Alma Mater Studiorum – Università di Bologna

DOTTORATO DI RICERCA IN

BIOLOGIA CELLULARE E MOLECOLARE

Ciclo XXVI

Settore Concorsuale di afferenza: CHIM/03

Settore Scientifico disciplinare: 03/B1

Structural and biochemical studies of *Sporosarcina pasteurii* UreE: a nickel-chaperone involved in the urease activation process

Presentata da: **Dott.ssa Anna Merloni**

Coordinatore Dottorato **Relatore**

Prof. Vincenzo Scarlato **Prof. Stefano Ciurli**

Esame finale anno 2014

Index

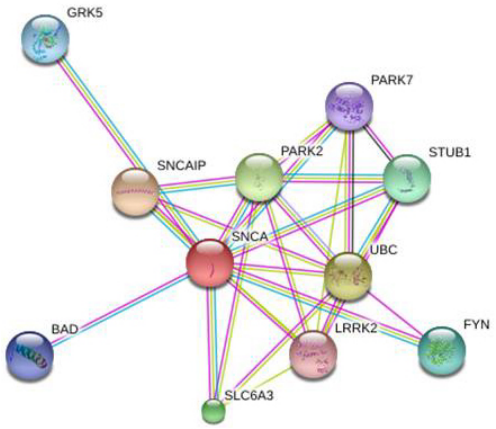
1. Introduction

1. Introduction

1.1. Overview on protein-protein interactions

Decades of research into cell biology, molecular biology, biochemistry, structural biology, and biophysics have produced a remarkable compendium of knowledge on the function and molecular properties of individual proteins. This knowledge is well recorded and kept updated into major protein databases, like UniProt (http://www.uniprot.org/). However, proteins rarely act alone. Many times they team up into "molecular machines" and have intricate physicochemical dynamic connections to undertake biological functions at both cellular and systems levels [1]. Thus, rather than individual proteins, protein complexes are the functioning biochemical entities in a cell [2]. A critical step towards understanding the complex molecular relationships in living systems is the mapping of protein-to-protein physical "interactions". The complete map of protein interactions that can occur in a living organism is called *interactome*.

Interactome mapping has become one of the main scopes of current biological research, similar to the way genome projects were a driving force of molecular biology 20 years ago. Unlike the genome, the interactome is dynamic (Fig. 1.1). Many interactions are transient, and others occur only in certain cellular contexts or at particular times in development. The interactome may be tougher to solve than the genome, but the information is crucial for a complete understanding of biology [3].



[4]

Figure 1.1: Interactome of α-synuclein: an intrinsically disordered hub. The network of predicted associations for a particular group of proteins is generated using STRING database (accessible at http://string.org). The network nodes are proteins, whereas the edges represent the functional associations evaluated based on the experiments, search of databases and text mining. Different line colours of edges represent different types of evidence for the association, such as experimental/biochemical data (*pink* lines), association in curated databases (*blue* lines), and co-mentioned in PubMed abstracts (*yellow* lines). From Uversky, V.N. 2014.

Protein-protein interactions - hereinafter PPIs - enable and exert stringent control over DNA replication, RNA transcription, protein translation, macromolecular assembly and degradation, and

signal transduction; essentially all cellular functions involve PPIs. Thus, alteration of critical PPIs is thought to be involved in the development of many diseases, such as neurodegenerative disorders, cancers, and infectious diseases too. Therefore, examination of when and how PPIs occur and how they are controlled is essential for understanding diverse biological processes as well as for elucidating the molecular basis of diseases and identifying potential targets for therapeutic interventions.

The regulation of cell function due to the interactions of proteins is delicately balanced by the relative affinities of the various protein partners and the modulation of these affinities by the binding of ligands, other proteins, nucleic acids, metal ions, and covalent modifications, such as specific phosphorlyation or acetylation reactions. Specificity and strength of signal transduction are encoded by the exact amino acid sequence of the domain, and it is this relationship between sequence, structure, dynamics, energetics and function that constitutes the fundamental issue for the biophysics of PPIs. This aim therefore requires a structural characterization of the domains and if possible their dispositions within the complete protein and their complexes with their specific protein partners. In addition to a structural characterization, understanding the basis for specificity in these systems necessitates very careful and thorough comparative studies of similar interacting partners or mutated domains in order to bring to light the energetic properties linked to a particular sequence/structure. Noteworthy, even differences as small as 1 kcal/mol in interaction energy between pairs of protein partners can lead to profound differences in cell growth and development [5]. While biochemists have successfully identified these protein interaction domains and demonstrated interaction preferences and, by mutational studies, certain of the necessary sequence determinants of the interactions, in most cases, thorough quantitative thermodynamic and kinetic studies of protein interactions in these systems remain to be carried out. One reason for this lack of information is that priority has been given over the past few years to the identification of these important proteins and to the determination of their three-dimensional structure, both of which are prerequisites to a full understanding of function. Nonetheless, understanding the mechanisms underlying the function of these proteins likewise requires the characterization of their energetic and dynamic properties [5].

The first major stumbling block encountered by the experimentalist interested in characterizing PPIs is that of having access to enough purified stable protein, and if possible some interesting constructs, homologous family members or functional mutants in order to carry out a thorough comparative structure/function study. In cases where the NMR or crystal structure of the protein has been determined, this first obstacle can likely be overcome, since the production procedure has been

worked out in such cases. The second is that of which of the many approaches employ to better analyze the sistem. As a general rule, the characterization will be all the more solid, the more complementary techniques are employed, since there are perturbations and uncertainties associated with all of them [5].

Proteins interact through their interfaces, which consist of interacting residues belonging to different chains, together with some isolated residues in spatial proximity. In order to evaluate PPIs, is necessary to understand the chemical and physical features of their associations, to consider the shape complementarity, the relative contributes of each component to complex stability and other parameters such as the size and the polar/hydrophobic character of the contact area, and the occurrence of protrusions and flatness. In general, PPIs are frequently mediated by hydrophobic effects, but also hydrogen bonds, electrostatic interactions and van der Waals attractions play a considerable role. In particular, it has been proposed that hydrophobic forces drive PPIs with large energetic contributions due to the desolvation of non-polar groups, whereas H-bonds and salt bridges confer specificity [6].

It is important to mention that disordered regions in proteins may be fundamental for a variety of interactions. A large fraction of cellular proteins are estimated to be natively disordered, such as the intrinsically unstructured proteins (IUPs), which lack a well-defined 3D structure in their native state [7], [8]. These proteins fulfil important functions that are often associated with signal transduction, gene expression and chaperone action [9]. The flexibility and structural adaptability of this kind of proteins represent a radical deviation from the classical structure-function paradigm [10]. Natively unstructured proteins undergo a disorder-to-order transition driven by binding their physiological partner, but this local structural rearrangement do not change the global fold of disordered proteins [8].

A basic mechanism by which individual proteins can increase interactions network complexity is moonlighting, which is, by definition, the ability of a protein to fulfil more than one, apparently unrelated, function. Moonlighting proteins might serve at distinct points of metabolic networks and might, thus, increase network complexity without increasing the number of underlying proteins [9]. Traditionally, this phenomenon is attributed to separate binding surfaces of globular, folded proteins, but recently it has been reported that unstructured proteins might provide unprecedented cases of moonlighting. In fact the underlying structural malleability enables an IUPs to elicit both inhibiting and activating effects on different partners or even the same partner molecule [9].

Lastly, since protein-protein interactions govern the structure and the function of multi-component assemblies involved in all sorts of biological processes, they must be subject to strong Darwinian

constraints [11]. The interplay between the opposite requirements of conservation and diversification can be analyzed in the light of phylogeny of residue contact networks, a concept that is common to the interior of proteins and to the interfaces of protein-protein complexes [12] (Fig. 1.2). Evolutionary approach is the natural guideline through the complexity of life, therefore, as this subject is so relevant to all aspects of biology, it will definitely be useful also in the PPIs field.

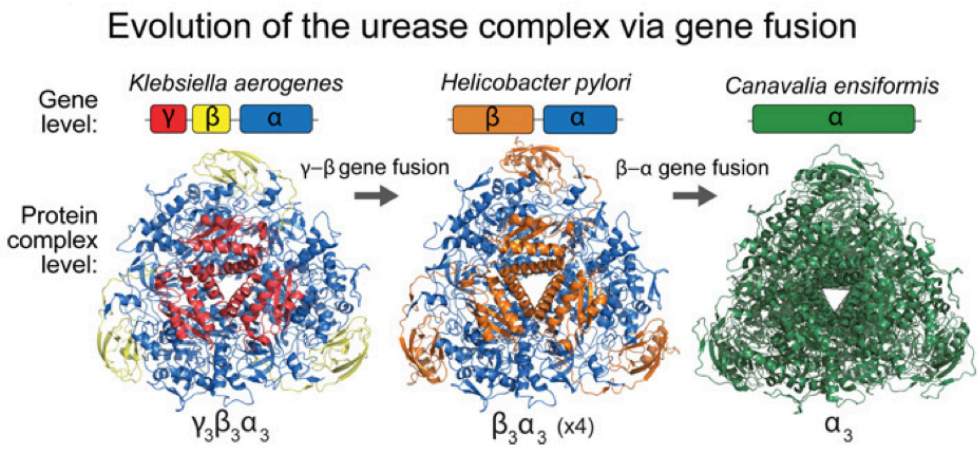


Figure 1.2: Gene fusion leads to three different forms of the urease complex in different species: *Klebsiella aerogenes* (PDB code 1KRA), *Helicobacter pylori* (PDB code 1E9Y) and *Canavalia ensiformis* (PDB code 3LA4). Note that only one-quarter of the full tetrahedral *Helicobacter pylori* complex is shown, which corresponds to the C_3 symmetry of the other two [13].

