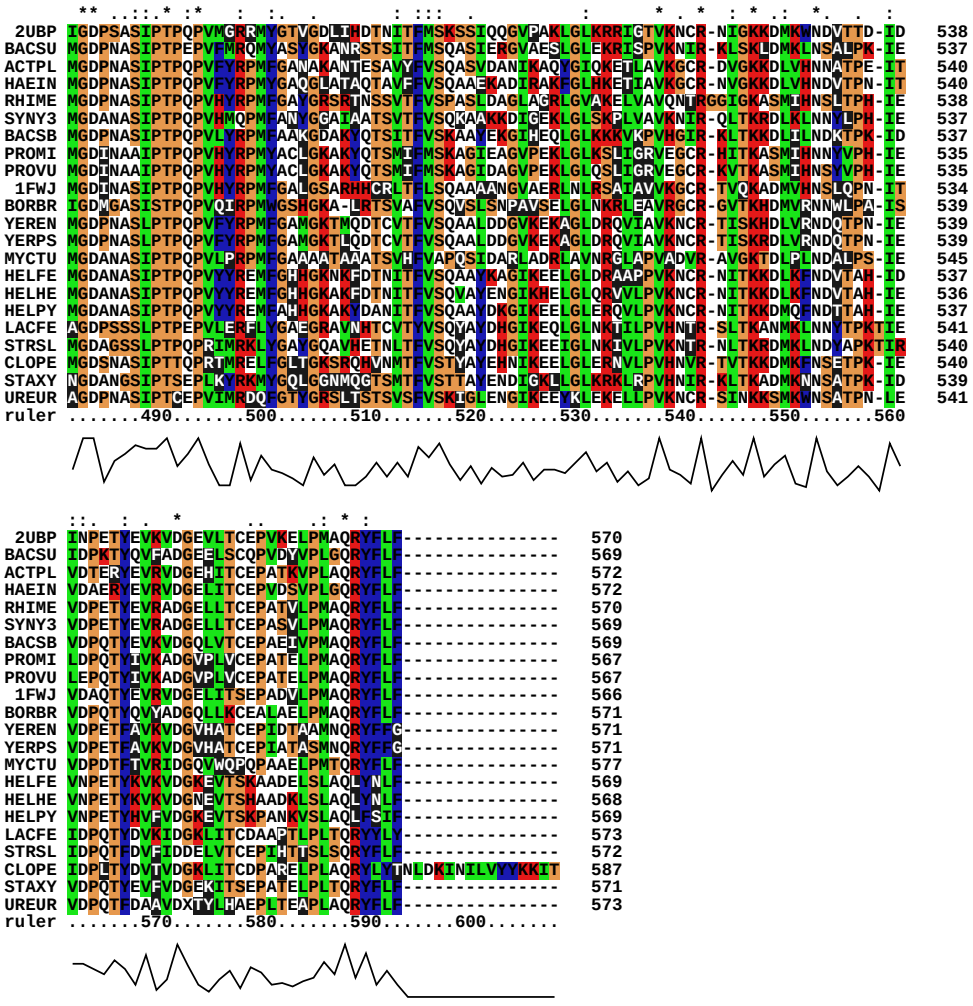


Figure 2.18: Urease sequence alignment (α subunit 461-570).

2.10 Modified Residues

The presence of a carbamylated lysine binding the nickel ions in urease, shown for the first time by Jabri *et al.*, [42] confirmed the discovery that carbon dioxide is required for the *in vitro* formation of an active enzyme from apo-urease [17]. Lys α220* (the * indicating the carbamylation) is located on a β-strand part of the $(\alpha\beta)_8$ barrel domain. The long side chain of Lys α220* allows the binding of the two nickel ions which are too far to be reached by an aspartic acid or a glutamate. The possibility for the carbamate to make resonance isomers that can generate a negative charge on its terminal oxygens, higher than the negative charge of a carboxyl group, could modulate the chemistry not only of nickel binding but also, indirectly, of urea hydrolysis.

A second modification found in *B. pasteurii* urease is the carbamylation of the N-terminus of the γ subunits. The N-terminus carbamylation seems to have a structural role, stabilising the trimer by interactions between the γ subunit modified N-terminus and residues from the γ' and the α' subunits (Fig. 2.19). The OT1 of the carbamylated Met γ1 is hydrogen bonded to His γ2 Nδ1 (at 2.44 Å) and Tyr γ'32 Oη (at 2.65 Å) while OT2 is hydrogen bonded to Gln α'472 Nϵ2 (at 2.85 Å) and with wat'-636 (at 2.87 Å). Gln α'472 Oϵ1 is bonded to Met γ1 N, further stabilising the carbamate. The angles between hydrogen bonds are close to 105° (as for the H–O–H angle in water). The angle is 102.7° between His γ2 Nδ1–Met γ OT1–Tyr γ'32 Oη and 100.7° between Gln α'472 Nϵ2–Met γ OT2–wat'-636.

Carbamate is a rather unstable group whose stabilisation can be achieved either by coordination to metals or by hydrogen bond networks. In haemoglobin, carbon dioxide is transported from the muscles back to the lungs by formation

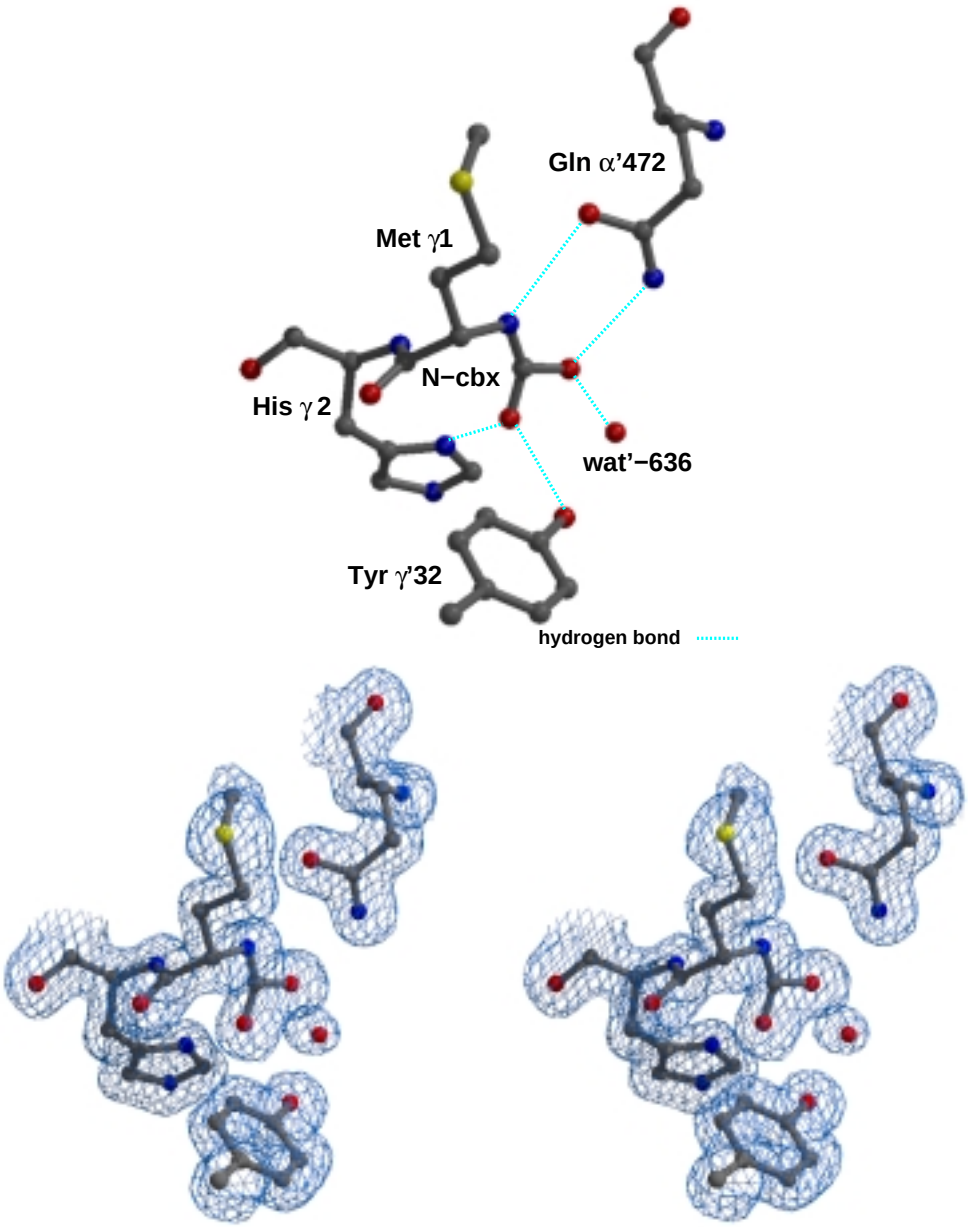


Figure 2.19: Top: model of the carbamylated γ subunit N-terminus of *B. pasteurii* urease, bottom: stereo picture of the $2mF_{obs} - DF_{calc}$ electron density map contoured at 1 σ. The figure was made with BOBSCRIPT [80] and Raster3D [81].

of carbamate at the N-termini, because the un-ionised form of haemoglobin's α-amino groups can react reversibly with CO_2. The bound carbamates form salt bridges that stabilise the "T" form lowering haemoglobin's affinity for oxygen [86].

A carbamylated N-terminus has been found in *E. coli* thymidylate synthase (pdb code 1TYS) [87]. In this protein the carbamate is protected from the bulk solvent by a network of hydrogen bonds involving the main chain nitrogens of two threonines, their Oγ1 and a water molecule. Mutation of one or both threonines to valine leads to loss of enzymatic activity and of the carbamate group, suggesting the importance of this modification in generating a functional active site [87].

Chapter 3

The Active Site

3.1 Native Active Site

The electron density maps ($2mF_{obs} - DF_{calc}$ and the omit $mF_{obs} - DF_{calc}$) are easily interpretable providing information not only about the protein ligands but also about the solvent structure (Fig. 3.1). The two Ni^{2+} ions are located deep in the active site cavity at a distance of 3.7 Å from one another. Lys α220* bridges the two nickels *via* its carboxylate group by coordinating Ni-1 through Oθ1 and Ni-2 through Oθ2. Ni-1 is coordinated to His α249 Nδ1 and His α275 Nϵ2, Ni-2 to His α137 Nϵ2, His α139 Nϵ2 and Asp α363 Oδ1 (Fig. 3.2). Completing the coordination sphere of the nickels is a tight cluster of four waters clearly shown in Figure 3.1 as a tetrahedral cloud of electron density.

Several attemps to interpret this density involved the fitting of molecules such as sulphate (present in the precipitant), phosphate (present in the purification buffers) and lithium tetrahydrate. A very strong negative peak resulted in the $mF_{obs} - DF_{calc}$ electron density map when refining with sulphate or phosphate

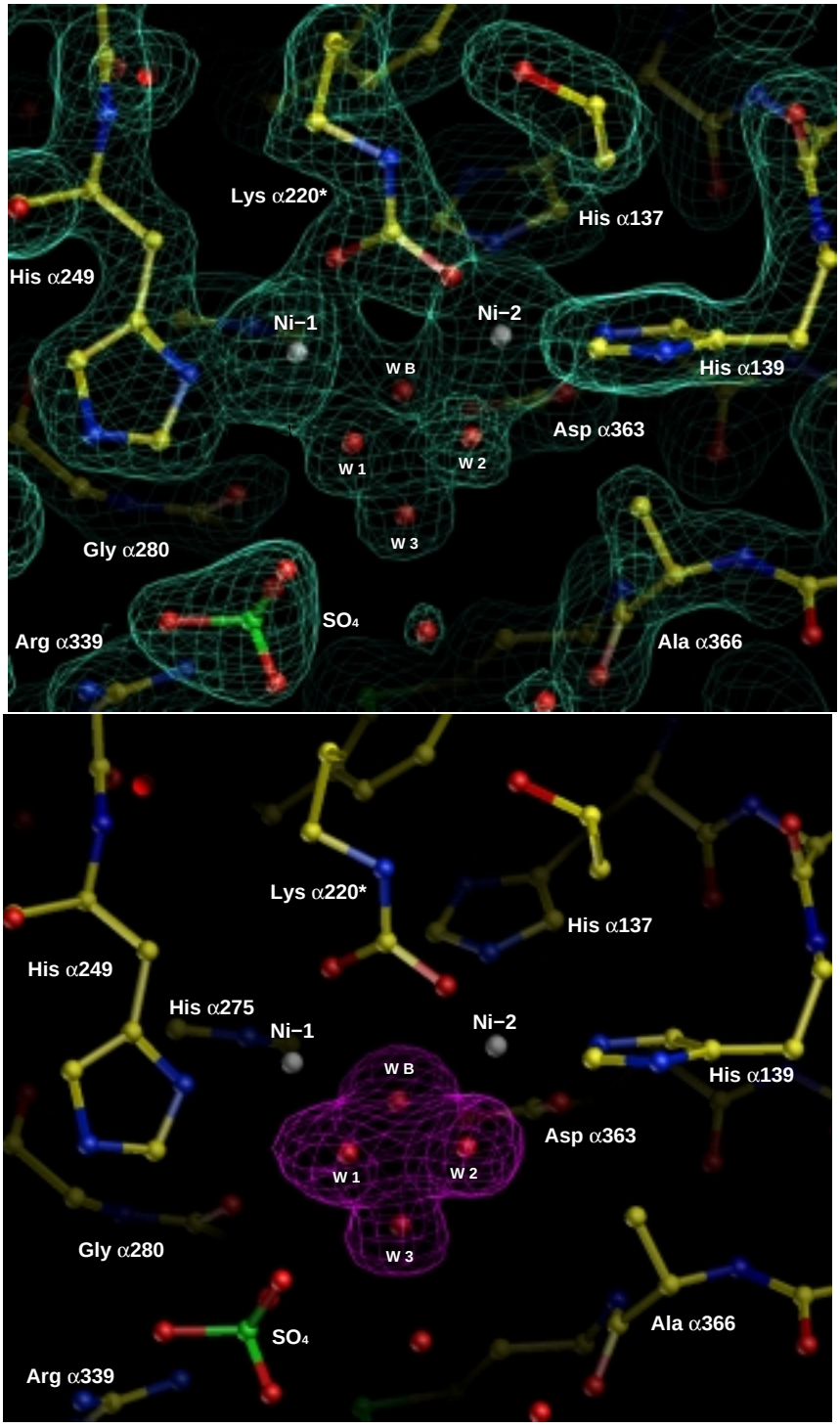


Figure 3.1: The model of native *Bacillus pasteurii* urease superimposed on the $2mF_{obs} - DF_{calc}$ contoured at 1 σ (top) and omit $mF_{obs} - DF_{calc}$ contoured at 3 σ electron density maps (bottom). The picture was made using "O"[71].

in the position corresponding to the sulphur or phosphor atom. When using the tetra-aquo lithium the negative density disappeared but the resulting Li-O distances had to be restrained to very short values (1.5 Å) compared to those found in small-molecule structures for the same compound (1.85-2.05 Å).

Three of these waters are directly bound to the Ni ions either in a monodentate (wat-1 to Ni-1, wat-2 to Ni-2) or a bridging mode (wat-B) while the fourth one (wat-3) is hydrogen bonded to the first three (wat-1 $\longleftrightarrow$ wat-3 = 2.2 Å, wat-2 $\longleftrightarrow$ wat-3 = 2.4 Å, wat-B $\longleftrightarrow$ wat-3 = 2.3 Å).

The coordination geometry of the two nickels is distorted square-pyramidal for Ni-1 and distorted octahedral for Ni-2. The apical ligand for the square-pyramidal Ni-1 is Oθ1 of Lys α220* while the basal plane is made by His α249 Nδ1, His α275 Nϵ2, wat-1 and wat-B.

The protonation state of the water cluster in the active site is crucial for the understanding of the hydrolysis of urea which requires the presence of an hydroxide ion acting as a nucleophile during the reaction [88]. It is known from model compounds, that the pK_a of a water bound to a hexacoordinated $[Ni(H_2O)_6]^{2+}$ ion is 10.6 [89]. At the optimal pH for urease activity (8.0), wat-1 and wat-2 are waters. At pH 8.0 wat-B is proposed to be in the hydroxo form because in water bridged bimetallic complexes the first pK_a assumes acidic values while the pK_a for a bridging hydroxide is $\sim$9-10, slightly lower than the first pK_a of a single bound water. The protonation scheme for the water cluster is shown in Figure 3.3. Wat-1 is at 2.9 Å from His α222 Nϵ2 which is protonated, as deduced from the interaction between His α222 Nδ1 (hydrogen acceptor) and the main chain nitrogen of Asp α224 (hydrogen donor). Wat-2

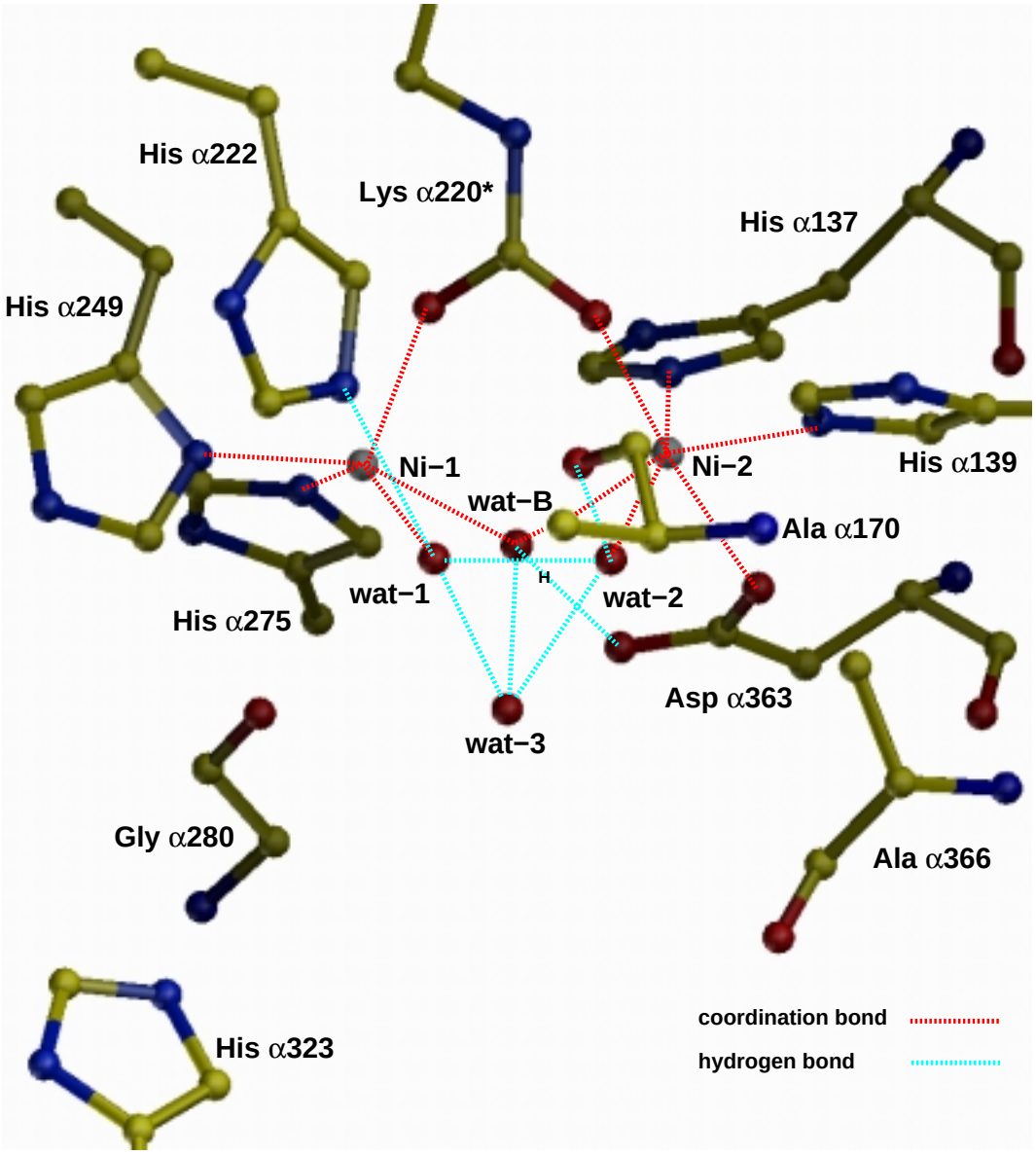


Figure 3.2: Model of the native *B. pasteurii* urease active site. The picture was made using "O"[71]. Only the H on the catalytic OH^- is shown.

donates a hydrogen to Ala α170 O located 2.9 Å apart.

The formation of a four-centered hydrogen bonding network provides the ex-

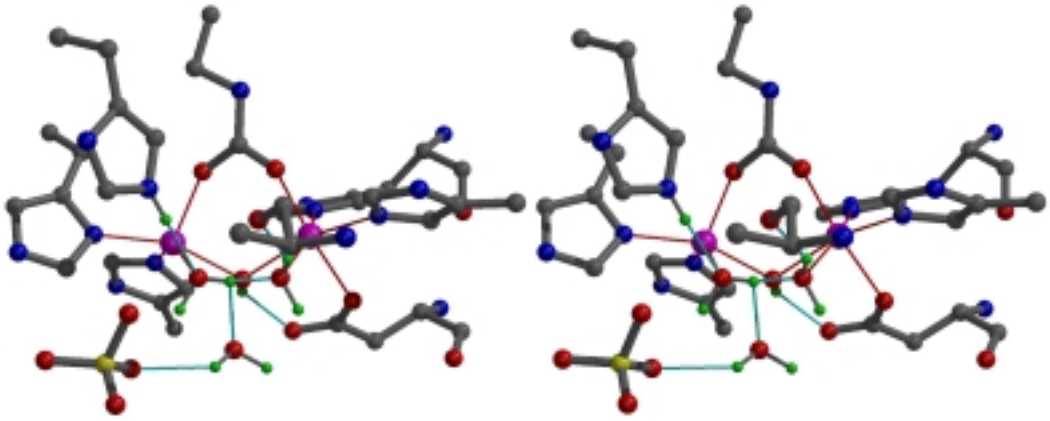


Figure 3.3: Stereoview of the hydrogen bonding network in the native *B. pasteurii* urease active site.

planation for the very short distances between the waters (wat-1 $\longleftrightarrow$ wat-2 = 2.1 Å, wat-1 $\longleftrightarrow$ wat-B = 2.1 Å, wat-2 $\longleftrightarrow$ wat-B = 2.2 Å) in the tetrahedral cluster. The central hydrogen is provided by wat-1 which points its lone pairs towards the protonated His α222 Nϵ2. The orientation of the water molecules and their hydrogens is imposed by the interactions with the two nickels for wat-1, wat-2 and wat-B, with Asp α363 Oδ2 for wat-B, and by the hydrogen bond partners His α222 Nϵ2 for wat-1 and Ala α170 O for wat-2.

The orientation of wat-3 is determined by its lone pairs pointing towards the hydrogen located in the centre of the tetrahedron and by the hydrogen bond with a sulphate ion (wat-3 $\longleftrightarrow$ O2 = 3.0 Å) present in the active site because of the high concentration of AMS in the precipitant (52-56% (saturated AMS + 1.2 M LiCl) in 20 MM citrate pH 6.3). The sulphate ion is stabilised in its position by the formation of a salt bridge with Arg α339 and of a hydrogen bond with His α323 (Arg α339 Nη $\longleftrightarrow$ O1 = 2.85 Å and His α323 Nϵ2 $\longleftrightarrow$ O4 =2.73 Å). No sulphate molecules are found in *K. aerogenes* urease despite

its presence in the precipitant (crystallisation conditions 1.8 M Li_2SO_4 in 50 mM HEPES pH 7.5).

In *K. aerogenes* urease the cluster of waters was described but wat-1 and wat-B were considered partially occupied because the distances between them were considered too short to be simultaneously present (wat-1 $\longleftrightarrow$ wat-2 = 2.3 Å, wat-1 $\longleftrightarrow$ wat-B = 2.5 Å, wat-2 $\longleftrightarrow$ wat-B = 2.0 Å, wat-1 $\longleftrightarrow$ wat-3 = 2.6 Å, wat-2 $\longleftrightarrow$ wat-3 = 2.4 Å) [32].

The mobile flap covering the catalytic cavity is in the open conformation as opposed to the native *K. aerogenes* urease whose flap is closed. The flap in native *B. pasteurii* urease is stabilised in the open conformation by the sulphate ion bridging between Arg α339 located at the hinge and His α323 located on the flap itself. The other *B. pasteurii* urease structure do not contain a bound sulphate ion. Table 3.1 shows the coordination distances for native *B. pasteurii* and *K. aerogenes* urease for comparison.

B. pasteurii urease			
Nickel	⟷	Ligand	Distances Å
Ni-1	⟷	Lys α220* Oθ1	2.1
Ni-1	⟷	His α249 Nδ1	2.2
Ni-1	⟷	His α275 Nϵ2	2.2
Ni-1	⟷	wat-B	2.1
Ni-1	⟷	wat-1	2.2
Ni-2	⟷	Lys α220* Oθ2	2.1
Ni-2	⟷	His α137 Nϵ2	2.2
Ni-2	⟷	His α139 Nϵ2	2.2
Ni-2	⟷	Asp α363 Oδ1	2.2
Ni-2	⟷	wat-B	2.2
Ni-2	⟷	wat-2	2.1
K. aerogenes			
Nickel	⟷	Ligand	Distances Å
Ni-1	⟷	Lys α217* Oθ1	2.1
Ni-1	⟷	His α246 Nδ1	2.0
Ni-1	⟷	His α272 Nϵ2	2.3
Ni-1	⟷	wat-B	2.1
Ni-1	⟷	wat-1	2.0
Ni-2	⟷	Lys α217* Oθ2	2.1
Ni-2	⟷	His α134 Nϵ2	2.2
Ni-2	⟷	His α136 Nϵ2	2.1
Ni-2	⟷	Asp α360 Oδ1	2.2
Ni-2	⟷	wat-B	2.0
Ni-2	⟷	wat-2	2.1

Table 3.1: Nickel coordination distances in native *B. pasteurii* and native *K. aerogenes* urease active sites

3.2 β-mercaptoethanol Inhibited Active Site

In β-mercaptoethanol (BME) inhibited urease the cluster of waters seen in the native active site is displaced by the inhibitor (Fig. 3.4 and 3.6). BME bridges with its sulphur atom between the two nickel ions and further coordinates to Ni-1 with its oxygen. The binding of BME to the active site in a

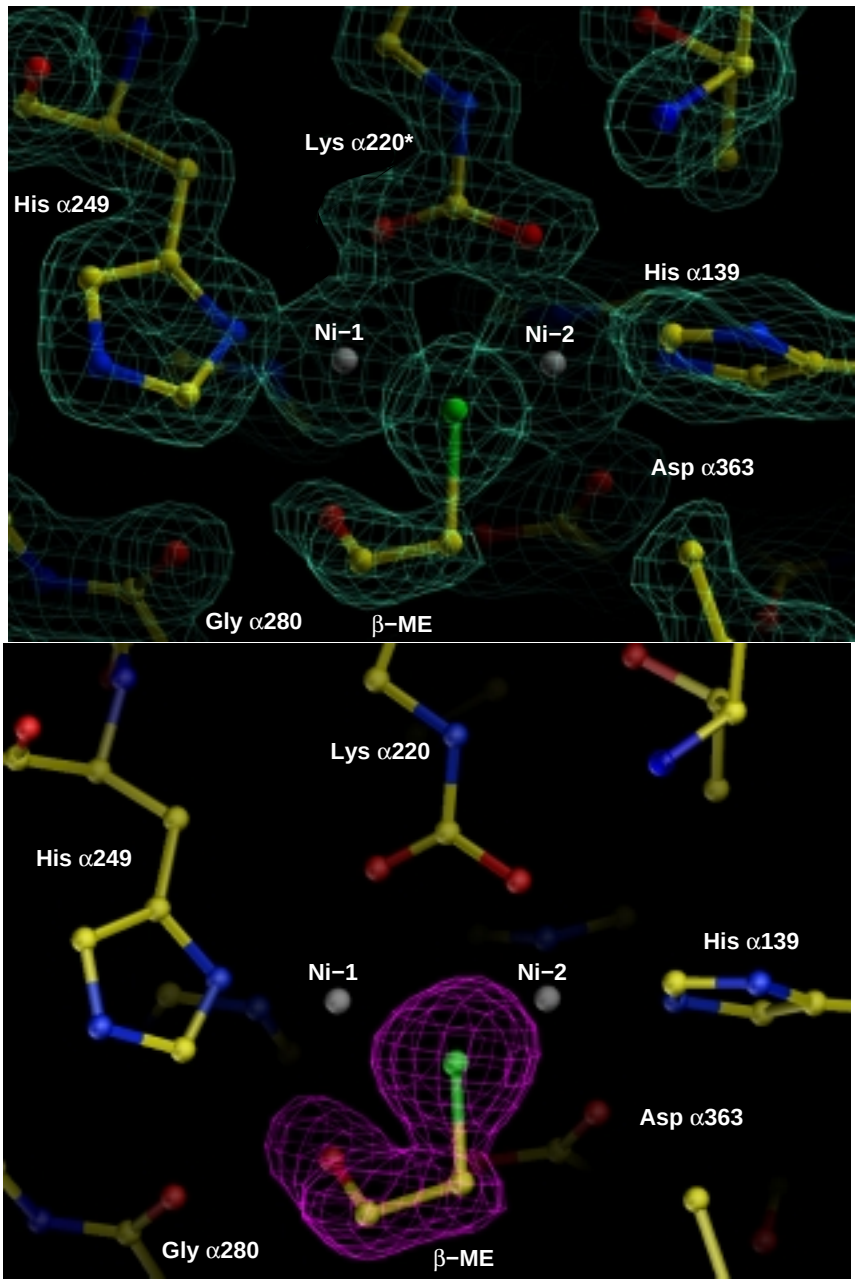


Figure 3.4: The model of BME inhibited *Bacillus pasteurii* urease superimposed on the $2mF_{obs} - DF_{calc}$ contoured at 1 σ (top) and $mF_{obs} - DF_{calc}$ contoured at 3 σ electron density maps (bottom). The picture was made using "O"[71].

bridging/chelating mode was unexpected because of the low Lewis base capacity of the BME alcoholic group. The basicity of the BME OH group is enhanced by the interaction with Gly α280 O through hydrogen bond donation. Both nickel ions are pentacoordinated with Ni-1 square-pyramidal and Ni-2 distorted trigonal-bipyramidal. The basal plane for the square-pyramidal Ni-1 is formed by the S and O atoms of the BME, by the Oθ1 of Lys α220* and the His α275 Nϵ2. The apical ligand is His α249 Nδ1. The equatorial plane of the distorted trigonal-bipyramidal Ni-2 is composed of BME S atom, His α137 Nϵ2 and His α139 Nϵ2. The apical ligands are Lys α220* Oθ2 on one side and Asp α363 Oδ1 at the opposite side. The dihedral angle betweeen the two coordination equatorial planes is 63°.

Compared to the native enzyme the distance between Ni-1 and Ni-2 is shortened from 3.5 to 3.1 Å in agreement with EXAFS studies on *K. aerogenes* urease [44]. The shortened Ni-Ni distance found in the BME inhibited enzyme confirms the ability of the active site to rearrange upon substrate/inhibitor binding. Three different interactions could be the possible causes for the shortening: the sulphur bridging between the two metals, the chelating mode of binding of BME, and the small dihedral angle between the two equatorial planes of the coordination geometry of the two nickel ions.

A second molecule of BME in the active site participates in urease inhibition by making a mixed disulphide bond with Cys α322 Sγ. Cys α322 is located on the flap covering the active site cavity here found in the open conformation. The involvement of the Cys α322 Sγ bound BME molecule with its hydroxyl group in a hydrogen bond with Ala α366 O reduces the flexibility of the flap (Fig. 3.5 and 3.6). Ala α366 O is involved in urea orientation and activation

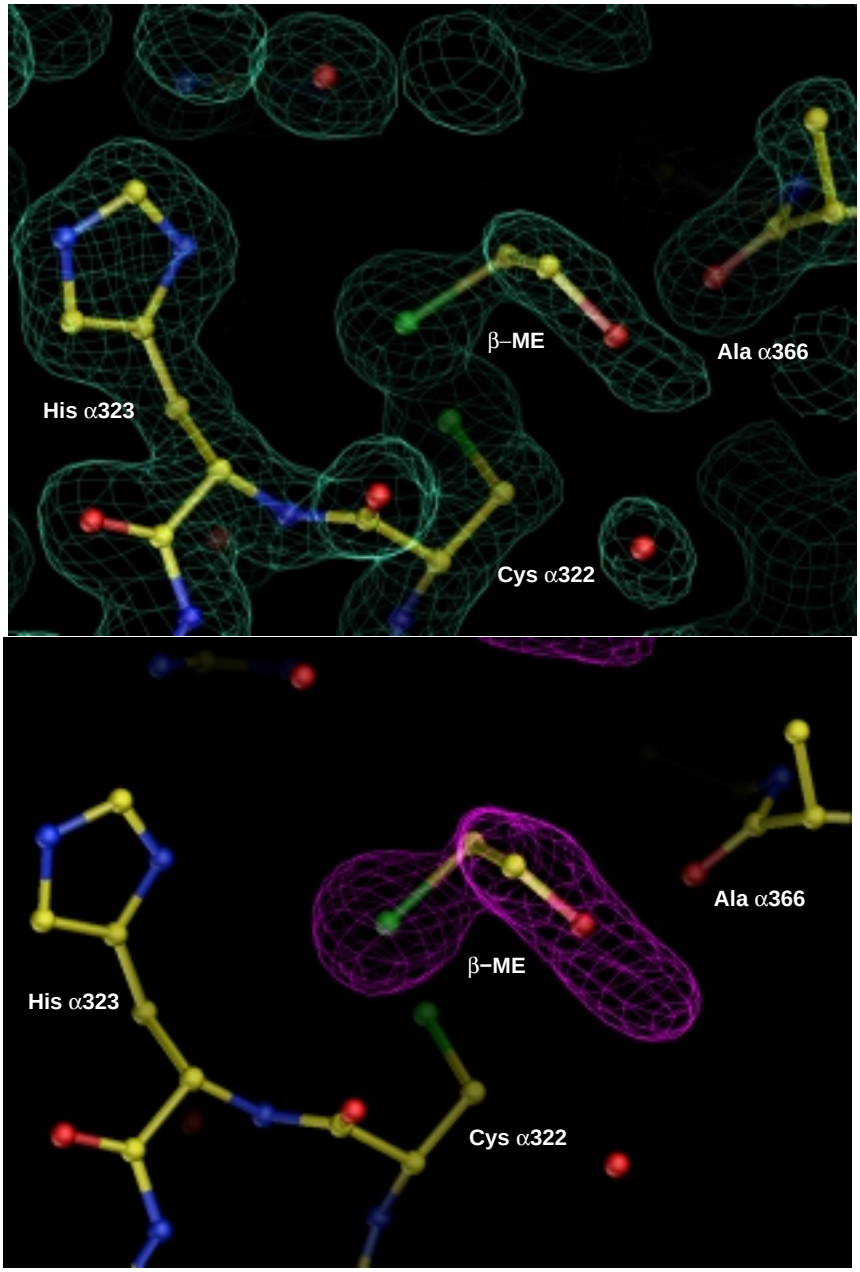


Figure 3.5: The model of BME inhibited *Bacillus pasteurii* urease superimposed on the $2mF_{obs} - DF_{calc}$ contoured at 1 σ (top) and $mF_{obs} - DF_{calc}$ contoured at 3 σ electron density maps (bottom) in the region around Cys α322. The picture was made using "O"[71].

during the reaction. Its engagment with the BME OH group prevents urea

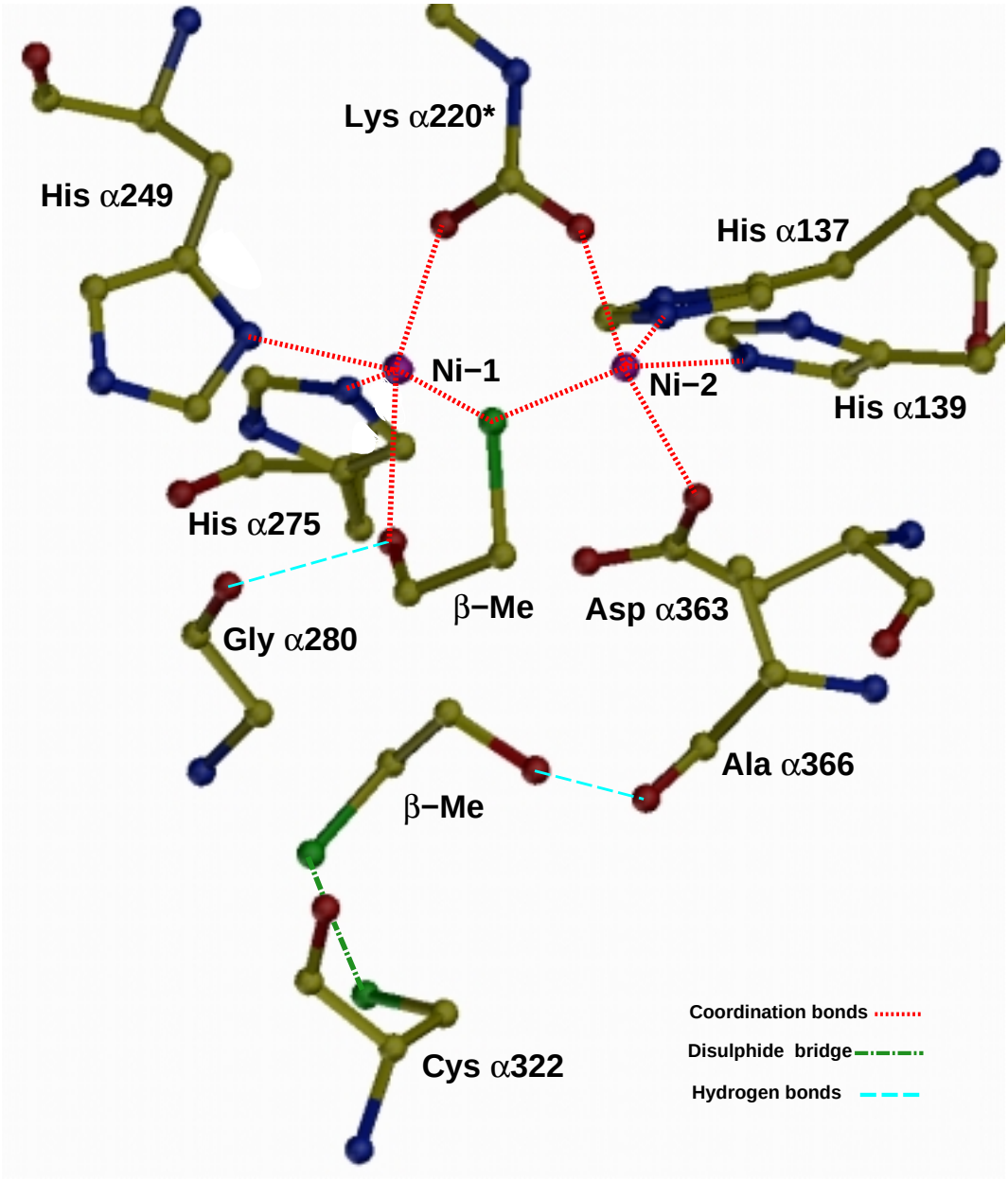


Figure 3.6: Model of the BME inhibited *B. pasteurii* urease active site. The picture was made using "O"[71].

hydrolysis from occuring. Inhibition *via* BME involves two mechanisms one with direct binding of the inhibitor to the metal centre and a second one with binding/bridging to Cys α322 Sγ and Ala α366 O. Table 3.2 lists the coordination distances for the nickel ions in BME inhibited urease.

Nickel	$\longleftrightarrow$	Ligand	Distances Å
Ni-1	$\longleftrightarrow$	Lys α220* Oθ1	2.1
Ni-1	$\longleftrightarrow$	His α249 Nδ1	2.2
Ni-1	$\longleftrightarrow$	His α275 Nϵ2	2.2
Ni-1	$\longleftrightarrow$	BME-S	2.3
Ni-1	$\longleftrightarrow$	BME-O	2.2
Ni-2	$\longleftrightarrow$	Lys α220* Oθ2	2.1
Ni-2	$\longleftrightarrow$	His α137 Nϵ2	2.1
Ni-2	$\longleftrightarrow$	His α139 Nϵ2	2.1
Ni-2	$\longleftrightarrow$	Asp α363 Oδ1	2.1
Ni-2	$\longleftrightarrow$	BME-S	2.3

Table 3.2: Nickel coordination distances in the BME inhibited *B. pasteurii* urease active site

3.3 Diamidophosphate Inhibited Active Site

EXAFS studies on a phenylphosphorodiamidate (PPD)*B. pasteurii* urease complex confirmed binding of the inhibitor to the nickel centre, but did not provide details about the orientation and the nature of the molecule bound to the enzyme [60]. Kinetic studies suggested that in the presence of phosphoroamidates (e.g., PPD) the actual inhibitor is the tetrahedral diamidophosphate DAP [34] formed by enzymatic hydrolysis. It is worth recalling the formation during urea hydrolysis of a tetrahedral reaction intermediate immediately dissociated into carbamic acid and ammonia.

The crystal structure of the complex between urease and DAP shows for the first time a tetrahedral intermediate analogue bound between Ni-1 and Ni-2 (Figures 3.7 and 3.8). The distance between the two nickel ions is 3.8 Å. Ni-1 is pentacoordinated and Ni-2 hexacoordinated as in the native enzyme. Upon binding DAP replaces the cluster of waters/hydroxides in the active site, almost perfectly matching their position, and establishes a network of hydrogen bonds oriented with high specificitiy. The presence of hydrogen bond donors on one side and hydrogen bond acceptors on the other clearly indicates the orientation of the DAP molecule and suggests the way urea binds and is activated. One DAP oxygen atom (O-1) binds to Ni-1 and accepts a hydrogen from His α222 Nϵ2 (at 2.6 Å) which is protonated, as deduced from the interaction between His α222 Nδ1 (hydrogen acceptor) and the main chain nitrogen of Asp α224 (hydrogen donor) as in the native enzyme. The second DAP oxygen (O-B) bridges symmetrically between Ni-1 and Ni-2 and is at a distance of 2.5 Å from Asp α363 Oδ2 establishing a hydrogen bond with it.