1.2. Metals in proteins

Metal ions are involved in a variety of biological reactions and estimated to be required in one third of all proteins. Their essentiality, coupled with their limited environmental availability, has prompted all organisms to develop metal homeostasis networks that ensure the availability and the correct localization of metal ions in metallo-proteins and sub-cellular compartments [14]. Metals are essentials for catalysis in important biological processes, such as photosynthesis (Mg, Mn, Fe and Cu), respiration (Fe, Cu), water oxidation (Ca, Mn) and nitrogen fixation (Fe, Ni). On the other hand, the intrinsic toxicity of the majority of metal ions demands a tightly regulated intracellular trafficking that maintains intracellular metal ion concentration under the physiological limits and minimizes the amount of free metal ions [15].

Much effort has been devoted to understand the metals role and the structure and function of metallo-proteins. Different transition metal ions have different chemical properties, thus generally each metallo-protein uses a specific metal to carry out a precise function. The metal-specificity of

cellular responses indicates that metal-binding proteins are able to choose, in the mixture of metal ions in solution, the correct cofactor for a specific biological function.

A systematic bioinformatics survey of 1.371 different enzymes for which three-dimensional structures are known, estimated that 47% required metals, with 41% containing metals at their catalytic centres [16]. Metallo-enzymes occur in all six Enzyme Commission (EC) classes, accounting for 44% of oxidoreductases, 40% of transferases, 39% of hydrolases, 36% of lyases, 36% of isomerases and 59% of ligases [16]. Magnesium is the most prevalent metal in metallo-enzymes, although it is often involved in loose partnerships with phosphate-containing substrates such as ATP and it is sometimes interchangeable with manganese. The frequency of manganese in protein structures may overestimate its use *in vivo*, where magnesium is truly the cofactor (Fig. 1.4) [17]. Zinc is the second-most abundant metal ion in enzymes. This is due to its distinctive combination of properties, in fact zinc on one hand resembles the group II elements, in that it is also stable in the +2 oxidation state, and it is redox inert. In particular, the Zn^{2+} ion has a radius (0.74 Å) similar to that of Mg^{2+}, and the electrostatic binding to negatively charged species [16].

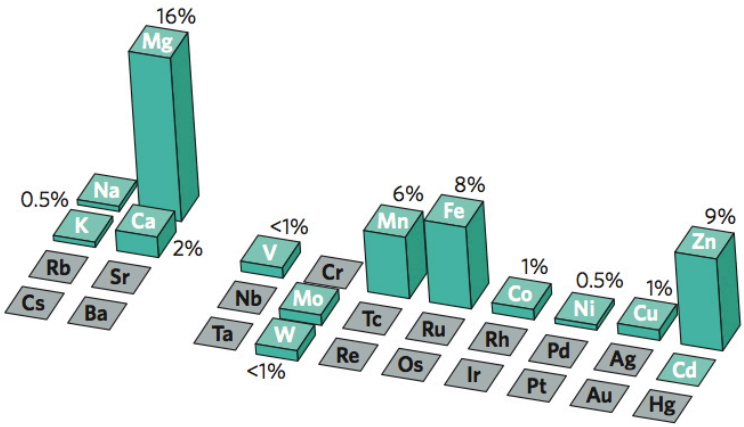


Figure 1.4: Metals used as cofactors in enzymatic catalysis are shown in *sea green*. The height of each column represents the proportion of all enzymes with known structures using the respective metal [17].

The simplest explanation to understand how metallo-proteins can selectively bind their specific metal ion among the intracellular metal ions pool is that each protein presents higher affinity for the cognate metal ions as compared to all others. Protein affinities for metals have a tendency to follow a universal order of preference, which for essential divalent metals is the Irving-Williams series [18] (Mg^{2+} and Ca^{2+} (weakest binding) < Mn^{2+} < Fe^{2+} < Co^{2+} < Ni^{2+} < Cu^{2+} > Zn^{2+}) [19]. However, in order to obtain a complete prediction on the selectivity, it is necessary to consider many other different factors, such as (i) stereo-electronic factors for preferred metal ion coordination geometry,

(ii) metal ion availability in the context of the overall intracellular metal ion content, (iii) metal-induced conformational changes of the protein structure, and (iv) kinetics [14].

The coordination geometry (e.g., octahedral, tetrahedral, square pyramidal, square planar, trigonal or linear) (Fig. 1.5) is determined by the kind and the position of protein residues in the metal binding pocket that allow limited coordination geometries. The presence of specific donor ligands (S, O or N from protein residues), the shape and the charge of the metal binding cavity can impart bias in favour of the correct metal. The metal-binding pocket can exclude ions with the wrong charge [20].

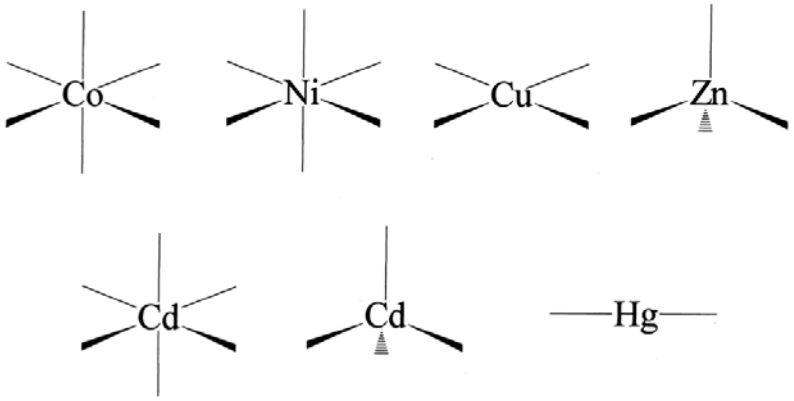


Figure 1.5: The prefered coordination geometries of the selected divalent (2+) metal ions: octahedral (Co^{2+}, Ni^{2+}, Cd^{2+}), tetrahedral (Zn^{2+}, Cd^{2+}), square pyramidal trigonal (Cu^{2+}) or linear (Hg^{2+}) [20].

A further way to ensure that the correct metal is acquired by a metallo-protein is to exploit delivery proteins, namely metallo-chaperones. For example, nickel is inserted into bacterial hydrogenase and urease by dedicated nickel metallo-chaperones. Metal is passed from metallo-chaperones to cognate apo-proteins by means of ligand-exchange reactions [17]. Thus, the specificity of protein-protein contact, and of the subsequent ligand-exchange reactions, determines which of the proteins can correctly incorporate metals supplied by metallo-chaperones. Cells are not ideal solutions and kinetic factors can dominate the distribution of metals, for example where metals are delivered by metallo-chaperones. Nonetheless, excluding the wrong metals from proteins may be more challenging than acquiring the right ones [17]. Metal availability is highly controlled in the cytoplasm through compartmentalization, which can be used to keep competitive metals out of the wrong nascent proteins [21]. Metal importers, metal exporters and metal stores maintain a limited supply of competitive metals, such as copper. Thus, proteins compete with other molecules for these metals rather than metals competing with other metals for proteins. To achieve this state, the amount of each metal is somehow sensed to adjust the actions of transporters (at plasma membranes or internal compartments) and storage proteins for each element [19]. Sensors also adjust

metabolism to minimize demand for scarce metals or to exploit abundant ones. These sensors are thus pivotal to metal selectivity because their capacity to distinguish the correct metal affects metal occupancy of other metallo-proteins.

Metal ion binding often does not modulate a direct allosteric structural change in the backbone conformation, but rather a change in protein flexibility. In this sense, it is possible that the kinetics of metal binding to the protein, and in particular which metal ion is presented to the protein in a precise stage of its biosynthesis and/or function determines which is the metal ion that eventually leads to the metabolic response [14].

2. Structural information techniques

2. Structural information techniques

2.1. Importance of structural information on proteins

At the beginning of the XXI century, after the sequencing of many other genomes in addition to the human one, it became clear that the simple knowledge of the sequence of DNA and even the protein primary structures would not been enough for a complete understanding of the tree-dimentional (3D) shape and function of proteins in the cell. Rapidly the gap between the number of protein sequences and protein structures data increased. To fill this gap, structural proteomics aroused with the purpose of giving a complete characterization of proteome of living organisms. Proteomics focuses on identifying when and where proteins are expressed in a cell so as to establish their physiological roles in an organism. Its aim is indeed ambitious and important since proteins are a huge type of biomolecules which can work as enzymes, transporters, transmembrane channels, gene regulators as well as structural components in the cell. A complete understanding of protein function and mechanism of action can only be accomplished with the knowledge of its 3D structure at atomic resolution. The assumption that lies behind this statement is the so called "structure-function hypothesis" that roughly states that each macromolecule coded by the genome has a function, and that the function can be understood using the chemical structure, interactions and dynamics of the macromolecule. Evolution forms the foundation of this reductionist view, since functionality is the basis of natural selection. Except for cases of QWERTY effect [22] - sometimes recognised - this assumption is widely accepted in biochemistry.