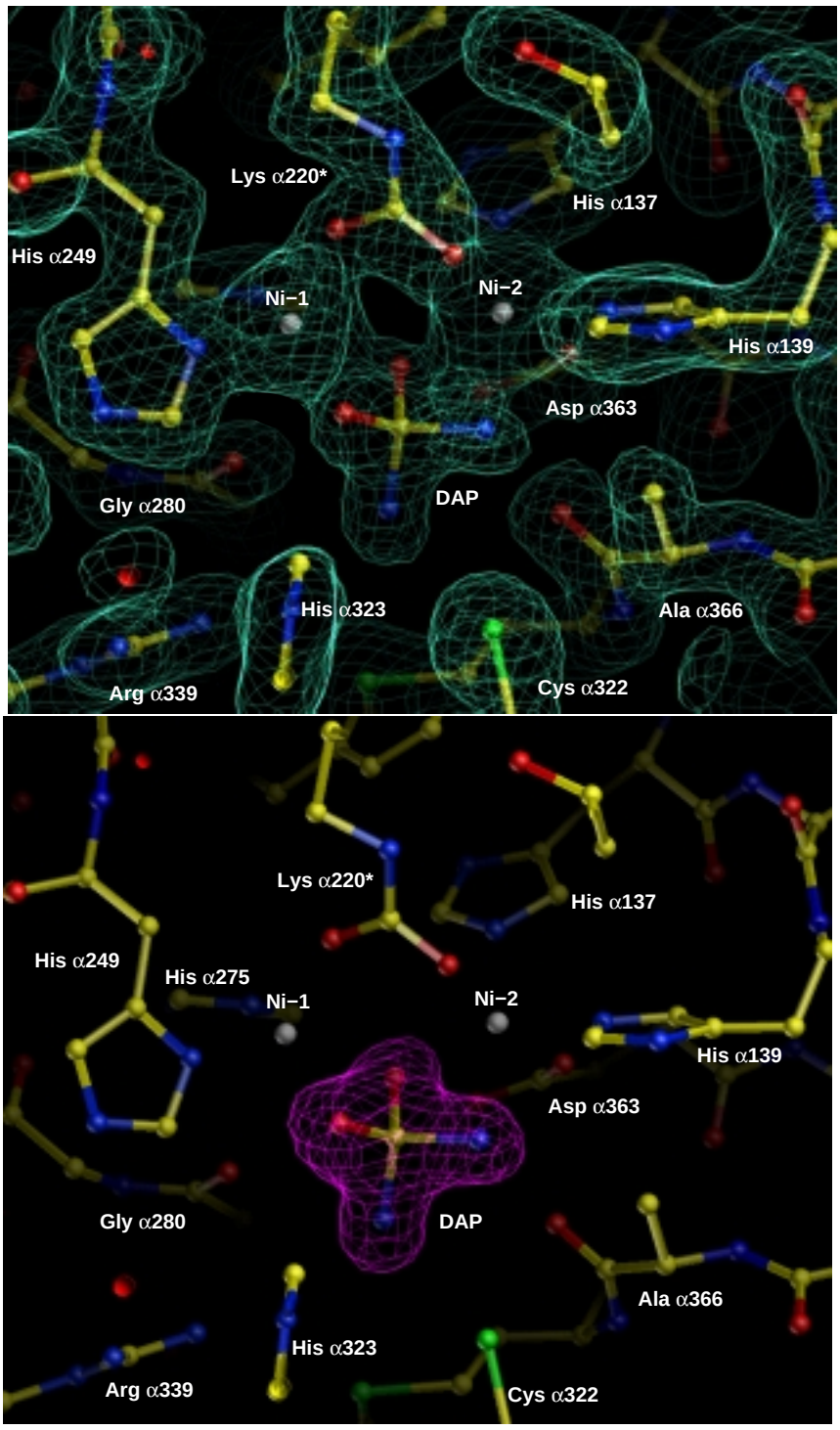


Figure 3.7: The model of DAP inhibited *Bacillus pasteurii* urease superimposed on the $2mF_{obs} - DF_{calc}$ contoured at 1 σ (top) and $mF_{obs} - DF_{calc}$ contoured at 3 σ electron density maps (bottom). The picture was made using "O"[71].

This interaction suggests that the inhibitor is in its neutral form, that is diamidophosphoric acid $(NH_2)_2PO(OH)$, and donates a hydrogen bond to the Asp α363 Oδ2. One DAP nitrogen (N-2) coordinates Ni-2 and donates the hydrogens for two hydrogen bonds: one with Ala α170 O (at 2.9 Å) and one with Ala α366 O (at 2.8 Å). The second DAP nitrogen (N-D) is distally located with regard to the two nickels and establishes a bifurcated hydrogen bond with Ala α366 O (at 3.1 Å) and with Gly α280 O (at 3.1 Å) while His α323 Nϵ2 is located at 3.3 Å. N-D is also positioned within contact distance of Asp α363 Oδ2 (at 3.5 Å).

The biggest conformational change upon DAP binding is the movement of the flap from the open to the closed conformation (see Figure 4.1) This change involves some rearrangement of the flap residues. The catalytic His α323 for instance moves by about 5 Å compared to its position in native urease, to interact with DAP N-D. The new position of His α323 supports direct binding to the tetrahedral intermediate during urea hydrolysis. The terminal guanidinium group of Arg α339, which is located on the hinge of the flap, moves by about 1 Å toward the active site in the DAP-urease complex compared to its position in the native enzyme. The flip of Ala α366 O toward the active site, establishing the hydrogen bond with the DAP N-2 confirms its role in orienting the substrate/inhibitor molecule (for comparison see Figures 3.2 and 3.8).

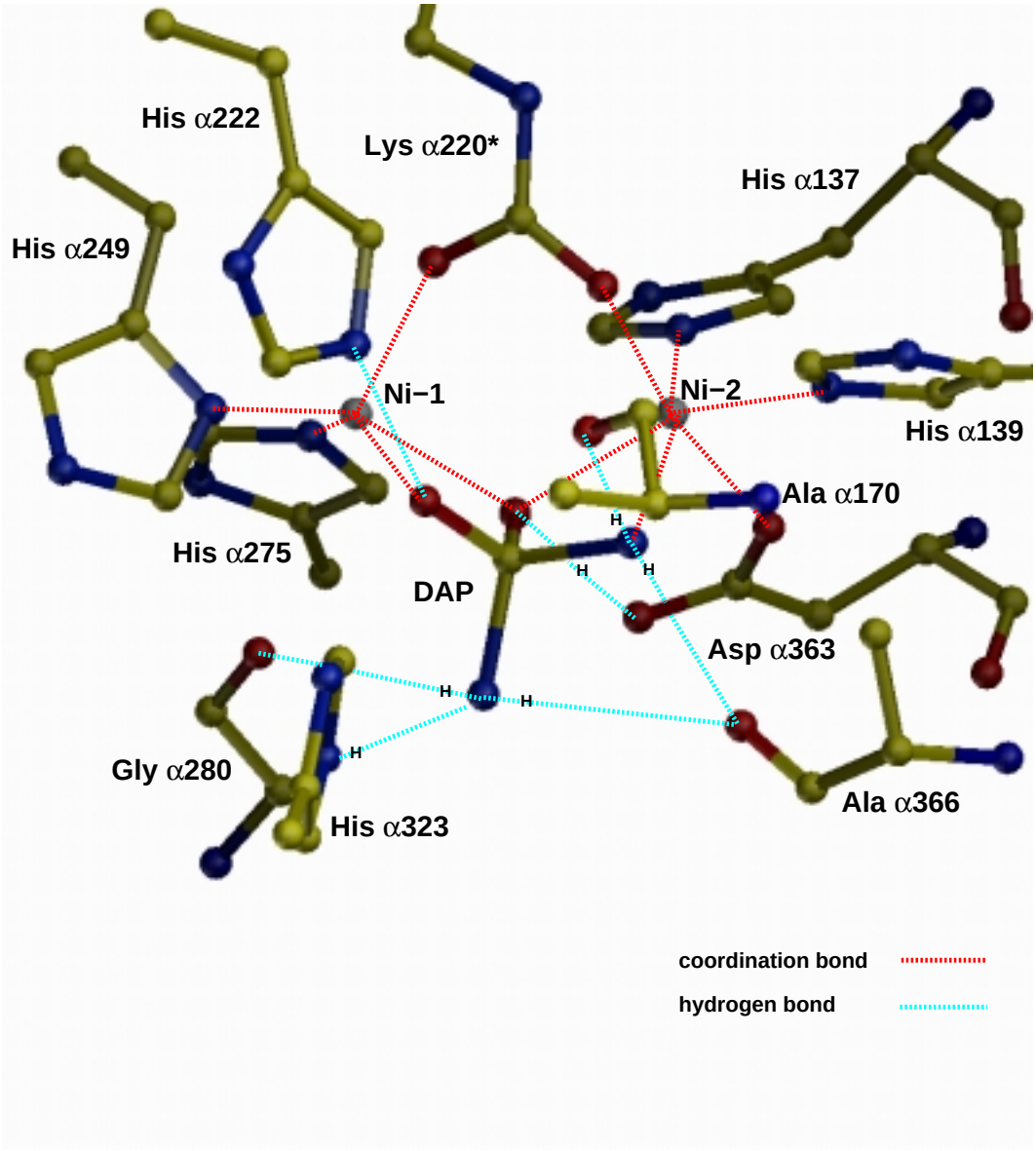


Figure 3.8: Model of the DAP inhibited *B. pasteurii* urease active site. Only those hydrogen atoms essential for catalysis are shown. The picture was made using "O"[71].

Table 3.3 lists the coordination distances for the nickel ions in DAP inhibited urease.

Nickel	$\longleftrightarrow$	Ligand	Distances Å
Ni-1	$\longleftrightarrow$	Lys α220* Oθ1	2.1
Ni-1	$\longleftrightarrow$	His α249 Nδ1	2.0
Ni-1	$\longleftrightarrow$	His α275 Nϵ2	2.1
Ni-1	$\longleftrightarrow$	O-B	2.3
Ni-1	$\longleftrightarrow$	O-1	2.2
Ni-2	$\longleftrightarrow$	Lys α220* Oθ2	1.9
Ni-2	$\longleftrightarrow$	His α137 Nϵ2	2.1
Ni-2	$\longleftrightarrow$	His α139 Nϵ2	2.2
Ni-2	$\longleftrightarrow$	Asp α363 Oδ1	2.1
Ni-2	$\longleftrightarrow$	O-B	2.3
Ni-2	$\longleftrightarrow$	O-2	2.3

Table 3.3: Nickel coordination distances in DAP inhibited *B. pasteurii* urease active site

3.4 Acetohydroxamic Acid Inhibited Active Site

Acetohydroxamic acid (AHA) binds to the urease active site in a chelating mode, as shown in Figures 3.9 and 3.10. Upon inhibitor binding the water/hydroxide cluster present in the native active site, is replaced by one molecule of AHA. The distance between the two nickel ions is slightly shortened to 3.5 Åcompared to the neative enzyme. One AHA oxygen (O-1) binds to Ni-1, while the second AHA oxygen (O-B) bridges symmetrically between the two nickel ions (Ni-1 $\longleftrightarrow$ O-B=Ni-2 $\longleftrightarrow$ O-B=2.0 Å) as already observed in the native structure for the bridging hydroxide.

His α222 Nϵ2, protonated, as deduced from the interaction of Nδ1 with the main chain nitrogen of Asp α224, donates a hydrogen bond to O-1 (O-1 $\longleftrightarrow$ His α222 Nϵ2 2.7 Å). The two nickel ions are penta-coordinated with Ni-1 distorted square-pyramidal and Ni-2 distorted trigonal-bipyramidal.

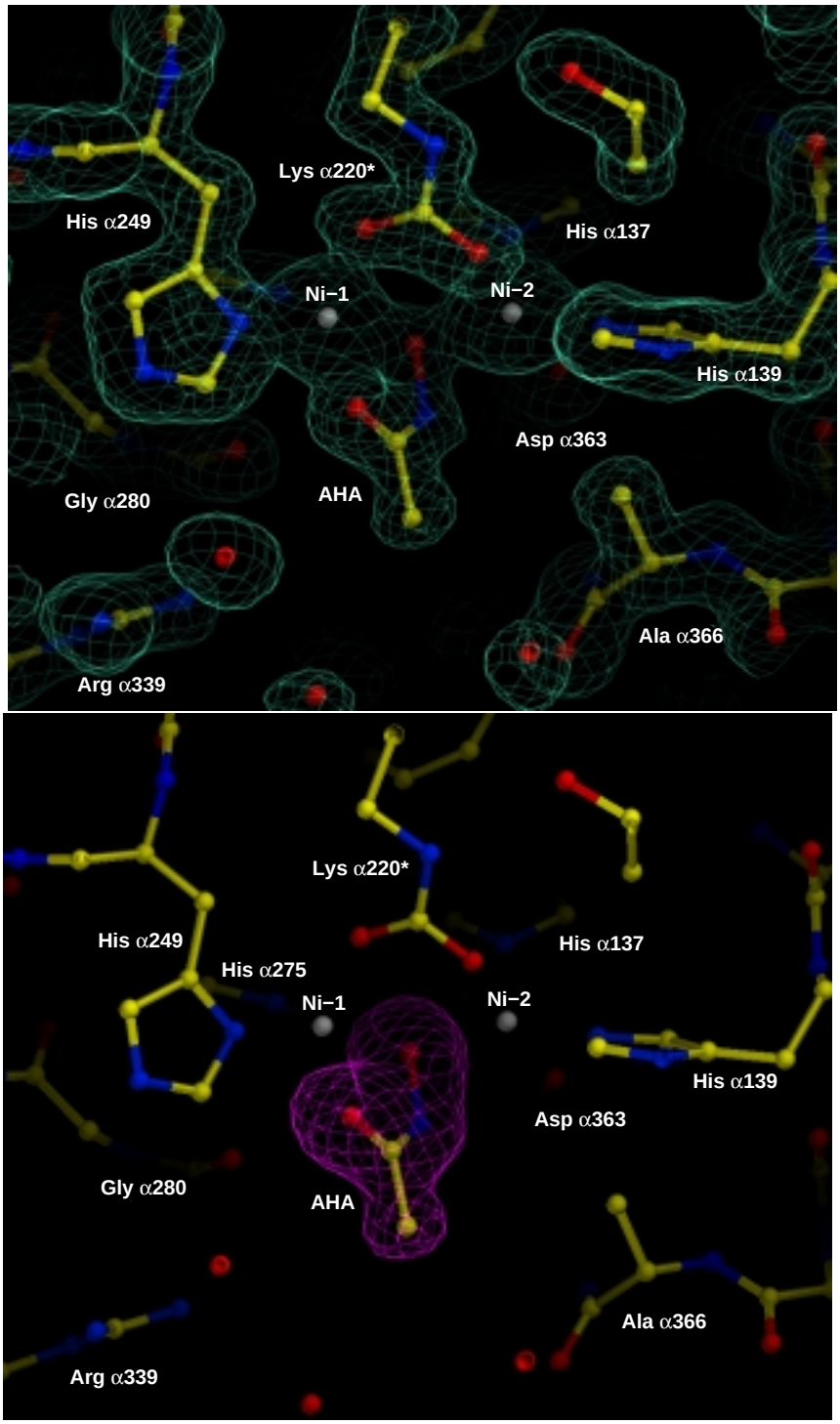


Figure 3.9: The model of AHA inhibited *Bacillus pasteurii* urease superimposed on the $2mF_{obs} - DF_{calc}$ contoured at 1 σ (top) and $mF_{obs} - DF_{calc}$ contoured at 3 σ electron density maps (bottom). The picture was made using "O"[71].

The equatorial plane for the square-pyramidal Ni-1 is formed by AHA O-B, AHA O-1, His α249 Nδ1 and His α275 Nϵ2 while the apical ligand is Lys α220* Oθ1. The equatorial plane for the trigonal-bipyramidal Ni-2 is formed by His α137 Nϵ2, His α139 Nϵ2 and AHA O-B, the apical ligands are Lys α220* Oθ2 and Asp α363 Oδ1.

In the AHA inhibited *K. aerogenes* C319A mutant, O-B asymmetrically coordinates the two nickel ions (Ni-1 $\longleftrightarrow$ O-B 2.6 Å, Ni-2 $\longleftrightarrow$ O-B 1.8 Å) differing in this respect from model compounds [90]. This asymmetry may be due to the formation of a hydrogen bond between AHA N-H and the carbonyl oxygen of Ala α363 (Ala α366 in *B. pasteurii* urease). In *B. pasteurii* urease the carbonyl atom of Ala α366 points away from the active site and does not make any hydrogen bond with the inhibitor whose N-H forms a hydrogen bond with Asp α363 Oδ2 (at 2.6 Å). The Asp α363 carboxylate group rotates about its Cβ−Cγ bond by 35° with respect to the native structure in order to establish a hydrogen bond with the AHA N-H. This behaviour confirms again the importance of the movements of the Asp α363 side chain for the reaction and inhibition mechanism [60]. The mobile flap is in the open conformation and has high B-factors. The side chain of the catalytic His α323 is not visible because of disorder. Table 3.4 lists the coordination distances for the nickel ions in AHA inhibited urease.

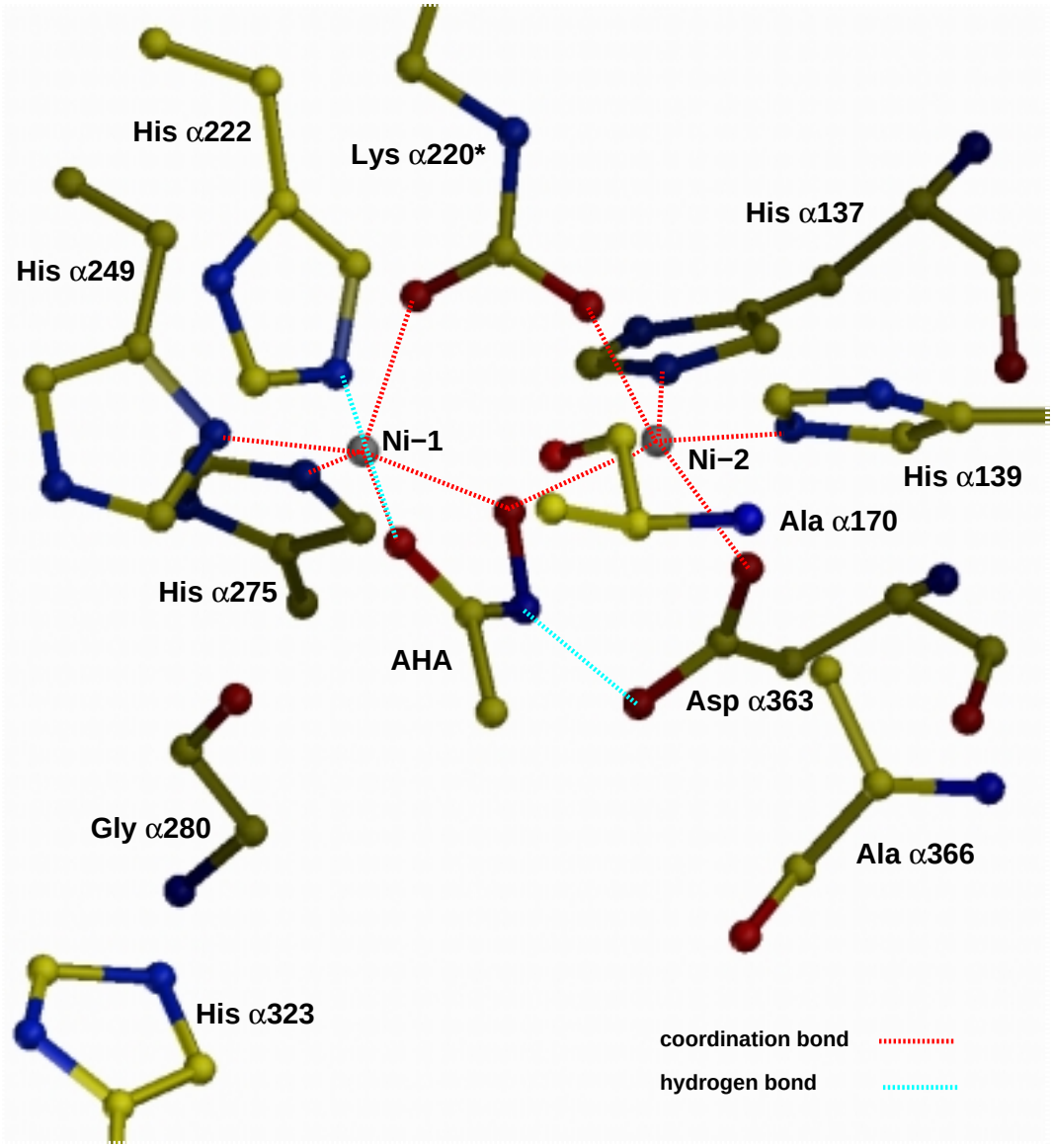


Figure 3.10: Model of the AHA inhibited *B. pasteurii* urease active site. The picture was made using "O"[71].

Nickel	$\longleftrightarrow$	Ligand	Distances Å
Ni-1	$\longleftrightarrow$	Lys α220* Oθ1	2.0
Ni-1	$\longleftrightarrow$	His α249 Nδ1	2.0
Ni-1	$\longleftrightarrow$	His α275 Nϵ2	2.0
Ni-1	$\longleftrightarrow$	O-B	2.0
Ni-1	$\longleftrightarrow$	O-1	2.2
Ni-2	$\longleftrightarrow$	Lys α220* Oθ2	2.0
Ni-2	$\longleftrightarrow$	His α137 Nϵ2	2.0
Ni-2	$\longleftrightarrow$	His α139 Nϵ2	2.0
Ni-2	$\longleftrightarrow$	Asp α363 Oδ1	2.1
Ni-2	$\longleftrightarrow$	O-B	2.0

Table 3.4: Nickel coordination distances in the AHA inhibited *B. pasteurii* urease active site

3.5 Mercaptoethanolamine Inhibited Active Site

The mode of binding of MEA through bridging between the nickel ions is clearly shown. The bridging of the MEA sulphur atom shortens the distance between Ni-1 and Ni-2 to 3.3 Å similar to BME in which the distance is 3.1 Å. No other non protein ligands are coordinated to the metal centre. Ni-1 is tetracoordinated and Ni-2 is pentacoordinated with a distorted trigonal-bypiramidal coordination geometry. Ni-1 is coordinated to protein ligands Lys α220* Oθ1, His α249 Nδ1 and His α275 Nϵ. Completing the coordination sphere for both nickel is the MEA sulphur atom. The Ni-2 equatorial plane is constituted by the MEA sulphur, His α137 Nϵ2 and His α139 Nϵ2. The opposite apical ligands are Lys α220* Oθ2 and Asp α363 Oδ1. Unfortunately the electron density maps ($2mF_{obs}-DF_{calc}$ and $mF_{obs}-DF_{calc}$) do not show the NH_2 group of the inhibitor mercaptoethanolamine (MEA) (Fig. 3.11) possibly due to multiple conformations or to the low completeness of the data (84%).

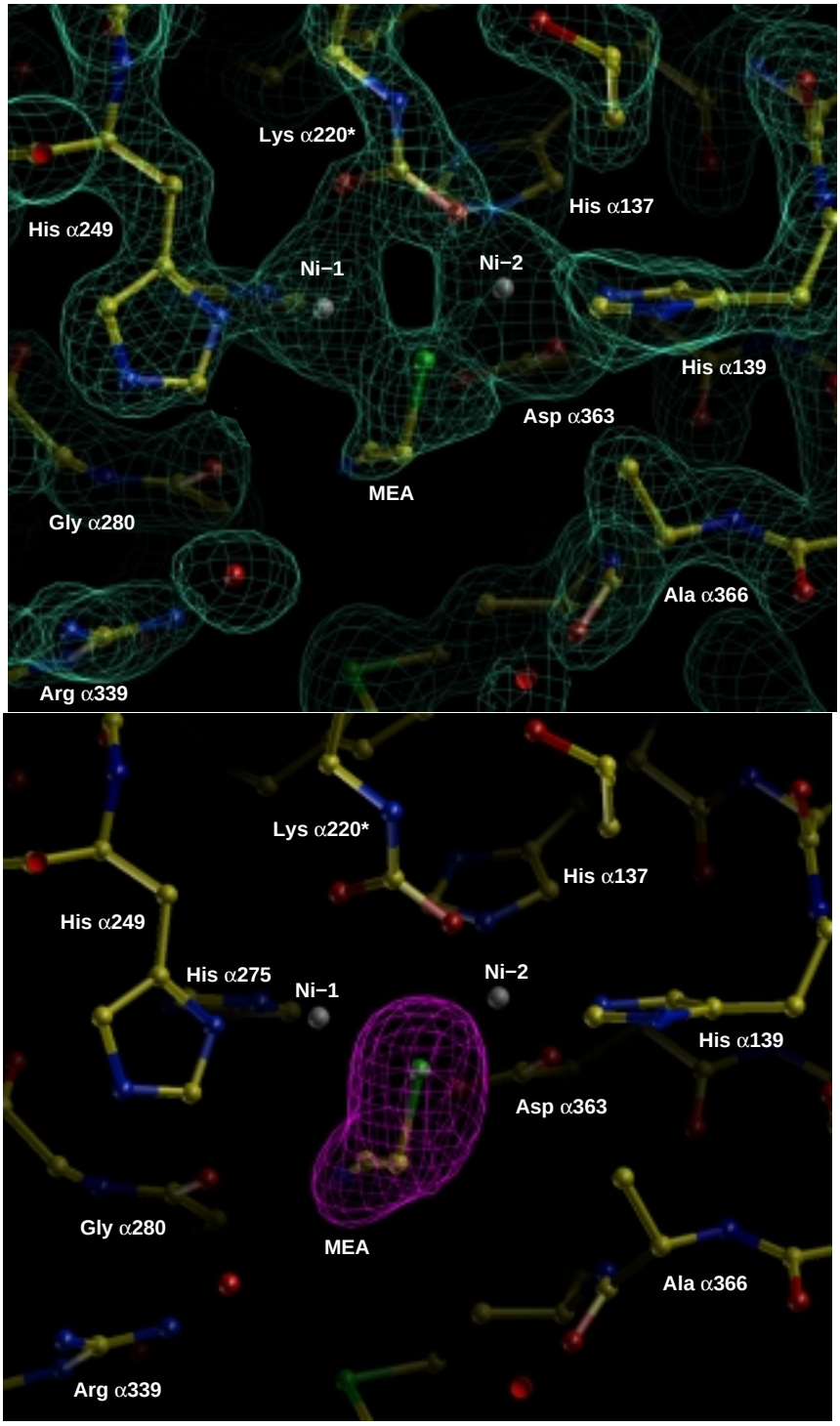


Figure 3.11: The model of MEA inhibited *Bacillus pasteurii* urease superimposed on the $2mF_{obs} - DF_{calc}$ contoured at 1 σ (top) and $mF_{obs} - DF_{calc}$ contoured at 3 σ electron density maps (bottom). The picture was made using "O"[71].

Amongst thiol compounds MEA is the more efficient in inhibiting urease with a $K_i \approx 0.01$ mM [31]. The high affinity of MEA for the urease active site is due to the presence of the positively charged amino group [31]. The MEA NH_2 can probably donate hydrogen bonds to the acceptors Gly α280 carbonyl oxygen and Asp α363 Nδ2 both located underneath the nickel centre.

Figure 3.12 shows the binding of MEA in the active site. The putative hydrogen bonds between the MEA amino group and the partners are marked with question marks. The mobile flap is again in the open conformation with the His α323 side chain not visible because of disorder. Cys α322, although in the presence of a thiol compound, does not form a mixed disulphide bridge as in the case of the BME inhibited enzyme. A better dataset might show electron density for the MEA NH_2 group providing a clearer explanation for the inhibition. Table 3.5 lists the coordination distances for the nickel ions in MEA inhibited urease.

Nickel	$\longleftrightarrow$	Ligand	Distances Å
Ni-1	$\longleftrightarrow$	Lys α220* Oθ1	2.3
Ni-1	$\longleftrightarrow$	His α249 Nδ1	2.2
Ni-1	$\longleftrightarrow$	His α275 Nϵ2	2.2
Ni-1	$\longleftrightarrow$	MEA-S	2.3
Ni-2	$\longleftrightarrow$	Lys α220* Oθ2	2.1
Ni-2	$\longleftrightarrow$	His α137 Nϵ2	2.0
Ni-2	$\longleftrightarrow$	His α139 Nϵ2	2.1
Ni-2	$\longleftrightarrow$	Asp α363 Oδ1	2.0
Ni-2	$\longleftrightarrow$	MEA-S	2.3

Table 3.5: Nickel coordination distances in MEA inhibited *B. pasteurii* urease active site

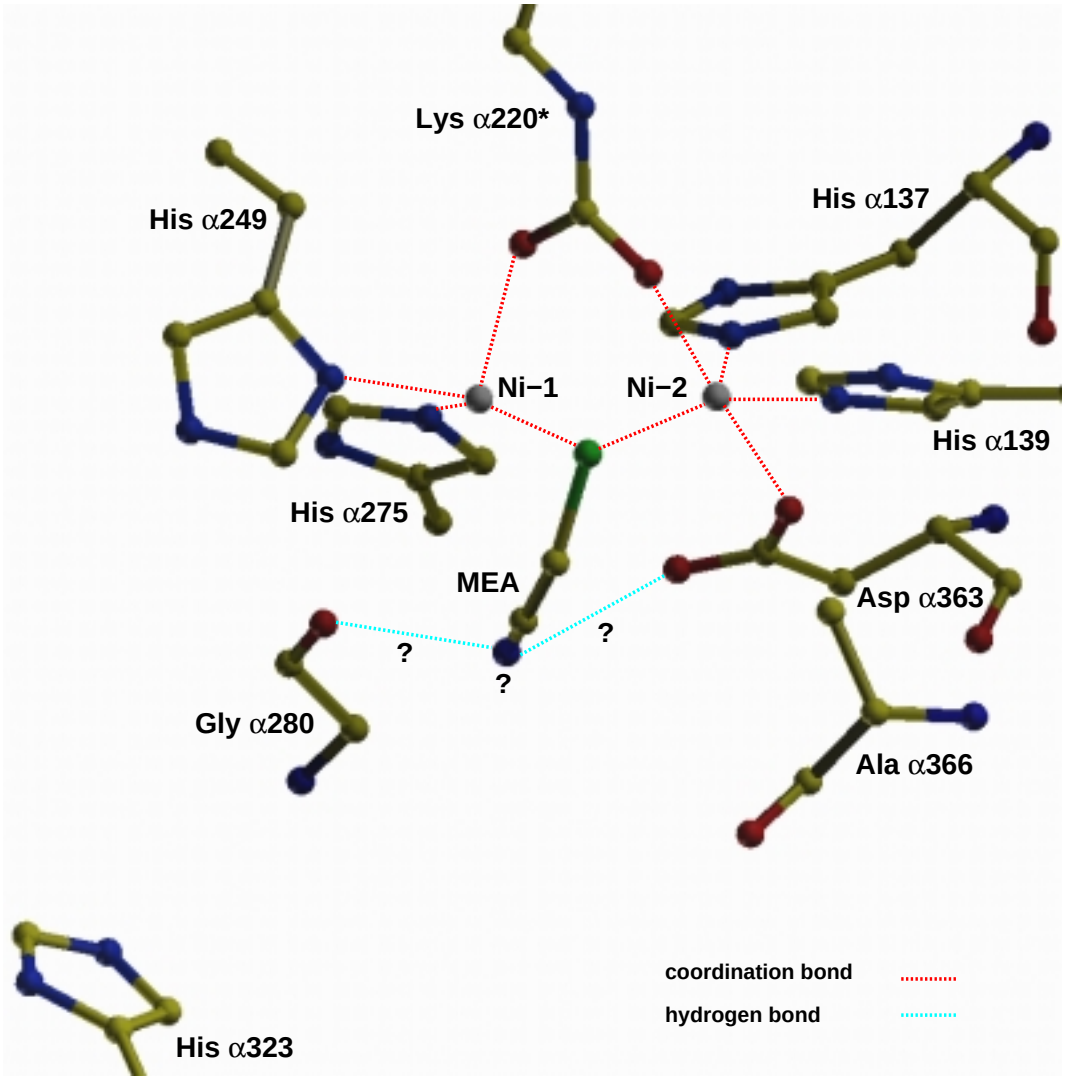


Figure 3.12: Model of the MEA inhibited *B. pasteurii* urease active site. The picture was made using "O"[71].

3.6 Phosphate Inhibited Active Site

Phosphate binds in the active site between the two nickel ions in a similar way to DAP (Fig. 3.13 and 3.14). The hydrogen bonding network established by the phosphate group and the pH of crystallisation (6.3) suggest that the bound

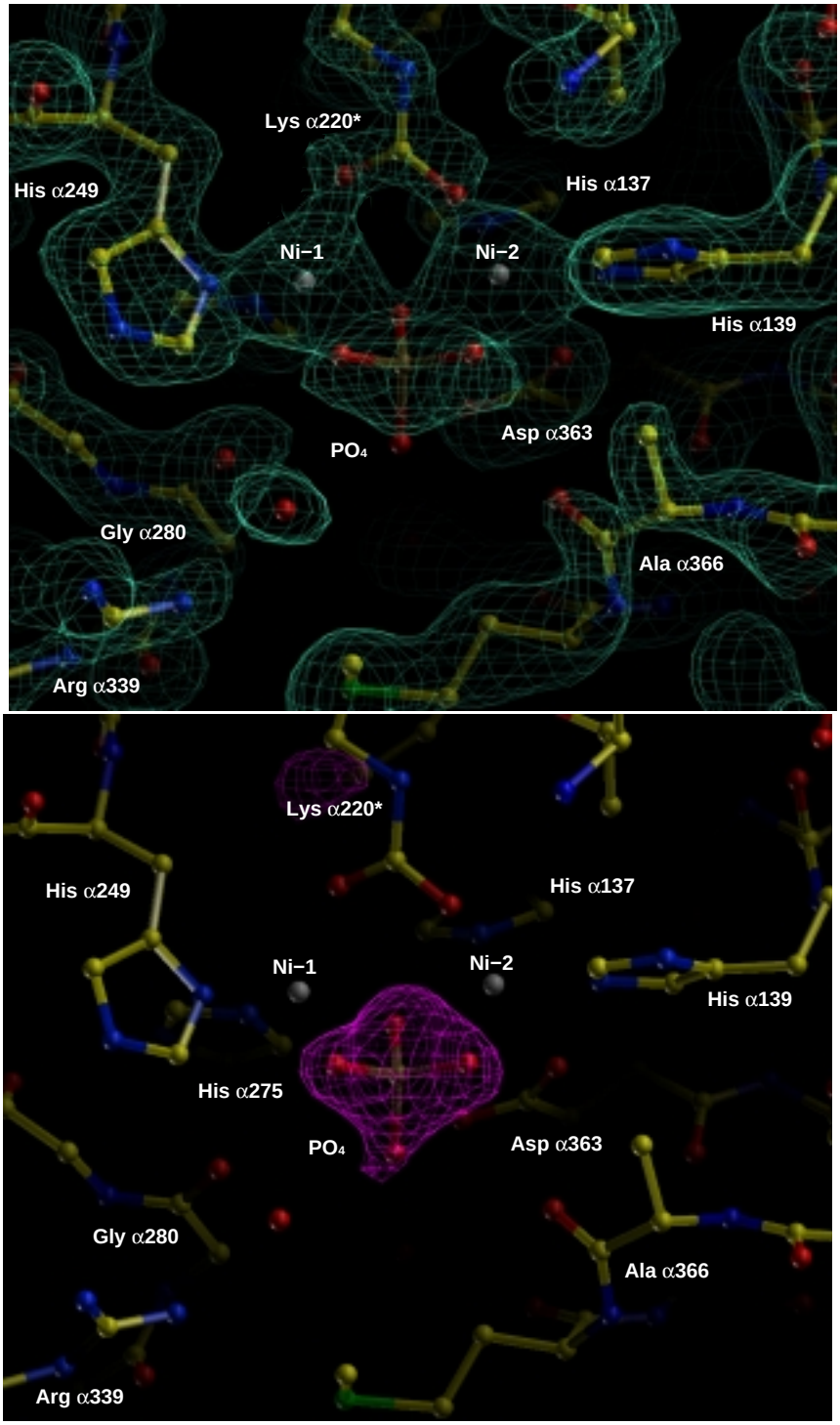


Figure 3.13: The model of phosphate inhibited *Bacillus pasteurii* urease superimposed on the $2mF_{obs} - DF_{calc}$ contoured at 1 σ (top) and $mF_{obs} - DF_{calc}$ contoured at 3 σ electron density maps (bottom). The picture was made using "O" [71].

molecule is $H_2PO_4^-$. The distance between Ni-1 and Ni-2 (3.5 Å) is very close to the distance observed for the native and the DAP inhibited structure. Ni-1 is pentacoordinated and Ni-2 hexacoordinated as in the native and the DAP inhibited enzyme. The coordination geometry is distorted square-pyramidal for Ni-1 and distorted octahedral for Ni-2. The oxygens of the phosphate are bound to the two nickels, while the phosphorous does not interact directly with the metal. O-1 coordinates to Ni-1 and accepts a proton from His α222 Nϵ2 (at 2.6 Å), protonated as deduced from the interaction of Nδ1 with the main chain nitrogen of Asp α224. O-2 coordinates to Ni-2 and donates an hydrogen bond to Ala α170 O (at 2.8 Å). O-B bridges symmetrically between Ni-1 and Ni-2 and is located at 2.8 Å from Asp α363 Oδ2 suggesting the presence of an hydrogen bond. In this context the hydrogen is located on the Asp α363 Oδ2. Meanwhile Asp α363 Oδ2 is located 2.9 Å from O-D. The distance of 3.0 Å between the Ala α366 carbonyl oxygen and O-D suggests a hydrogen bond interaction. Although the flap is in the open conformation the Ala α366 carbonyl points towards the active site as in the DAP inhibited urease whose flap is closed.

The phosphate inhibited structure reconfirms that a tetrahedral molecule can be accomodated between the two nickels and shows the importance of the protonation state of the inhibitor for the binding to the active site. Table 3.6 lists the coordination distances for the nickel ions in phosphate inhibited urease.

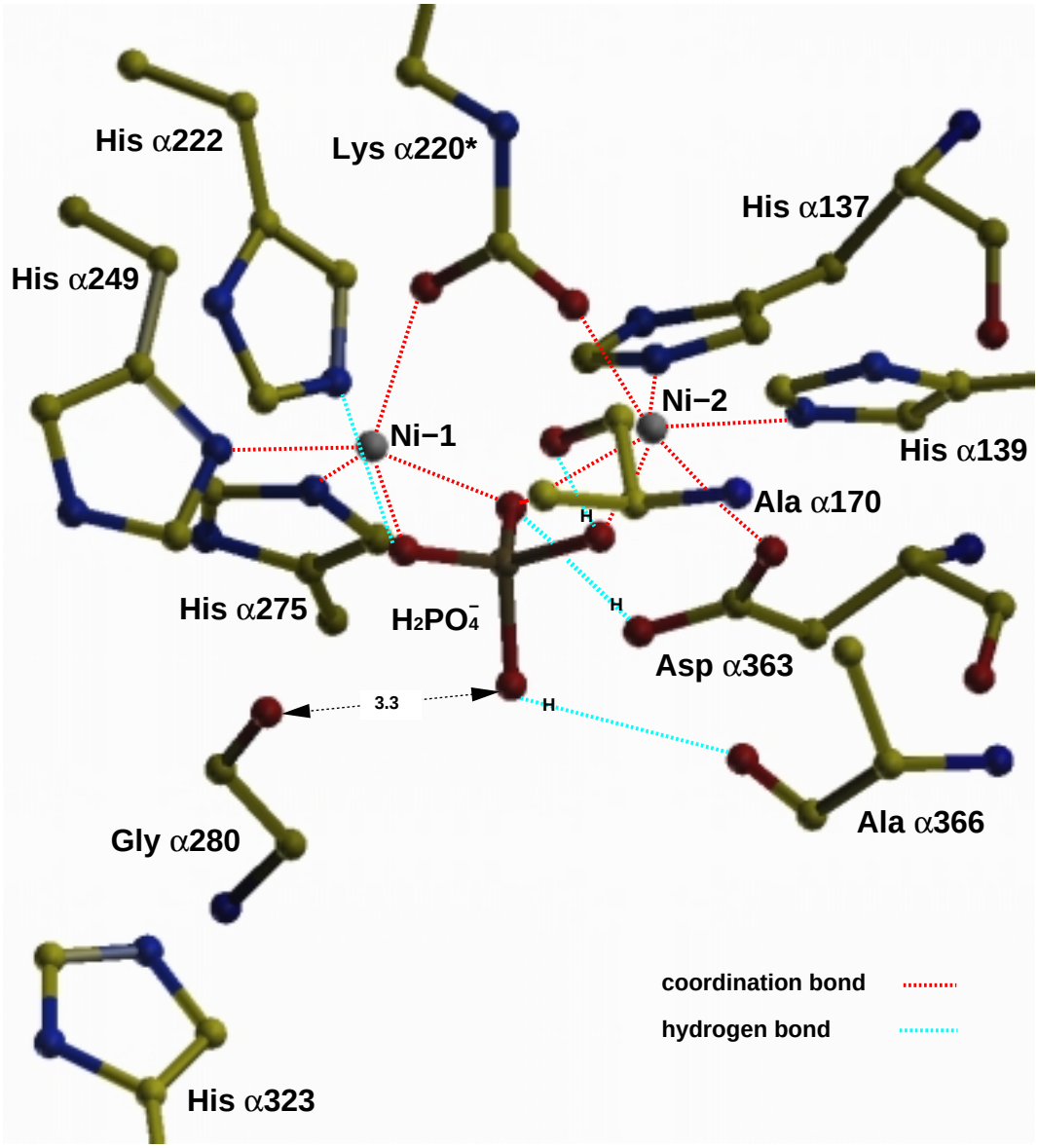


Figure 3.14: Model of the phosphate *B. pasteurii* urease active site. Only those hydrogen atoms essential for phosphate binding are shown. Note that the contact distance for Gly α280 CO no longer implies a hydrogen bond. Picture made with the package "O" [71].

Nickel	$\longleftrightarrow$	Ligand	DistanceÅ
Ni-1	$\longleftrightarrow$	Lys α220* Oθ1	2.0
Ni-1	$\longleftrightarrow$	His α249 Nδ1	2.0
Ni-1	$\longleftrightarrow$	His α275 Nϵ2	2.0
Ni-1	$\longleftrightarrow$	O-B	1.9
Ni-1	$\longleftrightarrow$	O-1	2.3
Ni-2	$\longleftrightarrow$	Lys α220* Oθ2	2.0
Ni-2	$\longleftrightarrow$	His α137 Nϵ2	2.1
Ni-2	$\longleftrightarrow$	His α139 Nϵ2	2.1
Ni-2	$\longleftrightarrow$	Asp α363 Oδ1	2.2
Ni-2	$\longleftrightarrow$	O-B	1.9
Ni-2	$\longleftrightarrow$	O-2	2.3

Table 3.6: Nickel coordination distances in phosphate inhibited *B. pasteurii* urease active site

Chapter 4

Urease: Discussion and Conclusions

4.1 The Mobile Flap

The very mobile stretch of residues α311-α339 acts as a lid for the active site cavity possibly participating in the catalytic mechanism. It opens to allow the entrance of the substrate and closes to allow the reaction to proceed. After urea hydrolysis the flap would open again to release the reaction products, ammonia and carbon dyoxide, into the bulk solution. In all *B. pasteurii* urease crystal structures determined so far the flap is in the open conformation except when the enzyme is in complex with the competitive inhibitor diamidophosphate (DAP) (Fig. 4.1 and 4.2).