In the present state of technology there is no single technique which provides an overall picture of 3D structure of proteins, in particular in the case of large molecules. Only through the combined use of complementary techniques it is possible to obtain an acceptable level of knowledge about the macromolecules. The importance to combine different approaches is well represented in the foreword of the book *Methods in molecular biophysics* citing the Indian story of *The six blind men and the elephant* : «Each of the blind men touched a different part of the elephant, and concluded on its nature: a big snake said the man who touched the trunk, the tusks were spears, its side a great wall, the tail a paint brush, the ears huge fans, the legs were tree trunks. We could add a seventh very short-sighted man to the story who can see the whole elephant but as a blurred grey cloud to illustrate diffraction methods.» [23].

At the present time X-ray diffraction and High-resolution NMR spectroscopy are the only techniques providing atomic-level structural information on proteins. However none of these experimental techniques is capable to render the position of atoms in macromolecules in vivo and, at the same time, to render the atomic motions and conformational changes that occur while the proteins are involved in the chemical and physical reactions associated with their biological function, regardless of the time scale involved. Each of this two techniques provides a partial field of the whole view.

2.2. X-ray diffraction crystallography

The first method which has been established for many years is X-ray diffraction of protein single crystals. From the late 1950s crystal structures of proteins began to be determined using X-ray crystallography and nowadays this technique has progressed to the point that several steps in the process can be automated. For example robots have been designed for cloning and expression trials, and software can index diffraction intensities, refine and build structural models which fit the X-ray data. Moreover they offer the possibilities to have short data collection times associated with synchrotron radiation and to study hundreds of proteins in parallel. Finally, the determination of every possible protein fold provides crucial empirical knowledge for structure-from-sequence modelling approaches. On the other hand, the X-ray crystallography has some limitation: first of all the fact that many proteins do not crystallize, and - even when they do - the crystals may diffract poorly or there may be difficulties in solving the phase problem; second, significant and possibly important functional differences may exist between structures in the crystal state and in solution; third, dynamic processes ranging from the picosecond to second time scales cannot be studied with the rigid organization of crystals.

It is well known that biological macromolecules take up their active structures only in a suitable solvent environment. The forces that stabilise them are weak forces which arise in part from interactions with the solvent. The study of biological macromolecules, therefore, cannot be completely separated from the study of their aqueous solutions and it is also understandable why the development of an alternative techniques was needed.

2.3. Nuclear Magnetic Resonance spectroscopy in structural proteomics

Nuclear magnetic resonance (NMR) spectroscopy is based on the absorption of electromagnetic radiation in the low-frequency end of the spectrum, i.e. the radio-frequency region (10 MHz - 1 GHz). In contrast to UV, visible and IR absorption spectroscopy, which involves outer-shell atomic electrons, NMR arises from the magnetic properties of atomic nuclei, which, when placed in an intense magnetic field, develop the energy states required for absorption to occur. The frequency of absorbed radiation can be measured as a function of time with an accuracy better than one part in a million. The precise nature of the signal depends on the chemical environment of the nucleus; hence structural information is obtained. NMR spectroscopy has sensitivity near to atomic resolution and it has the unique capability of allowing researchers to investigate the internal dynamics of protein in solution on a wide range of time scales. By coupling structural and dynamic aspects, NMR spectroscopy thus affords a complete picture of the behaviour of proteins. Another powerful aspect of NMR spectroscopy is its ability to characterize protein complexes under conditions nearly physiological at atomic detail, even if the interactions are weak and transient [24].

2.3.1. NMR basics

Magnetic properties of nuclei:

NMR spectroscopy is a manifestation of existence of *angular moment of nuclear spin,* $\vec{I}$, which is:

$$\left|\vec{I}\right| = \frac{h}{2\pi}\sqrt{I(I+1)} \qquad (1)$$

Where ***I*** is the nuclear spin quantum number and ***h*** is the Plank constant.

All nuclei with $\vec{I} \neq 0$ are NMR active and can be studied through NMR spectroscopy

- nuclear spin

The spin quantum number (***I***) of a nucleus is determined by the number of unpaired protons and neutrons it contains. A number of nuclei of particular importance to structural biology (^{1}H, ^{13}C, ^{15}N, ^{19}F, and ^{31}P) have nuclear spin values of 1/2 that is the preferred for solution NMR.

- magnetic quantum number

An important quantum mechanical property of a spinning nucleus is that the average value of the component of its magnetic moment vector along a defined direction takes up specific values described by a set of magnetic quantum numbers ***m*** **= 2*I* +1**, in integral steps between **+*I*** and **-*I***.

Nuclear magnetization

Nuclei having angular moment of spin $\vec{I} \neq 0$, also possess a magnetic moment ($\vec{\mu}$) and they behave themselves as magnetic dipoles.

The magnetic moment ($\vec{\mu}$) of a nucleus is a vector parallel (or antiparallel, for nuclei with negative **γ**) to angular moment of spin ($\vec{I}$). The gyromagnetic ratio **(γ)** is a proportionality constant between magnetic and angular moments.

The spin needs to experience a magnetic field and be excited to be observable. A charged spinning nucleus creates a magnetic field that is analogous to the field produced when electricity flows through a coil of wire. The resulting nuclear magnetic moment ($\vec{\mu}$) is orientated along the axis of spin and is directly proportional to the angular momentum vector ($\vec{I}$):

$$\vec{\mu} = \gamma \vec{I} \tag{2}$$

In the absence of a magnetic field, there is no preference for one or other of the two possible states, for a nucleus with spin equal to 1/2, so that in a large assemblage of such nuclei there are exactly equal numbers with ***m*** equal to +1/2 and ***m*** equal to -1/2 (Fig. 2.2).

The value of **μ** component on z-axes can therefore take the following values:

$$\mu_z = \gamma I_z = \gamma \hbar m \tag{3}$$

$$m = (-I, -I+1, ..., I-1, I) \qquad \text{and} \qquad \hbar = h/2\pi \tag{4}$$

When an *external magnetic field* (**B_0**) is applied, positive **γ** nuclei tend to assume the magnetic quantum number +1/2, which represents alignment with the field; the energy of a magnetic moment **μ** in a magnetic field **B_0** is:

$$E = -\vec{\mu} \cdot \vec{B}_0 \tag{5}$$

In the presence of a strong field, the quantization z-axis is no longer arbitrary, but coincides with the field direction Z. Therefore:

$$E = -\mu_z \cdot B_0 \tag{6}$$

where $\boldsymbol{\mu_Z}$ is the **z** component of $\boldsymbol{\mu}$ (the projection of $\boldsymbol{\mu}$ onto $\mathbf{B_0}$) and $\mathbf{B_0}$ is the strength of the field (Fig. 2.1).

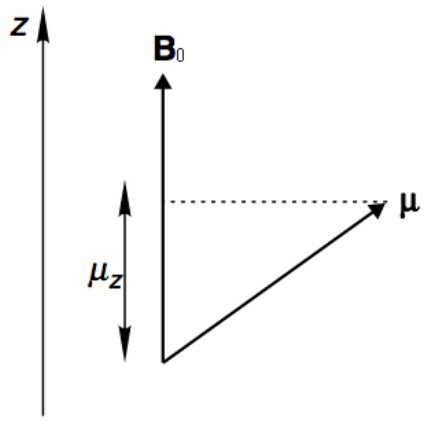


Figure 2.1: The relationship between the magnetic field $\mathbf{B_0}$, the nuclear magnetic moment $\boldsymbol{\mu}$, and its component along the field direction, $\boldsymbol{\mu_Z}$ (the scalar product of $\mathbf{B_0}$ and $\boldsymbol{\mu}$).

Substituting Eq. 3 into Eq. 6 it is possible to calculate the energy associated to each different energetic level, termed as *Zeeman levels*:

$$E = -\mu_z B_0 = -\gamma I_z B_0 \tag{7}$$

$$Em = -m\hbar\gamma B_0 \tag{8}$$

Spin 1/2 nuclei (e.g., ^{1}H, ^{13}C, ^{15}N) give rise to only two states corresponding to $\boldsymbol{m}$ **= +1/2** and **-1/2** (Fig. 2.2). The energy spacing between them is given by:

$$\Delta E = \gamma\hbar B_0 = (h/2\pi)\gamma B_0 = h\nu_0 \tag{9}$$

The speed at which the nuclei precess around the $\mathbf{B_0}$ axis is called *Larmor frequency* and its NMR absorption frequency is expressed as $\boldsymbol{\omega_0 = 2\pi\nu_0 = \gamma B_0}$ in rad^{-1} or $\boldsymbol{\nu_0 = \gamma B_0/2\pi}$ in Hz. The rotation is clockwise or anticlockwise depending on the sign of the gyromagnetic ratio for any particular nucleus. If $\boldsymbol{\gamma}$ it is positive (e.g., for ^{1}H and ^{13}C), then the +1/2 state lies lower in energy (α level in Fig. 2.2), and vice versa for negative $\boldsymbol{\gamma}$ values (e.g., ^{15}N).