Upon binding to the Ni^{2+} ions, DAP activates a network of hydrogen bonds involving also His α323 which is located in the flap. The shift of this residue by about 5 Å moves the flap to the closed conformation. Figure 4.1 shows the flap of the DAP-inhibited enzyme superimposed on the native structure. Based on

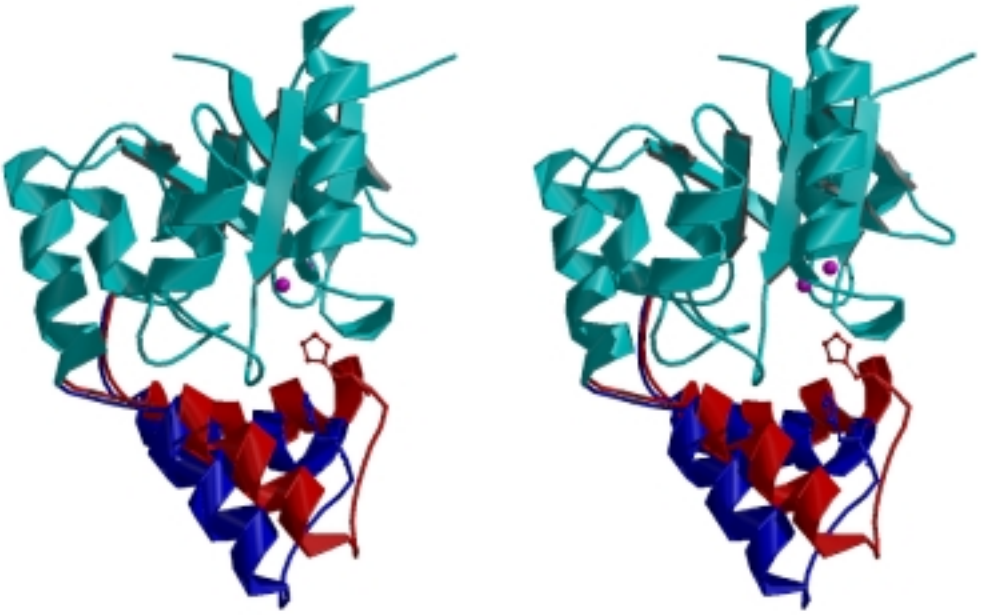


Figure 4.1: Stereo representation of the region around the active site in *B. pasteurii* urease. The region α130-α160, α205-α306 and α343-α376 is shown in cyan. The flap, α306-α343, is open the native structure (blue) and closed in the DAP inhibited structure (red). The catalytic His α323 is represented in ball and stick. The two nickel ions are represented as magenta spheres. The figure was made with BOBSCRIPT [80] and Raster3D [81].

biochemical and site-directed mutagenesis studies it has been proposed that the flap residues play a role in the catalytic mechanism [40, 41]. In particular His α323 is proposed to be the general base during the hydolysis [40, 41, 91].

Figure 4.2 shows the structural superposition of the *B. pasteurii* urease flap in the native form and in complex with the inhibitors, with the native *K. aerogenes* urease flap for comparison. All *K. aerogenes* urease structures de-

termined so far have the flap in the closed conformation apart from the mutant

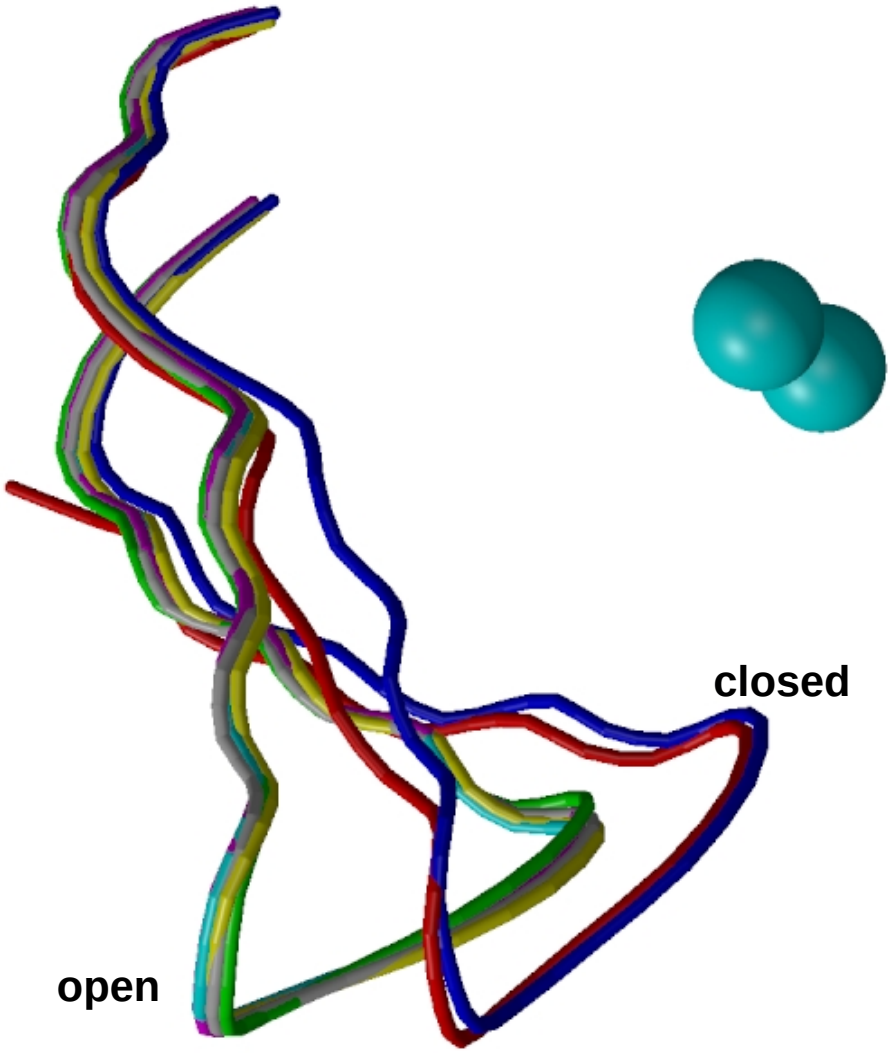


Figure 4.2: Superposition of the *B. pasteurii* urease and native *K. aerogenes* urease flaps. In cyan the native structure, in green the BME inhibited, in blue the DAP inhibited, in yellow the AHA inhibited, in grey the MEA inhibited, in purple the phosphate inhibited, in red native *K. aerogenes*. Only the native nickels are shown for clarity. Figure made with MOLSCRIPT [83] and Raster3D [81].

Cα322Y in the open conformation because of tyrosine steric hindrance.

4.2 The Reaction Mechanism

"Cosa studia la chimica?
L'intima struttura della materia e le reazioni"[1]

The comparison between the structures of the native and the DAP inhibited *B. pasteurii* urease provides information about the enzyme both in the resting state and with a transition state analogue bound to the active site. The observed movement of the flap from the open conformation (native enzyme) to the closed conformation (inhibited enzyme) together with the involvement of His α323 in DAP binding, demonstrate the important role of the flap during urea hydrolysis.

4.2.1 Selectivity of Urea Binding

The urea molecule, upon entering the active site groove, is steered to a defined position at the two nickel ions by a highly specific hydrogen bonding network. The orientation of the molecule is stabilised on one side by the presence of the hydrogen bond donor His α222 Nϵ driving urea oxygen towards Ni-1. After urea coordination to Ni-1 *via* its carbonyl oxygen the active site conformational changes bring one of the urea NH_2 groups into the vicinity of Ni-2 with which it eventually binds (Fig. 4.3).

The three water molecules (wat-1, wat-2 and wat-3) present in the active site are replaced by the urea molecule which donates hydrogen bonds through its NH_2 groups to the acceptors Ala α170 and Ala α366 carbonyl oxygens. Worthy of note is the shift of Ala α366 O from its original position, pointing away

[1] Angiolina Cristofori, Professor of Chemistry

from the active site, to a new position pointing towards Ni-2. The closing of the flap brings His α323 into a position to act as a base accepting a proton from the distal urea NH_2 group.

The chelating binding of urea to the nickel ions through coordination to Ni-1 with its carbonyl oxygen and to Ni-2 with one of its amino groups is facilitated by the enhanced basicity of the NH_2 group because of the hydrogen bonds established with Ala α170 and Ala α366 carbonyl oxygens.

4.2.2 Urea Activation and Hydrolysis

This section discuss the mechanism proposed in our *Structure* paper [91]. The binding of urea to the two Ni^{2+} ions together with the hydrogen bond network causes the polarisation of the C-O and C-NH_2 bonds with activation of the otherwise inert urea molecule. The urea carbon is prone to nucleophilic attack by the bridging hydroxide because the electron cloud around it is withdrawn by the more electrophilic nickels. The hydroxide is proposed to act either as a nucleophile, giving rise to the tetrahedral intermediate, or as an acid, donating the hydrogen necessary to protonate the distal urea NH_2 group giving ammonia. The transfer of the proton from the bridging hydroxide to the NH_2 is achieved *via* Asp α363 Oδ2 with a rotation of its carboxyl group about the Cβ−Cγ bond.

After ammonia formation, the C-NH_3 bond breaks and NH_3 is released from the active site upon flap opening, assisted by its hydrogen bond partner His α323. The resulting carbamic acid is then released from the active site by the repulsive interaction between negatively charged oxygens (one carbamate oxygen and Asp α363 Oδ2) assisted by the positively charged guanidinium group of

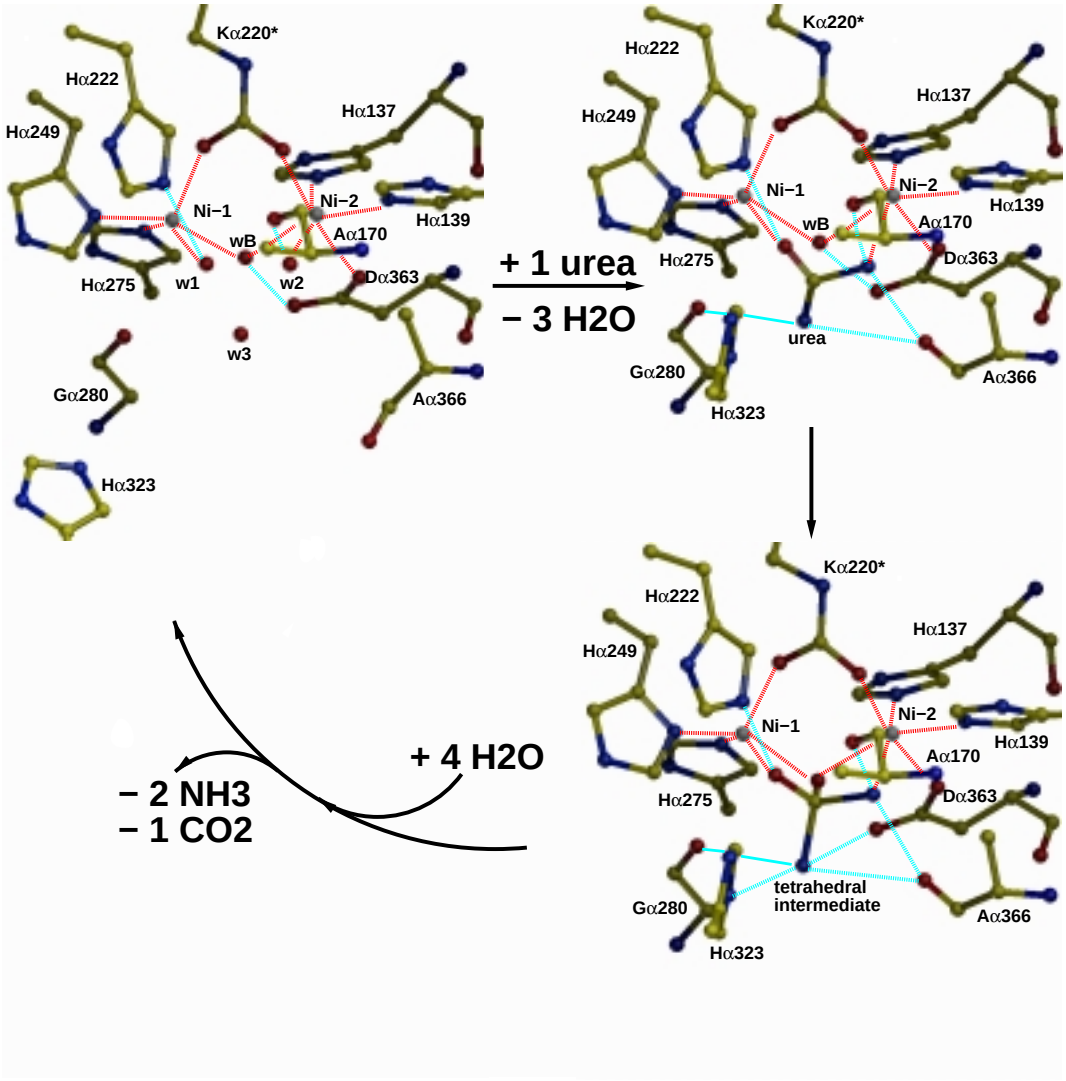


Figure 4.3: Urease reaction mechanism.

Arg α339 during flap opening. For the reaction to occur the presence of a group with a p$K_a \approx 6.5$ is required which would act as a catalytic base, and of a group with a p$K_a \approx 9$ acting as an acid [39]. In the previous urease mechanism proposed by Karplus *et al.* [48] based on the crystal structures of native *K. aerogenes* and its mutants, urea binds to the more electrophilic Ni-1 in a

monodentate mode while a hydroxide ion binds to Ni-2 and plays the role of the nucleophile. His α320 (corresponding to His α323 in *B. pasteurii*) would act as a general acid providing the proton for the nascent ammonia. However His α320 has a $pK_a \approx 6.5$ and at the optimal pH for the reaction (8 [13]) cannot be protonated. The water bound to Ni-2 must be deprotonated to hydroxide by a general base but the general base is not found at a reasonable distance in the active site. To explain the presence of the protonated His α320 (with $pK_a \approx 6.5$) the reverse protonation hypothesis was introduced which would result in only 0.3% of all urease molecules in the correct protonation state for the catalysis. The authors did not provide any explanation about the missing general base [48]. In *B. pasteurii* His α323 is proposed to be the catalytic base while the bridging hydroxide would play both the role of the nucleophile and of the general acid providing the proton needed in the reaction [91].

4.2.3 But Why Nickel?

The peculiarity of urease is the presence of a binuclear nickel containing active site while hydrolytic enzymes normally require the presence of zinc ions for their reactions to occur.

In the case of urease, the mode of binding the substrate, the presence of the bridging hydroxide and the multiple coordination sites needed for metal-protein binding, requires an octahedral coordination geometry, as formed for nickel, while zinc is mostly tetrahedral. Furthermore nickel ions are characterised by higer affinity for nitrogen-based ligands than zinc ions [92], thereby stabilising the binding of urea NH_2 group.

4.2.4 The Mechanism Revisited

After submission of the thesis and also as a result of intensive discussion with my examiners [2], I have revisited the question of the mechanism.

In the *B. pasteurii* DAP inhibited enzyme, the key residue His α323 makes an H-bond interaction with Asp α224 and is close (2.89 Å) to the Arg α339, as in native *K. aerogenes* urease. As shown in Figure 4.4 the hydrogen bond network established by His α323 with its neighbours will modulate the protonation state of the His α323 nitrogens and the pK_a ($\approx$6.5). These interactions are established however only when the flap is in the closed conformation suggesting a dynamic variation, not only in the position of the residues but also in their protonation state and contacts. When the flap is in the open conformation, as found in most *B. pasteurii* crystal structures, His α323 is often disordered and "free to swim" in the solvent. Upon flap closure His α323 Nδ1 establishes a strong hydrogen bond with Asp α224 Oδ2 (at 2.88 Å). Interactions between carboxyl groups generally raise the pK_a of histidines, as in the serine proteases such as trypsin in which the enzymatic function of the catalytic triad is based on the interactions between Asp-102, His-57 and Ser-195 [93]. In urease, Arg α339 Nη2 makes an unusual interaction with His α323 with one of its NH_2 protons pointing to the π-orbital of His α323 with Nϵ2 located at the very short distance of 2.89 Å. This interaction will reduce the pK_a of His α323. Thus in the active enzyme His α323 experiences two competing effects. Arg α339 Nη2 also donates a proton in a hydrogen bond to Ala α279 O (at 3.01 Å) and its Nη1 donates a hydrogen to a water molecule (W-213 at 2.82 Å). W-213 points its hydrogens towards Glu α223 Oϵ2 and Asp α224 Oδ1 located at 2.86 and 2.75 Å respectively to establish strong hy-

[2] Prof. Guy Dodson and Prof. Philip Evans

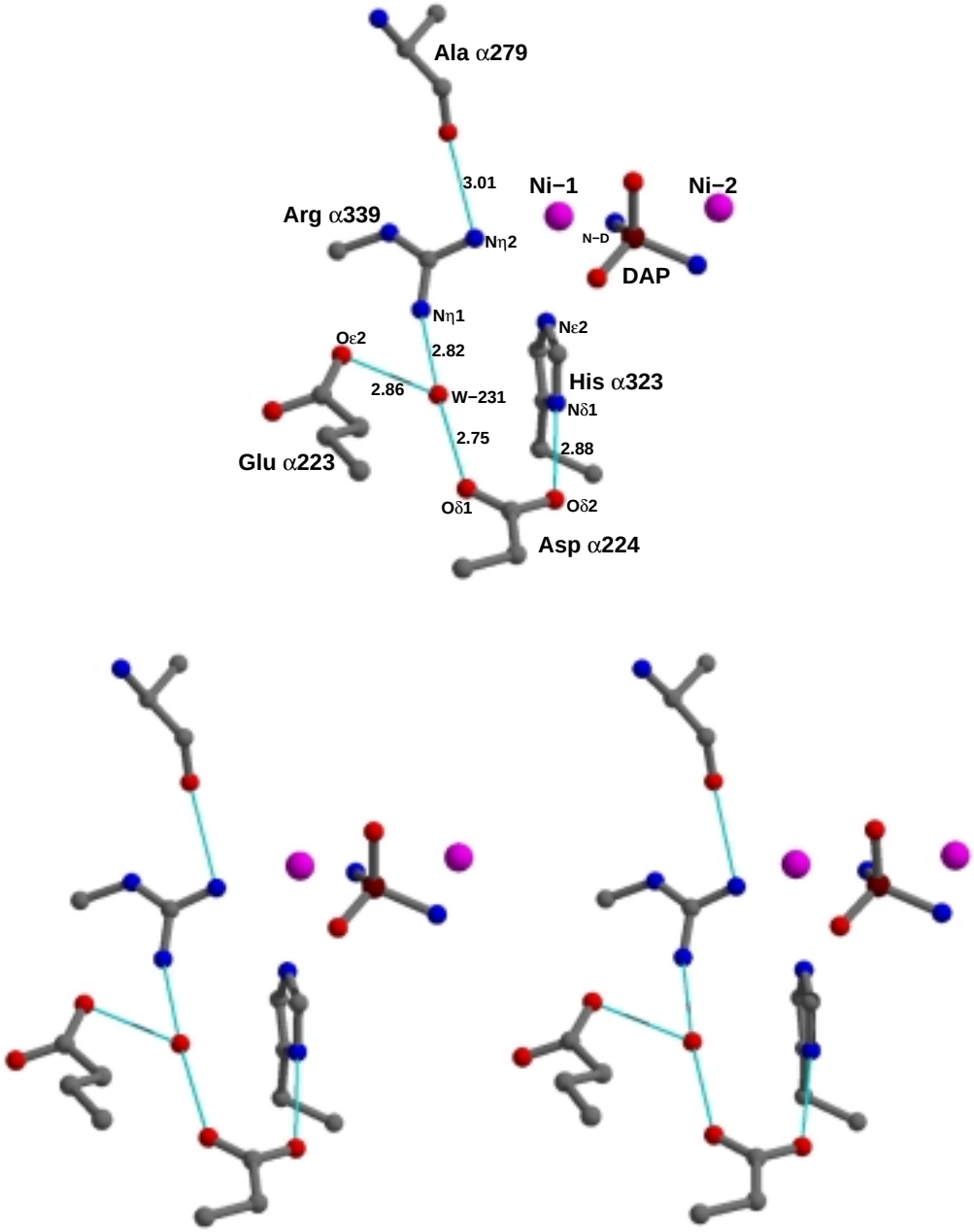


Figure 4.4: Interactions between His α323 and its partners in DAP inhibited *B.pasteurii* urease. In this context His α323 Nϵ2 is located 3.25 Å from DAP N-D.

drogen bonds. In this context His α323 Nϵ2 is located 3.25 Å from DAP N-D. The proposal for a role of Arg α339 is based on a dynamic view of protonation and deprotonation of His α323. The action of this histidine, if catalytic, requires movement of the flap to modulate the protonation state by temporarily raising and decreasing its pK_a. While the flap is closing His α323 picks up a proton from the bulk solvent because its pK_a is raised by the interaction with Asp α224. The interaction with Arg α339 is such that Arg α339 Nη2 kicks away the proton from His α323 Nϵ2 because of the very peculiar and unfavourable interaction.

Accepting the role of this "catalytic organisation" of Asp α224, His α323 and Arg α339 as the proton donor machinery, and the bridging hydroxide as the nucleophile, a urease reaction mechanism can be proposed as follows. The urea molecule, upon entering the active site cavity is oriented by a developing hydrogen bonding network. The presence of the hydrogen bond donor His α222 Nϵ2 drives the urea oxygen towards coordination to Ni-1. The enhanced basicity of the NH_2 group because of the hydrogen bonds established with the Ala α170 and Ala α366 carbonyl oxygens results in its coordination to Ni-2. The binding of urea to the two Ni^{2+} ions together with the hydrogen bond network causes the polarisation of the C-O and C-NH_2 bonds with resulting activation of the otherwise inert urea molecule. The urea carbon is now prone to nucleophilic attack by the bridging hydroxide activated by the electrophilic nickels. The hydroxide acts as the nucleophile, giving rise to the tetrahedral intermediate, its proton remaining in the now very strong hydrogen bond with Asp α363. The proton needed to form NH_3 from the terminal urea NH_2 group is provided by the catalytic arrangement Asp α224, His α323 and Arg α339 *via* His α323 Nϵ2. At this stage the tetrahedral intermediate collapses giving

rise to one molecule of ammonia and, bridging between the two nickel ions, a molecule of carbamic acid in its protonated form ($H_2N-COOH$). The opening of the active site flap causes the exit of the first ammonia molecule and the weakening of the hydrogen bond network allowing the entrance of the bulk solvent with subsequent hydrolysis of the carbamate.

In a recent paper Pearson *et al.* [94] report an extended kinetic and structural characterisation of *K. aerogenes* urease active site mutants in order to ascertain the role of each residue thought to be involved in catalysis. The results of this study confirm the role of His α219 as a proton donor involved in orienting and assisting substrate binding as demonstrated by the increase in K_m shown by its mutants. His α320 is shown to be directly involved in the catalysis with values of k_{cat} for its mutants close to 0 (see Table 4.1). Unfortunately the activity assay was performed in such a way as to make the comparison between the data obtained more difficult. The activity of the urease variant was measured on raw cellular extract and the pH of the assay was not always the same. However these data and the structural evidence together demonstrate the importance of Asp α221 and Arg α336.

In the Pearson paper our previous proposal has been criticised regarding the proton transfer from the bridging hydroxide to the terminal NH_2 urea group. The reason is that the proton transfer *via* Asp α363 carbonyl is thought to be very unlikely. A second ruinous (!) problem relates to the difficulty of removing the hydrogen atom from the bridging hydroxide of the tetrahedral intermediate. The revisited mechanism provides a reasonable model without being in conflict with the biochemical evidence.

mutant	K_m			k_{cat}			Buffer & pH
wild-type	2.4	±	0.1	2970	±	100	HEPES, 7.75
H320A	10.9	±	0.6	0.068	±	0.003	MES 6.2
H320N	7.4	±	0.5	0.060	±	0.001	MES 6.2
H320Q	10.6	±	1.0	0.033	±	0.001	MES 6.2
H219A	2090	±	630	194	±	42	HEPES 7.75
H219N	175	±	8	322	±	6	MES 6.2
H219Q	227	±	21	1860	±	80	MES 6.2
D221N	3.6	±	1.0	60	±	5	HEPES 7.75
D221A	24	±	2.4	1.7	±	0.03	MES 6.2
R336Q	2.7	±	0.3	0.62	±	0.07	HEPES 7.75

Table 4.1: Kinetic parameters for *K. aerogenes* urease mutants as determined in cell extract [94]. The K_m is measured in mM and k_{cat} in s^{-1}.

4.3 Evolutionary Convergence

The proposed reaction mechanism for urease is analogous to the one proposed for other bimetallic hydrolases in which a bridging hydroxide acts as nucleophile suggesting a possible common mechanism for these enzymes. The structures of bimetallic enzymes like arginase [95] and proline aminopeptidase [96] containing manganese, phospholipase C [97], leucine aminopeptidase [98] and aryldialkylphosphatase [99] containing zinc and inorganic pyrophosphatase [100] containing magnesium, despite the presence of different metal ions in the bimetallic centre all have a bridging hydroxyde. The comparison of the quaternary structure of urease with proteins in the PDB, carried out by Holm and Sander [101] using the program DALI [102], demonstrated that proteins belonging to the same hydrolase family but with very different primary structure share a similar structural architecture.

The structural alignment of urease, phosphotriesterase and adenosine deaminase reveals an evolutionary convergence toward the formation of a catalytic

$(\alpha\beta)_8$ barrel domain hosting a conserved metal binding site, with Zn substituted by Ni in urease. The metal binding residues conserved among the three proteins are four histidines and an aspartic acid. Phosphotriesterase and urease feature a bridging carbamylated lysine, which is peculiar to these two proteins since no other enzyme of the same family have conserved lysines as metal ligand. Also for these enzymes the reaction mechanism involves the presence of a metal bound hydroxide acting as a nucleophile to attack the substrate. This structural and functional similarity suggests a common catalytic mechanism and the proposed urease reaction mechanism may represent a general model for all amidohydrolases and related metallo hydrolases with similar catalytic sites [103].

4.4 Urease as a Target for Drug Design

Structure based inhibitor design represents the best approach to find compounds that can block or modulate the activity of enzymes with high specificity and low health and environmental impact. However the most common approach to discover new molecules is still based on extended screening of chemicals. The reason for these "*shoot first ask later*" methods are not only due to economical reasons but quite often to the lack of reliable models. The information provided by the structure of an enzyme is usually not enough to explain how the reaction occurs or how different inhibitors can bind to the native active site. Only the accurate comparison between different complexes can show the different conformations adopted by the residues involved in binding/catalysis. The molecular analysis of an active site can be performed with programs that calculate the force field taking into account hydrogen bonding and protein backbone parameters. The calculations by docking programs are time consuming, limited to static systems and require good starting models.

Urease from *B. pasteurii* represents an ideal model for structure based drug design. The structure of the enzyme native and in complex with five compounds with very distinct chemical characteristics provide models for efficient docking studies. The results obtained in this thesis allow a map of the active site to be drawn in which hydrogen bond donors and acceptor are easily identifiable. The different conformations adopted by the residues involved in binding give a dynamic view of the active site moving from a "free" to a "bound" state.

4.5 On the Binding of Urease Inhibitors

Figure 4.5 shows the *B. pasteurii* urease active site in the closed conformation in the absence of ligands. The residues involved in substrate/inhibitors binding are marked with letters according to their role, allowing an easy localisation of putative binding sites for new molecules. Five hydrogen bond acceptors, two hydrogen bond donors and one sulphur atom have been identified, while three coordination sites are available on Ni-1 and two on Ni-2. Very interesting is the position of His α222 Nϵ2 right above Ni-1 and of Ala α170 O close to Ni-2. Acting the first as an hydrogen bond donor and the second as an hydrogen bond acceptor these two residues are primarily responsible for orienting urea in the active site. The small size of the active site together with the predominance of hydrogen bond acceptors makes it prefer tetrahedral inhibitors mimicking the transition state of the reaction. The influence of the pH on the inhibitor binding to the active site was shown by pH dependence biochemical studies [31, 36]. For thiol compounds the ability to compete with urea is strongly influenced by the presence of a protonated group. MEA and BME bridge between the two nickels with the sulphur atom, probably not influenced

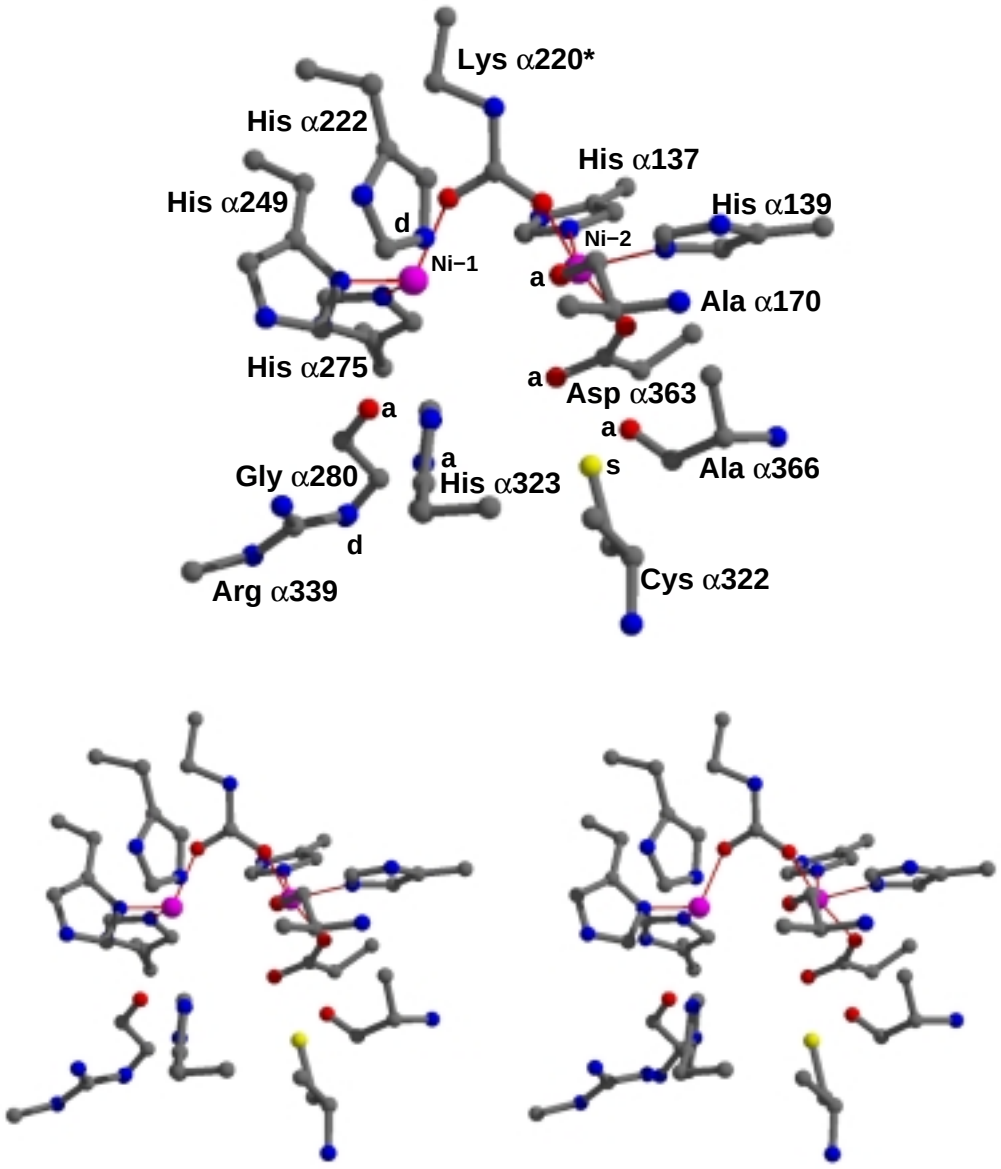


Figure 4.5: Urease binding sites. 'Top': H-bond donor are marked with 'd' H-bond acceptor with 'a' and the sulphur with 's'. 'Bottom': Stereo view of the same Figure.

much by changes in the pH. However both thiols show a drop in the inhibition of urease upon pH increase from 6 to 8 [31]. The reason for this behaviour is the presence of ionisable groups on the molecules. MEA has a β-amino group whose protonation state is important in the interaction with the hydrogen bond acceptors Gly α280 O and Asp α363 Oδ2. The alcoholic functionality of BME, activated by the hydrogen bond with the acceptor Gly α280 O, eventually coordinates to Ni-1. The second BME molecule involved in the inhibition reacts with the sulphur of Cys α322 while its protonated OH group donates the hydrogen to Ala α366 O. The structure of phosphate inhibited urease obtained at pH 6.3 shows the mode of binding of the $H_2PO_4^-$ ion to the nickel centre with the involvment of the hydrogen bond acceptors Ala α170 O, Ala α366 O. Increasing the pH would deprotonate the phosphate group to HPO_4^{2-} (phosphate $pK_{a1} \approx 2,\ pK_{a2} \approx 7$) and decrease the affinity for the urease active site by weakening the hydrogen bonding network.

Inhibition with AHA involves coordination of its anion (pK_a=8.7) to the metallocentre and two hydrogen bonds with Asp α363 Oδ2 and His α222 Nϵ2. No significant changes are observed in K_i from pH 6.75 to 8.25 [31] but the affinity decreases at lower pH as would be expected if a catalytic nucleophile were required for complex formation [31]. The decreased affinity is probably due to protonation of AHA that would disfavour the direct binding of the inhibitor to the nickel ions by establishing hydrogen bonds with the several hydrogen bond acceptors into the active site. Another reason for this could be the protonation of Asp α363 Oδ2 that would interact negatively with the AHA nitrogen.

The tetrahedral DAP, very closely matching the shape of the proposed reaction intermediate, is the best inhibitor among those used in this study. The

high affinity for the active site can be ascribed to the presence of the nitrogen atoms preferred to oxygen by the nickel atoms and to the presence of hydrogen bond donors able to satisfy all the hydrogen bond acceptors in the catalytic cavity. The binding of DAP, which has the possibility of being $(NH_2)_2PO_2^-$ or $(NH_2)_2PO_2H$ depending on the pH, is not influenced much by changes in the protonation state. In basic pH conditions DAP would be $(NH_2)_2PO_2^-$ with a unfavorable interaction with the negatively charged Asp α363 Oδ2 slightly lowering the affinity of the molecule. However the set of completely satisfied hydrogen bond interactions would tightly bind the inhibitor into the active site.

4.6 The Ideal Urease Inhibitor?

"Multi sunt vocati pauci vero electi"

A long list of compounds tested for their inhibitory activity on urease can be found in the Handbook of Enzyme Inhibitors [104]. Some of these molecules have been used successfully in this study but unfortunately none of them is the ideal candidate we are looking for. The kind of inhibition needed for urease depends on the target we want to reach. In the case of agricultural fertilisation, the high homology between the plant and the bacterial enzyme make the task almost impossible. When urea is given in complex with an inhibitor as a fertiliser the aim is just to slow down urea hydrolysis and not to block it completely otherwise this will result in inactivation, not only of the bacterial urease but also of the crop enzyme because of the lack of selectivity. For this purpose competitive inhibitors are the best compromise slowing down urea degradation but allowing the bacterial and plant enzymes to retain some ac-

tivity.

In contrast when the target for the drug is a bacterial urease involved in human or animal diseases, the characteristics of the inhibitor need to be completely different. In this case a complete and irreversible enzyme inactivation is required in order to eradicate the pathogen. In both cases the final product of the reaction between the enzyme and the inhibitor should be harmless to the health and to the environment.

The shape of the active site may suggest that tetrahedral molecules are the best for enzyme inhibition. This is partly true since tetrahedral molecules resembling the proposed reaction intermediate would be easily hydrolysed. Phenylphosphorodiamidate (PPD) and its hydrolysis product diamidophosphate (DAP) are very efficient in inhibiting urease since they resemble the proposed reaction intermediate but the enzyme slowly regains its activity as shown in table 4.2 [34]. The ideal compound must be harmless to the

Inhibitor	$10^4 k_{react} \cdot S^{-1}$
Phenylphosphorodiamidate	0.36±0.03
N-(3-methyl-2-butenyl)phosphoric triamide	0.34±0.04
Phosphoric triamide	0.36±0.03
Diamidophosphate	7±1
Phosphoramidate	8.2±0.5

Table 4.2: Urease reactivation.

patients, giving as the final result of enzymatic hydrolysis, products of low toxicity but with irreversible inactivation of the enzyme.

A new class of compound to be tried is represented by derivatives of ethyleneurea (imidazolone). 2-imidazolidone-2(N,N'-ethyleneurea) acting as a suicide in-

hibitor would represent a sort of "Trojan horse" because of its similarity to the substrate. The final products after hydrolysis would be carbon dioxide and ethylenediamine which would chelate the nickel ions.

Substitution of one NH group for an oxygen in the molecule might change the mode of binding in which the O in position 1 would bind to Ni-1 and the NH2 to Ni-2 while the "one" oxygen would replace the bridging hydroxide possibly blocking the reaction.

Further work must be done either from the structural point of view with new enzyme-inhibitor complexes or with computer modelling to simulate possible mode of binding and suggest modification to the molecules in order to improve their characteristics. They can be then screened eventually against the enzyme for inhibitory effects and finally *in vivo* against whole cell on organism targets.

4.7 Conclusions

"*Repetita juvant*"

The highly conserved urease sequences from the most primitive bacterium to the most evolved plant suggests an evolutionary pathway from a common ancestor. The essential features of the active site have been preserved but the quaternary structure has changed in terms of fusion of subunits and increase in overall size. The importance of urease, even though often forgotten, is enormous due to the influence on human and animal health, on ruminant's alimentation and on nitrogen turnover.

The central object of study of this thesis has provided new information about the structure and function relationship of *B. pasteurii* urease. The *B. pasteurii* urease crystal structure shows no difference in quaternary structure compared to *K. aerogenes* urease, apart from a 12 residues long helix at the C-term of the β subunit. In the native structure four waters are clustered in a tetrahedrally shaped arrangement between Ni-1 and Ni-2. One water bridges between them, while the other two bind in a monodentate way to Ni-1 and Ni-2. The nickel ions are five (Ni-1) and six coordinated (Ni-2). The fourth water is hydrogen bound with the other three.

In the BME complex two BME molecules are involved in the inhibition mechanism. One of them is bound to the nickel centre with its sulphur bridging between Ni-1 and Ni-2. Its OH group is involved in a hydrogen bond with Gly α280 carbonyl oxygen which makes it prone for coordinating Ni-1 in a chelating mode. The other BME molecule binds to Cys α322 (corresponding to Cys α319 in *K. aerogenes*) with a mixed disulphide while making a hydrogen

bond with Ala α366 carbonyl oxygen *via* its OH functionality [105]. The second thiol inhibitor, MEA bridges again between Ni-1 and Ni-2 with its sulphur atom. Its amino group is not visible because of the lack of electron density perhaps due to the relatively low completness of the data (84%). The DAP inhibited enzyme provides the first evidence for a tetrahedral intermediate of the reaction to be located between the two nickel ions [91]. In this structure the selectivity of binding is achieved *via* a network of hydrogen bonds. A consistent alignment of donors on one side and acceptors on the other suggests the mode of binding of urea in the active site during hydrolysis. The AHA inhibited structure confirmed what was already seen for the *K. aerogenes* Cα319A mutant but provided a more accurate model with a complex between inhibitor and wild-type enzyme. A second tetrahedral competitive inhibitor, $H_2PO_4^-$, again confirms the likelihood for the intermediate of the reaction to be located between the two nickels.

The presence of nickel, unusual for an hydrolytic enzyme, instead of zinc is explained by the requirement for a higher affinity for nitrogen ligands. The higher coordination provided by nickel with respect to zinc makes nickel preferred in urease because of the multiple binding sites required. With the possibility of being hexacoordinated nickel offers a stronger binding of the metal to the protein, compared with the mostly tetracoordinated zinc, while sufficient binding sites remain to coordinate waters and urea. Of the two nickel ions in the urease active site Ni-1 is the most electrophilic. Being coordinated to three protein ligands (His α249, His α275 and Lys α220*) Ni-1 offers one binding site more than Ni-2 which is coordinated to four protein ligands (His α137, His α139, Lys α220* and Asp α363). The chemistry of the two nickels is further modulated by the bridging carbonate of the carbamylated Lys α220*.

The resonance form of the carbamate develops a higher negative charge on the two carbonate oxygens, that in turn attract strongly the two Ni^{2+} ions. The influence of the carbamate on the two nickels is probably important also for the catalysis.

The specificity of the active site in binding substrate/inhibitors is due mainly to the very small cavity lined with a very specific hydrogen bond network. In the native enzyme the cluster of four waters found in the active site has the shape and the volume of the proposed tetrahedral reaction intermediate suggesting for the first time the possibility for the hydrolysis to occur between the two nickel ions both involved in urea binding/activation.

The comparison between the native and the DAP inhibited enzyme has given information about the enzyme in the resting state and with a transition state analogue bound into the active site. In DAP inhibited urease the catalytic His α323 is moved from a "free" state position, with the flap in the open conformation, to a "bound" state in which His α323 Nϵ2 establishes a hydrogen bond with the distal DAP NH_2 group. In this bound state the flap is in the closed conformation. During the movement of the flap from the open to the closed conformation His α323 moves by about 5 Å, strongly indicating the importance of the position of the flap for the reaction or, as in this case, for the inhibition. The information strongly supports a reaction mechanism in which the bridging hydroxide its involved in the reaction as a nucleophile in the attack on the urea carbon.

The structural and functional similarity with other amidohydrolases and related metallo hydrolases suggests a common catalytic mechanism and the pro-

posed urease reaction mechanism may represent a general model for this class of enzymes.

The extensive study on complexes between urease and inhibitors belonging to different chemical classes gave information about their different modes of binding to the metal centre and the residues involved in the reaction/inhibition mechanism. The information provided could be used for designing new inhibitors which are more efficient and less health damaging than those currently available. The knowledge of the residues involved in inhibitor binding is fundamental for structure based drug design. When structure and function relationships are known the information provided can be succesfully used for a rational design [106]. New compounds could be synthesised that resemble the known inhibitors but modified in order to exploit all the binding sites present in the vicinity of the metal centre. The new structure based inhibitors could combine the best efficiency and the lowest health and environment damage.