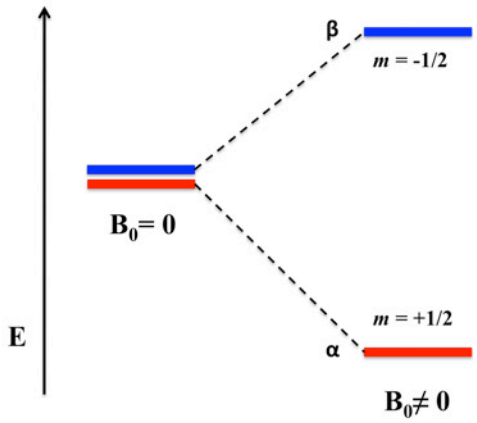


Figure 2.2: Energy-level diagram for nuclei with $I = 1/2$ and positive γ. When the external magnetic field (B_0) is applied the energy associated to each *Zeeman levels* is splitted in a number of different energetic levels determined by the magnetic quantum number ***m***.

The two energy state (α and β) will be unequally populated, so NMR absorption is a consequence of transitions between the levels stimulated by applied radiofrequency radiation. The ratio of nuclei in the upper- and lower-energy levels it is given by the *Boltzmann distribution* and is linearly related to the strength of magnetic field **(B_0)**; this explains why it is so important to produce increasingly powerful NMR instruments, able to generate higher magnetic field. Despite a tiny imbalance of populations there is a net magnetization that can be represented by a vector **M** pointing in the **z** direction and with a length proportional to the population difference (Fig. 2.3).

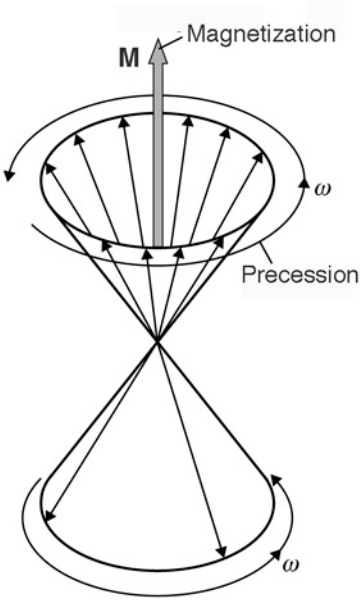


Figure 2.3: In the presence of an external magnetic field, the spins precess around their cones and there are also changes in the populations in the two spin states. As a result, there is a net magnetization along the *z*-axis.

Even if all the contributing spins have components precessing in the xy-plane there is no net magnetization in this plane. At the thermal equilibrium the magnetization is precessing around the field direction in a not coherent manner, which results in not observable magnetization (Fig. 2.4).

Magnetization becomes coherent (i.e. observable) when a Radio Frequency (RF) magnetic field $\mathbf{B_1}$ is applied into the xy-plane (Fig. 2.4). If the frequency of the field is equal to the *Larmor frequency* of a given nucleus, $\mathbf{B_1}$ can interact with the nuclear magnetic moment **(μ)** deviating the magnetization from its equilibrium position along the z-axsis towards the xy-plane where it is precessing (Fig. 2.4). This is the so called condition of *magnetic resonance* and it determines two important phenomena: 1) some nuclear spins jump from the state α to the state β, this causes the elimination of the population difference and so the net value of magnetization along z-axis decreases until zero; 2) nuclear magnetic moments are forced to precess in phase, i.e. to rotate all together on the cone surface, instead of random phase precessing. The two effects together produce a torsion of magnetization along z axes, in xy-plane (Fig. 2.4).

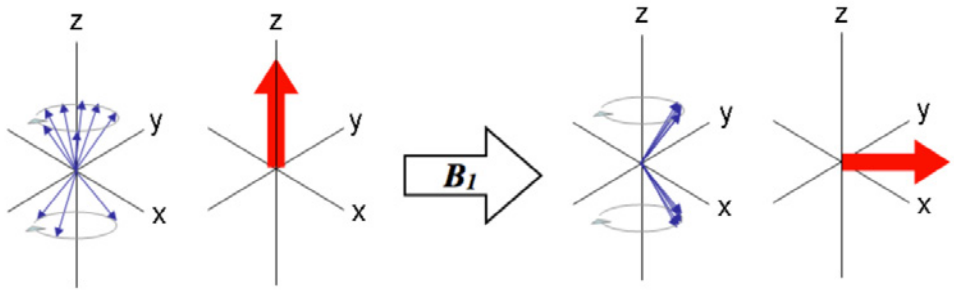


Figure 2.4: Magnetization torsion. In a resonance experiment, a radio-frequency magnetic field $\mathbf{B_1}$ is applied in the xy plane. If the frequency of the second field $\mathbf{B_1}$ is equal to the *Larmor* frequency of a nucleus in the sample, its magnetization vector begins to rotate around the direction of the $\mathbf{B_1}$ field [25].

The magnetization component in xy-plane **($\mathbf{M_{xy}}$)** can be measured, because it induces an alternating electric current in a coil (receiver) aligned with y axes; the signal in the coil can be amplified and processed.

If it is considered only one axis on the cartesian plane where magnetization precesses, the intensity of the latter can be measured as the sum of two vectors of equal intensity that rotate in opposite directions. Experimentally, the intensity of the magnetization **(M)** can be acquired by placing a loop detector on one of the axes of the plane and measuring the intensity of the current variation (Fig. 2.5).

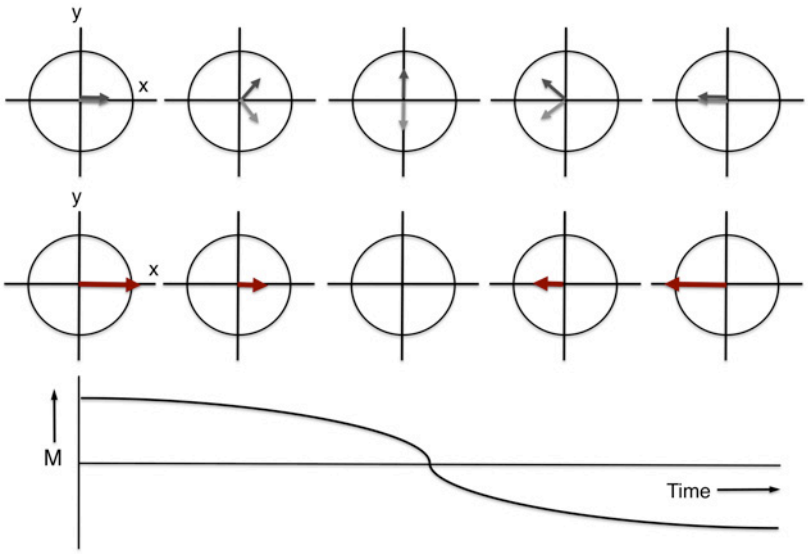


Figure 2.5: The intensity of the magnetization vector (**M**) is easy to describe if it is decomposed in two main components of equal intensity rotating in opposite direction. The sum of the two component gives the intensity of the magnetization on an axis, the x-axis in that case.

The oscillating signal decays as a function of time, as the phase coherence between the precessing magnetic dipoles is lost; it is called Free Induction Decay (FID) and represents the NMR signal, measured experimentally in the time domain. Time domain signals are converted into frequency domain signals (amplitude versus frequency) using the Fourier Transform (FT), thus acquiring an NMR spectrum. (Fig. 2.6)

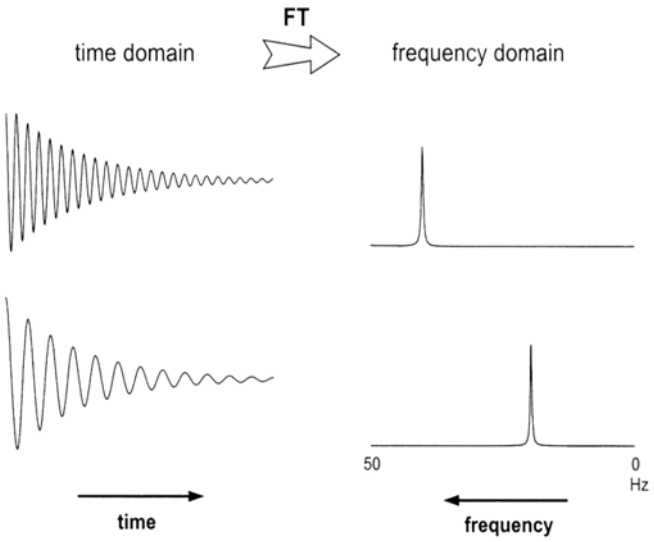


Figure 2.6: Representation of the NMR spectroscopy in Fourier transform.

In the case of Fourier Transform NMR all the nuclei of a species are excited simultaneously with a single electromagnetic radiofrequency pulse, produced using $\mathbf{B_1}$ field applied along x axes during a $\mathbf{t_p}$ time. This pulse is monochromatic, i.e. the frequency of the irradiation is unique and equal to *Larmor frequency* $(\boldsymbol{\omega_0})$. The velocity of magnetization precession around x axes, is defined by $\mathbf{B_1}$ direction and is expressed as:

$$\frac{d\alpha}{dt} = \omega_1 = \gamma B_1 \tag{10}$$

Consequently, the magnetization torsion angle is pulse-time dependent according to the equation:

$$\alpha = \gamma B_1 \cdot t_p \tag{11}$$

When the frequency of $\mathbf{B_1}$ is equal to $\boldsymbol{\omega_0}$, the resonant condition is achieved, so the alternating magnetic field ($\mathbf{B_1}$) and the magnetization ($\mathbf{M}$) interact generating a *torque* on $\mathbf{M}$ orientation. After the pulse, in the absence of the external $\mathbf{B_1}$, the system will try to go back to equilibrium conditions with the passage of time in a process that is called relaxation.