Part II

Structure and Function Relationship of Cytochrome *c*-553 from *Bacillus pasteurii* at 0.97 Å Resolution

Chapter 5

Cytochrome *c*-553 from *B. pasteurii*, an Electron Transfer Protein: Background and Material and Methods

5.1 The Respiratory Chain of Alkaliphilic Bacteria

Alkaliphilic bacteria are characterised by their ability to live at extremely high values of pH such as 11.4 as shown for the facultative akaliphile *Bacillus firmus* OF4. The optimal pH for growth is higher for obligate strains than for facultative strains but they are more sensitive to changes in pH with slower growth rate at pH values lower than the optimal [107]. Alkaliphiles need to maintain a cytoplasmatic pH more acidic than the external one by over 2 units in order to survive. Despite this requirement decreases the proton electrochemical gradient for oxidative phosphorylation no differences in respiration rates are observed between alkaliphiles and their own non-alkaliphilic mutants [108].

Since *Bacillus* species are Gram-positive and lack a true periplasmic space the respiratory chain for such alkaliphilic bacteria is localised on the cellular

membrane. Electron transfer and ATP synthesis take place on the cytoplasmic membrane and no other organelles or vesicles represent adaptation to the extremely high growth pH. A much higher cytochrome content for alkaliphiles when compared to normal bacteria or to their own non-alkaliphilic mutant derivatives has been observed and proposed as an adaptation to the growth conditions. Such overexpression of electron transfer proteins is required to facilitate an efficient energy transduction [108]. Facultative alkaliphilic bacteria have different contents of cytochromes when grown at different pH, the level of cytochromes increasing from neutral pH to very alkaline pH. A scheme of the respiratory chain as proposed for *B. firmus* OF4 [107] is given in Figure 5.1. In membranes of alkaliphilic bacteria a number of proteins involved in the res-

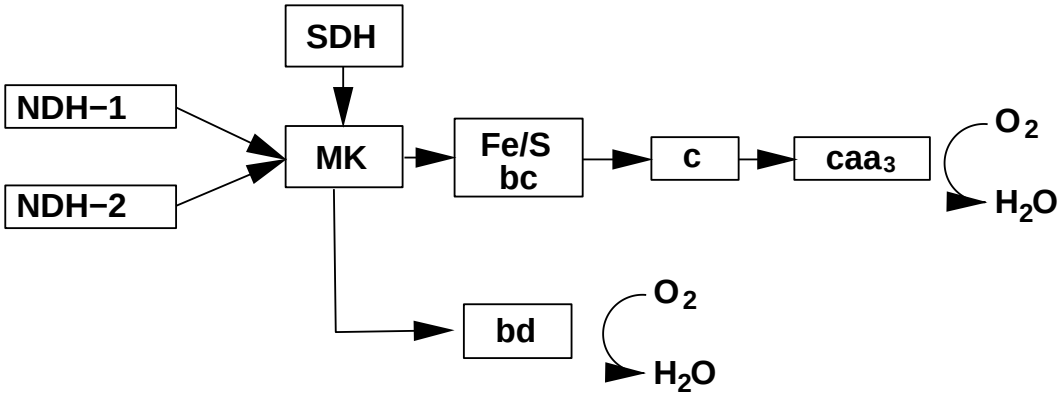


Figure 5.1: The Respiratory Chain of Alkaliphilic Bacteria. The proposed *B. firmus* OF4 model [107]. NDH-1 proton traslocating dehydrogenase, NDH-2 nonproton traslocating dehydrogenase, SDH succinate dehydrogenase, MK menaquinone, Fe/S and *bc* cytochrome *bc* complex containing the Rieske Fe/S centre, *bd* cytochrome *bd* terminal oxidase, *c* cytochrome *c*, caa_3 cytochrome caa_3 terminal oxidase. It is still unknown whether the cytochrome *c* is required for the electron transfer between the *bc* complex and the terminal caa_3 oxidase.

piration has been detected. NADH dehydrogenases, succinate dehydrogenase, menaquinone, cytochrome *bc* complex containing a Rieske (2Fe-2S) centre and two terminal oxidases that are expressed under particular growth conditions.

A terminal cytochrome *bd* oxidase is expressed in *B. firmus* cells grown exponentially at pH 7.5 while in alkaline growth conditions the normal cytochrome caa_3 terminal oxidase is produced. It is still unkown if the trasnfer of the electrons from the *bc* complex to the cytochrome caa_3 terminal oxidase requires the presence of a cytochrome *c*, which is always found in large quantities in alkaliphiles or if the electrons are directly transferred from one partner to the other.

5.1.1 Cytochromes *c* in Alkaliphilic Bacteria

The most striking characteristics of cytochromes *c* isolated from alkaliphilic bacteria are the relatively low midpoint redox potential and the related low isoelectric point. The low potential is supposed to facilitate the transfer of electrons from the outside to the inside of the cytoplasmatic membrane [107]. The values of the redox potential are grouped between about +50 mV and +100 mV while for non alkaliphilic bacteria they lie between about +180 mV and +250 mV.

All cytochromes of the *c*-type found in alkaliphiles are supposed to be membrane-bound since expressed in Gram-positive organisms lacking a true periplasmic space. The soluble forms of this proteins are usually cleaved forms of the bound ones anchored to the membrane with transmembrane-spanning segments or simply *via* a diacyl-glyceryl-cystein modified N-terminus.

5.2 The Respiratory Chain in *B. pasteurii*

This microorganism grows at an optimal pH of 9.2 in the presence of ammonium salts or urea readily hydrolysed to ammonia by a cytoplasmatic urease.

The growth conditions seems to influence the quantity of cytochromes in the cellular membrane. Cells grown in the presence of ammonium salts apparently give higher yield of haem proteins than cells grown in the presence of urea [109, 110]. Haddock *et al.* determined the presence of a set of membrane associated cytochromes by low-temperature reduced-minus-oxidised UV-Vis absorption spectroscopy on whole cells. At least five cytochromes were identified based on their UV-Vis absorption maxima at wavelengths of 548, 552, 556, 562 and 600 nm but the presence of additional haem proteins could not be excluded. The cytochrome components of the *B. pasteurii* respiratory chain were found to be similar to those found in other alkaliphilic bacilli with cytochromes of the *b, c* and *a* type [109].

The high external pH required by *B. pasteurii* for growth (optimal pH≈9.2) requires the creation of a gradient between the cytoplasmic pH (≈7.5) and the growth medium of about 1.5-2 units. Despite the lack of a primary Na^+ pumping activity, respiration occurs with normal extrusion of H^+ with high H^+/O ratio. The lack of primary Na^+ pumps might facilitate pH homeostasis [107]. The mechanism by which respiration energises oxidative phosphorylation in bacteria growing at high pH is of considerable interest because it does not conform to normal chemiosmotic models. For *B. pasteurii* the same bioenergetic problems are probably solved as demonstrated in *B. firmus* OF4 by pH dependence studies. Increasing the pH of the growth medium from 7.5 to 11.2 results in a decrease of the protonmotive force (Δp, expressed in mV) until it reaches a minimum value. The phosphorylation potential (ΔG_p) over the same range of pH slighlty increases.

The final result is a $\Delta G_p/\Delta p$ ratio increased for high values of pH. Typical val-

ues of the $\Delta G_p/\Delta p$ ratio are in the range of 3-4 for other bioenergetic systems, for *B. firmus* OF4 the $\Delta G_p/\Delta p$ ratio is increased from 3 at neutral pH to 13 at basic pH. The increased capacity for oxidative phosporylation at pH higher than 9 is correlated to an increased concentration of the caa_3 complex. Cells grown at pH 10.5 have a 2-3 fold higher concentration of the caa_3 complex than cells grown at pH 7.5, with a molar excess over the F_1F_0-ATP synthase content of 2-4 fold [107]. The increased concentration of the caa_3 complex also explains the requirement for a general increase of the cytochrome content in the cytoplasmic membrane of alkaliphilic bacteria [108]. Recently, a coupling mechanism of the ATP-generating system has been proposed to operate in *B. pasteurii* in association with urea hydrolysis. The alkalinisation of the cytoplasm would activate the ATPase and the flux of ammonium ions increase the membrane potential $\Delta\psi$ [111, 112].

5.3 *B. pasteurii* cytochrome *c*-553

During the purification of urease from *B. pasteurii* a soluble form of a *c* type cytochrome was identified and purified. The protein shows peaks in the UV-Vis oxidised spectrum at 552, 524 and 411 nm and in the reduced spectrum at 553, 523 and 416 nm. This cytochrome has been called *c*-553 because of the α absorption band present in the reduced spectrum at 553 nm [111].

The analysis of the sequence has revealed that the pure protein was a mixture of three isoforms with different aminoacidic chain length. This evidence supports the belief that the soluble form of the protein is the cleaved form of a membrane bound cytochrome as usually found in alkaliphilic Gram-positive bacteria. The sequence reveals that *c*-553 has the characteristic of a lypoprotein bound to

the cytoplasmic membrane through a N-terminal diacyl-glyceryl-cysteine anchor. The stretch of residues responsible for the anchoring is easily hydrolysed upon cell breakage [113]. Cytochrome *c*-553 has a chain length of 92 amino acids with the typical haem binding motif Cys-X-X-His of a class I cytochrome.

Entire *B. pasteurii c*-553 aminoacid sequence with indicated in bold the residues responsible for the haem binding and the iron axial ligands:

Gly Gly Gly Asn Asp Thr Ser Asn Glu Thr Asp Thr Gly Thr Ser Gly Gly Glu Thr Ala Ala Val Asp Ala Glu Ala Val Val Gln Gln Lys **Cys** Ile Ser **Cys** **His** Gly Gly Asp Leu Thr Gly Ala Ser Ala Pro Ala Ile Asp Lys Ala Gly Ala Asn Tyr Ser Glu Glu Glu Ile Leu Asp Ile Ile Leu Asn Gly Gln Gly Gly **Met** Pro Gly Gly Ile Ala Lys Gly Ala Glu Ala Glu Ala Val Ala Ala Trp Leu Ala Glu Lys Lys

The two axial ligand as shown by NMR spectroscopy are Met and His [114] located 35 residues apart as in subgroup S differing in this from the subgroup L in which 50 residues lie between the two axial ligands. Direct cyclic voltammetry has been used to determine the electrochemical properties of cytochrome *c*-553 revealing a very low reduction potential of +47 mV [114]. The low potential is consistent with the observation that cytochromes from alkaliphilic bacteria have low reduction potentials associated with the large negative membrane potentials. This feature has been proposed to facilitate electron transfer to the terminal oxidase of the respiratory chain [107].

The structural study of *B. pasteurii* cytochrome *c*-553 by protein crystallography is the subject of the following chapters.

5.4 Protein Purification

Bacillus pasteurii (DSM, type strain 33), was grown as previously described [35]. The cells were harvested by centrifugation at 30,000 g for 15 minutes and washed three times with 50 mM phosphate buffer pH 7.5 containing 1 mM EDTA (ethylenediaminotetraacetic acid), to chelate free metal ions and inhibit metalloproteinases, and 50 mM Na_2SO_3 (buffer A). About 100 g (wet weight) of cells were resuspended in buffer A and disrupted using a French press operating at 20,000 psi. The extract was first subjected to differential centrifugation at 30,000 g for 30 minutes and then ultracentrifuged at 150,000 g for 2 hours 30 minutes, in order to remove cell wall debris and aggregates.

The crude extract, about 400 mL, was dialysed overnight using a membrane with a molecular weight cut-off of 2,000 Dalton against buffer A. After centrifugation at 30,000 g for 15 minutes, to remove precipitate, the dialysed protein solution was loaded on a Q Sepharose XK 50/20 Pharmacia anionic exchange column, previously equilibrated with buffer A, and washed with buffer A, at a flow rate of 6 mL $\cdot$ min^{-1}, until the elution profile approached 0. A gradient procedure was used to elute the protein with steps of increasing ionic strength: NaCl up to concentrations of 150, 250, 450 mM (in buffer A) to elute with 1 L of solution for each step. Fractions containing c-553 detected in the 150 mM step were pooled and the ionic strength raised to 2 M $(NH_4)_2SO_4$ using solid ammonium sulphate.

After centrifugation at 30,000 g for 15 minutes, to remove precipitate, the solution was loaded onto a Phenyl Sepharose XK 26/20 Pharmacia hydrophobic interaction column equilibrated with 2 M $(NH_4)_2SO_4$. Use of 1 M AMS did not resolve the different isoforms of the protein and the protein eluted together

with the flow through: 2 M AMS tightly bound the protein to the resin and allowed a better separation of the bands. The column was washed extensively with conditioning buffer and the protein eluted with a linear gradient from 1 M to 0 M at a flow rate of 3 mL $\cdot$ min^{-1}. Cytochrome *c*-553, eluted with 1 M $(NH_4)_2SO_4$, was concentrated using an Amicon ultrafiltration cell with a membrane of 3,000 Dalton molecular weight cut-off and loaded onto a Superdex S75 XK 26/60 Pharmacia gel filtration column equilibrated with buffer A and eluted at a flow rate of 1 mL $\cdot$ min^{-1}. Fractions containing *c*-553 from the last step were concentrated to 16 mg mL$-$1 and stored in buffer A at 4°C for later use [111]. Figure 5.2 shows a flowchart of the purification.

Isoelectrofocusing (IEF) was performed with a Multiphor Pharmacia horizontal electrophoresis cell at 25°C and precast 7.5% acrylamide gels: the estimated pI was 3.3. Native and SDS gel electrophoresis were performed using 4-20% polyacrylamide gradient gels. When the gels were stained with Coommassie R250 Blue no bands were detected because of the low affinity of *c*-553 for the dye. A single band was detected in the native and in the SDS gel when Silver-stain was used. The molecular weight for the protein as determined by SDS gel electrophoresis was estimated to be about 9.5 kDa using the Bio-Rad broad-range standards. Native molecular mass determination was performed by comparing the elution profile of proteins of known molecular weight onto a gel filtration column. Standard proteins (albumin 66.2 kDa, ovalbumin 45 kDa, chymotrypsinogen 25 kDa, ribonuclease 13.7 kDa and *Rubrivivax gelatinosus* high-potential iron sulphur protein 7.9 kDa) and cytochrome *c*-553 were loaded sequentially onto an FPLC Superdex 75 HR 10/30 column and eluted with 50 mM phosphate buffer pH 7.5 containing 150 mM NaCl in order to avoid non-specific interactions with the matrix. The apparent molecular weight in

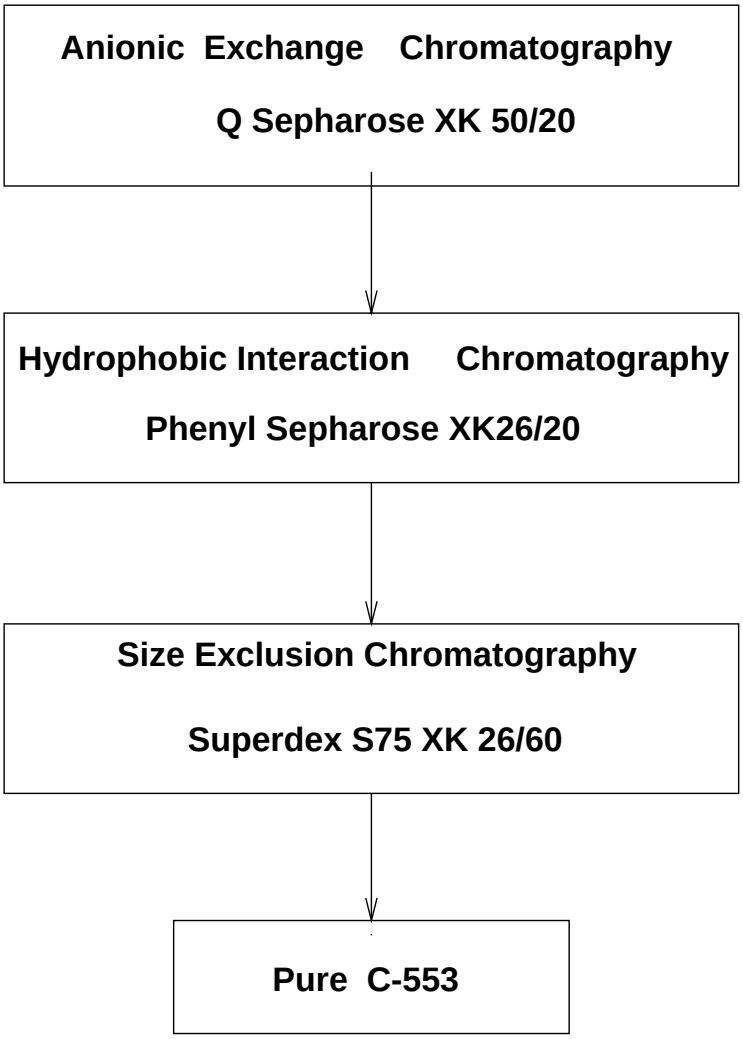


Figure 5.2: Purification flowchart of *B. pasteurii* cytochrome *c*-553.

solution was 19.5 kDa suggesting dimerisation of the protein in solution. This dimerisation might be due to the pairing of hydrophobic patches located on the protein surface.

5.5 Protein Crystallisation

For crystallisation the protein solution was thoroughly exchanged with 20 mM Tris-HCl pH 8.0 in a Centricon ultrafiltration unit with membrane of 3,000 Dalton molecular weight cut-off giving a final concentration of 16 mg/mL. The hanging drop method was used for all the crystallisation trials. Drops of 1 μL of 16 mg/mL protein solution were mixed with the precipitant and equilibrated by vapour diffusion against 1 mL precipitant solution in 24-well Linbro plates. An extended screening was performed at two different temperatures, 4 and 20°C, by a sparse matrix method [62] and different concentrations of ammonium sulphate (AMS), polyethylene glycol (PEG) 6000 and PEG 6000 plus 1 M LiCl at various pH. The buffers used to obtain the different pHs were 100 mM sodium acetate at pH 5.0, 100 mM sodium citrate at pH 6.3, 100 mM sodium citrate at pH 7.0, 100 mM Tris-HCl at pH 8.0, 100 mM Tris-HCl at pH 9.0. The concentration of precipitants ranged between 1.6 M and 3.2 M for AMS, from 15 to 30% for PEG 6000 and from 15 to 30% of PEG 6000 and from 0.5 to 1 M of LiCl for PEG 6000 plus LiCl. Succesful and reproducible crystallisation conditions were obtained when using 3.2 M AMS in 100 mM sodium acetate at pH 5.0 at 20°C. Under these conditions, regularly shaped rods of dimensions about $0.2 \times 0.2 \times 0.6$ mm^3 on average grew as stars from single nucleation points in the drops in about five days.

5.6 Collecting Diffraction Data

Although crystals of cytochrome *c*-553 did not show any apparent radiation damage when a preliminary 1.5 Å room temperature dataset was collected, the high resolution study was performed at 100 K on a vitrified single crystal to reduce thermal motion and increase data quality and resolution. Data were

collected using the synchrotron radiation from the BW7B wiggler line [63, 64] at the EMBL outstation Hamburg (Germany).

For cryoprotection, the crystal was soaked for a short time in a solution containing 20% glycerol in the mother liquor. The crystals were fished out with a nylon loop (Hampton Research) and rapidly exposed to a cold nitrogen stream (Oxford Cryosystems Cryostream). The dimensions of the crystal were about $0.3 \times 0.3 \times 0.6$ mm^3. Diffraction images were recorded using a 30-cm Mar Research imaging plate scanner in three sweeps at different exposure time, crystal to detector distance and plate size in order to measure accurately both high and low resolution diffraction intensities and obtain higher redundancy and good I/σ ratio. A representative diffraction image is shown in Figure 5.3. The Bravais lattice was identified using the autoindexing procedure implemented in the program DENZO [65, 66] as primitive orthorombic $P222$ with cell dimensions of about $a = 37.14$ Å, $b = 39.42$ Å, $c = 44.02$ Å, with the $c =$ dimension coincident with the long crystal axis. The correct space group was identified as $P2_12_12_1$ by looking at the systematic absences. Crystals were mounted in the cryo-loop with the c axis nearly parallel to the spindle axis which minimised overlaps and allowed data collection to be completed in 90° of rotation. To avoid overlaps data were collected in ϕ slices of 0.8° oscillation, for the high resolution pass (2.5-097 Å), of 1.6° for the medium resolution pass (6.9-1.46 Å) and of 3° for the low resolution pass (20.0-2.65 Å).

Data were processed using the program DENZO, scaled, merged and postrefined with SCALEPACK [65, 66]. Table 5.1 reports a summary of data collection statistics and reduction. The quality of the data can be judged by looking at Figure 5.4 which shows a Wilson plot and a I/σ *vs.* resolution plot.

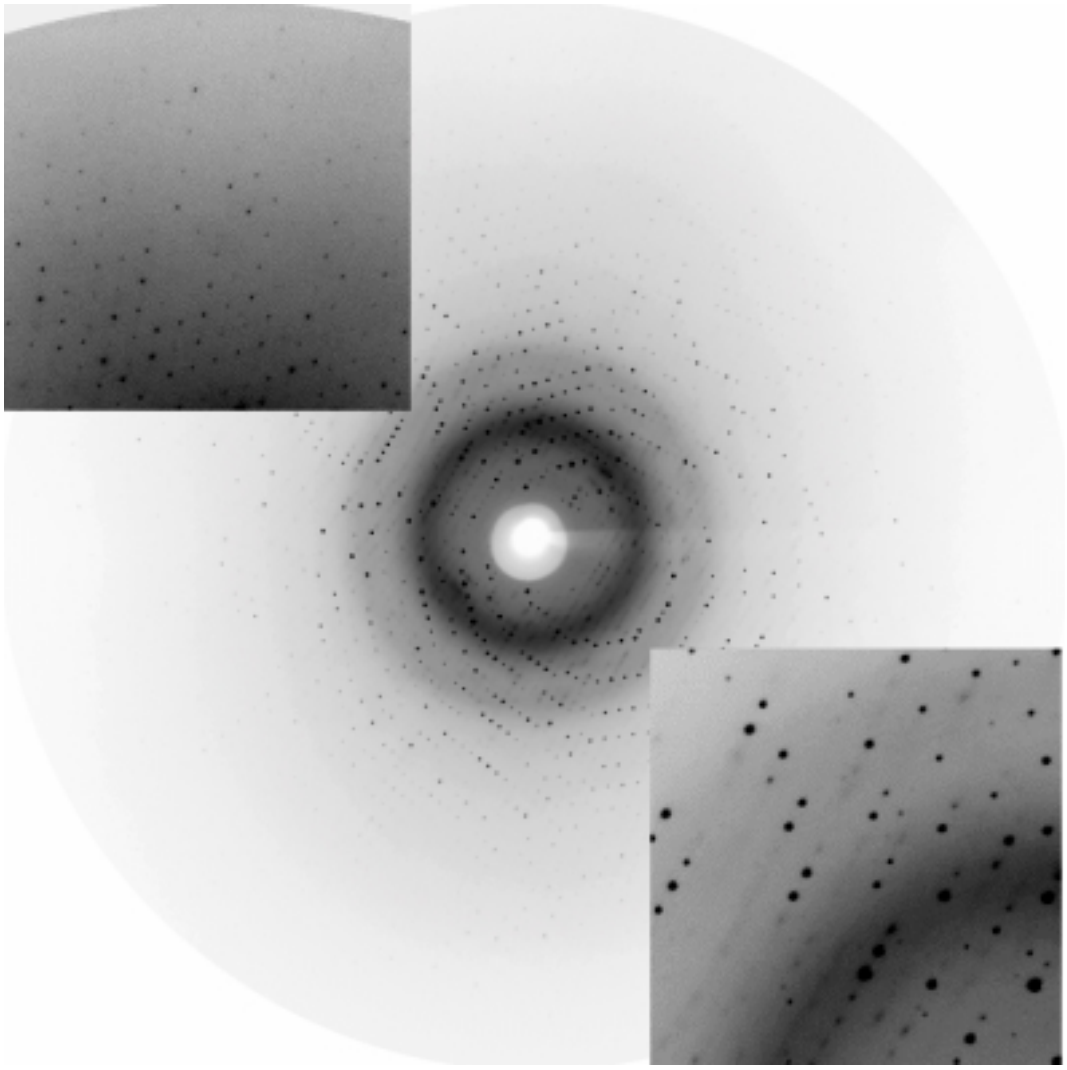


Figure 5.3: Diffraction image from a crystal of *B. pasteurii* cytochrome *c*-553 between 0.97 Å and 20 Å. The insets are magnification of the pattern, at the top the high resolution data, at the bottom the low resolution data.

The peculiar shape of the curve in the I/σ *vs.* resolution plot reflects the fact that the data were collected in three sweeps. The high I/σ ratio in the last resolution bin (0.99-0.97 Å) indicates that the maximum resolution was not reached and the crystal would have diffracted to higher resolution. Mechanical constraints did not allow the detector to be moved closer than 120 mm thus

Space group	$P2_12_12_1$
Unit cell parameters:	
a (Å)	37.14
b (Å)	39.42
c (Å)	44.02
Beam line at DORIS	BW7B
Resolution(Å)	20.0-0.97
Wavelength(Å)	0.885
Number of images	179
Oscillation range (°)	0.8-3
R_{merge}	0.074
Raw measurement	150,970
Unique reflections	38,956
Redundancy	3.87
% Completeness	99.9
% Greater than 3σ	86.3
I/σ in high res. bin(0.99-0.97)	4.18

Table 5.1: Summary of data collection for cytochrome c-553 from *B. pasteurii*.

limiting the resolution "only" to 0.97 Å. Assuming one molecule of molecular weight 9.5 kDa per asymmetric unit, the volume-to-mass ratio, V_M was 1.7 Å^3/Da, with a solvent content of 28%, indicating tight crystal packing.

5.7 Structure Solution

The position of the iron ion in the protein was identified from the anomalous Patterson map. Although data were collected at a wavelength far from the

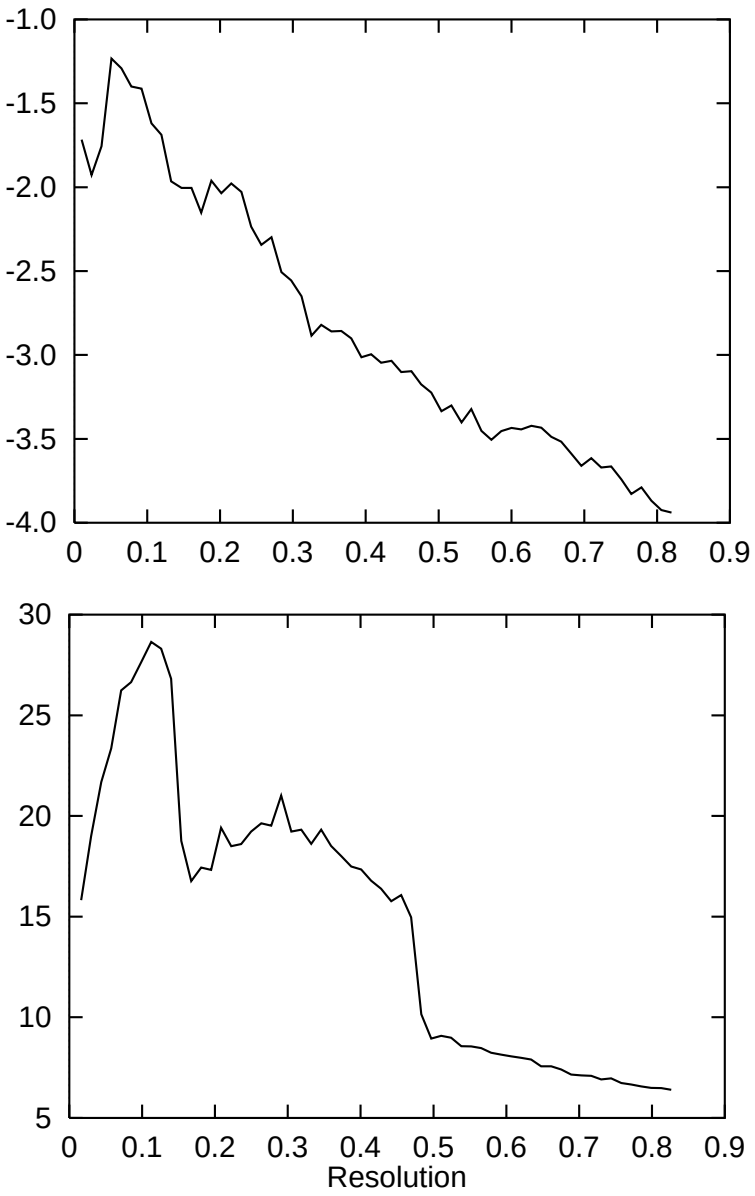


Figure 5.4: Wilson (top) and I/σ (bottom) plots of *B. pasteurii* cytochrome *c*-553. The resolution is reported as $1/d^2$. In the Wilson plot the 'Y' axis is: $ln\frac{\overline{I_{(hkl)}}}{\sum_i (f_i^0)^2}$

iron absorption edge (≈1.7 Å depending on the chemical environment), but optimised to collect atomic resolution data and not to maximise the anomalous

signal, the Patterson map calculated with coefficients $[F(+) - F(-)]^2$ clearly showed a peak at the 8 σ level, corresponding to the anomalous signal from the single iron atom. Figure 5.5 shows the $\omega = 1/2$ Harker section of the anomalous Patterson map calculated using all data between 0.97 and 20 Å. The inclusion of all the high resolution data made the localisation of the peak straightforward while decreasing the amount of high resolution data made the anomalous signal weaker and the peak fading away. This phenomenon is due to the fact that the anomalous contribution to diffraction intensity is almost constant from low to high resolution while the normal contribution to scattering is large at low resolution and decreases at high resolution. For this reason the ratio between anomalous and normal scattering increases from low to high diffraction angles and the anomalous signal is more significant at high resolution.

The position of the iron was used in an iterative procedure in which electron density map calculation, refinement of the data against the model and model update were made automatic by coupling the program ARP [75, 76] to the normal refinement programs in a unrestrained mode. The program ARP was used to interpret the electron density maps and find "dummy" water atoms in region of the $mF_{obs} - DF_{calc}$ (maximum likelihood refinement) or $F_{obs} - F_{calc}$ (least-squares refinement) maps in which the electron density was higher than 3σ and to remove atoms from the model if lying in region of the $2mF_{obs} - DF_{calc}$ (maximum likelihodd refinement) or $3F_{obs} - 2F_{calc}$ maps (least-squares refinement) and if located closer than 1.0 Å to each other merged as a single atom. New atoms were added if they were found within 1.1 to 3.3 Å from other atoms in the model. To monitor the progress of refinement and allow cross validation 5% of the reflections were flagged as *free*. First

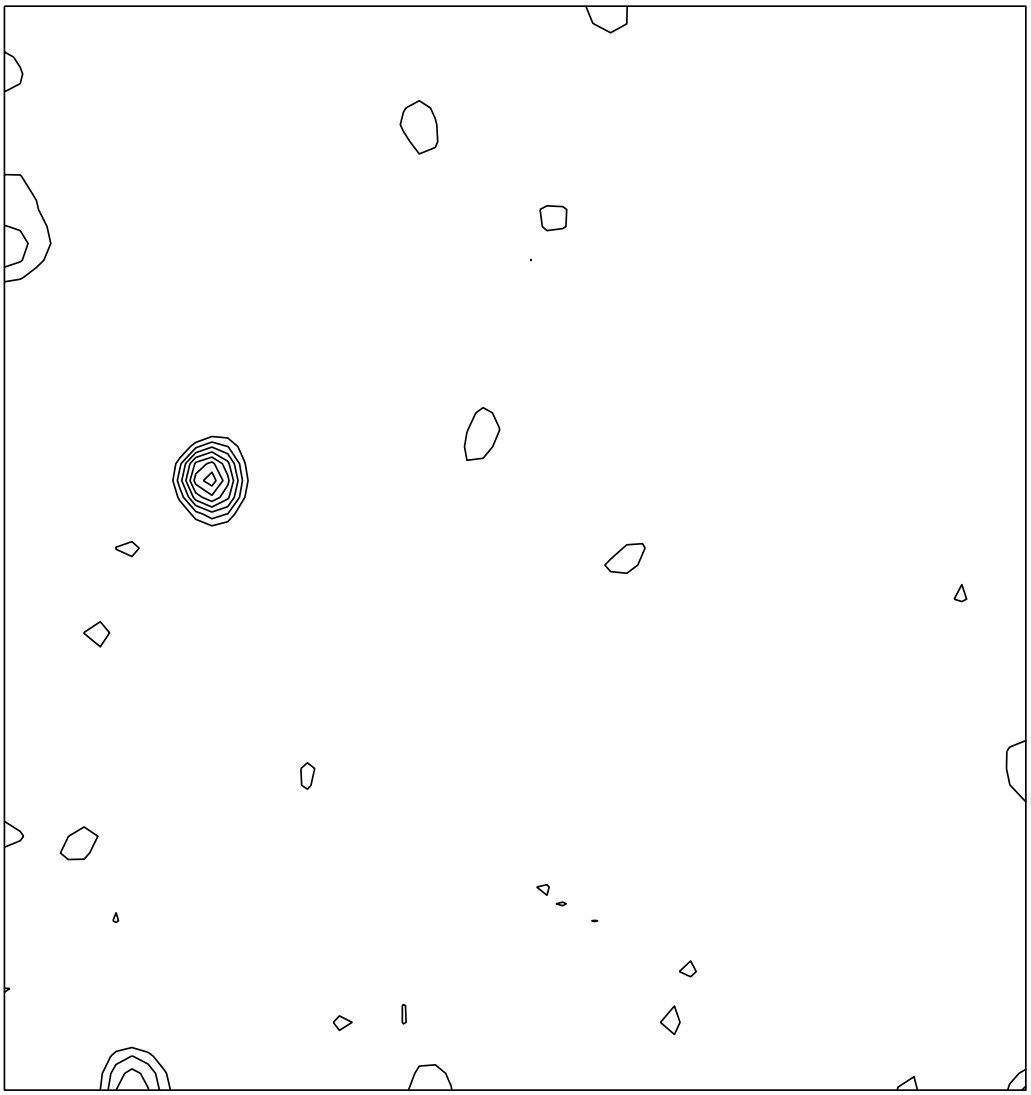


Figure 5.5: $\omega = 1/2$ Harker section of the anomalous Patterson map calculated using all data between 0.97 and 20 Å. The map is contoured at 1σ intervals, starting at 2σ.

attempts at solving the structure involving least-squares refinement (SFALL, PROTIN, PROLSQ [69]) required several hundreds of cycles of refinement and model update with a number of manual interventions using the graphic pro-

gram "O" [71]. A significant improvement in refinement and map quality was observed when maximum likelihood was used, as implemented in REFMAC [70], resulting in faster model building and refinement. The number of atoms

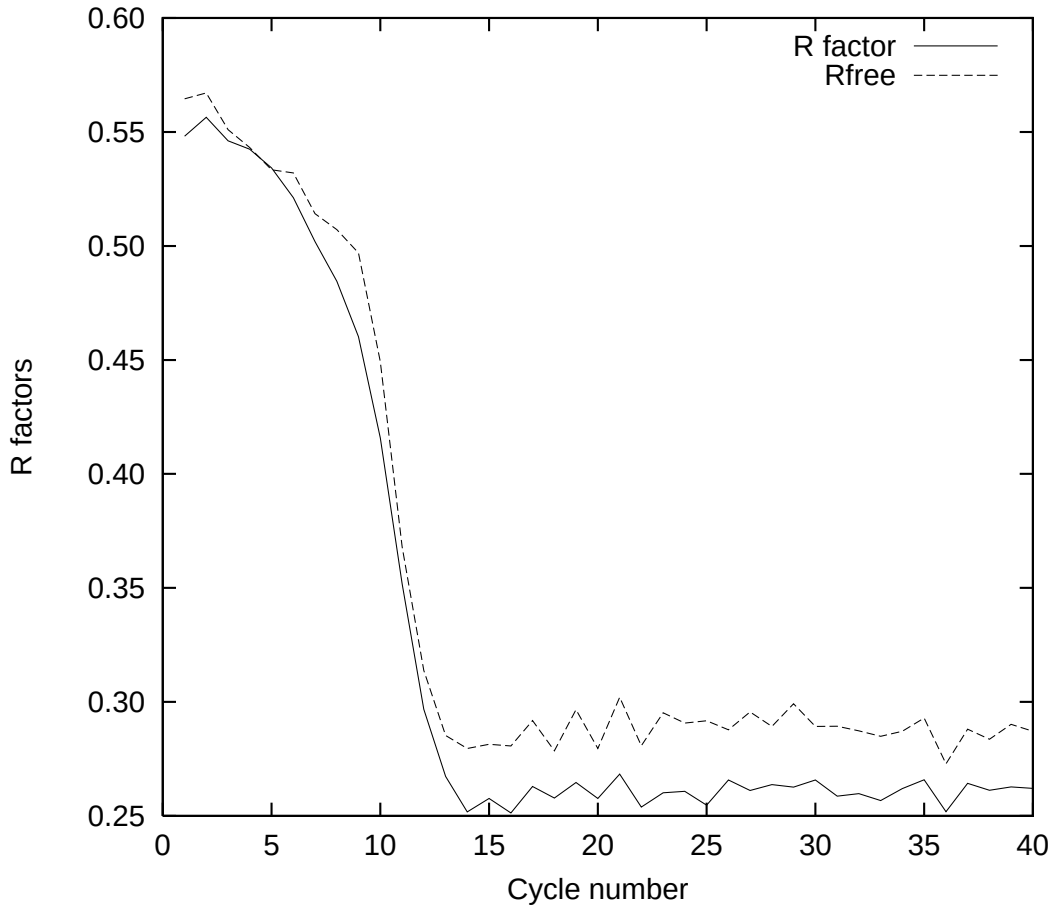


Figure 5.6: R and R_{free} factor plots of *Bacillus pasteurii* cytochrome *c*-553 structure solution using the only iron atom as starting model. The model was built with "dummy" atoms using REFMAC [70] for phase refinement.

to add and to remove per cycle was not limited and the structure was completely built in only thirteen cycles without manual intervention. On a Silicon Graphics Origin 200 4xR12000/270MHz the complete process took less than one hour (55'). The behaviour of the R and R_{free} factors is shown in Figure

5.6 and 5.7.

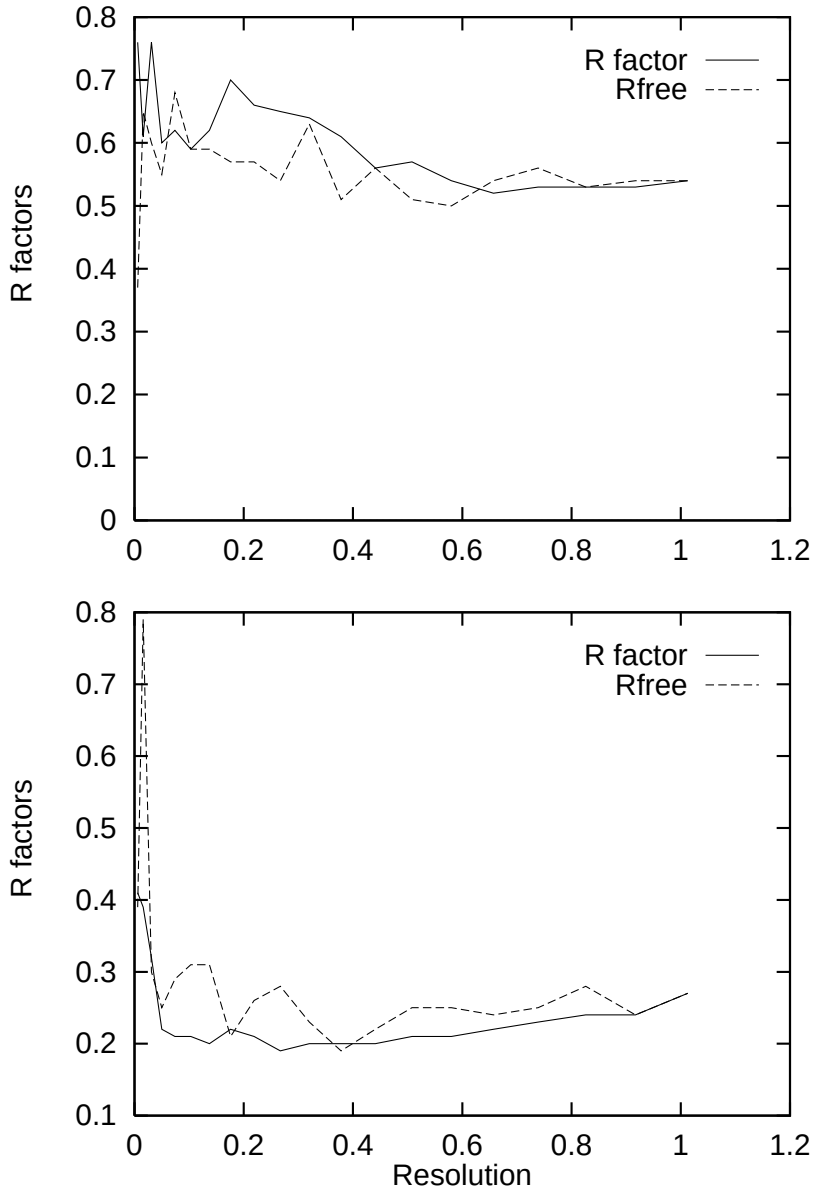


Figure 5.7: R and R_{free} *vs.* resolution plots. Top initial model. Bottom final model. The model was built with "dummy" atoms using REFMAC [70] for phase refinement. The resolution is reported as $1/d^2$.