Relaxation processes

There are two kinds of relaxation process in NMR that take place after the radio frequency pulse has been switched off. The first is related to the establishment of thermal equilibrium in a bulk of nuclear magnets with different energy. As the system reverts to thermal equilibrium exponentially, the $\mathbf{z}$ component of magnetization approaches its equilibrium value $\mathbf{M_z}$ with a time constant called the *longitudinal relaxation time* $\mathbf{T_1}$ (12) (Fig. 2.7).

$$M_z(t) = M_z(t_0) \cdot e^{-t/T_1} \tag{12}$$

The constant $\mathbf{T_1}$ (s rad^{-1}) reflects the efficiency of the coupling between a nuclear spin and its surroundings (lattice) and is also called the *spin-lattice relaxation time*. Spin-lattice relaxation is an energy effect. A shorter $\mathbf{T_1}$ value means that coupling is more efficient and vice versa. Spin-lattice relaxation times lie between 10^{-3} and 10^2 s for liquids.

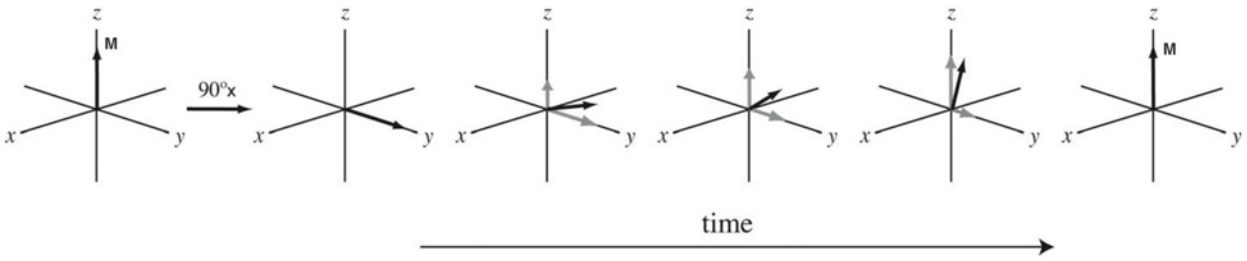


Figure 2.7: Longitudinal relaxation time $\mathbf{T_1}$ after switching off the radio frequency pulse. This relaxation represents a loss of energy (heat) from the spins to the surroundings.

The second kind of relaxation is illustrated in Figure 2.8. Consider a group of nuclei, precessing in phase about a common magnetic field along the z-axis, they produce a resultant rotating magnetic vector with the component in the xy-plane. If the nuclei, due to mutual exchange of spin energies, lose their phase coherence, there are as many positive as negative components in xy-plane and the resultant vector moves toward the z-axis (Fig. 2.8). The randomisation, i.e. the decay of the y or x component of magnetization to zero (equations 13 and 14), occurs exponentially with a time constant called the *transverse relaxation time*, $\mathbf{T_2}$ (s rad^{-1}).

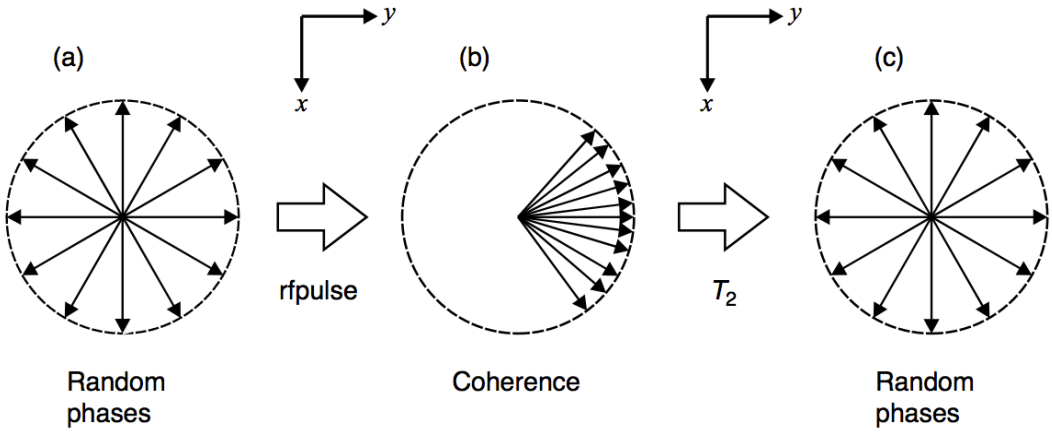


Figure 2.8: The effect of a radio frequency (rf) pulse on the magnetic moments of the individual spins in a NMR sample (looking down the z-axis). Starting from the equilibrium state with random phases (a), a pulse along the x-axis in the rotating frame causes the spins to precess in phase (b), producing a net y magnetization in the sample (coherence). After switching the field the randomization occurs with transverse relaxation time $\mathbf{T_2}$ (c).

From the equation 11 it is clear that, knowing the γ of the nuclei of a species and the intensity of $\mathbf{B_1}$ it is possible to obtain the values of $\mathbf{t_p}$ required to get specific torsion angles. After a 90° pulse, for example, the decrease of magnetization component in the xy-plane will be observed (13, 14); this is

due to the phase shift of the spins, and the simultaneous recovery of magnetization in the z-axis (12) (Fig. 2.4), determined by redistribution of the spins between the two energy levels due to the presence of the field $\mathbf{B_0}$.

$$M_x(t) = M_x(t_0) \cdot e^{-t/T_2} \tag{13}$$

$$M_y(t) = M_y(t_0) \cdot e^{-t/T_2} \tag{14}$$

Longitudinal relaxation is always slower than transverse relaxation ($\mathbf{T_1 \geq T_2}$), and therefore is the $\mathbf{T_2}$ relaxation time that determines the width of spectral lines according to the relationship:

$$\Delta\nu_{1/2} = \frac{1}{\pi T_2} \tag{15}$$

where the $\Delta\nu_{1/2}$ is the line width at half the maximum height, assuming that the magnet is perfect and the field is homogeneous in any locations of the sample. This means that the shorter the transverse relaxation time, the lower the height of the peak, with a consequent the broadening of the lines that lead an undesirable lowering of signal-to-noise (Fig. 2.9).

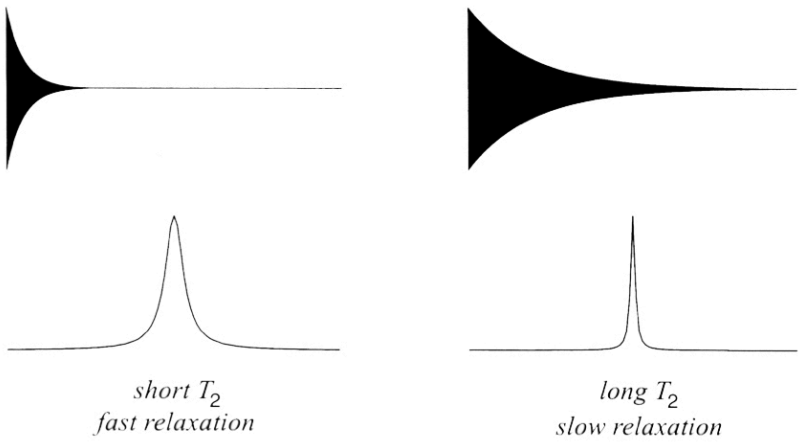


Figure 2.9: Transverse relaxation (T_2) magnetization loss in the xy-plane.

The Fourier transform NMR spectrometer

The sensitivity and resolution of the spectrometer depend upon the strength and quality of the magnet, which is thus the key component of the instrument. It is advantageous to operate at the highest possible field strength. In addition, the field must be highly homogenous and reproducible. The radio-frequency coil acts as both a transmitter and a detector of the resonance frequency. The measured signal processed by the computer is a low-frequency line resulting from the difference

between the transmitted and detected frequencies. The sample is placed in the centre of the cylindrical magnet to ensure that all the magnetic nuclei experience the same average field. Although a superconducting magnet operates at liquid helium temperature (4 K), the sample itself is normally at room temperature. In order to perturb the spin system with radio-frequency energy, the spectrometer contains sophisticated pulse programmer and transmitter units, which allow the application of complex pulse sequences to the sample of interest. The pulse/acquisition/delay Sequence is repeated *N* times (*NS*) until Signal-to-Noise ratio (*S/N*) is satisfactory.

$$S/N \propto \sqrt{NS} \tag{16}$$

During an NMR experiment, the FID that it is obtained is the sum of all the FIDs of singles nuclei that are present in the sample. The Fourier transform extracts all the individual FID and assigns to each nucleus a peak, with intensity dependent on its $\mathbf{T_2}$, centered on a particular frequency. The *Larmor* frequency of a given nucleus is strongly affected by its chemical environment that is to say that every nucleus of the same species in a sample experienced a different magnetic field (**B**). The variability in the electron density surrounding a nucleus, depending on the group to which it is bound, generates specifics local magnetic fields $\mathbf{B'}$ in the opposite direction of external field $\mathbf{B_0}$. As a consequence, NMR signals from molecules provide a wealth of spectral information that can serve to elucidate their chemical structure.

$$B = B_0 - B' \tag{17}$$

The shift in absorption frequency of a nucleus depending on the group to which it is bound is called the *chemical shift*. The *chemical shift* (***δ***) is proportional to the applied magnetic field (B_0) and it is commonly indicated in parts per million (***ppm***) relative to a reference compounds.

$$\delta(ppm) = \frac{\nu(\mathrm{Hz})}{\nu_0(M\mathrm{Hz})} \cdot 10^6 \tag{18}$$

Another important effect to take into account during the analysis of a NMR spectrum is the *coupling constant* (**J**) for the magnetic interaction between two nuclei that have spins and are connected by chemical bond (through-bond effect). **J coupling** introduces a local field on the bound nuclei that either is against or along the local field observed at the nuclei. This field difference will in turn lead to a *spin-spin splitting* of the peak that does not depend on the applied magnetic field and is commonly indicated in Hz.