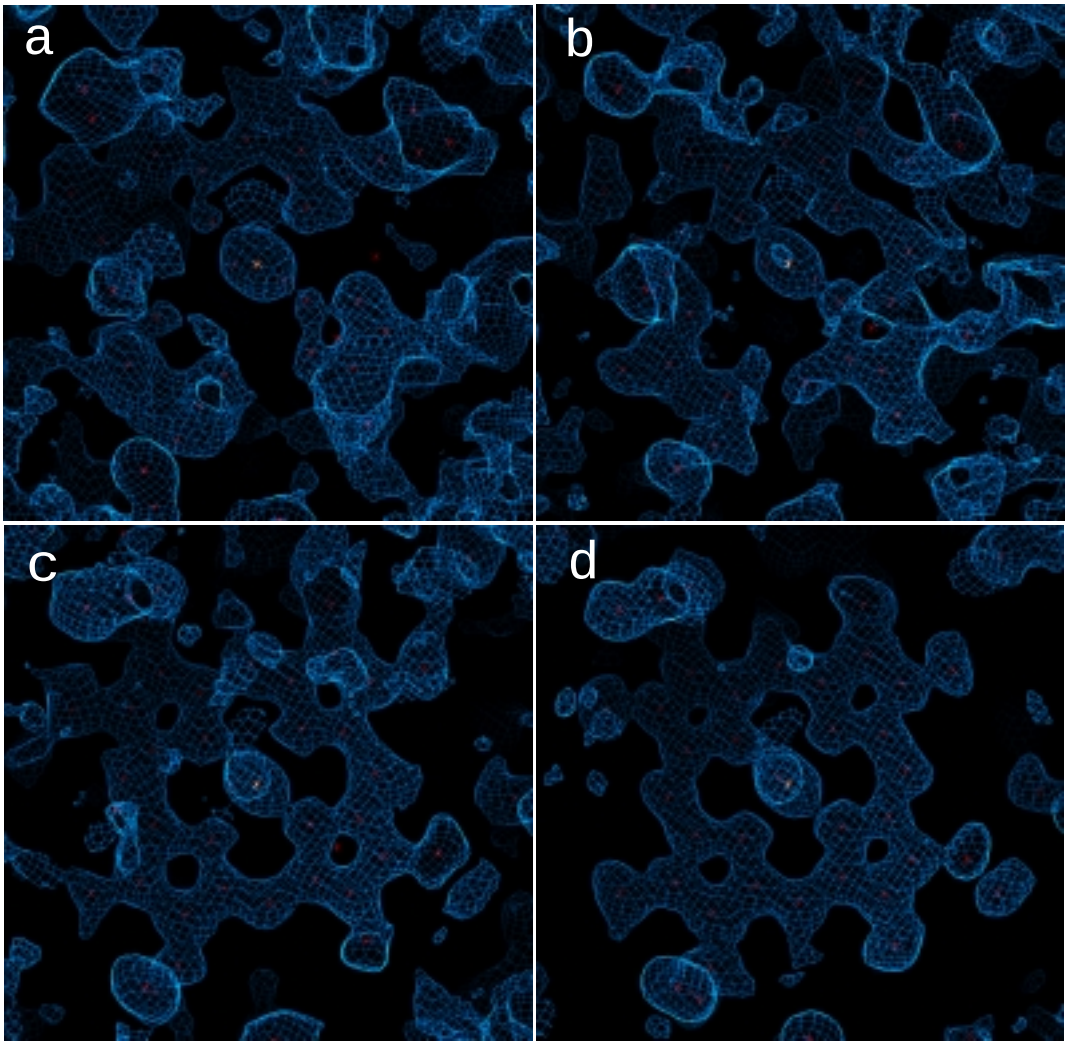


Figure 5.8: Electron density maps and ARP "dummy" atoms model at different stages of refinement. a) after 5 cycles ($R = 52\%$), b) after 7 cycles ($R = 45\%$), c) after 9 cycles ($R = 39\%$) and d) after 13 cycles ($R = 26\%$). Picture made with the package "O" [71].

5.8 Model Building, Refinement and Validation

The good quality of the final maps and the presence of the "dummy" atoms in the position corresponding to the protein atoms helped in tracing the polype-

tide chain of the final model based on the amino acid sequence. A *c* type cytochrome haem was taken from the PDB and modelled in the electron density map. After the model was complete the protein chain was found to be shorter than the sequence obtained by conventional Edman degradation of overlapping peptides and mass spectrometry by 21 residues. Since the molecular weight of the molecule in the asymmetric unit was now 7.7 kDa, the volume-to-mass ratio V_M had to be corrected to 2.1 Å^3/Da, with a solvent content of 41%, in the normal range for proteins [67]. Further refinement with REFMAC and automatic solvent building with ARP were carried out until the R factor reached 18.0% and the R_{free} factor 19.3%.

The atomic resolution of the data allowed anisotropic refinement to be carried out with the program SHELXL96 [115]. Data were prepared with the subroutine SHELXPRO (SHELX interface for protein applications) while the stereochemical parameters for the haem had to be inserted manually since they are not implemented in SHELXPRO library. For the first cycles of refinement the bond distances and angular distances for the haem were restrained to fixed values while in later stage of refinement they were allowed to refine with the only restraints that chemically equivalent bond and angular distances should be equal. Hydrogens were inserted in calculated positions by SHELXL96 automatically and updated at every cycle. The position of the hydrogens of the methyl groups of the haem were allowed to refine their position by searching the electron density map for the best orientation.

The model was refined against diffraction intensity rather than structure factor amplitudes eliminating the problem of truncating the data when negative intensities for weak data were present. SHELXL96 uses a least-squares refine-

ment procedure and the conjugate gradient algorithm. This requires very good quality data, an almost final model and long computing time. After several cycles of isotropic refinement the program was allowed to refine the anisotropic temperature factors for all non-hydrogen atoms. The final statistics are given in Table 5.2. The estimated standard uncertainties (e.s.u.)were calculated

Resolution(Å)	20.0-0.97
Raw measurement	150,970
Unique reflections	38,956
Redundancy	3.87
% Completeness	99.9
Protein atoms	497
Hydrogen atoms	486
Heme atoms	42
Metal ion	1
Solvent atoms	125
Temperature factors (Å ^{2}):	
Overall	14.7
Main chain	8.7
Side chain	13.3
Haem	6.9
solvent	33.4
Fe^{+2}	5.1
B-factor from Wilson plot (Å ^{2})	10.1
RMSD for bond lengths (Å)	0.016
RMSD for bond angles (Å)	0.034
RMSD for chiral volumes (Å)	0.117
R factor	11.5 %

Table 5.2: Summary of the final c-553 model

using the co-variance matrix in a special cycle of block-matrix least-squares minimisation without restraints or damping factors. The model was refined in blocks of 20 residues each, with a one-residue overlap between blocks, in order to decrease the size of computation. The overall estimated standard uncertainties for all atomic position are given in Table 5.3. The structure was well

Overall		0.051
Protein		0.040
	carbon	0.041
	nitrogen	0.039
	oxygen	0.039
	sulphur	0.004
Main chain		0.024
	carbon	0.028
	nitrogen	0.023
	oxygen	0.017
Haem		0.014
	carbon	0.015
	nitrogen	0.010
	oxygen	0.013
	iron	0.002
Solvent		0.085

Table 5.3: Overall estimated standard uncertainties for the final *c*-553 model

ordered with few residues in double conformations or disordered. The multiple conformations were refined keeping the sum of the occupancy constrained to unity. Figure 5.9 shows the isotropic temperature factors for the main chain atoms.

The stereochemistry of the protein was checked using the validation programs PROCHECK [77] and WHATIF [78, 79]. The Ramachandran plot with 89.1% of the residues in the most favored region, 9.1% in the additional allowed region and 1.8% in the generously allowed region is shown in Figure 5.10.

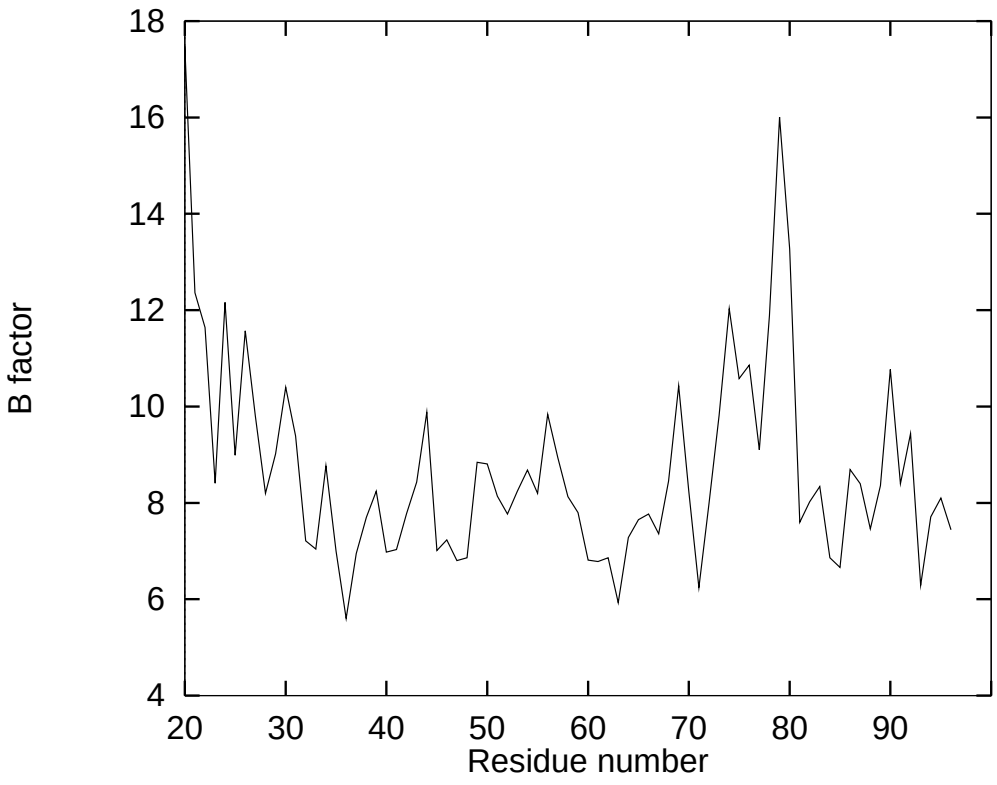


Figure 5.9: B-factor plot for the main chain of cytochrome *c*-553 from *Bacillus pasteurii*.

The final *B. pasteurii* cytochrome *c*-553 model and the diffraction data have been deposited in the Protein Data Bank (PDB) with accession code 1C75.

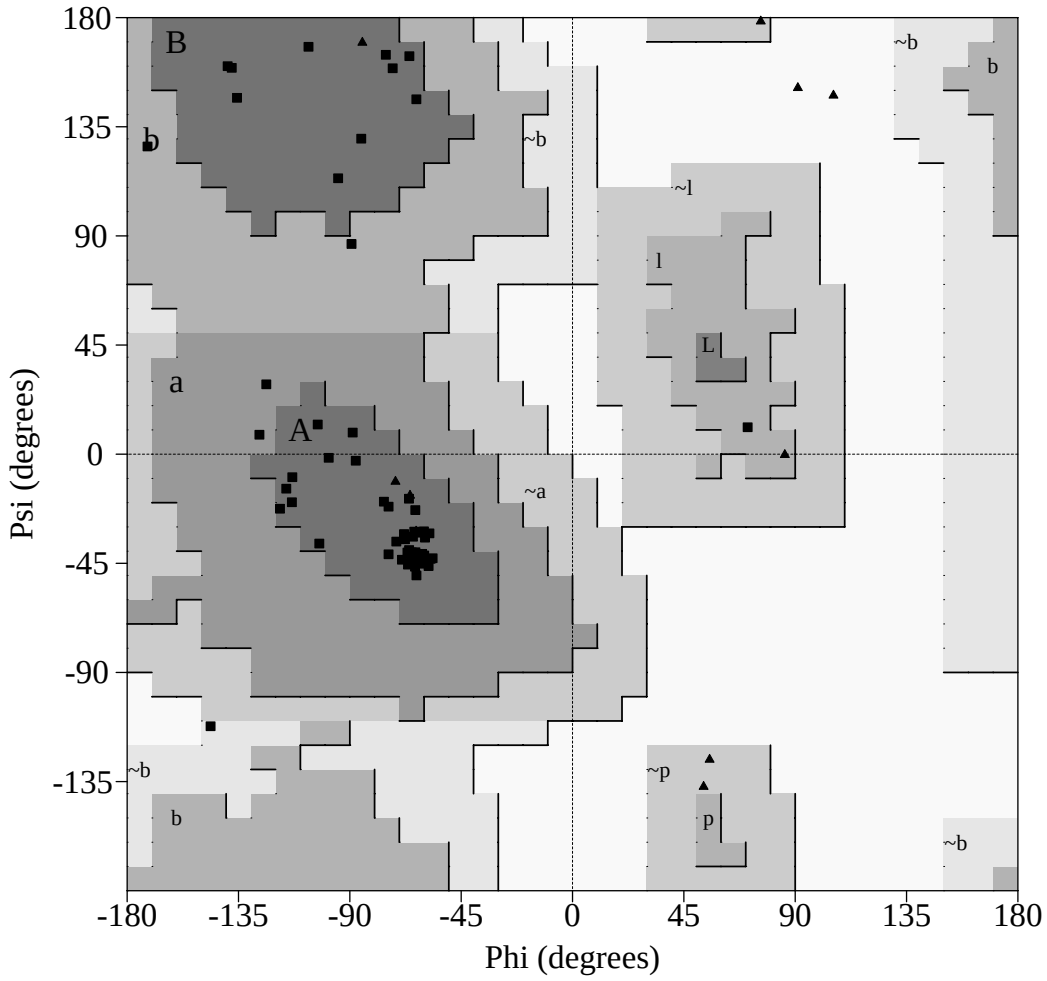


Figure 5.10: Ramachandran plot of cytochrome *c*-553 from *Bacillus pasteurii*.

Chapter 6

Cytochrome *c*-553: Results and Discussion

6.1 The Overall Structure

According to DSSP [82] nomenclature, the compact fold of *B. pasteurii* cytochrome *c*-553 is characterised by three α-helices (47.9% of the total residues), a short 3_{10} helix (4.2%), several turns (25.4%) and random coils (22.6%), surrounding the haem prosthetic group (Figure 6.1). The polypeptide chain is well ordered without flexible or disordered loops.

The residues involved in haem covalent binding, Cys32 and Cys35, as well as the iron axial ligand His36 are located on the N-terminal α-helix which is followed by a Ω-shaped loop composed of a type-I reverse turn (Gly37 - lys40), a sharp turn (Ala43 - Ala45) and a second type-I reverse turn (Ile48 - Ala51). These secondary structure elements make up the first side of the haem pocket. The stabilisation of the sharp turn is achieved by a hydrogen bond between Ala45 NH and Ala45 O, with the Ser44 side chain protruding towards the out-

side of the turn. Ala45 NH is further involved in a hydrogen bond interaction with Cys35 O. This interaction links the loop to the haem group. The short 3_{10} helix following the loop (Ala51 - Ala53) is located before a four-residue β-hairpin (Gly52 - Tyr55) and a longer helix (Glu57 - Asn66). The other

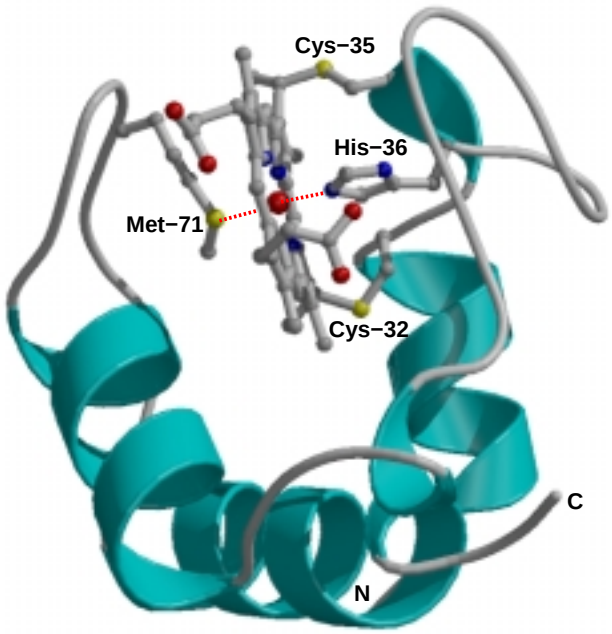


Figure 6.1: Ribbon representation of cytochrome *c*-553 from *Bacillus pasteurii*

side of the haem pocket is formed by a short two stranded antiparallel β-sheet with only two interchain hydrogen bond interactions (Glu68 and Met71), and by a type-II' β-turn. Met71 is also involved in the coordination of the haem iron as the second axial ligand *via* its Sδ. The polypeptide chain continues

from Pro72 to Lys77 without any hydrogen bond interactions with any other residue, being stabilised in its conformation by the interaction with a solvent molecule (W95) which donates two hydrogen bonds to the carbonyl group of Ile64 and Pro72 while receiving a hydrogen bond from Ile75 NH. W95 is the only buried water molecule in the structure being shielded from contact with the bulk solvent by a small hydrophobic pocket formed by the residues from Pro72 to Ala76. These residues connect Met71 to the C-terminal helix (Gly78 - Ala89) which makes close contacts (3.6 - 3.8 Å) with all the other helices through hydrophobic interactions.

In the region of the generously allowed conformation of the Ramachandran plot Ala43 lies with unusual values for its dihedral angles ($\phi = -147°,\ \psi = -112°$). The reason for such a conformation presumably resides in satisfying structural needs like the very tight turn formed by the residues Ala43 - Ala45 part of the loop facing the haem group on the side of His36.

Large deviations from peptide bond planarity (up to 20°) are observed with highly significant values of standard deviations of approximately 1°, in agreement with previous observations of atomic resolution structures [116]. Figure 6.2 shows a plot of the deviation from planarity *vs* residue number.

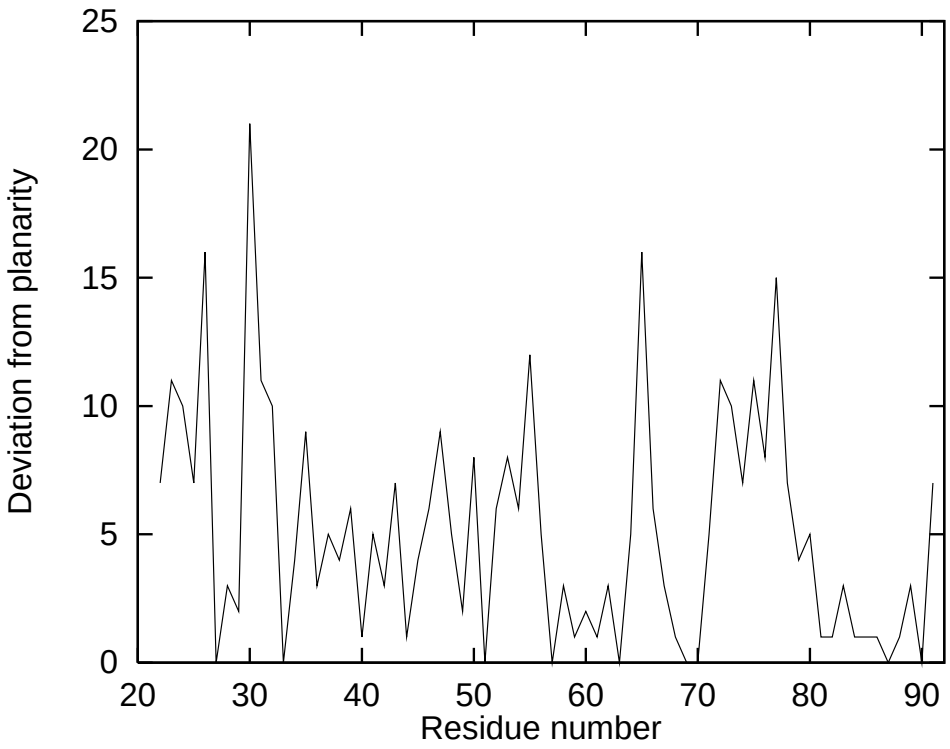


Figure 6.2: Deviation from planarity *vs* residue number (°).

6.2 The Haem Group

The haem group is covalently bound to the protein by two thioether bonds between Cys32 and Cys35 Sγ atoms and the Cα atoms of the haem in positions 2 and 4. The iron atom, almost in the plane of the haem, is coordinated by the four nitrogens of the pyrrole groups while the two axial ligands are His36 Nϵ and Met71 Sδ (see Table 6.1 for relevant distances and angles between the iron and the ligands). Figures 6.3 and 6.4 show the model of the haem superimposed on the electron desity $3F_{obs} - F_{calc}$ map contoured at 3σ. The four nitrogen atoms of the pyrrole rings are in close contact with the iron axial ligands His36 Nϵ (2.86, 2.79, 2.78 and 2.79 Å) and Met71 Sδ (2.93, 3.11, 3.20 and 3.02 Å).

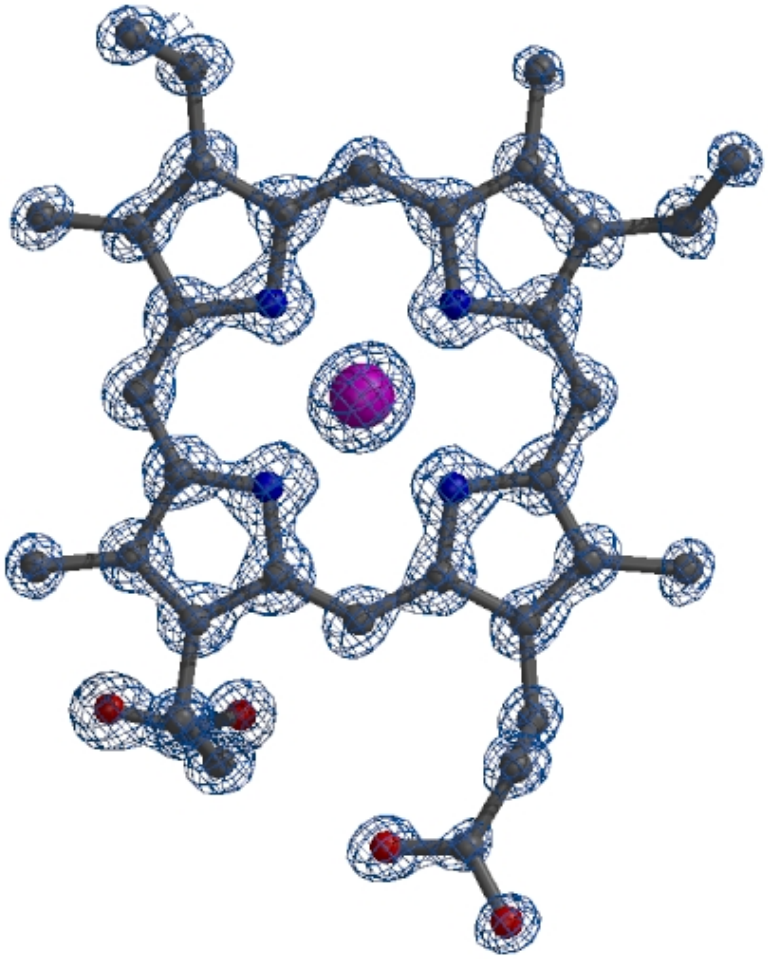


Figure 6.3: Model of cytochrome *c*-553 haem from *Bacillus pasteurii* superimposed on the $3F_{obs} - 2F_{calc}$ electron density map contoured at 3 σ (1 σ = 0.65 electrons/Å^3). Front view.

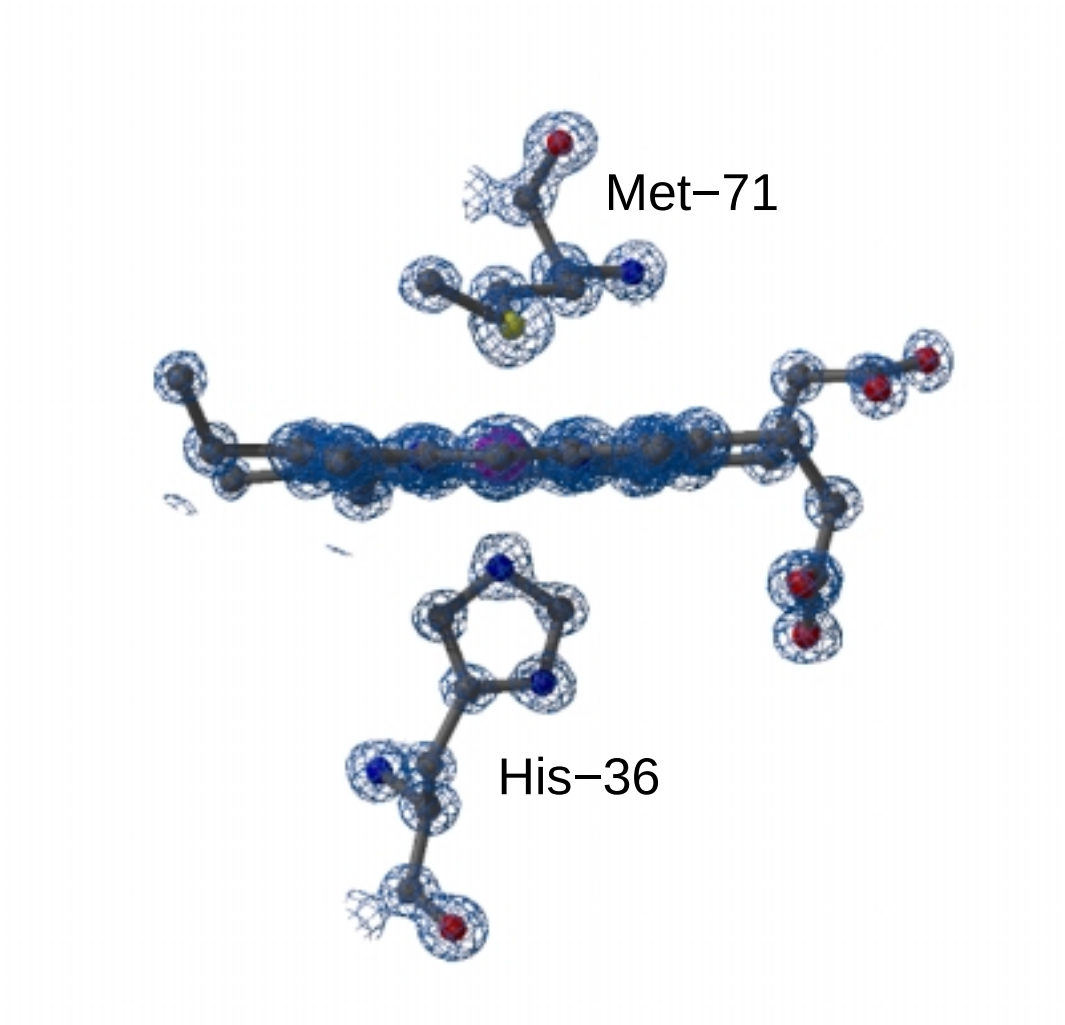


Figure 6.4: Model of cytochrome *c*-553 haem from *Bacillus pasteurii* superimposed on the $3F_{obs} - 2F_{calc}$ electron density map contoured at 3 σ (1 σ = 0.65 electrons/Å^3). Side view.

These distances are shorter than expected by definition of the van der Waals radii, because the nitrogen atoms electron density is reduced by being drawn towards the electrophilic iron atom. As in all known cytochromes, in *c*-553 a hydrophobic pocket hosts the haem prosthetic group. Two walls face the haem, one on the Met71 side composed of the hydrophobic side chains of Ile63, Ile64,

Pro72, Met71 and Ile75 and the other on the His36 side composed of Cys32, Ser34, Cys35, Ala45, Pro46 and Ile48. The pyrrole ring A of the haem group is located in a cavity composed of the hydrophobic side chains of Val84 and Leu88. The other pyrrole rings B, C, D are very exposed to the solvent as shown in Figure 6.1. Although the resolution extends to 0.97 Å no electron density is visible for the hydrogen atoms on the propionate groups whose protonation state can however be deduced based on the observed C-O distances and the hydrogen bonding pattern. Propionate-7 is bent towards the protein core and its Oδ1 forms a strong hydrogen bond donating a hydrogen to Ala47 carbonyl oxygen (at 2.64 Å). The protonation of propionate-7 Oδ1 is confirmed by the length of 1.29 Å of the Cγ - Oδ1 bond, statistically consistent with the canonical value of 1.3 Å for the C-OH length in a neutral carboxylic group [74]. The propionate-7 Cγ - Oδ2 length is 1.25 Å close to the expected value for C=O double bond (1.21 Å) but slightly larger because of the strong hydrogen bond with a solvent molecule (W97 at 2.80 Å). Propionate-6 extends its side chain towards the protein exterior establishing an hydrogen bond between its Oδ1 and Glu82 Oϵ2 of a symmetry related molecule (at 2.50 Å). The lengths Cγ - Oδ1 and Cγ - Oδ2 of propionate-6 are equivalent (1.26 and 1.27 Å) suggesting that the group is deprotonated. The lengths of the symmetry related Glu82 Cδ - Oϵ2 are also equivalent (1.26 and 1.27 Å) suggesting that the proton involved in the hydrogen bond is equally shared between the two oxygen atoms (propionate-6 Oδ1 and Glu82 Oϵ2). Propionate-6 Oδ2 is hydrogen bonded to a solvent molecule (W94at 2.84 Å) situated at the centre of a tetrahedral hydrogen bonding network involving Tyr55 Oη (at 2.72 Å), Gln68 Oϵ1 (at 2.77 Å) and Gly78 N (at 2.87 Å) from a symmetry related molecule. The lone pairs of the water molecule are directed towards Tyr55 Oη (which is protonated) and the symmetry related Gly78 N, while the protons

Iron $\longleftrightarrow$ Ligand	Distance (Å)
Fe $\longleftrightarrow$ His36 Nϵ	1.99
Fe $\longleftrightarrow$ Met71 Sδ	2.33
Fe $\longleftrightarrow$ N_A	2.00
Fe $\longleftrightarrow$ N_B	1.98
Fe $\longleftrightarrow$ N_C	1.99
Fe $\longleftrightarrow$ N_D	1.97
Ligand-Iron-Ligand	**Angles (°)**
N_A-Fe-N_B	90.2
N_B-Fe-N_C	89.4
N_C-Fe-N_D	89.9
N_D-Fe-N_A	90.4
His36 Nϵ-Fe-N_A	91.4
His36 Nϵ-Fe-N_B	89.4
His36 Nϵ-Fe-N_C	88.3
His36 Nϵ-Fe-N_D	89.7
Met71 Sδ-Fe-N_A	85.0
Met71 Sδ-Fe-N_B	92.2
Met71 Sδ-Fe-N_C	95.3
Met71 Sδ-Fe-N_D	88.7
Met71 Sδ-Fe-His36 Nϵ	176.0

Table 6.1: Relevant distances and angles for the iron atom coordination in the final *c*-553 model.

are donated to Gln68 Oϵ1 and to the propionate-6 Oδ2. A conserved water close to the propionate groups is observed in all known cytochromes but its structural or functional role has still to be clarified.

A RASTEP [81] representation of the anisotropicity of the haem is given in Figure 6.5.

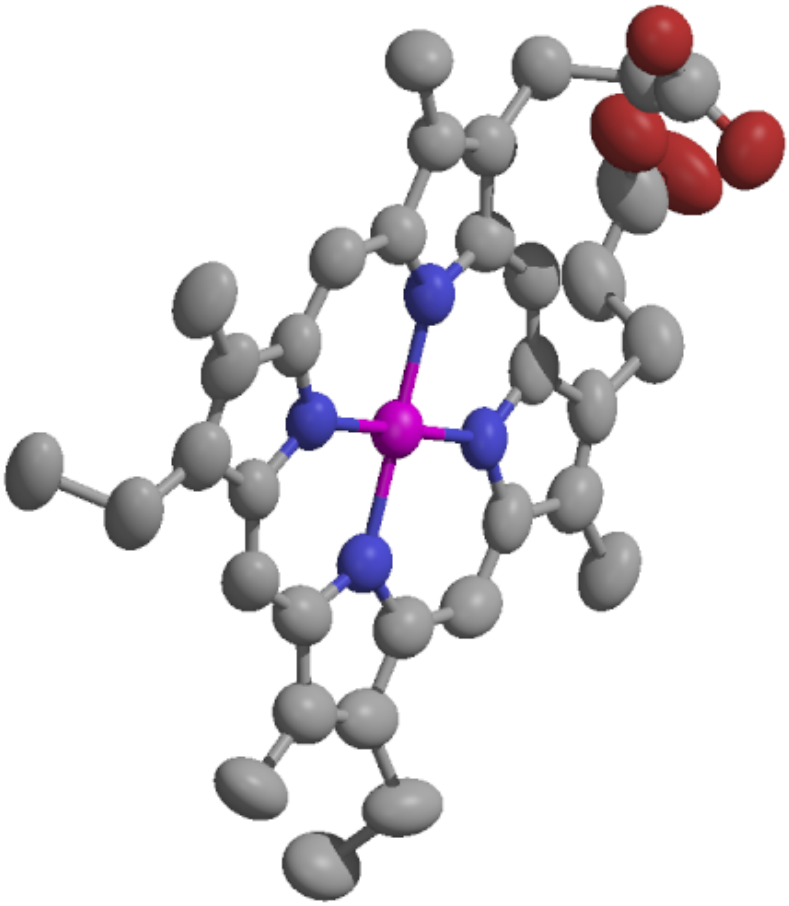


Figure 6.5: Representation of the haem of cytochrome *c*-553 from *Bacillus pasteurii* in which each atom is represented by an ellipsoid enclosing an isosurface of the probability density function. These are commonly known as thermal ellipsoids. The angle of view is automatically adjusted to spread atoms out as much as possible in the XY plane of the image. Figure made with Raster3D [81].

Overall the haem prosthetic group shows a moderate anisotropy higher for the propionate exposed to the solvent and very low for the centre of the structure

with the iron and the nitrogen ligands of the pyrrole rings almost isotropic. Tables **??** and 6.2 report the aniotropic B-factors for the heam group atoms. NMR and cyclic direct voltammetry studies revealed that a change in the protonation state of propionate-7 occurs at pH 5.5 with deprotonation of its Oδ1 [114]. The observed interaction between propionate-7 Oδ1 and Ala47 O in the crystal structure obtained at pH 5.0 (final pH of the crystallisation drop was $\approx$ 5.1 while the cryobuffer was pH 5.0) would disrupt at a pH higher than 5.5 with major confomational changes involving the side chain of propionate-7. The comparison between the geometry of the haem group in *Bacillus pasteurii* cytochrome *c*-553 with the atomic resolution structure of *Monoraphidium braunii* cytochrome *c*-6 (1.1 Å), the standard reference values and the Cambridge structural database (CSD) [117] showed that the distances are essentially identical within experimental error.

6.3 Surface Electrostatic Potential

The mapping of the electrostatic potential surface is shown in Figure 6.6 for different orientations of cytochrome *c*-553. The molecular surface map was calculated with GRASP [118], using a probe radius of 1.4 Å. The electrostatic potential was calculated by GRASP using a simple Poisson-Boltzmann solver and dielectric constants for the solvent and the protein interior of 80 and 2 respectively.

A clear asymmetry of charges is evident. The region around the haem hosting pocket is devoid of charge since it is mainly constituted by hydrophobic residues while most charges are located at the side of the protein opposite to exposed haem edge. The algal cytochrome *c*-6 [119] as well as cytochrome

Atom	B11	B12	B13	B22	B23	B33
CHA	721	1025	475	-42	-1	63
CHB	723	857	538	109	-14	16
CHC	746	832	909	38	-15	213
CHD	873	902	671	116	-211	-209
NA	576	861	491	14	-2	1
C1A	800	859	488	-2	-83	-25
C2A	727	854	603	-2	-120	60
C3A	744	808	578	32	-58	-23
C4A	695	801	496	-5	-41	-14
CMA	1020	858	926	7	39	-184
CAA	821	924	644	-20	-49	81
CBA	805	1021	751	-109	-28	54
CGA	940	1081	757	54	-129	-29
O1A	1035	1393	915	196	-185	-11
O2A	994	1085	879	-6	-192	75
NB	580	749	592	77	-31	91
C1B	590	904	599	91	52	19
C2B	768	1041	551	232	64	238
C3B	777	1055	678	133	20	251
C4B	593	851	770	6	-96	109
CMB	927	1298	553	284	191	109
CAB	781	1172	978	117	89	381
CBB	754	1365	1172	196	96	437
NC	688	810	640	52	-123	-46
C1C	803	813	838	-44	-230	16
C2C	923	886	1008	-25	-242	-55
C3C	938	822	866	6	-373	-96
C4C	818	879	747	79	-163	-91
CMC	1204	880	1413	-279	-254	72
CAC	1449	765	863	119	-365	-146
CBC	1358	1049	1142	-106	-438	-240
ND	500	889	598	14	-41	-113
C1D	558	1108	622	125	-112	-142
C2D	724	1072	673	150	-117	-111
C3D	545	1243	592	11	-42	-103
C4D	562	1116	508	46	-93	-151
CMD	883	1386	673	290	36	-151
CAD	850	1363	616	-55	38	-178
CBD	900	1424	612	-204	-43	-89
CGD	1117	1600	556	-310	-15	55
O1D	1881	1603	867	-608	320	-116
O2D	1354	1379	905	-251	80	-16
FE	611	790	531	43	-29	-33

Table 6.2: Individual anisotropic B-factors for the heam group.

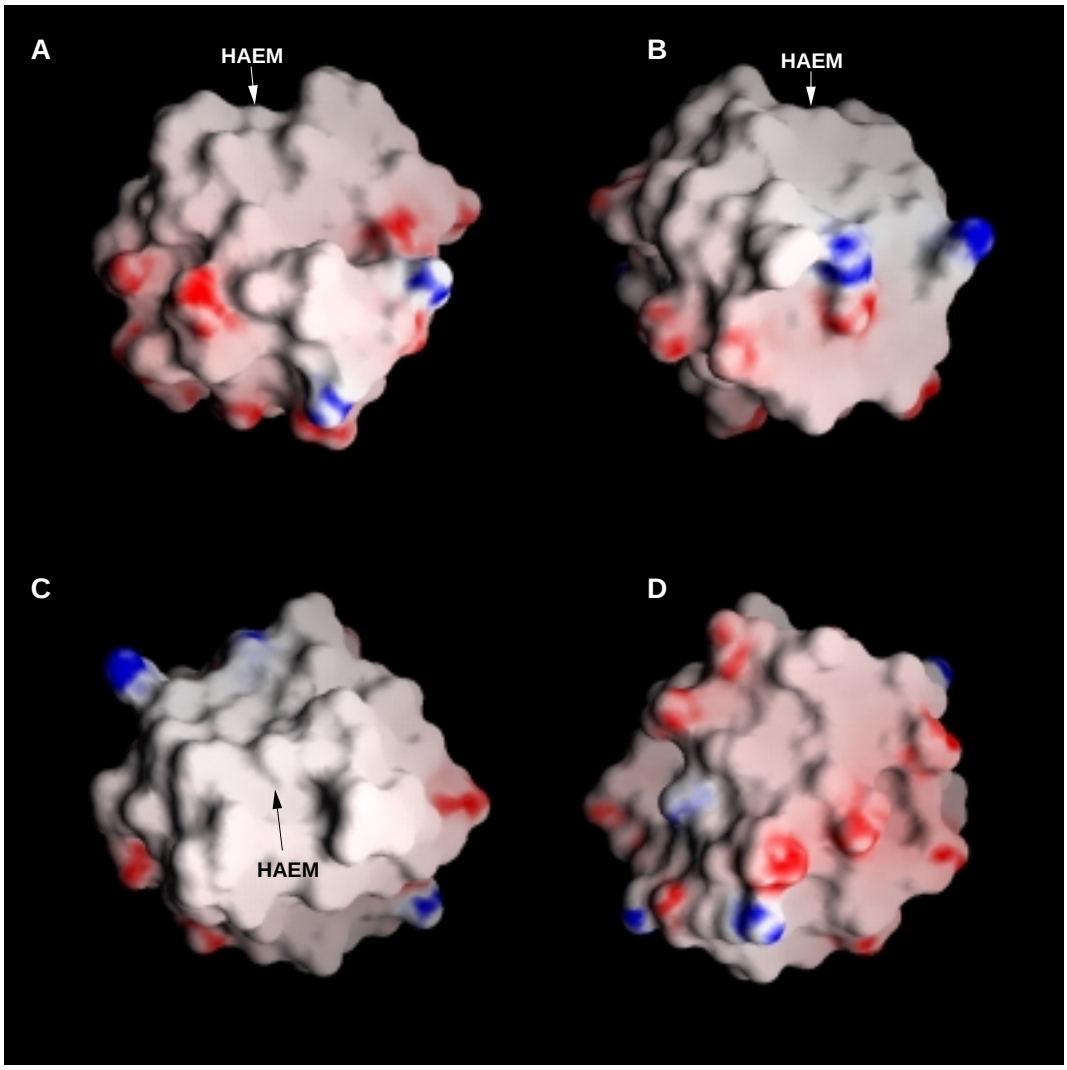


Figure 6.6: *Bacillus pasteurii* cytochrome *c*-553 electrostatic surface potential representation. In the A view the protein is in the same orientation as in Figure 6.1, in the B view rotated along the vertical axis of 180°, in the C view the protein is shown from the top and in the D view from the bottom.

c-552 from thermophiles [120] have the same charge distribution, while for cytochrome *c* [121] and *c*-2 [122] a positively charged front surface, composed of a group of lysines, is found around the haem pyrrole C. The surface around the haem crevice is generally considered the molecular recognition site during

electron transfer between the different partners in the pathway.

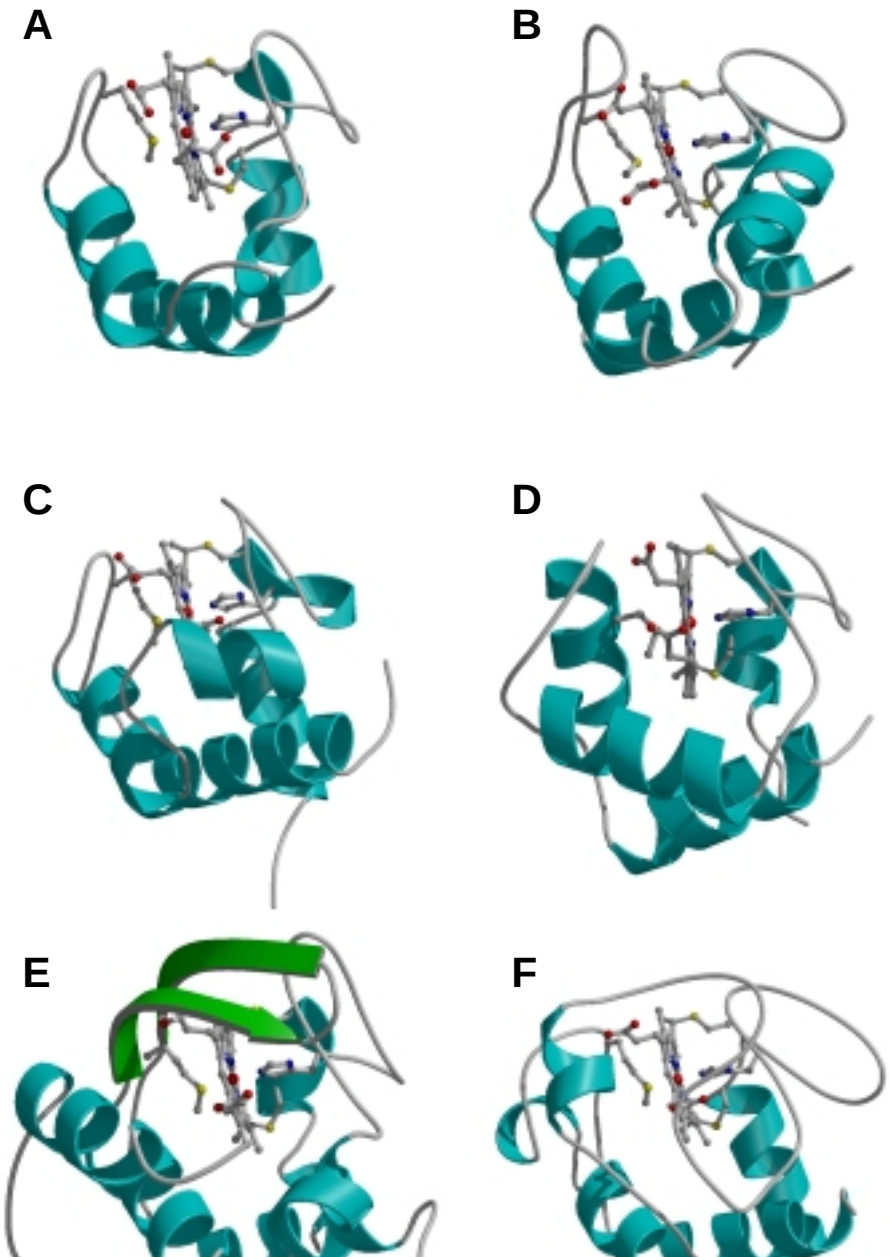
6.4 Structural Alignment and Comparison

To date no other structure of Gram-positive cytochromes has been solved making a general structural comparison difficult. While the amino acid sequences of Gram-positive cytochromes are similar to *B. pasteurii* *c*-553 with a sequence identity of 53% for *Bacillus licheniformis* *c*-552, 45% for *Bacillus sp.* strain PS3 *c*-551 and 41% for *Bacillus subtilis* *c*-551, the similarity with cytochromes from Gram-negative bacteria and eukaryotes is very low [113]. For this reason a structural alignment rather than a sequence alignment has been carried out. The program DALI [102] searched the PDB for structural analogues to *B. pasteurii* cytochrome *c*-553 and found good similarity between cytochromes with very little sequence identity (see Table 6.3). Despite the differences in the primary structure the overall fold and the basic secondary structure elements are highly conserved. Figure 6.7 shows a "rogues" gallery of the aligned cytochrome structures while the structure based sequence alignment carried out with CLUSTALX [85] is reported in Figure 6.8.