Both the *chemical shift* and *spin-spin splitting* are very important in structural analysis. Experimentally, the two effects are easily distinguished. The peak separation resulting from a

chemical shift is directly proportional to the field strength, while spin-spin splitting is independent of the strength of the external magnetic field.

Nuclear Overhauser Effect

A saturating radio-frequency field is applied to the high-γ spins ***S***. The resulting population redistribution leads to a polarization enhancement of the ***I*** spins, provided the relaxation processes are favourable. This transfer of polarization through-space is called **NOE**; in other words **NOE** is a change of the intensity of an NMR signal when the transitions of another one, close in space, are perturbed (Fig. 2.10).

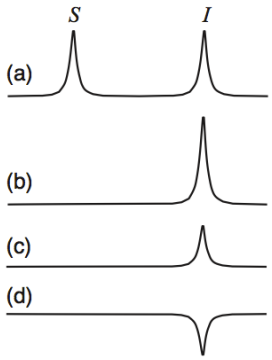


Figure 2.10: Schematic presentation of NOE. A molecule contains two inequivalent spins, *I* and *S*, with no scalar coupling, so that the NMR spectrum consists of a singlet at each of the chemical shifts. (a) Conventional spectrum of two neighbouring spins *S* and *I*. (b), (c) and (d) Possible spectra resulting from saturation of *S* resonance: the *I* peak gets stronger (b), weaker (c), or inverts (d) depending on the conditions.

NOE is a consequence of the modulation of *dipole-dipole coupling* between the two spins by the molecular Brownian motion. The equation describing NOE has the general form:

$$\text{NOE} \propto f(r_{IS})\, f(\tau_c) \tag{19}$$

where $\boldsymbol{f(r_{IS})}$ is a function of the distance $\boldsymbol{r_{IS}}$ between the protons and $\boldsymbol{f(\tau_c)}$ is a function of the molecular rotational correlation time $\boldsymbol{\tau_c}$, which accounts for the influence of the motional averaging process on the observed NOE. NOE is applicable only for very close in space nuclei, typically closer than 5 Å to each other.

2.3.2. The determination of protein structures

As it has been already mentioned, NMR spectroscopy technique is used for solution determination of protein structures. In this process structures are calculated by using computer algorithms, starting from the primary sequence of the molecules and a set of measured interactions as well as dihedral angles that are structure determining. The first procedure for determining NMR constraints is the assignment of signal, meaning that the protein atoms responsible for every signal are determined. The original procedure implies the two dimension ^{1}H-^{1}H NMR experiments in which only protons are used to resolve the spectra. This is applicable only to small and medium sized molecules, whereas, when the protein is large and the spectral resolution is decreased, it is necessary to adjust the approach. Such proteins have larger peak widths and extensive signal overlap, which introduce ambiguity in chemical shift assignment making it difficult to analyze. In this case multidimensional NMR experiments and additional protein labeling are used.

Two-dimension (2D) NMR

The term one-dimensional NMR refers to experiments in which the transformed signal is presented as a function of a single frequency. By analogy, in a two-dimensional NMR experiment a double Fourier Transformation yields a two-dimensional spectrum in which the coordinate axes correspond to two frequency domains.

According to Ernst [26] the 2D-NMR experiments are classified into three groups:

1) Experiments designed to ***correlate*** transitions of coupled spins by transferring transverse magnetization or multiple-quantum coherence from one transition to another in the course of a suitably designed mixing process. This type of experiment is called Correlation Spectroscopy known under the acronym (COSY).
2) Experiments designed to ***separate*** different interactions (e.g., chemical shifts and spin-spin couplings) in orthogonal frequency dimensions, with the purpose of resolving one-dimensional spectra by spreading overlapping resonances in a second dimension. These experiments require conditions such that the spectra in the evolution and detection periods contain different information. The method is called homonuclear or heteronuclear two-dimensional J-resolved spectroscopy or spin-echo spectroscopy.
3) Nuclear Overhauser Enhanced Spectroscopy (NOESY), which is concerned with the study of ***dynamic processes*** such as chemical exchange, cross-relaxation, or transient Overhauser effects.

The first step in order to overcome the resolution problems of 2D ^{1}H-^{1}H NMR related to the protein size is the use of ^{15}N-labeled protein, that introduce the possibility to measure the signal relative to NMR frequencies of ^{15}N spins in a third axis. As a result, the NMR peaks of the 2D ^{1}H-^{1}H NMR spectrum are distributed among several ^{1}H-^{1}H planes. Each plane corresponds to a small ^{15}N chemical shift range and, on that plane, only ^{1}H-^{1}H peaks, from the H_N bound to nitrogen atoms having the proper ^{15}N chemical shift value to correlated protons, are present.

Nevertheless, with proteins larger than 18-20 kDa, ^{15}N editing is not enough to provide the peaks separation needed for spectral analysis. In this case it is necessary the use of doubly-labeled ^{13}C, ^{15}N protein samples in triple resonance NMR experiments. In these experiments three different nuclei (^{1}H, ^{13}C and ^{15}N) are correlated through heteronuclear scalar couplings **(J)**. Triple resonance spectra have many advantages: they contain only a few signals on each frequency, they have an high sensitivity as the magnetization is efficiently transferred, transfer times are shorter than for homonuclear couplings and, moreover, the signal intensity losses due to fast relaxation associated to high molecular weight are smaller than in ^{1}H-^{1}H experiments. (Fig. 2.11)

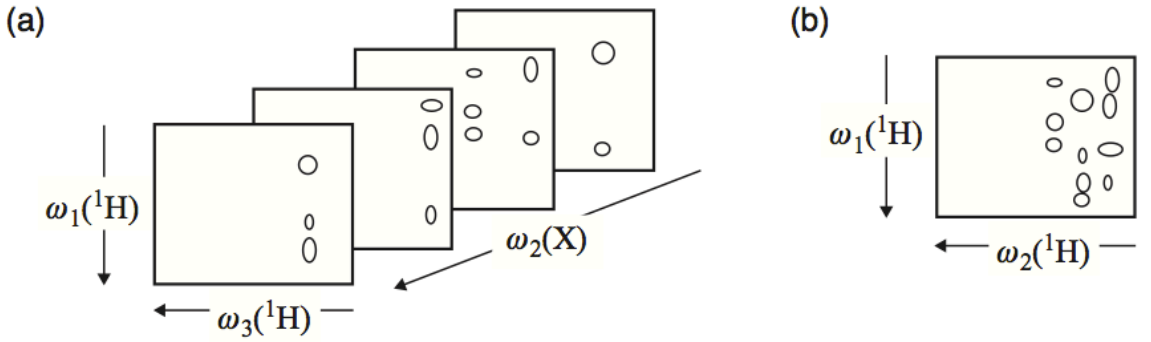


Figure 2.11: Scheme illustrating the improved peak separation in a 3D heteronuclear-resolved [^{1}H,^{1}H] NMR experiment (**a**), when compared with the corresponding 2D [^{1}H,^{1}H] NMR experiment (**b**). The two spectra contain the same number of peaks. In the 3D spectrum these are distributed among multiple $\omega_1(^{1}H)\omega_3(^{1}H)$ planes that are separated along the heteronuclear chemical-shift axis, $\omega_2(X)$ [27].

In general a *cross-peak* in a plane of a multidimensional spectrum results from a correlation between the signals of the nuclei along each of the axes at these value; it is the effect of a phenomenon called magnetization transfer between nuclei that are coupled. By looking for *cross-peaks* between various signals one can thus determine which atoms are connected to one another (within a small number of chemical bonds).

A large number of triple resonance experiment exist and the names of all nuclei which are used for magnetization transfer during the experiment are listed in the order of their use, bracketing the names of nuclei which are involved only for transfer and whose frequencies are not detected in the spectra. Below are reported some of the experiments used for assignment of the backbone of the protein; in the following figures, atoms that are contributing to the respective spectra by cross-peaks are colored in pink, whereas atoms that are involved in transferring magnetization but are not observed in the spectrum are colored in light blue (adapted from "Protein NMR - A Practical Guide" - http://www.protein-nmr.org.uk).

^{1}H-^{15}N HSQC (Heteronuclear Single Quantum Coherence) shows cross-peaks for each N-HN pair for a protein. The spectrum presents cross peaks from backbone N-HN pairs but also from side chains of certain aminoacids. Proline residues are not visible in the spectra.

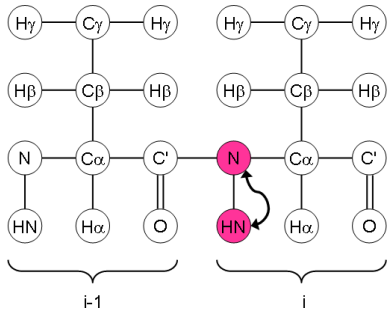


Figure 2.12: ^{1}H-^{15}N HSQC. Schematic representation of amino acids of a protein and atoms involved in the effects recorded in an HSQC-experiment.

This spectrum is rather like a fingerprint and is usually the first heteronuclear experiment performed on proteins. From it you can assess whether other experiments are likely to work and for instance, whether it is worth carbon labelling the protein before spending the time and money on it; or if your protein is reasonably large you might be able to judge whether deuteration might be necessary. Moreover all the experiments on ^{15}N and ^{13}C labeled proteins use HSQC type of magnetization transfer.

HNCO is the most sensitive of all the 3D spectra and it correlates the N-HN pair with the COi-1; together with **HN(CA)CO** (3D), that gives individual cross-peaks for COi and COi-1, **HNCO** makes it possible to distinguish which signals belong to which residues for each NH group.

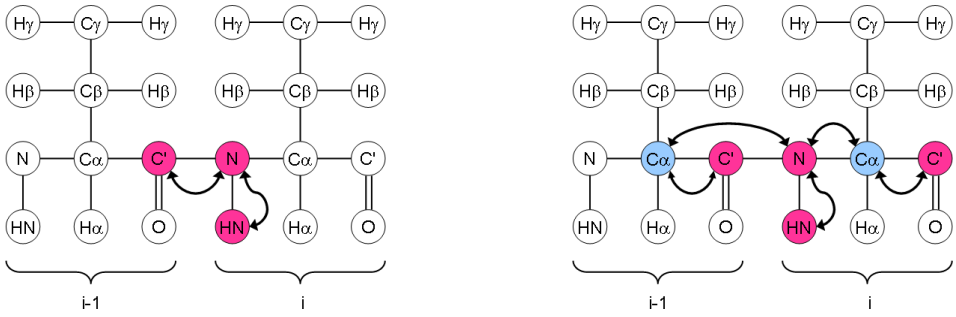


Figure 2.13: HNCO and **HN(CA)CO**. Schematic representation of amino acids of a protein and atoms involved in the effects recorded in an HNCO- and HN(CA)CO-experiments (left and right, respectively).

HNCA (3D) correlates the N-HN pair with Cαi and Cαi-1, thus giving two cross-peaks in the third dimension for each N-HN pair (Fig. 2.14). The cross-peaks for Cα is usually more intense.

HN(CO)CA (3D) is a spectrum which is like the **HNCA**, but it is selective for the Cα of the preceding residue (i-1).

Together these two spectra make it possible to distinguish which cross-peaks belong to which amino acid for each NH group.

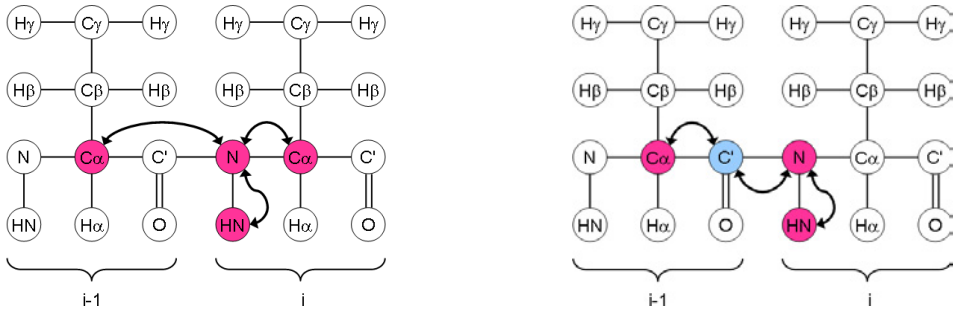


Figure 2.14: HNCA and **HN(CO)CA**. Schematic representation of amino acids of a protein and atoms involved in the effects recorded in an HNCA- and HN(CO)CA-experiments (left and right, respectively).

CBCANH gives individual cross-peaks for Cαi, Cβi, Cαi-1 and Cβi-1 for each N-HN pair.

CBCA(CO)NH gives individual cross peaks only for Cαi-1 and Cβi-1 and also in this case a comparison of the two spectra makes it possible to distinguish which cross-peaks belong to which amino acid in the backbone assignment.

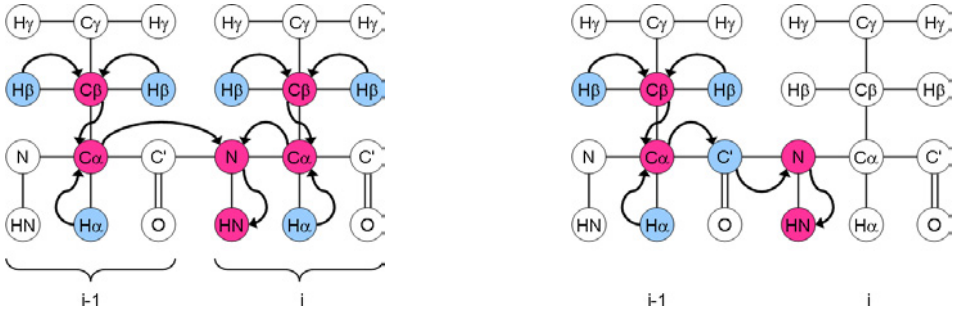


Figure 2.15: CBCANH and **CBCA(CO)NH**. Schematic representation of amino acids of a protein and atoms involved in the effects recorded in an CBCANH- and CBCA(CO)NH-experiment (left and right, respectively).

These two experiments, although with a lower sensitivity with respect to the previous two pairs, provide in addition the Cβ frequencies which are instrumental for the identification of residue type and secondary structure.

During the past 20 years the power of magnets used to generate the static magnetic field (B_0) has been increased with a gain in term of sensitivity and resolution. Nowadays the highest magnetic field available for NMR has a proton resonance frequency of 950 MHz. For commonly used heteronuclear experiments, however, the advantages of higher magnetic fields are offset partly by field-dependent line broadening due to increased transverse relaxation rates. The latter increases with molecular mass and has a dominant impact on the upper size limit for macromolecular structures that can be studied by NMR in solution. The **Transverse Relaxation-Optimised Spectroscopy (TROSY)** method was developed to overcome this limitation [28]. In order to study systems with a molecular weight above 30 kDa there has been a development of strategies for line sharpening that combine an at least partial ^2H labelling of non-exchangeable hydrogens with novel NMR pulse schemes. TROSY is a spectroscopic means that suppresses transverse nuclear spin relaxation, which is the direct cause of the deterioration of NMR spectra of large molecular structures [29] (Fig. 2.16). TROSY uses interference effects between different relaxation mechanisms and generally it works best with deuterated proteins at high magnetic fields. By applying TROSY, the spectra of proteins of molecular mass close to 100 kDa were obtained.

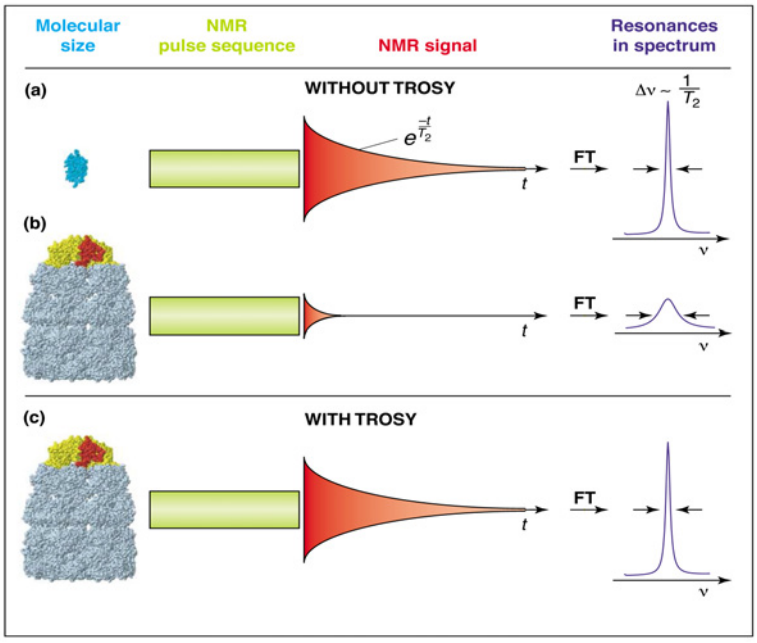


Figure 2.16: NMR spectroscopy with small and large molecules in solution. (**a**) The NMR signal obtained from small molecules in solution relaxes slowly; it has a long transverse relaxation time (T_2). A large T_2 value translates into narrow line widths in the NMR spectrum after Fourier transformation (FT) of the NMR signal. (**b**) By contrast, for larger molecules, the decay of the NMR signal is faster (T_2 is smaller). This results both in a weaker signal measured after the NMR pulse sequence and in broad lines in the spectra. (**c**) Using TROSY, the transverse relaxation can be substantially reduced, which results in improved spectral resolution and improved sensitivity for large molecules [29].

Isotope labelling of proteins

Proteins used for NMR studies can be labelled by using a number of different protocols to produce molecules with different patterns of ^{2}H, ^{13}C and ^{15}N incorporation. The majority of isotopically labelled proteins studied by NMR were obtained by heterologous protein expression in *E. coli*. This recombinant expression is achieved using defined media for bacterial growing. For larger proteins, uniform or fractional ^{2}H labelling is also used. Uniform or random labelling strategies result in ^{2}H incorporation throughout a protein in a roughly site-independent manner [30]. The use of deuteration allows to simplify the assignment procedure and it is of fundamental importance for structural and dynamics studies of proteins and protein complexes.