B. pasteurii cytochrome *c*-553 can be considered the archetype of all cytochromes in this structural alignment since it is composed of the minimum amount of secondary structure common to all the cytochromes from different sources.

Organism	Type	PDB Code	Z-Score	RMSD Å	Identity%
B. pasteurii	*c*-553	1B75	-	-	-
Ps. aeruginosa	*c*-551	451C	9.3	1.6	32
M. braunii	*c*-6	1CTJ	6.1	2.2	28
D. vulgaris	*c*-553	1C53	4.9	2.5	22
T. thermophilus	*c*-552	1C52	4.7	2.4	23
S. cerevisiae	*c*-(iso-1)	1YCC	4.5	2.2	16

Table 6.3: Summary of the structural alignment for cytochrome *c*-553 from *B. pasteurii* as obtained by DALI. The Z-score represent the strength of structural similarity as defined by DALI [102]. The % of identity refers to the structure based sequence alignment.



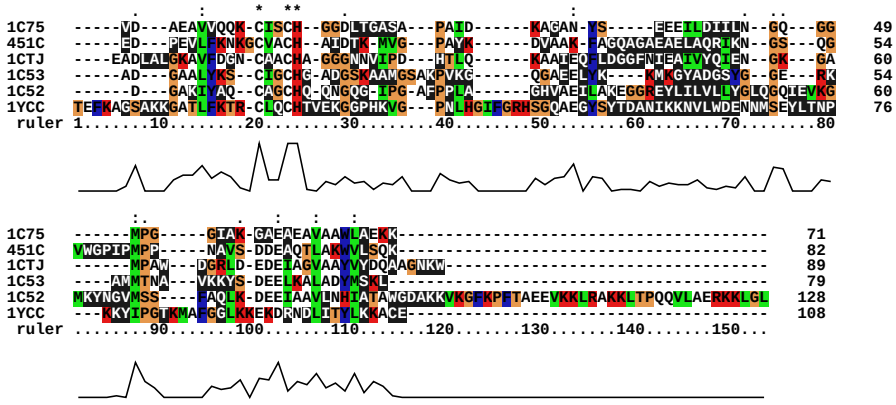


Figure 6.8: CLUSTAX [85] structure based sequence alignment. The proteins, found with a DALI search for structure similarity are indicated by their PDB code. Residues are colour coded according to CLUSTALX: AVFPMILW are represented in red (small plus hydrophobic, including aromatic but not Y), DE in blue (acidic), RHK in magenta (basic), STYHCNGQ in green (hydroxyl plus amine plus basic but not Q) and others grey. The symbols in the consensus line are:
"*" = identical or conserved residues in all sequences in the alignment
":" = indicates conserved substitutions
"." = indicates semi-conserved substitutions

During evolution the diversion from a common ancestor by insertion of secondary elements could have given rise to more complicated structures like *T. thermophilus* c-552 in which the short loop connecting the middle helix and the C-terminus helix observed in *B. pasteurii* c-553 is replaced by two β-strands or like in *S. cerevisiae* c-(iso-1) in which the same structural element is replaced by a long extended loop.

6.5 Conclusions

The method used to solve *B. pasteurii* cytochrome *c*-553 structure is relatively novel. *C*-553 was one of the first structures to be solved by this combination of REFMAC and ARP [123, 124]. The result obtained underlines the importance of collecting data to atomic resolution whenever possible making structure solution almost staightforward. However it should be kept in mind that the presence of the iron atom allowed easy detection of the first atom in the model because of its anomalous signal. It has been recently shown that phasing a protein is even possible using the anomalous signal provided by the sulphur atoms at resolutions up to 2.0 Å; high redundancy of the data was necessary to give an accurate estimate of the anomalous signal with the wavelength optimised to get the highest anomalous differences [125]. For *c*-553 no effort was made to maximise the anomalous signal. At the wavelength used for the data collection (0.885 Å) $f' = 0.28$ e$^-$ and $f'' = 1.26$ e$^-$ significantly lower than the optimal values which would be used in a MAD experiment. For *c*-553 the optimal wavelength for f'' would be 1.735 Å with a value of 3.91 e$^-$, for f' 1.741 Å with a value of -6.5 e$^-$ [1].

The atomic resolution crystal structure of *B. pasteurii* cytochrome *c*-553 provides the best model for a haem c type prosthetic group in terms of accuracy of distances and angles. The absence of the stretch of residues from Gly1 to Ala21 is probably due to the fact that this peptide might be easily hydrolysed upon cell breakage because it is very flexible and exposed to the solvent and to enzymatic proteolysis. The role of this extension might reside in the need for the protein to "swing" (Figure 6.9) from the *bc* complex in order to transfer the electron to the aa_3 complex which represents the terminal oxidase in

[1] values obtained with CROSSEC [69]

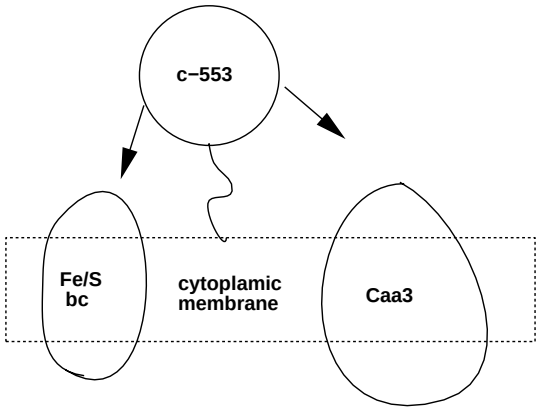


Figure 6.9: Cytochrome *c*-553 represented while "swings" between the putative oxi-redox partner *bc* complex and aa_3 complex with the aid of the membrane anchor. The proposed model is in agreement with the respiratory chain of alkaliphilic bacteria [107]. Figure not in scale.

the respiratory chain. Both *bc* and aa_3 complex are deeply embedded in the membrane and the only way to communicate might be *via* a small fast soluble cytochrome c type fixed to the membrane through a diacyl-glyceryl-cysteine anchor. The charge distribution on the protein surface is higly asymmetric with an extended hydrophobic area in the region around the haem crevice suggesting a probable role in molecular recognition during electron transfer.

The very low reduction potential characteristic of *c*-553 was determined from electrochemical experiments to be mainly due to a large negative reaction entropy, while a smaller effect was due to an enthalpic effect [114]. The large negative entropy observed for the reduction was suggested to derive from extrusion of water molecules from the protein hydration shell into the bulk solvent upon reduction [114]. This phenomenon was proposed to be accompanied by an increased rigidity of the cytochrome structure in the reduced state. Probably the increased rigidity is due to the extrusion of water molecules from the hydration shell or *vice versa*. The accessibility of the haem prosthetic group

to the solvent might have a big influence on the redox properties of cytochromes if we accept the important role of the solvent in the regulation of the reduction entropy. There seems to be a correlation between haem accessibility and entropy indicating a possible link between the electrochemical reduction potential (entropy) and the structure in terms of haem solvent exposure [126].

Last but not least is the influence of the amino acide composition and haem interaction with the "hosting" protein matrix on the redox properties. The amino acide composition is less conserved than the fold, as confirmed by the primary and quaternary structure comparison between this group of cytochromes. This observation suggests a general need for the microorganisms to express a functional electron transfer protein whose electrochemical parameters will be then modulated either by different degrees of haem exposure or by variation in the amino acide composition.

Fine della tesi!

Part III

Appendix

Appendix A

X-ray Diffraction and Protein Crystallography

"Una ragazza e' sempre un mistero:
non c'e' che da fidarsi al suo viso e
all' ispirazione del proprio cuore."[1]

The history of X-rays began at the end of last century (1895) with the discovery by Röngten of a new radiation with very peculiar characteristics. Its use in crystallography was first dicovered only in 1912 when Laue showed the wave nature of X-rays, and Friedrich and Knipping recorded the first diffraction photograph from a $CuSO_4 \cdot 5H_2O$ crystal. Only one year later Sir Lawrence Bragg solved the structure of several simple compounds like NaCl, KCl, KBr and KI. In 1926 Sumner obtained crystals from jack bean urease (*Canavalia ensiformis*) [1] the first crystals from an enzyme. This was followed soon by Northop and Herriot with the crystallisation of pepsin, trypsin, and chymotrypsin. The first attempts to get diffraction from protein crystals were unsuccesful because they

[1] E. De Amicis

were carried out on dried samples and only in 1934 were Bernal and Crowfoot able to get a diffraction pattern from a pepsin crystal mounted in a capillary. The first protein structure myoglobin was unravelled only in the late fifties. The solution of protein structures like myoglobin by Kendrew and haemoglobin by Perutz represent the first milestones in a field which still gives an enormous wealth of information about protein structure and function relationships. Protein crystallographers are interested in the interactions between X-rays and protein crystals, a magic world in which protein molecules are arranged together to form a three dimensional lattice. The scattered waves are recorded and used to reconstruct the electron density map of the protein contained in the crystal. The final goal is to understand *inter alia* how enzymes work, how proteins interact with each other, and how drugs dock in the active sites. The way to understand the function is through the knowledge of the structure and this is what protein crystallography is all about.

A.1 The Experiment

The minimal requirement to perform an X-ray diffraction experiment is a source of X-rays, optics to focus and monochromatise the beam, good quality diffracting crystals mounted on a goniometer and a detector to record the diffracted X-rays. Nowadays with the use of synchrotron radiation it is absolutely necessary to make use of cryogenics to avoid radiation damage and a cryosystem is usually used on home sources as well. Figure A.1 shows schematically the basic setup. Data collection, data analysis and calculations are performed with the aid of computers and dedicated software.

The primary X-ray beam, coming from the source, is monochromatised by

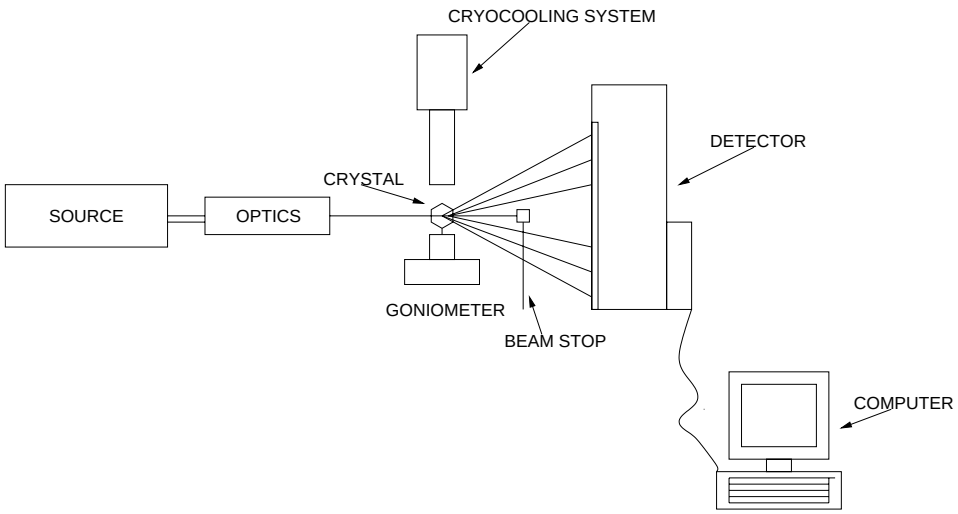


Figure A.1: Setup for a diffraction experiment

crystal monochromators. The beam passes through the crystal mounted on a goniometer which allows one to position and rotate the crystal in different orientations in the beam. The diffracted X-rays are usually recorded using image plate or CCD detectors [127, 128]. A beam stop is placed right after the crystal to prevent the direct beam reaching the detector so that only the diffracted X-rays are recorded and the detector is not saturated by the more intense direct beam.

The evolution of the technique, the constant increase of computer power and the availability of large amount of protein obtained with molecular biology, has made structure determination much easier and the limiting factor in protein crystallography is mostly the availability of good quality diffracting crystals.

A.1.1 Protein Crystallisation

Growing protein crystals is still regarded more like an art rather than a science because of the many factors involved that are still not completely understood. Crystallisation is usually achieved by a gradual decrease of solubility of the protein in solution. Variation in concentration, pH, ionic strength, temperature, cofactors or impurities can affect protein solubility and crystal growth or quality. The basic requirements for a crystallisation experiment are the purity and homogeneity of the sample (close to 100%) and availability in large quantities (of the order of at least some mg). A typical approach is based on increasing the concentration of salt (the classic salting-out with e.g. ammonium sulphate) or polyethyleneglycol (PEG) in the protein solution to decrease its solubility. Figure A.2 shows a protein solubility curve. Crystals grow in the metastable region of the diagram called the supersaturation area in which nuclei can be formed. After the formation of the first nuclei the concentration of the protein is slightly decreased, the nucleation stops and the crystals formed can usually grow to a bigger size. The factors involved in the aggregation of protein molecules are of various type: electrostatic forces, hydrophobic interaction, hydrogen bonding. Many factors can be varied, hydrophobic interaction increases with the temperature and the ionic strength, while the presence of organic solvent, such as alcohol or change in pH, influence electrostatic forces.

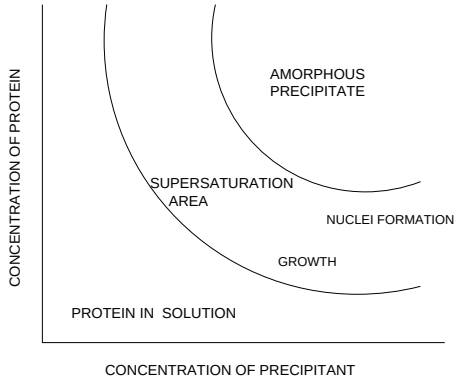


Figure A.2: Solubility curve for proteins. The area between the two curves is the "zone of growth".

Crystallisation techniques

Vapour diffusion is the most common method of crystallisation. In the hanging drop a small amount of protein solution is mixed with the precipitant (typical volumes range from 1 to 10 μL) and placed on a siliconised cover slip that is used to seal a well containing the precipitant solution. The drop is placed upside down and equilibrates its ionic strength with the precipitant by vapour diffusion. The sitting drop is a variation of the method which allows bigger drop volumes to equilibrate.

Dialysis can be performed for big volumes with conventional membranes while for small volumes microdialysis in capillaries or in dialysis buttons is preferred. This system allows an easy exchange of the precipitant solution as many times as desired and makes it possible to test crystallisation of a protein at low ionic strength.

The **batch method** is the oldest. The addition of precipitant agent is made directly into the protein solution until it reaches supersaturation. Nowadays, with the use of crystallisation robots, this technique has become fashionable in the microbatch variant. Good results are obtained with the microbatch technique using silicon oils in which the sample drops are immersed [129].

Liquid-liquid diffusion. In this technique the protein solution and the precipitant are layered on top of each other allowing a slow equilibration. It is usually perfomed in capillary tubes.

The approach to find the condition at which a protein would crystallise is still mostly a trial and error even with the "sparse matrix" search used to

quickly obtain information about the factors affecting solubility and crystallisation [62]. A Biological Macromolecular Crystallization Database is available to correlate protein characteristic and crystal growth conditions [130]. An overview of the process of crystallisation is found in [131].

Protein crystals are characterised by a relatively high water content ranging typically from 30 to 80% of the volume with a specific volume V_M, which is the crystal volume per unit of protein molecular weight. This ranges between 1.7 $Å^3 \cdot Da^{-1}$ and 3.5 $Å^3 \cdot Da^{-1}$ [67]. The ordered lattice of a protein crystal is far from perfect; it can be seen as an assembly of smaller crystal mosaics slightly misaligned with respect to each other. Typical values of mosaicity range between 0.2 and 0.5 degrees. In cryogenic data collection a slight increase in mosaicity is common, compensated by the reduction of radiation damage. Typical size for protein crystals are from 0.1 to 1.0 mm but smaller or bigger crystals are not unusual.

A.1.2 Crystals and Symmetry

The ordered arrangement of molecules that give rise to a crystal can be seen as a regular repetition of identical parallelepiped-shaped blocks called unit cells. The space lattice is defined as an arrangement of points with exactly the same environment and orientation with respect to every other point in the lattice. The dimensions of the unit cell are chosen such that the edges are as short as possible while allowing for the symmetry of the crystal. The vectors **a**, **b**, **c** define the length a, b, c, and the angles α, β, γ characterising the unit cell. The content of the unit cell is obtained by repetition of a single object through the symmetry elements. This part of the unit cell is called the asymmetric unit.

A special case of symmetry is non crystallographic symmetry (NCS) through which the molecules within one asymmetric unit are related by appropriate operations in addition to the normal crystallographic symmetry .

To define the planes in the crystal the so called Miller indices (h, k, l) have been introduced. The h, k, l terms define parallel planes with intercepts $a/h,\ b/k,\ c/l$ on the three a, b, c axes of the unit cell with h, k, l small integer numbers. The possible symmetry elements in a crystal are derived from external (based on the external appearance) and internal symmetry (revealed by X-ray difraction). The external symmetry elements are described as 32 classes of symmetry point groups:

- Mirror plane, which does not occur in crystals of proteins because they are enantiomorphic molecules built with L-amino acids.
- Rotation axis, characterised by a rotation about one axis of $360°/x$ where x can only be 1, 2, 3, 4 or 6.
- Inversion point, which does not occur in crystals of protein since they are enantiomorphic molecules.

Crystals can be divided into seven crystal systems with defined characteristics and parameters (Table A.1) by analysis of the rotational symmetry. The fourteen Bravais Lattices are obtained by combining crystal symmetry with lattices that still satisfy the space lattice but which instead of having only a point at each corner of the unit cell, have additional points. These internal symmetry elements give rise to $P,\ I,\ F$ and C lattices. In a cubic lattice only the vertex are enough to define the unit cell and the cell is called primitive (P), when a point is in the centre of the cell is called body centred $(I$, from

Innenzentrierte) and when all the faces have a point in the centre is face centred (F). Monoclinic and orthorhombic can also have a point on the (001) face giving a C lattice.

Other internal symmetry elements are:

- Glide plane. Obtained by a combination of a mirror plane and a translation, it is not possible for enantiomorphic protein crystals.
- Screw axis. A rotation is combined with a translation parallel to the rotation axis. The molecule is shifted by a fraction of unit cell and rotated.

The combination of the 32 point groups with the Bravais Lattice and the internal symmetry gives rise to 230 space groups (Table 1.1) of which only 65 are possible for enantiomorphic molecules.

The diffraction pattern of a crystal exhibits the same crystal symmetry but, in the absence of anomalous scattering, with an additional centre of symmetry. The diffraction pattern symmetries are grouped in 11 Laue classes. The presence of symmetry elements, like screw axes, can be detected since they give rise to systematic absences in the diffraction pattern.

Crystal system	Cell Geometry	Axes of symmetry	Bravais Lattices

Table A.1: The seven crystal systems

A.1.3 X-ray Sources

There are two different type of process that are used to produce X-rays for diffraction experiments. In conventional generators radiation is produced by bombarding a metal target with electrons, while for synchrotron radiation X-rays are obtained by accelerating particles (electrons or positrons) at relativistic speeds in storage rings.

Laboratory sources

Conventional sources used in laboratory are basically of two kinds: sealed tubes and rotating anodes. The process behind the emission of X-rays from these devices is essentially the same. A beam of electrons is accelerated in vacuum from a cathode until it reaches a metal target (the anode). When the electrons hit the metal the energy is partially converted to white radiation. Increasing the energy of the electrons striking the target to higher than a certain threshold (characteristic for every element) leads to electrons, being excited to higher atomic energy levels, ejected from one of the inner shells. Electrons from the higher levels fall back to fill the vacancy and emit the excess energy in the form of radiation of characteristic wavelength. The most common metals used to produce X-rays are copper with K_α radiation at 1.5418 and molybdenum with K_α at 0.7107 Å. The high vacuum is required to avoid collisions between gas molecules and travelling electrons in the tube, a phenomenon that would decrease the efficiency of the process. Only 0.1 % of the energy is transformed into X-rays; the rest is dissipated as heat. For this reason a water cooling system is always needed.

A **sealed tube** consists of a cathode with a filament that emits electrons when a high voltage current is applied. These electrons are then accelerated

under vacuum and hit the anode (copper or molybdenum). They do not require maintenance and they are easy to change but they do not last long.

Rotating anodes, differ from a sealed tube only in the construction, while the principle remains the same. The surface that electrons hit is much bigger and the fact that the anode is rotating helps in dissipating the heat so that a more intense beam can be produced. It requires more maintenance since it is more mechanically complicated.

Synchrotron Radiation

Synchrotrons were first built for particle physics studies and the radiation emitted was seen as an annoyng phenomenon resulting in loss of energy.

In 1971 Rosenbaum *et al.* explored the use of synchrotron radiation as a source for producing X-rays for crystallography at the DESY synchrotron site in Hamburg showing that after monochromatization and focussing the total flux density was two orders of magnitude greater than the CuK_α radiation coming from a rotating anode [132]. Nowadays the use of synchrotron radiation in protein crystallography is an important tool to answer biologically relevant questions in cases where the small size of the crystal or the big dimensions of the protein would be limiting factors on a conventional source [133].

Apart from the high intensity, an additional advantage of using synchrotron radiation is the tunability of the wavelength and potentially its pulse length for time resolved experiments (e.g. Laue diffraction). The tunability allows the use of shorter wavelengths to minimise absorption as well as to perform multiple wavelength anomalous dispersion (MAD) experiments selecting the

wavelength at the characteristic energy absorption edges. The X-ray emission from a synchrotron is characterised by the following parameters:

Flux = photons/sec/0.1%$\delta\lambda/\lambda$

Brightness = photons/sec/0.1%$\delta\lambda/\lambda$/mrad^2

Brilliance = photons/sec/0.1%$\delta\lambda/\lambda$/mrad^2/mm^2

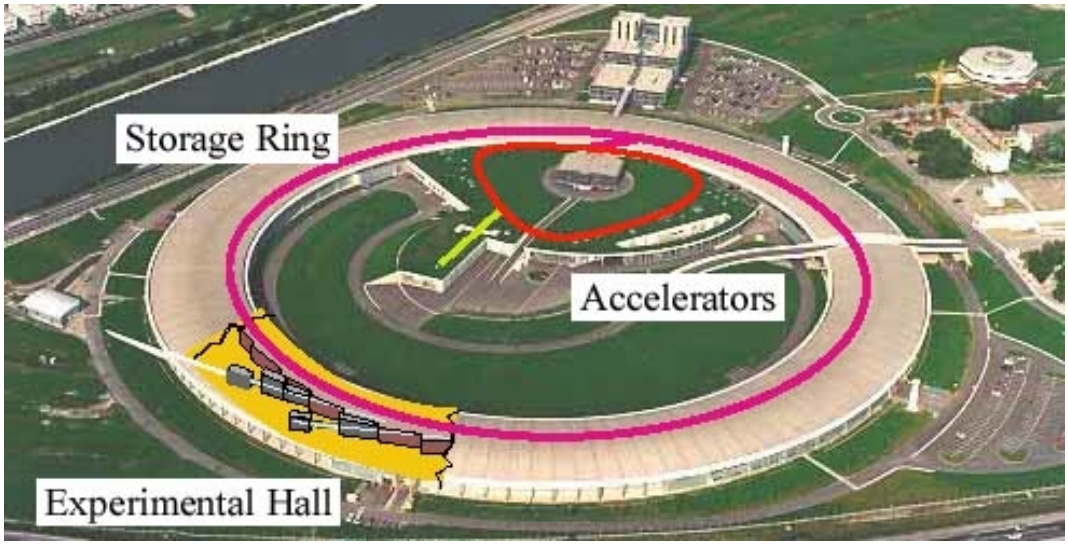


Figure A.3: Aereal scheme of a synchrotron. European Synchrotron Radiation Facility, ESRF, Grenoble, France

To produce radiation, electrons or positrons are first accelerated in a LINAC (linear accelerator). The particles are then injected in bunches directly into the storage ring or into a so called booster in which they are brought to the desired speed before being redirected in the storage ring (Fig. A.3). Radio cavity frequency devices are used to restore the energy that is lost because of the emission of radiation and to keep the particles grouped together. In the storage ring the bunches of particles travel in the vacuum focused horizontally and vertically by quadrupole magnets and bent by bending magnets. The vac-

uum must be very high to prolong the beam lifetime because particles are lost in collisions with residual gas molecules in the ring. In this respect positrons allow for longer lifetime than electrons because their repulsive interaction with the residual gas ions in the vacuum reduces the collisions. The choice between electrons and positrons is based essentially on costs and quality of the vacuum. When the vacuum is good enough there is no real need for using positrons since they are more expensive to produce than electrons.

The emission of radiation occurs when the particles deviate in their trajectory and bending magnets are the first source of synchrotron radiation (Fig. A.4). Because they are travelling close to the speed of light the radiation is emitted in a forward cone rather than isotropically.

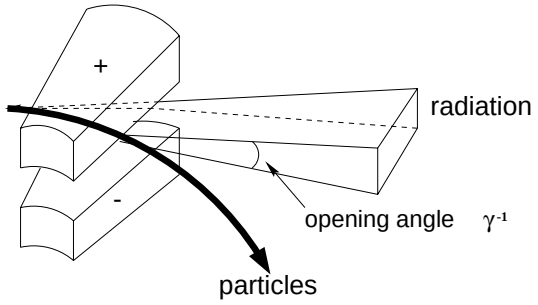


Figure A.4: Radiation from a bending magnet

Insertion devices (ID) can be placed in the straight sections of the ring to produce very intense and focused X-ray beams. Wigglers and undulators are essentially based on the same principle. An array of magnets makes the particles "wiggle" in a periodic magnetic field (Fig. A.5 and A.6). IDs are used to increase the intensity (wiggler) or increase the brilliance *via* interference (undulator).

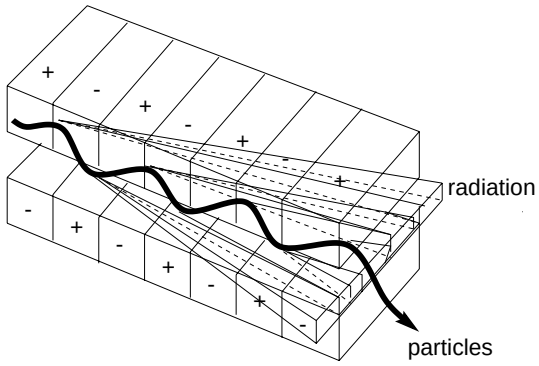


Figure A.5: Radiation from an insertion device

Varying the gap between the two magnet arrays allows the deflection angle (δ) of the particle beam to be changed because the magnetic field decreases as the distance increases. Being γ^{-1} the opening angle of the radiation, ($\gamma^{-1} = \frac{mc^2}{E}$ with mc^2 the electron rest energy and E the machine energy) when $\delta \gg \gamma^{-1}$ the ID is a wiggler because the radiation coming from one pole of the magnet does not interfere with the radiation coming from the next one. When $\delta \approx \gamma^{-1}$ the ID becomes an undulator since interference is obtained. The emission from a wiggler is a continuous spectrum while the radiation coming from an undulator is characterised by the presence of peaks at discrete wavelengths.

A.1.4 Optics

The aim of the optical elements is to provide a narrow and focused monochromatic beam which is of vital importance in a single wavelength diffraction experiment. For conventional sources the production of X-rays is characterised by bands at defined wavelength (energies) superimposed on a white radiation continuum. The use of filters for radiation coming from conventional sources (e.g. rotating anode) allows the filtering out of undesired radiation since they are made with material that selectively absorbs unwanted wavelengths. Nickel

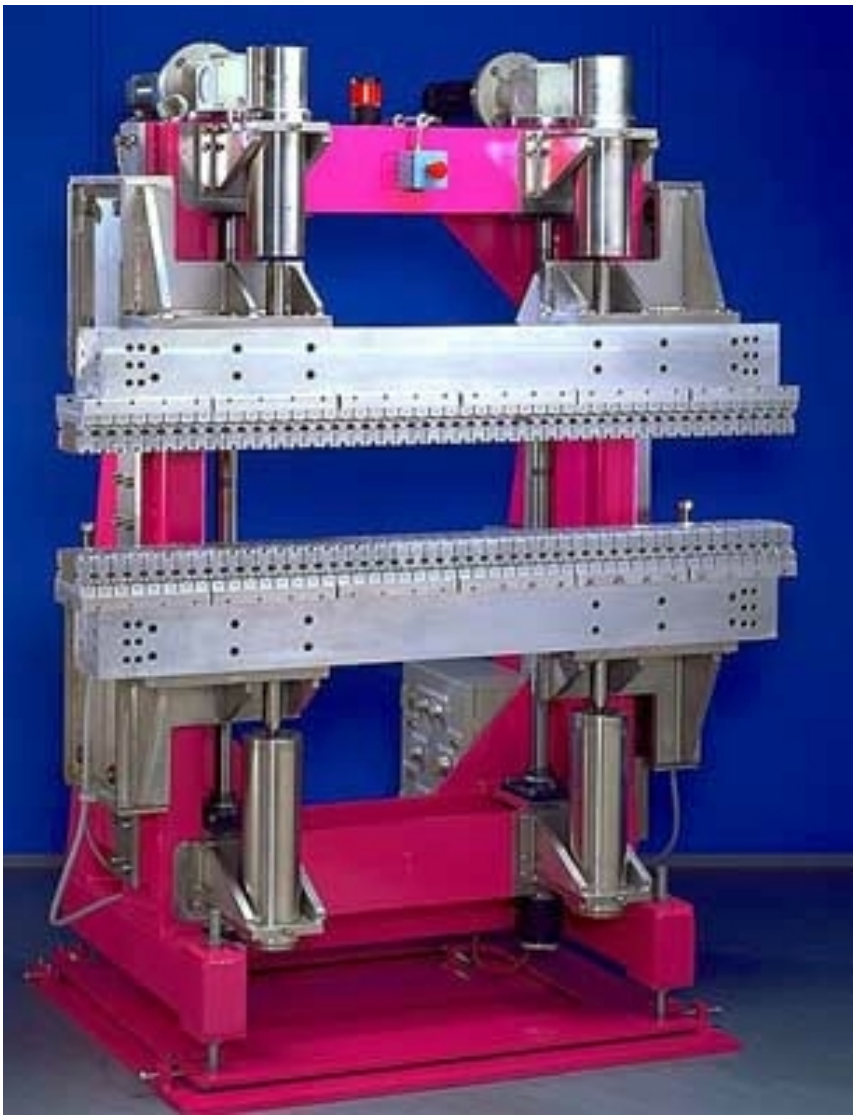


Figure A.6: An insertion device.

filters are used to eliminate the K_β emission band when using copper K_α radiation. The mass absorption coefficient (μ_m) for Ni K_α has its maximum at the copper K_β edge thus the X-ray photons hitting the Ni filter have the energy necessary to eject an electron from the Ni K_α orbital with absorption of radiation and emission of energy as fluorescence.

A better way of selecting a specific wavelength (λ) is the use of crystal monochromators that exploit Bragg's law ($n\lambda = 2dsin\theta$). When radiation (i.e., white radiation coming from a synchrotron source) reaches the crystal it is scattered at angles θ that depend on the wavelength of the incident beam. Selecting a diffraction angle θ selects a specific (λ). This type of monochromator consists either of a single crystal, on which the beam is diffracted once, or double crystal (more selective but more sensitive to small changes of θ) on which the beam is diffracted twice.

As part of the optic elements, mirrors are used to partially monochromatise the beam (eliminating harmonics), since they reflect only when the incidence angle is smaller than a critical angle function of the wavelength. Curved mirrors made of a single bent piece or segmented mirrors are used to focus the beam. When a toroidal shaped mirror is used the beam is focused horizontally and vertically. For this purpose an alignement of two mirrors, one focusing horizontally and the other one vertically, can be used instead of the expensive toroidal mirror.

The size of the beam can be adjusted to match crystal size by mean of two sets of vertical and horizontal collimation slits. A typical aperture for the front end slits is $0.3 \times 0.3mm^2$.

A.1.5 Detectors

The development of detectors for X-ray crystallography has been very fast in the last few years, stimulated by the increasing use of synchrotron radiation which requires not only good quality but also very fast detectors. An

ideal detector should have [127]: high detective quantum efficiency ($DQE = (\frac{S_o}{N_o})^2/(\frac{S_i}{N_i})^2$, where S is the signal and N the noise with the subscript o and i meaning the output and input), wide dynamic range (the ability to measure both strong and weak signals), linearity of response, high spatial resolution (ability to resolve very close spots), large active area (allows collection of many reflections at the same time), uniformity of response over all the surface of the detector active area, and high count rate capability.

For protein crystallography 2D detectors are the most used because of the large numbers of reflections to be collected as opposed to single photon counters which measure one reflection at a time. The old **photographic film**, despite its very high resolution (due to its fine grains), has been replaced by image plates (IP) and Charged Coupled Devices (CCD) because of the long and tedious work required by X-ray films and their limited dynamic range (1:200) [134].

Image plates are based on the excitation by diffracted X-ray photons of inorganic storage phosphors deposited on a flat surface. The energy provided by the diffracted X-rays is stored in colour centres as excited electrons. The release of the stored energy is obtained by a red laser and the blue light emitted by the colour centres is measured by a photomultiplier. Before a new exposure the plate has to be erased with white light. This process brings the phosphor back to the ground state. Data collection with a image plate is divided in three steps of different length. The exposure, whose length depends on the sensitivity of the phosphor, on crystal quality and intensity of the source. The read out and the subsequent erasing depend on the size of the plate to scan. They are becoming faster with the progress of technology.

In the MAR Research IP a disk of 345 mm of diameter is used. The read out is obtained by scanning the plate following a spiral obtained by translating laser and photomultiplier vertically while the plate rotates. The typical dynamic range for an IP is in the order of $1{:}10^4 - 10^5$ with a sensitivity at least 10 times more than X-ray film [134].

Charged Coupled Device area detectors have been developed from a silicon memory device [128]. The capacity of charge accumulation induced by light exposure made a very good imaging sensor. The basic design is composed of essentially three parts: a front end surface covered by fast X-ray sensitive phosphor converting the incident photons to visible light, a fibre optic taper demagnifying and transfering the light photons to the CCD chip, and the chip in which the photons induce a charge generation. The charges are then transferred and detected. The need for fast phosphors (light emission drops to less than 0.1% in about 10 msec) is given by the fact that CCD reads out the photon transformed as visible light almost instantaneously and immediately after is ready for a new exposure.

In the MAR research CCD an X-ray sensitive surface of 165 mm diameter is coupled via the fibre optic taper with a 61 ×61 mm CCD chip with 4096×4096 pixels. For the CCD chip the cooling system is of vital importance to lower the noise level caused by accumulation of thermal charge, the so called "dark image" [128].

A.2 The wiggler beamlines at the EMBL outstation Hamburg

The European Molecular Biology Laboratory (EMBL) outstation in Hamburg is located at the Deutsche Elektron Synchrotron (DESY) site. The storage ring for EMBL beamlines is DORIS III operating at an energy of 4.5 GeV with a current in excess of 100 mA immediately after the injection of positrons every 12 hours in optimal conditions.

BW7 wiggler is a 4 m long 1 Tesla magnet with 56 poles installed in one of the seven straight sections of the storage ring. It has a critical energy of 13.6 keV at a minimum gap size (distance between the magnetic arrays) of 30 mm. The first optical component after the wiggler is a Rh-coated SiC mirror removing X-rays with energies above 20 keV. A water-cooled tungsten beam splitter divides the beam in two parts, one tunable used for MAD experiments (BW7A) and one for high flux fixed wavelength (BW7B with $\lambda \approx$0.9 Å). The tunability for BW7A is achieved by mean of a double crystal monochromator built with two independent Si (111) crystals mounted on a wheel. The first crystal can be accurately positioned with respect to the second one which defines the fixed exit of the monochromatic beam.

For BW7B there is a single crystal monochromator. The single triangular Si crystal is bent by holding its base and pushing the opposite apex. The result is a cylindrically shaped surface of varying radius exploited for horizontal focusing of the beam together with its monochromatisation. The vertical focusing is obtained for both branches by segmented quartz mirrors. They consist of twelve 10 cm long segments mounted on a wedge shaped aluminum bed bent to form a focusing ellipsoid. Two sets of slits define the dimension

of the beam ($0.3 \times 0.3mm^3$). Both beam lines are equipped with optical tables on which image plate detectors and cryosystem are installed. The tables are aligned to the X-ray beam by mean of three height and two horizontal translation stages controlled by a computer [63, 64]

A.3 The Principles of X-ray Diffraction

X-rays are electromagnetic waves characterised by wavelength (λ) in the range of 0.1-1000 Å, amplitude (A), and frequency ($\nu = c/\lambda$). The electric and magnetic fields characterising an electromagnetic wave, oscillate in planes perpendicular to each other along the direction of propagation of the waves as a cosine function. The interaction between the travelling waves and the electrons in the crystal gives rise to scattering.

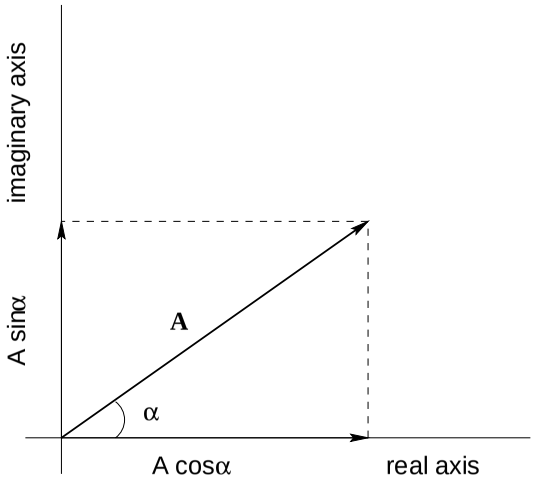


Figure A.7: A wave represented in an Argand diagram. The vector **A** represent the wave and its two components are $Acos\alpha$ and $Asin\alpha$.

Two kinds of scattering take place: Thomson and Compton. The first phenomenon is due to an energy transfer from the incident wave to the electron causing the latter to oscillate at the same frequency as the incident wave. The energy is then released as a scattered wave with the same wavelength as the incident radiation but differing in phase of 180° from it. The energy is conserved, all the scattered waves have the same phase relationship with the incident beam (180°) and the scattering is called coherent.

Compton or incoherent scattering is characterised by a loss of energy with change of wavelength of the scattered radiation with respect to the incoming radiation. In X-ray crystallography coherent scattering gives rise to diffraction while incoherent scattering contributes to the diffuse background.

A wave can be represented as $Acos(\omega t + \alpha)$ with $\omega = 2\pi\nu$, t the time and α the phase angle. The wave $Acos(\omega t + \alpha)$ can be regarded as the summation of two components: a real $Acos\alpha$ and an imaginary $Asin\alpha$ parts as represented in an Argand diagram (Fig.A.7). The total scattering from a crystal is given by summation of all the diffracted waves. Vectorial summation is not convenient for big numbers and the waves are expressed in mathematical form as $Acos\alpha + iAsin\alpha$ or $A\exp[i\alpha]$ to allow mathematical summation. Given a system of two electrons located one in 0 and the other one in 1 (Fig. A.8) the relation between the scattered waves is related by the path difference of the wave passing through 1 (A1+B1).

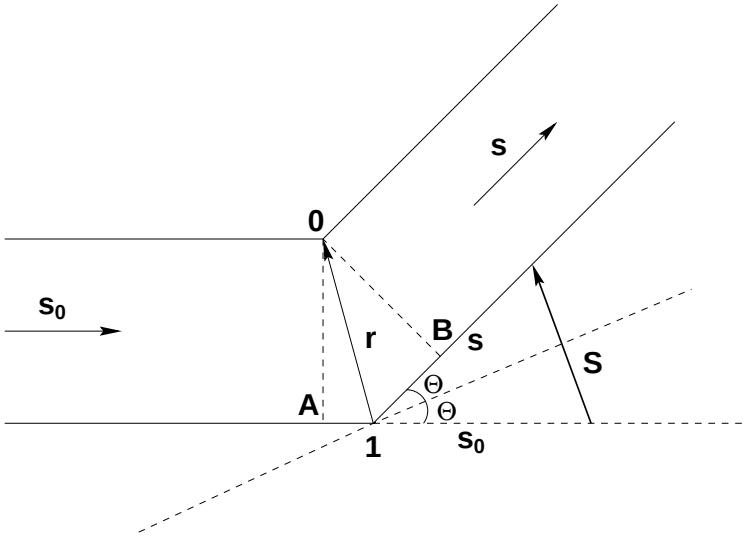


Figure A.8: Scattering from a system of two electrons.

The phase difference between the two unit vector $\mathbf{s}$ and $\mathbf{s}_0$ is:

$$2\pi\mathbf{r}\cdot(\mathbf{s}-\mathbf{s}_0)=2\pi\mathbf{r}\cdot\mathbf{S}$$

($\mathbf{s}-\mathbf{s}_0=\mathbf{S}=2sin\theta/\lambda$ with $\mathbf{s}$=$\mathbf{s}_0$=$1/\lambda$). The plane (dashed line) passing through 0 and $\perp$ to $\mathbf{S}$ can be considered as a reflecting plane, and the diffraction as a specular reflection against this plane. If A_0 is the wave scattered by the electron located in 0 the wave A_1, scattered by the electron located in 1, will be: A_1=$A_0\exp(2\pi i\mathbf{r}\cdot\mathbf{S})$.

For a system with n electrons the total scattering will be:

$$\mathbf{F(S)}=\sum_{j=1}^{n}A_j\exp[2\pi i\mathbf{r}_j\cdot\mathbf{S}]$$

Considering an atom as scatterer instead of isolated electrons the scattering will be related to the number and position of the electrons in its orbitals. The scattering from an atom will be thus called the atomic scattering factor f expressed as integration over the entire space (with $\mathbf{r}$ the electron cloud radius):

$$f=\int_{\mathbf{r}}\rho(\mathbf{r})\{\exp[2\pi i\mathbf{r}\cdot\mathbf{S}]+\exp[-2\pi i\mathbf{r}\cdot\mathbf{S}]\}d\mathbf{r}$$

further simplified to:

$$f=2\int_{\mathbf{r}}\rho(\mathbf{r})cos[2\pi i\mathbf{r}\cdot\mathbf{S}]\}d\mathbf{r}$$

The atomic scattering factor depends on the length of $\mathbf{S}$ (since $\mathbf{S}=\frac{2sin\theta}{\lambda}$) but is independent of the direction of the vector $\mathbf{S}$. For a carbon atom for instance the scattering factor will be equal to 6 (the number of carbon's electrons) for $\theta=0$. The bigger the angle θ (thus higher the resolution) the smaller the scattering factor.

Since the atoms scattering the X-rays are not fixed in their position but vibrate around an equilibrium position their scattering factor is affected. This motion, dependent on the temperature, is called B or temperature factor. The scattering factor diminishes because of thermal vibration especially at high diffraction angles. When the vibration is the same in every direction is called isotropic B factor and is expressed as:

$$B = 8\pi^2 \times \overline{u^2}$$

where $\overline{u^2}$ represent the mean square atomic displacement and is thus expressed in Å^2. For structures with normal or low resolution only the isotropic temperature factor can be refined while for high (near to atomic) resolution structures the anisotropic B factor can be refined.

If we consider a more complicated system like the unit cell the total scattering will be given by the summation of all the partial contributions. Every atom in the unit cell will diffract according to its scattering factor. Thus the structure factor representing the summation of all the waves scattered by all atoms in the unit cell is expressed as:

$$\mathbf{F(S)} = \sum_{j=1}^{n} f_j \exp[2\pi i \mathbf{r}_j \cdot \mathbf{S}]$$

The total scattering from a crystal $\mathbf{K(S)}$ is obtained by adding the scattering coming from all its unit cells. This is done by considering the crystal as formed by translation of the unit cell along the vectors $\mathbf{a}$, $\mathbf{b}$, $\mathbf{c}$ at positions $t{\cdot}\mathbf{a}{+}u{\cdot}\mathbf{b}{+}v{\cdot}\mathbf{c}$ with t, u, v whole numbers. The total scattering is then:

$$\mathbf{K(S)} = \mathbf{F(S)} \times \sum_{t=0}^{n1} \exp[2\pi i t \mathbf{a} \cdot \mathbf{S}] \times \sum_{u=0}^{n2} \exp[2\pi i u \mathbf{b} \cdot \mathbf{S}] \times \sum_{v=0}^{n3} \exp[2\pi i v \mathbf{c} \cdot \mathbf{S}]$$

Scattering is only observed when $\mathbf{a} \cdot \mathbf{S} = h,\ \mathbf{b} \cdot \mathbf{S} = k,\ \mathbf{c} \cdot \mathbf{S} = l$ being h, k, l whole numbers representing the indices of reflections.

Diffraction can be understood as reflection of the waves from a set of planes, defined by the three indices h, k, l with a distance d between them (Fig. A.9). The phase difference between two consecutive waves is given by their path difference $2dsin\theta$. Therefore the phase difference between waves scattered from successive (hkl) planes is 360°, meaning a coherence of the scattering.

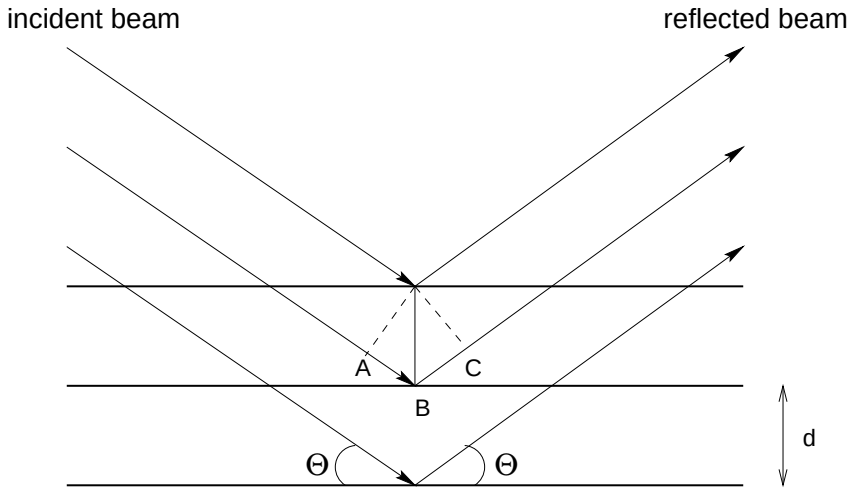


Figure A.9: The Bragg's law

The waves are in phase when they fulfil the condition: $\lambda = 2dsin\theta$ (Bragg's law) thus giving rise to diffraction.

From Bragg's law an important tool for constructing the direction of the X-rays diffracted by a crystal can be derived (Fig. A.10). The basic principle is founded on the fact that each reflection arising from interaction between X-rays and lattice planes in the crystal (the direct lattice) gives rise to a diffraction pattern which is reciprocal to the direct lattice. Being $\mathbf{a} \cdot \mathbf{S} = h,\ \mathbf{b} \cdot \mathbf{S} = k,\ \mathbf{c} \cdot \mathbf{S} = l$ and in reciprocal space $\mathbf{S} = h/\mathbf{a},\ \mathbf{S} = k/\mathbf{b},\ \mathbf{S} = l/\mathbf{c}$, this becomes $\mathbf{S} = h \cdot \mathbf{a^*},\ \mathbf{S} = k \cdot \mathbf{b^*},\ \mathbf{S} = l \cdot \mathbf{c^*}$ (with $\mathbf{a^*} = 1/\mathbf{a},\ \mathbf{b^*} = 1/\mathbf{b},\ \mathbf{c^*} = 1/\mathbf{c}$).

If we take the incoming X-ray vector $\mathbf{s}_0$ of a length $1/\lambda$ and we draw a sphere with the same radius $(1/\lambda)$, the diffracted beam $\mathbf{s}$ will intercept the surface of the sphere at a point P. The vector between the origin O and P is the reciprocal space vector $\mathbf{S}$ originated by difference between the incoming $\mathbf{s}_0$ and the diffracted $\mathbf{s}$ vectors. $\mathbf{S}$ is perpendicular to the reflecting plane hkl and its length is $2sin\theta/\lambda = 1/d$. Diffraction will only occur when a reciprocal lattice point crosses the surface of the Ewald sphere.

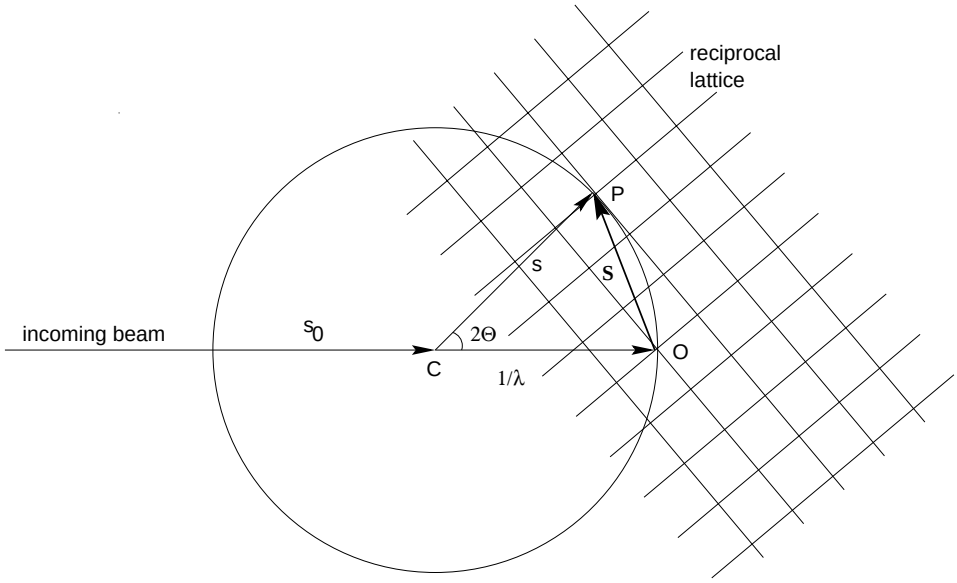


Figure A.10: The Ewald's sphere

The intensity ($\mathbf{I}(hkl)$) recorded during a diffraction experiment is the square of the structure factor $\mathbf{F}(hkl)$. The structure factor is a function of the electron density distribution in the unit cell and is therefore:

$$\mathbf{F}(hkl) = V \int_{x=0}^{1} \int_{y=0}^{1} \int_{z=0}^{1} \rho(xyz) \exp[\pi i(hx + ky + lz)] dxdydz$$

To calculate the electron density we need to write:

$$\rho(xyz) = \frac{1}{V} \sum_{h} \sum_{k} \sum_{l} |F(hkl)| \exp[-2\pi i(hx + ky + lz) + i\alpha(hkl)]$$

From the diffracted intensity only the amplitude of the structure factor can be derived but the phase information $\alpha(hkl)$ is lost making it impossible to calculate the electron density ρ.

A.3.1 The Patterson Function

If we write the Fourier summation to calculate the electron density map as a self convolution, with all the phase angle $\alpha = 0$ we obtain:

$$P(uvw) = \frac{1}{V} \sum_{hkl} |F(hkl)|^2 \cos[2\pi(hu + kv + lw)]$$

that is the so called Patterson function with the coordinates of the cell called u, v, w instead of x, y, z to distinguish the Patterson cell from the real cell. We can consider $P(\mathbf{u})(\mathbf{u} = uvw)$ as the convolution between the electron density $\rho(\mathbf{r})$ at the position $\mathbf{r}$ with the electron density at the interatomic distance vector $\mathbf{u}$ ($\rho(\mathbf{r}+\mathbf{u})$. The result of this convolution will be non zero only when $\rho(\mathbf{r}) \neq 0$ with $\mathbf{u}$ corresponding to an interatomic distance vector. Figure A.11 A represents a point atom structure, in which the atoms are positioned at points $\mathbf{a}$, $\mathbf{b}$, $\mathbf{c}$ in a two dimensional unit cell.

If we take the corners of the cell as origins of the Patterson map, a Patterson cell is obtained. The interatomic distance vectors between the atoms ($\mathbf{a} \rightarrow \mathbf{b},\ \mathbf{b} \rightarrow \mathbf{a},\ \mathbf{a} \rightarrow \mathbf{c},\ \mathbf{c} \rightarrow \mathbf{a},\ \mathbf{b} \rightarrow \mathbf{c},\ \mathbf{c} \rightarrow \mathbf{b}$) are then drawn from the origin peaks ($\mathbf{u}$=0) giving the corresponding Patterson map (Fig. A.11 B) The number of peaks in a Patterson map is N^2, reduced to $N \times (N-1)$ because N of them are located at the origin since the self distance vector $\mathbf{u}$ ($\mathbf{a} \rightarrow \mathbf{a},\ \mathbf{b} \rightarrow \mathbf{b},\ \mathbf{c} \rightarrow \mathbf{c}$) is $= 0$. For very small structures, with few atoms the Patterson map can be interpreted easily to reveal the structure but for proteins the map will be featureless because of the superposition of peaks. For

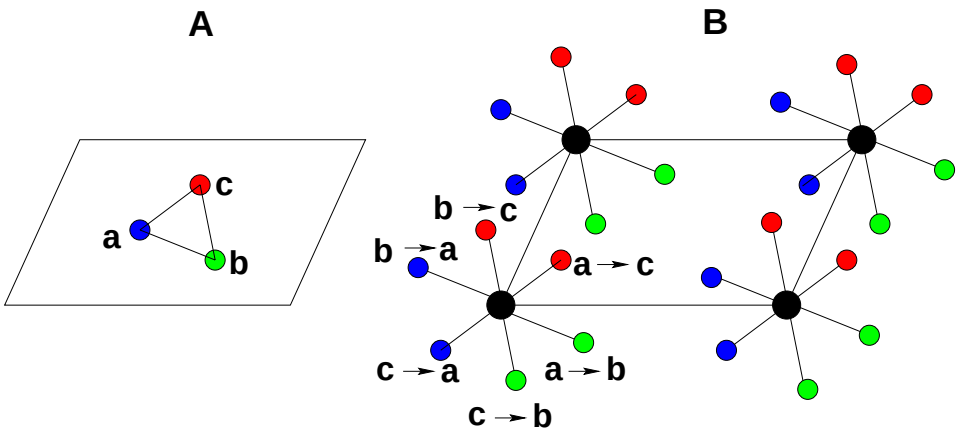


Figure A.11: A: The real cell, B: The Patterson map.

protein crystallography the Patterson map represents an important tool for structure determination as it is the basis for finding the heavy atom positions in the Isomorphous Replacement or Multiple Anomalous Dispersion Methods. It also represents the essence of the Molecular Replacement.

A.4 Data Collection

Diffraction from a crystal is obtained when Bragg's law is fulfilled. The recorded diffraction coming from the intersection between the Ewald sphere and the reciprocal lattice points is a series of elliptically arranged rings of spots arising from different planes in the crystal lattice. Since in this way only few points would fulfil Bragg's law (giving a so called "still" image) the rotation method has been introduced [135]. When the crystal is rotated different points in the reciprocal lattice cross the Ewald sphere and give rise to a diffraction pattern which is now given by all the reflections from two consecutive reciprocal lattice planes.

The diffraction spots fill the area between two rings. The two rings correspond to the start and the end of the rotation and the area between them is called "lune" (see Fig.A.12). During the oscillation some reflections will

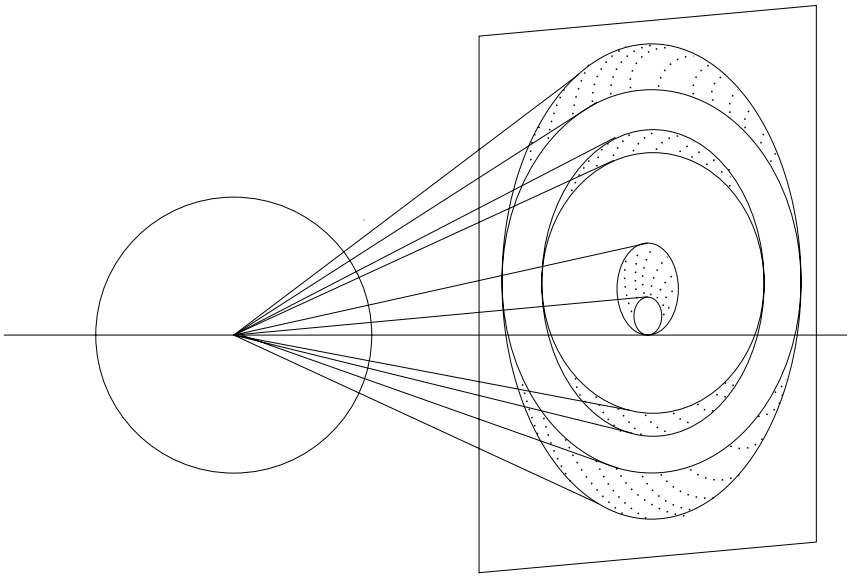


Figure A.12: The rotation method and, dashed the "lunes"

have passed completely the Ewald sphere and are called fully recorded reflections while some others will have just started diffracting in one image and finish diffracting in the following image. These partially recorded reflections are called "partials" and their total intensity is integrated by summation of the intensities recorded in sequential images.

The maximum angle of oscillation ($\Delta\phi$) is chosen as large as possible but avoiding the overlaps. Since the width of the lunes is proportional to the oscillation range and the gaps between lunes correspond to the spacing of the reciprocal lattice along the beam direction, increasing the $\Delta\phi$ would result in wider lunes that eventually would start overlapping. A simple formula to calculate the maximum $\Delta\phi$ between images is:

$$\Delta\phi = (180/\pi)(d/a)$$

where d is the resolution limit and a is the cell dimension (the one along the beam during the exposure). This formula does not take into account the mosaicity of the crystal and the beam divergence, thus the value of oscillation must be smaller than the calculated one. The best crystal orientation is with the longest cell dimension approximately along the spindle axis because it allows a larger oscillation angle to be used.

The completeness of a dataset may be affected by the so called blind region. This is the part of reciprocal space close to the spindle axis where even after a rotation range of 360° the reciprocal lattice points never cross the Ewald sphere. The blind region is smaller at low resolution and at shorter wavelength.

For a triclinic crystal (no symmetry) the only way to obtain a complete dataset is to use a goniometer arc to change the orientation of the crystal (skewing the rotation axis of more than 2θ) after having collected a first dataset and collect in the new orientation a second one. The missing reflections will be now far away from the blind region and will cross the Ewald sphere. When the crystal has symmetry it is usually easy to get the crystal oriented in such a way that the reflections have at least symmetry related reflections outside the blind region and are possible to measure.

Even if often people judge a dataset only by looking at the high resolution limit it is very important for the structure solution and refinement that the low resolution limit should be as low as possible (at least 10 Å) and to a completeness close to 100%. This is because low resolution data give information about the overall shape of the molecule and the solvent while high resolution

data provide information about the details.

The high resolution limit should also be as high as possible while still keeping a good signal to noise ratio. A good quality indicator is the I/σ ratio in which the intensity of the reflections is compared with the associated error. A general criteria is to keep all the reflections in the high resolution bin if at least 50% of them have their I/σ higher than 2. Less indicative of data quality is R_{symm} ($\Sigma[I^i - < I >]/\Sigma < I >$, where I^i is the individual intensity measurement and $< I >$ is the average intensity for this reflection with summation over all the data) because it depends on the redundancy of the data and crystal symmetry, being higher for higher symmetry. Well scaled data should have their χ^2 as close as possible to 1.

$$\chi^2 = \sum \frac{\left((I - < I >)^2\right)}{(\sigma^2 \cdot N/(N-1))}$$

It is also important to have good redundancy or multiplicity to get better statistics and estimates of intensities and errors.

Several computer programs have been developed in order to process diffraction data. Data analysis is divided in the following steps.

Preliminary analysis of the data. A visual inspection of the diffraction pattern tells if there are ovelaps, if the oscillation angle is too big or too small, if the beam stop has to be moved closer or further away, the diffraction limit, and if there are overloads allowing to choose the correct exposure time. It gives a general impression about crystal quality, if it is single, twinned, mosaic.

Dedicated software allows analysis of the diffraction pattern and the peak profile to visualise overloads and to determine the resolution limits. Experi-

ence is necessary to evaluate if the crystal is worth collecting and to determine all the data collection parameters ($\Delta\phi$, detector to crystal distance, exposure time).

Indexing of the diffraction pattern. The determination of the crystal lattice is usually done automatically (DENZO [65, 66], MOSFLM [136]). The program DENZO, part of the HKL package, allows automatic or interactive indexing of the diffraction pattern. One image is recorded and analysed by DENZO, some spots are selected and used to calculate the unit cell dimensions. This unit cell does not satisfy the standard lattice so a subroutine called "cell reduction" finds the parameters that according to the Tables for Crystallography are closest to the 14 Bravais lattices. In order to choose the best solution the program provides a distortion index indicating how much the cell parameters must be changed to satisfy the theoretical lattice and space group symmetry.

Refinement of data collection parameters. After cell reduction all the crystal and detector parameters have to be refined. The best thing to do is to provide the program with the most accurate parameters (e.g. crystal to detector distance). The program refines them either separately or simultaneously. The process is repeated for each image to be analysed and a file is written with all the refined parameters (crystal, detector, wavelength) and all the spot profiles.

Peak integration. When the spot positions are correctly determined the integration of the peak profile is now possible. The measured intensity is given by the summation between the Bragg peak and the background contribution.

Corrections are applied in order to estimate the value of I and its standard deviation σ_I. The total intensity for a crystal rotated with an angular speed ω is indeed given by:

$$I(int.\,hkl) = \frac{\lambda^3}{\omega \cdot V^2} \times \left(\frac{e^2}{mc^2}\right)^2 \times V_{cr} \times I_0 \times L \times P \times A \times |F(hkl)|^2$$

where $I(int.\,hkl)$ is the integrated intensity of the recorded reflection depending on the wavelength λ of the incident beam of intensity I_0, on a Lorentz factor L depending on the data collection technique, on a polarisation factor P (very important for synchrotron sources) and on the absorption factor A. V_{cr} is the volume of the crystal with V the volume of the unit cell. The charge e the mass of the electron m and the speed of the light c are constant terms. The intensity $|F(hkl)|^2$ is obtained on a relative scale by calculating:

$$I(int.hkl)/(L \times P \times A)$$

The correction for the polarisation and the Lorentz effect are included in the software for data processing, while the terms $(\lambda^3/\omega V^2) \times (e^2/mc^2)^2 \times V_{cr} \times I_0$ are a constant term characteristic for the experiment. The correction for the absorption effect is usually not required in the rotation method [134]. Errors can be introduced when the spots are too close to each other (e.g. high resolution data in a highly mosaic crystal). In this case integration and background measurement area intrudes upon the adjacent peaks leading to inaccurate estimate of the intensities.

Postrefinement of crystal and detector parameters, scaling and merging. Scaling program like SCALEPACK [65, 66] or SCALA [69] perform scaling and merging of different datasets and postrefinement of crystal and detector parameters. Fluctuations of beam intensity are taken into account. Postrefinement is carried out on the whole set of data and repeated iteratively

producing at the end statistical measures allowing judgement of the quality of the data ($I/\sigma, R_{symm}, R_{merge}, \chi^2$, completness, redundancy), and the refined cell dimensions. All the scaled reflections with their h, k, l I and σ_I are written to a separate file. During this last step of data processing partially recorded reflections are summed together, outliers are rejected and symmetry related reflections are merged giving a list of unique reflections.

Cryogenic data collection

Cryogenic data collection has been introduced to decrease radiation damage during crystal exposure to X-rays. Upon irradiation two different degradation processes take place in the crystal, the first one dose dependent and the second one time dependent. The dose dependent damage is a consequence of the direct irradiation of the sample with ejection of electrons from the molecules while the time dependent damage comes from the diffusion of free radicals formed upon exposure to X-rays (hydroxyl, hydroperoxyl, oxygen). The large amount of water present in protein crystals makes them very sensitive to radiation damage because of the fast diffusion of the free radicals.

Cooling the crystal to 100 K decreases the time dependent damage by avoiding diffusion. Cryogenics generally allow collection of a complete dataset from a single crystal and increases the resolution because the atoms thermal motions are slowed down. Another advantage is the possibility of testing the quality of the crystals on a conventional source and keeping them for later use at a synchrotron site.

The technique involves the use of cryoprotectant solutions to avoid the formation of crystalline ice that would give rise to powder diffraction rings and

would in certain cases destroy the crystal because of the increased volume from liquid to crystalline water. The cooling process must be very fast so that the water in the crystal becomes amorphous "vitreous" ice. The crystal is first soaked in a solution of mother liquor plus the cryoprotectant agent and then "fished" out with a loop (usually of nylon) and transferred on the goniometer head in a cold nitrogen stream (100 K). Table A.2 reports some substances used for cryoprotection.

Compound	Concentration (%)
Glycerol	13-30(v/v)
Ethylene glycol	11-30(v/v)
Polyethylene glycol (PEG) 400	25-35(v/v)
Xylitol	22(w/v)
Butane-2,3-diol	8(v/v)
Erythritol	11(w/v)
Glucose	25(w/v)
2-methyl-2,4-pentanediol (MPD)	20-30 (v/v)

Table A.2: Compounds used as cryoprotectant in protein crystallography [137].

The quality of the data is usually improved using cryogenics if compared with crystals mounted in conventional glass capillaries in which scattering from the mother liquor and the glass wall are the major source of background scattering and absorption. Nevertheless a small increase in crystal mosaicity upon freezing is often observed. The typical range of mosaicity for frozen crystals is

of the order of 0.25-0.6°. Small and fragile crystals can be easily handled with this technique allowing data collection otherwise impossible in a capillary. An extensive overview of cryocrystallography is found in [138].

A.5 The Phase Problem, How to solve it?

The main problem in protein crystallography after having recorded good data is the lack of phase information. Since a diffraction pattern can be considered as a convolution of the diffraction pattern of the molecule (the molecular transform) with the diffraction pattern of the lattice (the reciprocal lattice) the problem of reconstructing the image (the electron density of the molecule) is solved by mathematical methods (with Fourier transforms) but the phases needed for the calculation have to be determined experimentally or mathematically for small molecules or small proteins at atomic resolution.
The possible ways to obtain the phases are:

- Multiple Isomorphous Replacement (MIR) or Single Isomorphous Replacement (SIR) in which heavy atom reagents are used to react with the protein in the crystal. The presence of heavier scatterers in the crystal is then exploited to determine the initial phases.

- Multiple Anomalous Dispersion (MAD) or Single Anomalous Dispersion (SAD) in which anomalous scatterers naturally present or inserted in the crystal are used to obtain the phases. The use of selenomethionine is becoming routine because it is relatively easy to incorporate with molecular biology techniques.

- Molecular Replacement (MR) in which a known protein structure can be used as a starting model because of its similarity (e.g. sequence or fold

identity) to get the set of starting phases.

- Direct Methods in which calculations are performed on a first set of random phases, then refined and updated repeatedly. It is not very succesful for big structures in the absence of heavy atoms and requires large computational facilities.

A.5.1 Isomorphous Replacement

The introduction of heavy atom containing compounds in the crystal is obtained either by soaking or by cocrystallisation.

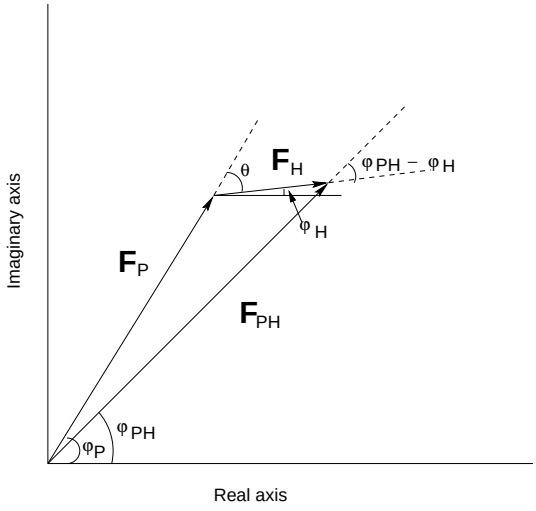


Figure A.13: Argand diagram representing the relationship between $\mathbf{F}_P$, $\mathbf{F}_{PH}$ and $\mathbf{F}_H$ for two perfecly isomorphous crystals

Provided the introduction of the heavy atoms does not alter the structure of the protein or the crystal packing the derivative is called isomorphous, since there are only small local changes in the crystal lattice. A change in the cell dimension between native and derivative up to 1.5 % can be accepted for data at 3 Å while for data at 2 Å makes the structure solution very difficult [139]. Another criteria of acceptance is that the change in cell dimension should be smaller than $d_{min}/4$ (d_{min} is the resolution limit) [134]. The difficulty of obtaining isomorphous derivatives with conventional means is circumvented when xenon derivatives are prepared. Under moderate pressure xenon can bind to proteins forming weak non covalent but specific interactions.

Its use in solving proteins has been recently reviewed [140].

In the isomorphous replacement method, which is based on the difference of magnitude between native and derivative structure factors, two data sets, one for the native and one for the derivative crystal are recorded. If $\mathbf{F}_P$ is the structure factor for the native crystal and $\mathbf{F}_{PH}$ the structure factor for the derivative $\mathbf{F}_{PH} = \mathbf{F}_P + \mathbf{F}_H$ where $\mathbf{F}_H$ is the heavy atom contribution to the scattering (Figure A.13).

The first step is the determination of the heavy atom positions. For this purpose a Patterson function is calculated using as coefficients $(|F_{PH}| - |F_P|)^2$ that means using the intensities coming from the derivative $(F_{PH})^2$ and the native $(F_P)^2$ data sets. This difference Patterson will give a map showing only the peaks coming from the heavy atom contribution (being $\mathbf{F}_H = \mathbf{F}_{PH} - \mathbf{F}_P$) and thus their position in the cell. The protein phase angle φ_P is obtained by trigonometric calculations:

$$F_{PH}^2 = F_P^2 + F_H^2 - 2F_P F_H \cos\theta$$

since θ is $\varphi_P - \varphi_H$

$$\varphi_P = \varphi_H + \cos^{-1}[(F_{PH}^2 - F_P^2 - F_H^2)/2F_P F_H]$$

The presence of a cosine function tells us that φ_P can have two values (e.g. $\cos^{-1} 0 = 90°$ or $270°$), and introduces a phase ambiguity (Fig. A.14 A). The phase ambiguity can be solved with a second derivative in which a different heavy atom compound is bound in positions different from those in the previous one. The combination of three datasets (one native and two derivatives) is shown in figure A.14 B. The "correct" phase is given mathematically by solving

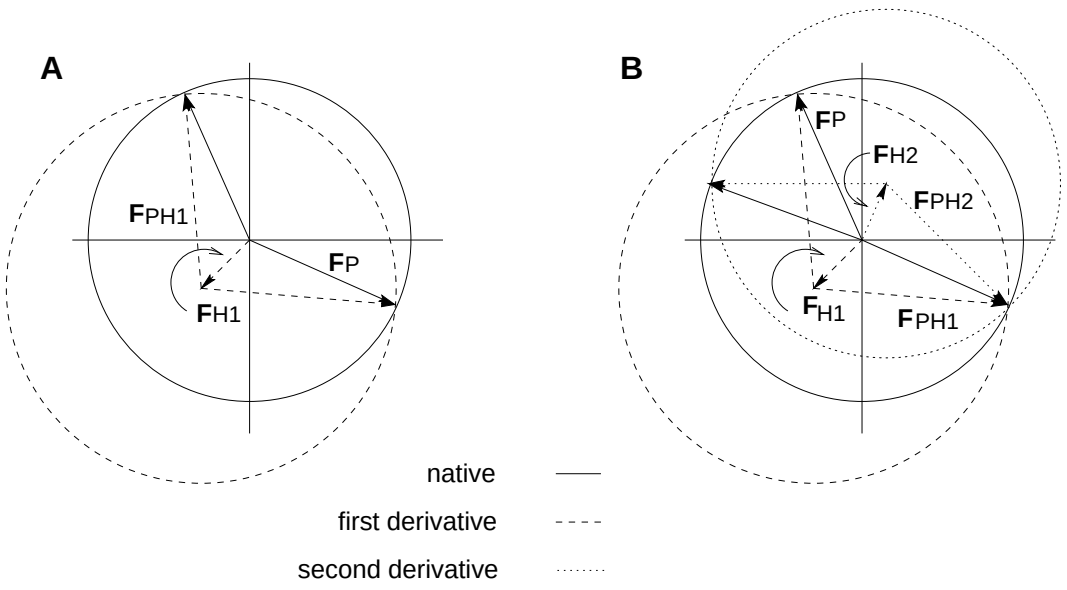


Figure A.14: A: SIR and phase ambiguity. B: Two derivatives to solve the phase ambiguity in MIR

the two equations:

$$\varphi_P = \varphi_{H1} + \cos^{-1}[(F_{PH1}^2 - F_{P1}^2 - F_{H1}^2)/2F_{P1}F_{H1}]$$

$$\varphi_P = \varphi_{H2} + \cos^{-1}[(F_{PH2}^2 - F_{P2}^2 - F_{H2}^2)/2F_{P2}F_{H2}]$$

The "correct" value of the phase angle φ_P is not obtained because there are many sources of errors that give rise to a "lack of closure" of the vector diagram (Fig. A.13, therefore only an approximate value for the phase can be determined. Possible sources of errors are: inaccuracy in measurement or data processing, leading to errors in intensities; errors in locating the heavy atom positions and in their occupancy and B factors, and last but not least lack of isomorphism between the native and the derivative crystals [139]. Successful variations of MIR and SIR are represented by MIRAS (Multiple Isomorphous Replacement and Anomalous Scattering) and SIRAS (Single Isomorphous Re-

placement and Anomalous Scattering). The phase information is not only provided by the heavy atom but also by its anomalous signal as in Multiwavelength Anomalous Dispersion (MAD) or Single Anomalous Dispersion (SAD).

A.5.2 MAD or SAD?

The possibility of getting the phases using Multiwavelength Anomalous Dispersion (MAD) or Single Anomalous Dispersion (SAD) experiments rely on the presence of anomalously scattering atoms in the crystal. The presence of metals in the protein, the insertion of heavy atoms, or of seleniomethionines by molecular biological techniques, allows exploitation of the so called "breakdown" of Friedel's law (Fig. A.15 B).

In the absence of anomalous scatterers the structure factor vectors $\mathbf{F}_{(hkl)}$ and $\mathbf{F}_{(-hkl)}$ have the same magnitude (thus $F_{(hkl)} = F_{(-hkl)}$ and $\alpha_{(hkl)} = -\alpha_{(-hkl)}$ with the recorded intensity $\mathbf{I}_{(hkl)} = \mathbf{I}_{(-hkl)}$ (Fig. A.15 A). When anomalous scatterers are present Friedel's law no longer applies and the structure factors differ because of the variation of the scattering factors:

$$f_{anom.} = f^{\circ}(\theta) + f'(\lambda) + if''(\lambda)$$

The normal scattering component f° is real, independent of the wavelength and decreases as the scattering angle 2θ increases. The anomalous scattering factor $f_{anom.}$ in contrast is complex since it includes a phase shift sharply dependent on the incident wavelength and independent of the scattering angle because it arises from interaction between X-rays and inner core electrons (K or L orbitals) [141]. Because of this dependence on the wavelength λ, Hendrickson [142] proposed to separate the contribution to the scattering in two components one due to normal and the other one due to anomalous scattering. The total

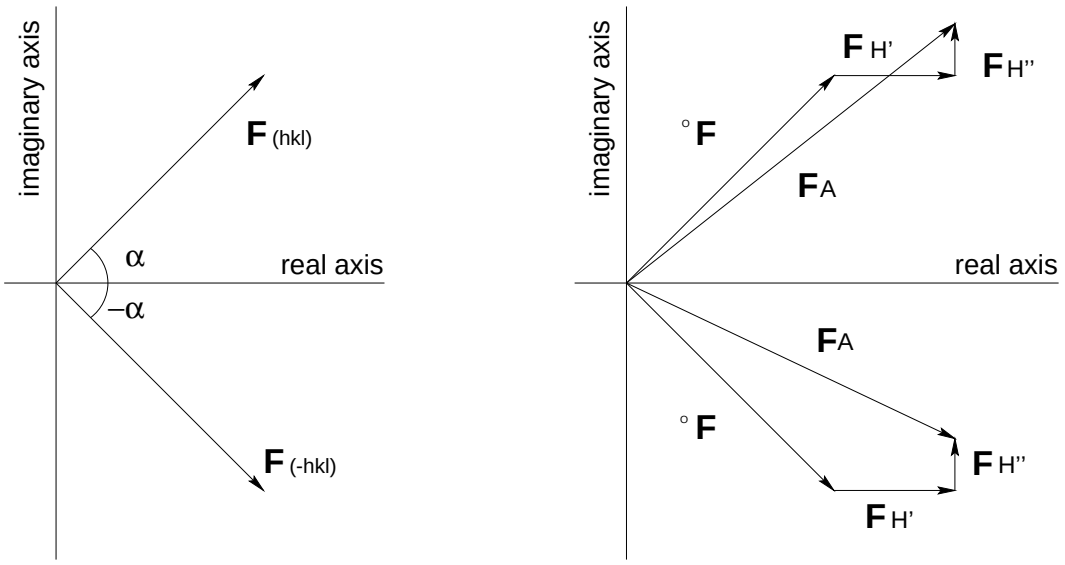


Figure A.15: A: the Friedel's law, $\mathbf{F}_{(hkl)}$ and $\mathbf{F}_{(-hkl)}$ are the Friedel pair. B: the "breakdown" of the Friedel law, $\mathbf{F}_{(hkl)}$ and $\mathbf{F}_{(-hkl)}$ are Bijvoet mates. $^{\circ}\mathbf{F}$ in presence of normal scattering, $\mathbf{F}_A$ and the two components $\mathbf{F}'_H$ and $\mathbf{F}''_H$.

scattering ${}^{\lambda}F_T$ is given by the contribution of all the scattering components at the wavelength λ.

$$|{}^{\lambda}F_T(\mathbf{h})|^2 = |{}^{\circ}F(\mathbf{h})|^2 + a(\lambda)|{}^{\circ}F_A(\mathbf{h})|^2$$

$$+b(\lambda)|{}^{\circ}F(\mathbf{h})||{}^{\circ}F_A(\mathbf{h})|\cos[{}^{\circ}\varphi(\mathbf{h}) - {}^{\circ}\varphi_A(\mathbf{h})]$$

$$+c(\lambda)|{}^{\circ}F(\mathbf{h})||{}^{\circ}F_A(\mathbf{h})|\sin[{}^{\circ}\varphi(\mathbf{h}) - {}^{\circ}\varphi_A(\mathbf{h})]$$

Where $\mathbf{h}$ denotes a reflection (hkl), and

$a(\lambda) = (f'^2 + f''^2)/f^{\circ 2}$
$b(\lambda) = 2(f' + f^{\circ})$
$c(\lambda) = 2(f''/f^{\circ})$

The suffix $^\circ$ denotes the normal wavelength independent scattering for the protein ($^\circ F$) and for the anomalous scatterer ($^\circ F_A$). The coefficients f' and f'' can be determined experimentally and with appropriate measurements at different wavelengths the values of $|^\circ F(\mathbf{h})|$, of $|^\circ F_A(\mathbf{h})|$ and $^\circ\varphi(\mathbf{h}) -^\circ \varphi_A(\mathbf{h})$ can be determined.

The availability of tunable synchrotron beam lines allows the users to vary the two components f' and f'' of the anomalous scattering factor by varying the wavelength of the incoming beam close to the absorption edge characteristic for each element. The most useful and accessible range of energies is from 7 to 17 keV (from 1.7 to 0.7 Å E(keV)= $12.3985/\lambda$). To optimise the dispersive signal f' the energy (thus the wavelength) is moved to the edge inflection point, while for the anomalous difference f'' right above the edge. A third wavelength, so called remote, is usually collected at higher energy (shorter wavelength) also for collecting higher resolution data. In order to speed up the data collection a MAD experiment can be carried out using only two wavelengths [143] of vital importance when collecting on radiation sensitive crystals.

The ultimate reduction on the number of wavelength to collect is the SAD case where a single dataset is recorded at a wavelength with a good f'' followed by density modification to solve the phase ambiguity [126, 143].

A.5.3 Molecular Replacement

The increasing number of known protein structures makes the Molecular Replacement (MR) technique very popular since it exploits the existence of a known model structure to solve the phase problem of an unknown structure.

The first requirement is the similarity between the unknown and the known structure. The choice criteria is based on sequence identity relying on the fact that homologous proteins with similar sequence share a common fold. The sequence similarity does not need to be very high and the model can be truncated as demonstrated for pig liver γ-aminobutyrate aminotransferase which only shares 23% sequence identity with human ornithine aminotransferase and for which the model included only 24% of the atoms of the two dimers contained in the asymmetric unit [144].

The Patterson function, the basic tool in MR, represents a vector map in which interatomic distance vectors are represented by peaks of positive density. When the interatomic distances are between atoms inside the molecule they are self-Patterson vectors while when between atoms belonging to different molecules in the unit cell they are cross-Patterson vectors. The distinction between self and cross vectors is fundamental to MR since similar or identical molecules will give similar or equal Patterson map respectively, apart from a rotational and a translational term. This is because the self-vectors give a representation of the molecule itself, while cross-vectors are useful in finding the position of the model in the unit cell because they are related to intermolecular atomic distances. Molecular replacement is therefore divided in two steps one to find the right orientation, solved with the rotation function, and the other one to find the correct position solved with the translation function.

The rotation function was first proposed by Rossman and Blow in 1962 as the convolution between the Patterson function calculated with the squared structure factor amplitudes of the unknwon structure and the Patterson function calculated from the model rotated around the origin. The radius of calculation

is related to the dimension of the model since it is limited to the self Patterson vectors and identifies the volume of integration.

$$R(\mathbf{C}) = \int_U P_{obs}(\mathbf{u})\, P_{calc}(\mathbf{Cu})\, \mathrm{d}u$$

where P_{obs} and P_{calc} are the Patterson function for the observed data and for the model respectively, $\mathbf{C}$ is the operator rotating the coordinate system of P_{calc} with respect to P_{obs} and centre in the origin of the Patterson cell, U is the volume of integration. When $R(\mathbf{C})$ is maximised a potential solution is found since the two Patterson functions superimpose their maxima. When the rotation function has found its maximum the model has to be positioned in real space in the unit cell. This is done by translating the model through the cell in the x, y, z directions. The translation function as proposed by Crowther and Blow in 1967 is the convolution between the Patterson function of the observed data (P_{obs}) at the point $\mathbf{u}$ and the Patterson function of the model (P_{calc}) now translated according to the cross-vectors (intermolecular atomic vectors) at the point $\mathbf{u}$+$\mathbf{t}$.

$$T(\mathbf{t}) = \int_U P_{obs}(\mathbf{u})\, P_{calc}(\mathbf{u}+\mathbf{t})\, \mathrm{d}u$$

The program package AMoRe (Automated Molecular Replacement) developed by Navaza [68] is divided into four parts, the first to prepare the data for the calculation (SORTING and TABLING) the second performing the rotation function (ROTING), giving the results sorted according to the R-factor (R).

$$R = \frac{\sum_{hkl} ||F_{obs}| - k|F_{calc}||}{\sum_{hkl} |F_{obs}|}$$

The best solutions are then input to the translation function (TRAING) which gives a list of peaks with both R factor and correlation coefficient (C).

$$C = \frac{\sum_{hkl}(|F_{obs}|^2 - \overline{|F_{obs}|^2}) \times (|F_{calc}|^2 - \overline{|F_{calc}|^2})}{\sqrt{[\sum_{hkl}(|F_{obs}|^2 - \overline{|F_{obs}|^2})^2 \times \sum_{hkl}(|F_{calc}|^2 - \overline{|F_{calc}|^2})^2]}}$$

When a solution is found the last part of the package perfomes a rigid body fitting refinement (FITING). The rotation and translation are applied to the model which is then used for the phase calculation. It should be kept in mind that the initial phases are biased from the starting model, bias usually being removed during successive cycles of refinement and model building.

A.5.4 Direct Methods

Although very succesful for small molecule crystallography [134] their application in solving protein structures is not yet usually possible either for computer time requirement or for the low ratio of measurement to parameters due to limited resolution.

A successful application of direct methods in protein crystallography has been carried out by Hauptman with the program *Shake and Bake* (SnB) [145], with the assumption of the atomicity of the model, the positivity of the electron density and phase relations based on probability theory. Even though the method succeeds, it is so far limited to structures with no more than 600 atoms and with very high resolution data (1.2 Å or better). Structure factors are calculated from a starting model based on random atoms placed in the unit cell at a minimum distance from one other of 1.2 Å. Their phases are refined and the resulting electron density map is analysed. The highest peaks are kept and used as atoms to generate new phase values for the next cycle. The convergence is reached in about $n/2$ cycles with n the expected number of atoms in the structure.

A.6 Refinement and Model Building

Once the initial phases (e.g. from MIR or MAD) or an initial model (e.g. from MR) are available they have to be refined against the experimental data and the final structure has to be built. The aim of refinement is to find a model in which the position of the atoms gives calculated structure factors (F_{calc}) as close as possible to the observed structure factors (F_{obs}). The general criteria from least-squares refinement is to minimise the difference:

$$\sum_{hkl} w(hkl)\Big(|F_{obs}(hkl)| - |F_{calc}(hkl)|\Big)^2$$

where w is a weight assigned to each (hkl) and is inversely proportional to the square of its standard deviation σ.

Protein structure refinement has peculiar problems: big unit cell, large number of reflections with low signal to noise ratio, and limited resolution. All these problems are due to the characteristic nature of proteins crystals in which the molecules are held together by weak interactions with very high solvent content and with large motility of their atoms and flexible loops. Most protein crystals do not diffract to atomic resolution thus the ratio of observation to parameters is low and allows the refinement of only the positional x, y, z atomic parameter and a single isotropic temperature factor. For data extending to about 2.8 Å the ratio between number of reflections and parameters to be refined (x, y, z, B=4 × the number of atoms) is close to 1 which is almost the border-line for the refinement. Since the X-ray data from protein crystals are limited, stereochemical information such as bond lengths, bond angles, planar groups and contact distance [74] can be used as additional sources of information [146] in the form of constraints or restraints (a constraint being an exact condition while a restraint is an approximation to that condition).

Least-squares refinement has been used for many years [147, 148]. The main limitation is the possibility of getting trapped in local minima, when the model is not very good or not complete. To overcome this problem different approaches have been chosen. The program X-PLOR includes a molecular dynamics algorithm which exploits a simulated annealing technique. In the simulation the temperature is increased and the atoms are allowed to move freely from their original position, the temperature is then slowly cooled down allowing the structure to rearrange and eventually find a global minimum [149]. Murshudov introduced Maximum Likelihood, first proposed by Bricogne [150], as the main approach in the program REFMAC [70] using stereochemical restraints as prior knowledge. Maximum Likelihood refinement applies Bayes theorem and since the ratio between the experimental data and the number of parameters to be estimated is small it is essential to use prior information knowledge in the form of stereochemical restraints derived from models (e.g. bond length and angles). Bayes theorem can be written as:

$$P(x;|F|) = p(x)P(|F|;x)/P(|F|) = p(x)L(x;|F|)$$

where P is the posterior probability distribution of parameters when the data are known, $|F|$ are the experimental data, x is the parameter to be estimated, p is the prior knowledge (the parameters known before the experiment) and L is the likelihood function. To estimate the parameter x the posterior knowledge P needs to reach its maximum [70].

The progress of refinement can be monitored with the use of R and R_{free} factors usually defined as a percentage.

$$R = \frac{\sum_{hkl} ||F_{obs}| - k|F_{calc}||}{\sum_{hkl} |F_{obs}|}$$

The use of R_{free} was proposed by Brünger [72, 73] as a tool to check whether the refinement is correct or if the improvement of the R factor is only an artefact due to overfitting of the data (e.g. noise in the map fitted by solvent molecules). The idea is to exclude a fraction of data (typically from 5 to 10%) from the refinement process as a cross validation tool. Typical values of R factor for protein structures range between 10 to 25 % with R_{free} usually higher by about 3-5% depending of resolution. Higher discrepancy between R and R_{free} are often sign of overfitting.

When atomic resolution data are available anisotropic temperature factor refinement can be carried out. The parameters to refine are now three positional (x, y, z), and six defining the thermal ellipsoids described by the displacement of the atom about its average position. Anisotropic refinement is usually performed with SHELX [115], a program originally for small molecule crystallography adapted to work with proteins. It is based on least-squares refinement using I instead of F. The advantage of using intensities are the possibility of using all the reflections, strong weak and even negative and the easier estimate of σ_I compared to σ_F. The main disadvantage is the need for a very good starting model because the least-squares minimisation is not able to improve poor quality models. The limited speed is another reason why SHELX is used by protein crystallographers only at the latest stage of refinement, only when isotropic refinement has converged. Anisotropic refinement has been recently introduced in REFMAC [151] which uses fast Fourier transforms for the required structure factor calculations.

Graphics programs [71, 152] are used to visualise the protein model, the electron density maps and for manual building. Automatic programs are being

developed to shorten the time of manual intervention. The ARP-WARP suite [75, 76, 153, 154] exploits the difference Fourier map $mF_{obs} - DF_{calc}$ to find and position dummy water atoms, and checks in the $3F_{obs} - 2F_{calc}$ or $2mF_{obs} - DF_{calc}$ if the atoms are placed correctly in good electron density regions. The following steps are automatic model interpretation and chain tracing.

A.7 Validation and Deposition

Structure deposition at the Protein Data Bank (PDB [155, 156] is the final and fundamental step once a structure is refined. The access to the data at the PDB is provided to the scientific community which can get structural information as atomic coordinates and structure factors. The availability of the data is of enormous importance not only for the crystallographers interested in solving a similar protein with MR or comparing structures but also for biochemists and biologists investigating the structure and function relationship of proteins. The importance of the validation of structures before deposition is of course fundamental. Mistakes introduced in the PDB can lead to biased statistical distribution of protein parameters. Validation is essentially necessary either to assess the quality of the refined structure and to allow users to discriminate between good and bad data and to which degree of reliability they can trust the model. The target values for the geometric restraints are based on analysis of X- ray structures of amino acids and peptides [74] deposited in the Cambridge Structural Databank (CSD) but they do not take into account the natural variability found in proteins in which the amino acids have specific different environment. Validation packages such as PROCHECK [77] and WHATIF (the last used by the PDB during the deposition/validation procedure) utilise statistical information derived from the PDB itself. The danger

is to force protein parameters within "normal" limits in order to be within the known criteria of acceptance, obscuring in this way novel conformational features [116]. Validation software analyses in detail the stereochemistry of a protein structure producing a series of easily interpretable plots, checks for the correct naming of atoms (according to IUPAC convention), for "bumps" between atoms and for correct peptide geometry. A Ramachandran plot [157] is always included showing the distribution of ϕ and ψ torsional angles. The values for these angles tend to group in zones of the plot called respectively most favourable, allowed and disallowed conformational regions. A good model is supposed to have over 90% of the residues in the most favourable regions. The presence of outliers is sometimes a sign of a new conformation rather than a mistake and the information should be carefully examined. WHATIF suggests side chain flips in order to satisfy all hydrogen bond donors and acceptors whenever possible.

After all protein crystallography can be FUN!

Appendix B

Publications

1. S. Benini, S. Ciurli, W. R. Rypniewski, K. S. Wilson, and S. Mangani. Crystallization and preliminary high-resolution X-ray diffraction analysis of native and β-mercaptoethanol-inhibited urease from *Bacillus pasteurii*. *Acta Cryst.*, D54:409-412, 1998.

2. S. Benini, W. R. Rypniewski, K. S. Wilson, S. Ciurli, and S. Mangani. The complex of *Bacillus pasteurii* urease with β-mercaptoethanol from X-ray data at 1.65-Å resolution. *J. Biol. Inorg. Chem.*, 3:268-273, 1998.

3. S. Benini, W. R. Rypniewski, K. S. Wilson, S. Miletti, S. Ciurli, and S. Mangani. A new proposal for urease mechanism based on the crystal structure of the native and inhibited enzyme from *Bacillus pasteurii*: why urea hydrolysis costs two nickels. *Structure*, 7:205-216, 1999.

4. S. Benini, W. R. Rypniewski, K. S. Wilson, S. Miletti, S. Ciurli, and S. Mangani. The complex of *Bacillus pasteurii* urease with acetohydroxamate anion from X-ray data at 1.55 Å resolution. *J. Biol. Inorg. Chem.*, 1:110-118, 2000.

5. S. Benini, S. Ciurli, W. R. Rypniewski, and K. S. Wilson. Crystals of cytochrome *c*-553 from *Bacillus pasteurii* show diffraction to 0.97 Å. *Proteins: Struc. Func. Genetics*, 28:580-585, 1997.

6. S. Benini, A. Gonzalez, W. R. Rypniewski, K. S. Wilson, J. J. Van Beeumen, and S. Ciurli. Crystal structure of oxidised *B. pasteurii* cytochrome *c*-553 determined by *ab initio* and by MAD methods: structure-function relationships and molecular evolution at 0.97 Å resolution. *submitted*

Appendix C

List of Abbreviations

Å	Ångström
a, b, c	unit cell axial lengths
a, **b**, **c**	unit cell vectors of the direct lattice
a*, **b***, **c***	unit cell vectors of the reciprocal lattice
AHA	acetohydroxamic acid
B-factor	Temperature factor
BME	β-mercaptoethanol
Da	Dalton
DAP	Diamidophosphate
e.s.u.	estimated standard uncertainties
f	atomic scattering factor
$\mathbf{F}_{hkl}$	structure factor for the reflection hkl
F_{calc}	calculated structure factor amplitudes
F_{obs}	observed structure factor amplitudes

$\mathbf{F}_{PH}$	structure factor for protein and heavy atom
$\mathbf{F}_{P}$	structure factor for protein
$\mathbf{F}_{H}$	structure factor for heavy atom
hkl	Miller indices
I	Intensity of diffraction
kDa	kiloDalton
K_i	inhibition constant
L_p	Lorentz polarisation factor
MAD	multiple anomalous dispersion
MEA	mercaptoethanolamine (cysteamine)
MIR	multiple isomorphous derivatives
NCS	non crystallographic symmetry
P_{uvw}	Patterson function
PDB	Protein Data Bank
PEG	polyethylene glycol
RMSD	root mean square deviation
R factor	reliability factor
R_{free}	reliability factor for the $free$ set of reflection
R_{merge}	R factor between merged symmetry related reflections
V	unit cell volume
V_m	specific volume
x, y, z	atomic coordinates
α_{hkl}	phase angle for the hkl reflection
λ	wavelength of the radiation
θ	angle of incidence
ρ_{xyz}	electron density at the point x, y, z
σ	standard deviation

Bibliography

[1] J. B. Sumner. The isolation and crystallization of the enzyme urease. *J. Biol. Chem.*, 69:435–441, 1926.

[2] N. E. Dixon, C. Gazzola, J. J. Watters, R. Blakeley, and B. Zerner. Inhibition of jack bean urease (E.C. 3.5.1.5) by acetoxydroxamic acid and by phosphoramidate. An equivalent weight for urease. *J. Am. Chem. Soc.*, 97:4130–4131, 1975.

[3] N. E. Dixon, C. Gazzola, R. Blakeley, and B. Zerner. Jack bean urease (E.C. 3.5.1.5). A metalloenzyme. A simple biological role for nickel? *J. Am. Chem. Soc.*, 97:4131–4133, 1975.

[4] R. Von Burg. Toxicology update. *J. of Apll. Toxicol.*, 17:425–431, 1997.

[5] S. W. Ragsdale. Nickel biochemistry. *Curr. Opin. in Chem. Biol.*, 2:208–215, 1998.

[6] L. F. Wu. Putative nickel binding sites of microbial proteins. *Res. Microbiol.*, 143:347–351, 1992.

[7] R. P. Hausinger. Nickel utilization by microorganisms. *Microbiol. Rev.*, 51:22–42, 1987.

[8] L. Wolfram, B. Friedrich, and T. Eitinger. The *Alcaligenes eutrophus* protein HoxN mediates nickel transport in *Escherichia coli*. *J. Bacteriol.*, 177:1840–1843, 1995.

[9] H. L. T. Mobley, R. M. Garner, and P. Bauerfeind. *Helicobacter pylori* nickel-transport gene *nix*A: synthesis of catalytically active urease in *Escherichia coli* independent of growth conditions. *Mol. Microbiol.*, 16:97–109, 1995.

[10] R. P. Hausinger. Metallocenter assembly in nickel-containing enzymes. *J of Biol. Inorg. Chem.*, 2:279–286, 1997.

[11] H.L.T. Mobley and R.P. Hausinger. Microbial urease: significance, regulation and molecular characterization. *Microbiol. Rev.*, 53:85–108, 1989.

[12] R.L. Blakeley and B. Zerner. Jack bean urease: the first nickel enzyme. *J. of Mol. Cat.*, 23:263–292, 1984.

[13] S. Ciurli, C. Marzadori, S. Benini, S. Deiana, and C. Gessa. Urease from the soil bacterium *Bacillus pasteurii*: immobilization on capolygalacturonate. *Soil Biol. Biochem.*, 28:811–817, 1996.

[14] J. C. Polacco and M. A. Holland. Roles of urease in plant cells. *Int. Rev. Cytol.*, 145:65–103, 1993.

[15] H. L. T. Mobley, M. D. Island, and R. P. Hausinger. Molecular biology of microbial ureases. *Microbiol. Rev.*, 59:451–480, 1995.

[16] I. Park and R. P. Hausinger. Evidence for the presence of urease apoprotein complexes containing UreD, UreF and UreG in cells that are competent for *in vivo* enzyme activation. *J. Bacteriol.*, 177:1947–1951, 1995.

[17] I. Park and R. P. Hausinger. Requirement of carbon dioxide for in vitro assembly of the urease nickel metallocenter. *Science*, 267:1156–1158, 1995.

[18] M. B. C. Moncrief and R. P. Hausinger. Purification and activation properties of UreD-UreF-urease apoprotein complexes. *J. Bacteriol.*, 178:5417–5421, 1996.

[19] M. B. C. Moncrief and R. P. Hausinger. Characterisation of UreG, identification of a UreD-UreF-UreG complex, and evidence suggesting that a nucleotide-binding site in UreG is required for *in vivo* metallocenter assembly of *Klebsiella aerogenes* urease. *J. Bacteriol.*, 179:4081–4086, 1997.

[20] R. P. Soriano, A. Hausinger. GTP-dependent activation of urease apoprotein in complex with the UreD, UreF and UreG accessory proteins. *Proc. Natl. Acad. Sci. USA*, 96:11140–11144, 1999.

[21] S. M. T. Taha, T. G Brayman, P. A. Karplus, and R. P. Hausinger. *Transition metals in microbial methabolism*, chapter 14, pages 391–413. Harwood Academic Publishers, Amsterdam, 1997.

[22] G. J. Colpas, T. G. Brayman, L. Ming, and R. P. Hausinger. Identification of metal binding residues in the *Klebsiella aerogenes* urease nickel metallochaperone UreE. *Biochemistry*, 38:4078–4088, 1999.

[23] J. You, B. Song, J. Kim, M. Lee, and S. Kim. Genetic organization and nucleotide sequencing of the *ure* gene cluster in *Bacillus pasteurii*. *Mol. Cells*, 5:359–369, 1995.

[24] G. Mörsdorf and H. Kaltwasser. Ammonium assimilation in *Proteus mirabilis, Bacillus pasteurii* and *Sporosarcina ureae*. *Arch. Microbiol.*, 152:125–131, 1989.

[25] C. M. Collins and S. E. F. D'Orazio. *Mol. Microbiol.*, 9:907–913, 1993.

[26] M. Marchetti, B. Arico, D. Burroni, N. Figura, R. Rappuoli, and P. Ghira. Development of a mouse model of *Helicobacter pylori*. *Science*, 267:1655–1658, 1995.

[27] T. Chiyoda, Iida K., Takatori K., and M. Kajiwara. Screening system for urease inhibitors using 13C-NMR. *Chem. Pharm. Bull.*, 46:718–720, 1998.

[28] R.B. Blakeley, J. A. Hinds, H. E. Kunze, E. C. Webb, and B. Zerner. Jack bean urease (EC 3.5.1.5). Demostration of a carbamoyl-transfer reaction and inhibition by hydroxamic acids. *Biochemistry*, 8:1991–2000, May 1969.

[29] R.B. Blakeley, E. C. Webb, and B. Zerner. Jack bean urease (EC 3.5.1.5). A new purification and reliable rate assay. *Biochemistry*, 8:1984–1990, 1969.

[30] M. W. Weatherburn. Phenol-hypochlorite reaction for determination of ammonia. *Anal. Chem.*, 39:971–974, July 1967.

[31] M. J. Todd and R. P. Hausinger. Competitive inhibitors of *Klebsiella aerogenes* urease. mechanism of interaction with the nickel active site. *J. Biol. Chem.*, 264:15835–15842, 1989.

[32] M. A. Pearson, L. O. Michel, R. P. Hausinger, and P. A. Karplus. Structures of Cys319 Variants and acetohydroxamate-Inhibited *Klebsiella aerogenes* Urease. *Biochemistry*, 36:8164–8172, 1997.

[33] N. E. Dixon, R. L. Blakeley, and B. Zerner. Jack bean urease (E.C. 3.5.1.5). III. the involvement of active-site nickel ion in inhibition by βb-mercaptoethanol, phosphoramidate, and fluoride. *Can. J. Biochem.*, 58:481–488, 1980.

[34] R. K. Andrews, A. Dexter, R. L. Blakeley, and B. Zerner. Jack bean urease (E.C. 3.5.1.5). VIII. On the inhibition of urease by amides and esters of phosphoric acid. *J. Am. Chem. Soc.*, 108:7124–7125, 1986.

[35] S. Benini, C. Gessa, and S. Ciurli. *Bacillus pasteurii* urease: an heteropolymeric enzyme with a binuclear nickel active site. *Soil Biol. Biochem.*, 28:819–821, 1996.

[36] G. W. McCarty, J. M. Bremner, and J. S. Lee. Inhibition of plant and microbial ureases by phosphoroamides. *Plant and soil*, 127:269–283, 1990.

[37] A.D. Larson and R.E. Kallio. Purification and properties of bacterial urease. *J. Bacteriol.*, 68:67–73, 1954.

[38] G. R. D. Jones. More on the inhibition of the urease reaction by boric acid. *Ann. Clin. Biochem.*, 34:430–432, 1997.

[39] M. J. Todd and R. P. Hausinger. Purification and characterization of the nickel-containing multicomponent urease from *Klebsiella aerogenes*. *J. Biol. Chem.*, 262:5963–5967, 1987.

[40] I. Park and R. P. Hausinger. Diethylpyrocarbonate reactivity of *Klebsiella aerogenes* urease: effect on pH and active site ligands on rate of enzyme inactivation. *J. Protein Chem.*, 12:51–56, 1993.

[41] I. Park and R. P. Hausinger. Site-directed mutagenesis of *Klebsiella aerogenes* urease: identification of histidine residues that appear to function in nickel ligation, substrate binding, and catalysis. *Protein Science*, 2:1034–1041, 1993.

[42] E. Jabri, M. B. Carr, R. P. Hausinger, and P. A. Karplus. The crystal structure of urease from *Klebsiella aerogenes*. *Science*, 268:998–1004, 1995.

[43] E. Jabri and P. A. Karplus. Structures of the *Klebsiella aerogenes* urease apoenzyme and two active-site mutants. *Biochemistry*, 35:10616–10626, 1996.

[44] S. Wang, M. H. Lee, R. P. Hausinger, P. A. Clark, D. E. Wilcox, and R. A. Scott. Structure of the dinuclear active site of urease. X-ray Absorption Spectroscopic study of native and 2-mercaptoethanol-inihibited bacterial and plant enzymes. *Inorg. Chem.*, 33:1589–1593, 1994.

[45] M. J. Todd and R. P. Hausinger. Identification of the essential cysteine residue in *Klebsiella aerogenes* urease. *J. Biol. Chem.*, 266:24327–24331, 1991.

[46] P. R. Martin and R. P. Hausinger. Site-directed mutagenesis of the active site cysteine in *Klebsiella aerogenes* urease. *J. Biol. Chem.*, 267:20024–20027, 1992.

[47] C. Csiki and M. Zimmer. A molecular mechanical analysis of the active site of urease with a special emphasis on determining the binding con-

formations available to oxygen-bound urea. *J. Biomol. Struc. Dynam.*, 17:121–131, 1999.

[48] P. A. Karplus, M. A. Pearson, and R. P. Hausinger. 70 years of crystalline urease: What have we learned? *Acc. Chem. Res.*, 30:330–337, 1997.

[49] J. Alper. Ulcers as an infectious disease. *Science*, 260:159–160, 1993.

[50] C. Montecucco, E Papini, M. de Bernard, and Zoratti M. Molecular and cellular activities of helicobacter pylori pathogenic factors. *FEBS Lett.*, 452:16–21, 1999.

[51] B.E. Dunn and S.H. Phadnis. Structure, function and localization of helicobacter pylori urease. *Yale J. Biol. Med.*, 71:63–73, 1998.

[52] N. E. Dixon, C. Gazzola, C. J. Asher, D. S. W. Lee, R. L. Blakeley, and B. Zerner. Jack bean urease (E.C. 3.5.1.5). II. the relationship between nickel, enzymatic activity, and the 'abnormal' ultraviolet spectrum. the nickel content of jack beans. *Can. J. Biochem.*, 58:474–480, 1980.

[53] N. E. Dixon, C. Gazzola, R. Blakeley, and B. Zerner. Metal ions in enzymes using ammonia or amides. *Science*, 191:1144–1150, 1976.

[54] E. Jabri, H. L. Lee, R. P. Hausinger, and P. A. Karplus. Preliminary crystallographic studies of urease from jack bean and from *Klebsiella aerogenes*. *J. Mol. Biol.*, 227:934–937, 1992.

[55] K. Takashima, T. Suga, and G. Mamiya. The structure of jack bean urease. the complete amino acid sequence, limited proteolysis and reactive cysteine residues. *Eur. J. Biochem.*, 175:151–165, 1988.

[56] R. L. Blakeley, N. E. Dixon, and B. Zerner. Jack bean urease. VII. Light scattering and nickel(II) spectrum. Thiolate-nickel(II) charge transfer peaks in the spectrum of the β-mercaptoethanol-urease complex. *Biochim. Biophys. Acta*, 744:219–229, 1983.

[57] N. E. Dixon, R. L. Blakeley, and B. Zerner. Jack bean urease (E.C. 3.5.1.5). I. a simple dry ashing procedure for the microdetermination of trace metals in proteins. the nickel content of urease. *Can. J. Biochem.*, 58:469–473, 1980.

[58] S. Christians and H. Kaltwasser. Nickel-content of urease from *Bacillus pasteurii*. *Arch. Microbiol.*, 145:51–55, 1986.

[59] G. Mörsdorf, P. Weinmann, and H. Kaltwasser. Accession number X78411. Technical report, EMBL GenBank DDBJ databases, 1994.

[60] S. Benini, S. Ciurli, H. F. Nolting, and S. Mangani. X-ray absorption spectroscopy study of native and phenylphosphorodiamidate-inhibited *Bacillus pasteurii* urease. *Eur. J. Biochem.*, 239:61–66, 1996.

[61] S. Ciurli, S. Benini, W. R. Rypniewski, K. S. Wilson, S. Miletti, and S. Mangani. Structural properties of the nickel ions in urease: novel insights into the catalytic and inhibition mechanism. *Coord. Chem. Rev.*, 190:331–355, 1999.

[62] J. Jancarik and S. H. Kim. Sparse matrix sampling: a screening method for crystallization of proteins. *J. Appl. Cryst.*, 24:409–411, 1991.

[63] R. G. van Silfhout and C. Hermes. X-ray instrumentation for a focused wiggler beamline at the EMBL Outstation Hamburg. *Rev. Sci. Instrum.*, 66(2):1818–1820, 1995.

[64] R. G. van Silfhout. A beam tracking optical table for synchrotron X-ray beamlines. *Nucl. Instr. and Meth. in Phys. Res.*, A403:153–160, 1998.

[65] Z. Otwinowski and W. Minor. Processing of X-ray diffraction data collected in oscillation mode. *Methods Enzymol.*, 276:307–325, 1997.

[66] Z. Otwinowski. Oscillation data reduction program. In L. Sawyer, N. Isaacs, and S. Bailey, editors, *Proceedings of the CCP4 Study Weekend. Data Collection and Processing*, pages 56–62, Warrington, England, 1993. SERC Daresbury Laboratory.

[67] B. W. Matthews. Solvent content of protein crystals. *J. Mol. Biol.*, 33:491–497, 1968.

[68] J. Navaza. AMoRe: an automated package for molecular replacement. *Acta Cryst.*, A50:157–163, 1994.

[69] Collaborative Computational Project, Number 4. The CCP4 suite: Programs for protein crystallography. *Acta Cryst.*, D50:760–763, 1994.

[70] G. N. Murshudov, A. A. Vagin, and E. J. Dodson. Refinement of macromolecular structures by the maximum-likelihood method. *Acta Cryst.*, D53:240–255, 1997.

[71] G. J. Kleywegt and T. A. Jones. Crystal structures of cellular retinoic acid binding proteins I and II in complex with all-trans-retinoic acid and a synthetic retinoid. *Structure*, 2:1241–1258, 1994.

[72] A. T. Bruenger. Free R-value - a novel statistical quantity for assessing the accuracy of crystal structures. *Nature*, 355:472–475, 1992.

[73] G. J. Kleywegt and A. T. Bruenger. Checking your imagination: applications of the free R value. *Structure*, 4:897–904, 1996.

[74] R. A. Engh and R. Huber. Accurate bond and angle parameters or X-ray protein structure refinement. *Acta Cryst.*, A47:392–400, 1991.

[75] V. S. Lamzin and K. S. Wilson. Automated refinement of protein models. *Acta Cryst.*, D49:129–149, 1993.

[76] V. S. Lamzin and K. S. Wilson. Automated refinement for protein crystallography. *Methods in Enzymology*, 277:269–305, 1997.

[77] R. A. Laskowski, M. W. MacArthur, D. S. Moss, and J. M. Thornton. PROCHECK: a program to check the stereochemical quality of protein structures. *J. Appl. Cryst.*, 26:283–291, 1993.

[78] G. Vriend. WHATIF; a molecular modelling and drug design program. *J. Mol. Graph.*, 8:52–56, 1990.

[79] G. Vriend and C. Sander. Quality control of protein models: directional atomic contact analysis. *J. Appl. Cryst.*, 26:47–60, 1993.

[80] B. M. Esnouf. An extensively modified version of MOLSCRIPT that includes greatly enhanced coloring capabilities. *J. Mol. Graph.*, 15:133–138, 1997.

[81] E. A. Merrit and D. J. Bacon. Raster3D photorealistic molecular graphics. *Methods Enzymol.*, 277:505–524, 1997.

[82] W. Kabsch and C. Sander. Dictionary of protein secondary structure: Pattern recognition of hydrogen-bonded and geometrical features. *Biopolymers*, 22:2577–2637, 1983.

[83] P. J. Kraulis. MOLSCRIPT: a program to produce both detailed and schematic plots of protein structures. *J. Appl. Crystallogr.*, 24:946–960, 1991.

[84] T. Etzold, A. Ulyanov, and P. Argos. SRS: Information retrieval system for molecular biology data banks. *Methods in Enzymol.*, 266:114–128, 1996.

[85] J. D. Thompson, D. G. Higgins, and T. J. Gibson. Improving the sensitivity of progressive multiple alignment through sequence weighting, positions-specific gap penalties and weight matrix choice. *Nucleic Acids Res.*, 22:4673–4680, 1994.

[86] L. Stryer. *Biochemistry.* W.H. Freeman and Company, fourth edition, 1995.

[87] E.B. Fauman, E.E. Rutenber, G.F. Maley, F. Maley, and M. Stroud, R. Water-mediated substrate/product discrimination: the product complex of thymidylate synthase at 1.83 A. *Biochemistry*, 33:1502–1511, 1994.

[88] N. E. Dixon, R. L. Blakeley, and B. Zerner. Jack bean urease (E.C. 3.5.1.5). V. on the mechanism of urease on urea, formamide, acetamide, N-methylurea, and related compounds. *Can. J. Biochem.*, 58:1335–1345, 1980.

[89] F. Basolo and R.G. Pearson. *Mechanism of Inorganic Reactions*, chapter Acid-Base Properties of Complex Ions, pages 31–33. Wiley, New York, 1967.

[90] S. Benini, W. R. Rypniewski, Wilson K. S., S. Miletti, S. Ciurli, and S. Mangani. The complex of *Bacillus pasteurii* urease with acetohydroxamate anion from X-ray data at 1.55 A resolution. *J. Biol. Inorg. Chem.*, 1:110–118, 2000.

[91] S. Benini, W. R. Rypniewski, Wilson K. S., S. Miletti, S. Ciurli, and S. Mangani. A new proposal for urease mechanism based on the crystal

structure of the native and inhibited enzyme from *Bacillus pasteurii*: why urea hydrolysis costs two nickels. *Structure*, 7:205–216, 1999.

[92] R. M. Smith and A. E. Martell. *Critical stability constant*. Plenum press New York, 1975.

[93] A. Fersht. *Enzyme Structure and mechanism*. Freeman, W. H. and Company, 1993.

[94] M. A. Pearson, I. S. Park, R. A. Schaller, L. O. Michel, P. A. Karplus, and R. P. Hausinger. Kinetic and structural characterisation of urease active site variants. *Biochemistry*, 39:8575–8584, 2000.

[95] Z. F. Kanyo, L. R. Scolnick, D. E. Ash, and D. W. Christianson. Structure of a unique binuclear manganese cluster in arginase. *Nature*, 283:554–557, 1996.

[96] M. C. J. Wilce and J. A. Wilce. Structure and mechanism of a proline-specific aminopeptidase from *Escherichia coli*. *Proc. Natl. Acad. Sci. USA*, 95:3472–3477, 1998.

[97] S. Sundell, S. Hansen, and E. Hough. A proposal for the catalytic mechanism in phospholipase C based on interaction energy and distance geometry calculations. *Protein Eng.*, 7:571–577, 1994.

[98] N. Strater and W. N. Lipscomb. Two-metal ion mechanism of bovine lens leucine aminopeptidase: active site solvent structure and binding mode of l-leucinal, a gem-diolate transition state analogue, by X-ray crystallography. *Biochemistry*, 34:14792–14800, 1995.

[99] J. L. Vanhooke, M. M. Benning, F. M. Rausel, and H. M. Holden. Three-dimensional structure of the zinc-containing phosphotriesterase with the

bound substrate analog diethyl 4-methylbenzylphosphate. *Biochemistry*, 35:6020–6025, 1996.

[100] P. Heikinheimo and R. Lathi. A site-directed mutagenesis study of *Saccharomyces cerevisiae* pyrophosphatase. functional conservation of the active site of soluble inorganic pyrophosphatases. *Eur. J. Biochem.*, 239:138–143, 1996.

[101] L. Holm and C. Snader. An evolutionary treasure: unification of a broad set of amidohydrolases related to urease. *Proteins*, 28:72–82, 1997.

[102] C. Sander and R. Schneider. Database of homology-derived protein structures and the structural meaning of sequence alignment. *Proteins*, 9:56–68, 1991.

[103] D. E. Wilcox. Binuclear metallohydrolases. *Chem. Rev.*, 96:2435–2458, 1996.

[104] Helmward Zollner. *Handbook of Enzyme Inhibitors*. VCH Verlagsgeselshaft, Weinheim Germany, 1993.

[105] S. Benini, W. R. Rypniewski, K. S. Wilson, S. Ciurli, and Mangani S. The complex of *Bacillus pasteurii* urease with β-mercaptoethanol from X-ray data at 1.65-A resolution. *J. Biol. Inorg. Chem.*, 3:268–273, 1998.

[106] S. Benini and M. A. Cremonini. Structure and function relationship of enzymes. Do we really learn something? *J. Chem. Bull.*, 1:154–235, 2000. Review.

[107] D. B. Hicks and T. A. Krulwich. The respiratory chain of alkaliphilic bacteria. *Biochim. Biophys. Acta*, 1229:303–314, 1995.

[108] J. R. Lewis, S. Belkina, and T. A. Krulwich. Alkalophiles have much higher cytochrome contents than conventional bacteria and their own non-alkaliphilic mutant derivatives. *Biochim. Biophys. Res. Comm.*, 95:857–863, 1980.

[109] B. A. Haddock and J. G. Cobley. Electron transport in the alkalophile *Bacillus pasteurii* ([n.c.i.b.] 8841). *Biochem. Soc. Trans.*, 4:709–711, 1976.

[110] M. H. Hoddinott, G. A. Reid, and W. J. Ingledew. The respiratory chain and proton electrochemical gradient int the alkalophile *Bacillus pasteurii*. *Biochem. Soc. Trans.*, 6:1295–1298, 1978.

[111] S. Benini, S. Ciurli, W. R. Rypniewski, and K. S. Wilson. Crystals of cytochrome *c*-553 from *Bacillus pasteurii* show diffraction to 0.97 A. *Proteins: Struct. Funct. Genetics*, 28:580–585, 1997.

[112] T. Jahns. Ammonium/urea dependent generation of a proton electrochemical potential and synthesis of [atp] in *Bacillus pasteurii*. *J. Bacteriol.*, 178:403–409, 1996.

[113] I. H. M. Vandenberghe, Y. Guisez, S. Ciurli, S. Benini, and J. J. Van Beeumen. Cytochrome c-553 from the alkalophilic bacterium *Bacillus pasteurii* has the primary structure characteristics of a lipoprotein. *Biochem. Biophys. Res. Comm.*, 264:380–387, 1999.

[114] S. Benini, M. Borsari, S. Ciurli, A. Dikiy, and M. Lamborghini. Modulation of *Bacillus pasteurii* cytochrome *c*553 reduction potential by structural and solution parameters. *J. Biol. Inorg. Chem.*, 3:371–382, 1998.

[115] G. M. Sheldrick and T. R. Schneider. SHELXL: High-resolution refinement. *Methods Enzymol.*, 277:319–343, 1997.

[116] EU 3-D Validation Network. Who Checks the Checkers? Four Validation Tools Applied to Eight Atomic Resolution Structures. *J. Mol. Biol.*, 276:417–436, 1998.

[117] F. H. Allen and D. G. Watson. The development of version 3 and 4 of the Cambridge structural database system. *J. Chem. Inf. Comp. Sci.*, 31:187–204, 1991.

[118] A. Nicholls, K. A. Sharp, and B. Honig. Protein folding and association: insights from the interfacial and thermodynamic properties of hydrocarbons. *Proteins*, 11:281–296, 1991.

[119] C. Frazao, C. M. Soares, M. A. Carrondo, E. Pohl, Z. Dauter, K.S. Wilson, M. Hervas, J.A. Navarro, M. A. De la Rosa, and G. M. Sheldrick. *Ab initio* determination of the crystal structure of cytochrome c_6 and comparison with plastocyanin. *Structure*, 3:1159–1169, 1995.

[120] J. Hasegawa, T. Yoshida, T. Yamazaki, Y. Sambongi, Y. Yu, Y. Igarashi, T Kodama, K. Yamazaki, Y. Kyogoku, and Y. Kobayashi. Solution structure of thermostable cytochrome c-552 from hydrogenobacter thermophilus determined by 1h-nmr spectroscopy. *Biochemistry*, 37:9641–9649, 1998.

[121] D.M. Tiede, A.C. Vashishta, and M.R. Gunner. Electron-transfer kinetics and electrostatic properties of the rhodobacter sphaeroides reaction center and soluble c-cytochromes. *Biochemistry*, 32:4515–4531, 1993.

[122] R.G. Meyer, T.E.and Bartsch, M.A. Cusanovich, and G. Tollin. Kinetics of photooxidation of soluble cytochromes, hipip, and azurin by the photosynthetic reaction center of the purple phototrophic bacterium rhodopseudomonas viridis. *Biochemistry*, 32:4719–26, 1993.

[123] S. Benini, W. R. Rypniewski, S. Ciurli, and K. S. Wilson. *Ab initio* structure solution of cytochrome *c-553* from *Bacillus pasteurii*. *Erice course: Direct Methods in solving macromolecular structures*, 1997.

[124] Z. Dauter, L. S. Lamzin, and K. S. Wilson. The benefits of atomic resolution. *Current Opinion in Structural Biology*, 7:681–688, 1997.

[125] Z. Dauter, M. Dauter, E. de la Fortelle, G. Bricogne, and G. M. Sheldrik. Can anomalous signal of sulfur become a tool for solving protein crystal structures? *J. Mol. Biol.*, 289:83–92, 1999.

[126] S. Benini, A. Gonzalez, W. R. Rypniewski, K. S. Wilson, J. J. Van Beeumen, and S. Ciurli. Crystal structure of oxidised *B. pasteurii* cytochrome *c*-553 determined by *ab initio* and by MAD methods: structure-function relationships and molecular evolution at 0.97 A resolution. *submitted*, 2000.

[127] Y. Amemiya. X-ray storage-phosphor imaging-plate detectors: High-sensitivity x-ray area detector. *Methods in Enzymol.*, 276:233–243, 1997.

[128] E. M. Westbrook and I. Naday. Charge-coupled device-based area detectors. *Methods in Enzymol.*, 276:244–268, 1997.

[129] N. E. Chayen. The role of oil in macromolecular crystallization. *Structure*, 5(10):1269–1274, 1997.

[130] G. L. Gilliland, M. Tung, D. M. Blakeslee, and J. E. Ladner. The Biological Macromolecule Crystallization Database, version 3.0: New features, data, and the NASA archive for protein crystal growth data. *Acta Cryst.*, D50:408–413, 1994.

[131] A. Ducruix and R. Giege. *Crystallization of Nucleic Acids and Proteins: A Practical Approach.* IRL Press/Oxford University Press, 1992.

[132] G. Rosenbaum, K. C. Holmes, and J. Witz. *Nature*, 230:434–437, 1971.

[133] J. R Helliwell. *Macromolecular Crystallography with Synchrotron Radiation.* Cambridge University Press, 1992.

[134] J. Drenth. *Principles of Protein X-ray Crystallography.* Springer, 1994.

[135] U.W. Arndt and A.J. Wonacott, editors. *The Rotation Method in Crystallography.* Amsterdam: North Holland., fourth, revised edition, 1977.

[136] A. Leslie. Data collection and processing. In L. Sawyer, N. Isaac, and S. Bailey, editors, *Proceedings of the CCP4 Study Weekend*, pages 44–51, 29-30 January 1993.

[137] D. W. Rodgers. Cryocrystallography. *Structure*, 2:1135–1140, 1994.

[138] E. F. Garman and T. R. Schneider. Macromolecular Cryocrystallography. *J. Appl. Cryst.*, 30:211–237, 1997.

[139] C Giacovazzo, editor. *Fundamentals of Crystallography.* International Union of Crystallography Oxford University Press, 1994.

[140] M. Schiltz, T. Prange, and R. Fourme. On the preparation and X-ray data collection of isomorphous xenon derivatives. *J. Appl. Cryst.*, 27:950–960, 1994.

[141] W. A. Hendrickson and C. M. Ogata. Phase determination from multiwavelength anomalous diffraction measurements. *Methods in Enzymol.*, 276:494–523, 1997.

[142] W. A. Hendrickson, J. L. Smith, and S. Sheritt. Direct phase determination based on anomalous scattering. *Methods in Enzymol.*, 115:41–55, 1985.

[143] A. Gonzalez, J.-D Pedelacq, M Sola, F. X. Gomis-Rueth, M. Coll, J.-P. Samama, and S. Benini. Two-wavelength MAD phasing: in search of the optimal choice of wavelengths. *Acta Cryst.*, D 55:1449–1458, 1999.

[144] P. Storici, G. Capitani, R. A. De Biase, and T. Shirmer. The crystal structure of GABA-aminotransferase, a target for antiepileptic drug therapy. Poster presented at the 18^{th} IUCr Congress Glasgow 1999.

[145] H. A. Hauptman. *Shake-and-Bake* an algorithm for automatic solution *ab initio* of crystal structure. *Methods in Enzymol.*, 277:3–13, 1997.

[146] H. L. Jensen. Refinement and reliability of macromolecular models based on X-ray diffraction data. *Methods in Enzymol.*, 277:353–366, 1997.

[147] J. H. Konnert. A restrained-parameter structure-factor least-squares refinement procedure for large asymmetric units. *Acta Cryst*, A32:614–617, 1976.

[148] J. H. Konnert and W. A. Hendrickson. A restrained-parameter thermal-factor refinement procedure. *Acta Cryst.*, A36:344–350, 1980.

[149] A. T. Bruenger, P. D. Adams, and L. M. Rice. New applications of simulated annealing in X-ray crystallography and solution NMR. *Structure*, 5:325–336, 1997.

[150] G. Bricogne. *Acta Cryst.*, A44:517–545, 1998.

[151] G. N. Murshudov, A. A Vagin, A. Lebedev, K. S. Wilson, and E. J. Dodson. Efficient anisotropic refinement of macromolecular structures using FFT. *Acta Cryst.*, D55:247–255, 1999.

[152] T. A. Jones. A graphics model building and refinement system for macromolecules. *J. Appl. Cryst.*, 11:268–272, 1978.

[153] A. Perrakis, T. K. Sixma, K.S. Wilson, and V. S. Lamzin. wARP: improvement and extension of crystallographic phases by weighted averaging of multiple refined dummy atomic models. *Acta Cryst.*, D53:448–455, 1997.

[154] E. J. Asselt van, A. Perrakis, K. H. Kalk, and V. S. Lamzin. Accelerated X-ray structure elucidation of a 36 kDa muramidase/trans glycosylase using wARP. *Acta Cryst.*, D54:58–735, 1998.

[155] F. C. Bernstein, T. F. Koetzle, G. J. B. Williams, E. F. Meyer, Jr., M. D. Brice, J. R. Rodgers, O. Kennard, T. Shomanouchi, and M. Tasumi. The protein data bank: a computer-based archival file for macromolecular studies. *J. Mol. Biol.*, 112:535–542, 1977.

[156] E. E. Abola, F. C. Bernstein, S. H. Bryant, T. F. Koetzle, and J. Weng. Crystallographic databases - information content, software systems, scientific applications. In F.H. Allen, G. Bergerhoff, and R. Sievers, editors, *Data Commision of the International Union of Crystallography*, pages 107–132. Bonn/Cambridge/Chester, 1987.

[157] G. N. Ramachandran, C. Ramakrishnan, and V. Sasisekharan. Stereochemistry of polypeptide chain configurations. *J. Mol. Biol.*, 7:95–99, 1963.