2.3.3. Protein-protein interaction using NMR

The information flow associated with essential biological events (e.g., gene transcription and translation, cell growth and differentiation, the immune response, and neurotransmission) is largely

mediated by cascades of protein-protein interactions, which lead to changes in shape, dynamics, and chemical or physical properties of the proteins involved. It is becoming clear that such detailed knowledge cannot be reliably obtained from the high-resolution structures of the individual components [31]. Taking advantage of a powerful aspect of NMR spectroscopy, namely its ability to examine these interactions at the atomic level and at near physiological conditions is solution, the latter has assumed a unique role in the investigation of protein complexes that cannot crystallize or that are not possible to crystallize in a biologically relevant conformation [24]. The development of NMR spectroscopy equipment, such as high field magnets and cryogenic probes, in combination with new NMR approaches, i.e. development of more efficient isotopic labelling schemes, especially perdeuteration, and the advent of TROSY, were accompanied to progress in methods for studying protein-protein interactions by NMR [31]. In some cases perturbation could also be an excellent indicator of allosteric processes induced by interactions, when its effects are extended slightly beyond the direct contact area [32].

Chemical Shift Perturbation (CSP) Mapping

Generally, one can use chemical shift changes to predict what exactly happens at the interface of interaction. Qualitatively important information about interacting surface of the labelled partner can be obtained by monitoring the chemical shift perturbation per residue. This analysis relies on assumption, that an HSQC-map of a particular protein represents a fingerprint of its structure. The chemical shift of ^{15}N and ^{1}H are particularly sensitive to any change in environment; thus, perturbation of these shifts as a result of complex formation provides a highly sensitive tool for the mapping of binding sites on a protein. In a CSP experiment the ^{1}H-^{15}N HSQC spectrum of one protein is monitored when the unlabeled interaction partner is titrated in, and the perturbations of the chemical shifts are recorded. Complex formation results in selective changes in chemical shift of the various nuclei, thus providing information on the structure of the complex in the solution. For a binary complex comprising protein A and B, labelled protein A is mixed with unlabeled protein B or vice versa. Usually the protein that has to be added has a much higher concentration than the other one and it is added stepwise in small volumes. The total volume of the mixture does not change during the titration by more than 5-10%. Ionic strength and pH values should not change much either. Combining pertubation observed in both dimensions of ^{15}N-^{1}H HSQC with the following equation:

$$\Delta\delta = \Delta H_N + \Delta N / 7 \qquad (20)$$

or for glycine residues

$$\Delta\delta = \Delta H_N + \Delta N/5 \quad (21)$$

Where ΔH_N and ΔN are the chemical shift differences (in ppm) of the amide proton (H_N) and nitrogen (N) resonances, respectively [33].

Moreover, the variation of intensity of the amide NH cross-peaks upon addition of titrant is estimated using the equation:

$$\Delta I = \frac{I - I_0}{I_0} \quad (22)$$

where I and I_0 are the amide cross-peak intensities in the absence and in the presence of titrant, respectively.

The binding interface hence will be defined by the residues exhibiting the largest shifts upon the complex formation, which in other words means that the residues that do not exhibit any change or exhibit changes within the range of error are those not-localized at the binding interface.

2.3.4. NMR dynamics and motility

Dynamics are crucial to protein function that depends on alteration in three-dimensional structure in response to specific molecular interactions. Structures analysis allow to derive many valuable information on the organization and interaction of protein molecules, but only a detailed knowledge about motion can provide a complete picture about the mechanisms of proteins function in living cell. A dynamic study at the atomic level of all timescales is, therefore, necessary, because a variety of motions take place in the same molecule and at the same time. Limited windows onto this expansive timescale are afforded by a number of specialised techniques, thus only allowing certain motions to be probed. Solution NMR permits to characterize protein internal dynamics, giving an accurate elucidation of structure-activity relationships, since it can provide site-specific information about protein motions over a large range of time scales (Fig. 2.17). NMR dynamics analysis is extremely important for studying atomic motion of protein regions, such as solvent-exposed loops, that is not possible to monitor with rigid crystals. Over the past decade, NMR relaxation experiments employing model-free analysis have become the standard used to characterize protein motions on a picosecond to nanosecond time scale. Any molecule in solution rotates as a whole,

with a tumbling rate that depends on its size and shape as well as on the viscosity of the solution. The isotropic overall tumbling correlation time ($\boldsymbol{\tau_m}$) is the time needed for the molecule to rotate by one radian, where the molecule is considered as a sphere, and sets in the range of units to hundred nanoseconds for proteins; its units are (s rad^{-1}).

To analyze protein backbone dynamics, a set of T_1, T_2, and ^{1}H-^{15}N NOE experiments for amide nitrogen are often recorded [34]; since there is one backbone amide in each amino acid, except for proline, these experiments provide information on backbone dynamics of an entire protein. Among them, ^{1}H-^{15}N NOE is sensitive to sub-nanosecond motion, and provides qualitative information on internal dynamics in proteins [35].

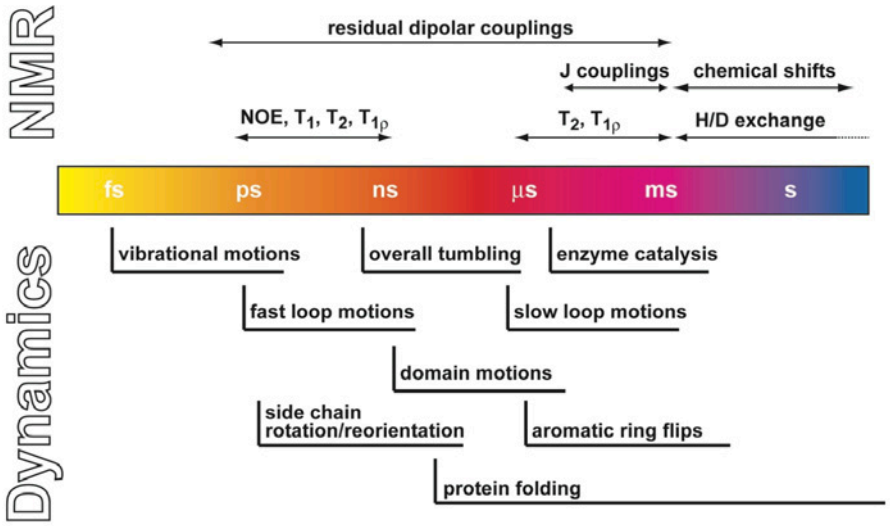


Figure 2.17: Proteins dynamics time scale with high resolution NMR spectroscopy.

The link between molecular dynamics and relaxation is due to the fact that local magnetic field fluctuations are caused by molecular motions. NMR relaxation data, consisting of the R_1 and R_2 relaxation rates (reciprocals of T_1 and T_2, respectively) and ^{1}H-^{15}N NOEs, contains information about the dynamics of individual atoms of the molecule in the timescale window from picosecond to nanosecond [36][37][38]. In an isotropic approximation, the correlation time for the overall tumbling (τ_m) of the protein can be derived from the ratio of R_2/R_1 that for spherical molecules is constant [39]:

$$\tau_m = \frac{1}{2\omega_N}\sqrt{6\frac{R_2}{R_1} - 7} \qquad (23)$$

where ω_N is the *Larmor frequency* of ^{15}N nucleus [40].

It should be mentioned that proteins often cannot be approximated to spherical molecules and therefore experience anisotropic molecular rotation. Since rotational diffusion is faster around a long axis than a short axis, the relaxation of a ^{15}N nucleus will be differentially affected depending on whether the associated NH bond vector is aligned with the long or the short axis of the molecule, causing the R_2/R_1 ratio to be site-specific even in the absence of dynamics [41]. In case the isotropic diffusion model cannot be assumed, but the global tumbling of the protein can rather be described by an axially symmetric or even anisotropic diffusion tensor, the components of the diffusion tensor must first be determined. The rotational diffusion tensor is specified by three principal components, $\mathbf{D_{xx}}$, $\mathbf{D_{yy}}$, and $\mathbf{D_{zz}}$ - the diffusion constants for rotation about the x, y, z principle axes - and by the orientation of the principal axes relative to that of the protein. The principal axes, x, y, z, are fixed in the protein, and defined as the coordinate axes in which the diffusion tensor assumes a simple, diagonal form. Assume we have a molecule with $D_{xx} < D_{yy} < D_{zz}$. For an internuclear vector parallel to the z-axis, its reorientations will be caused by molecular rotations around the x- and y-axes, but not z-axis, so that the apparent rotational diffusion rate for the vector will be determined by $D_{zz} = (D_{xx}+D_{yy})/2$ (assuming small degree of anisotropy). On the contrary, reorientations of a vector perpendicular to the z-axis (e.g. along the x-axis) will be determined by $D_{xx} = (D_{zz}+D_{yy})/2 > D_{zz}$ and hence will proceed faster than for the previous internuclear vector. These differences in the tumbling rates will lead to differences in the spin relaxation rates for the corresponding pairs of nuclei; the effect will increase with the anisotropy of the molecule. The rotational diffusion tensor must be characterised accurately using the spin-relaxation rate ratio (R_2/R_1), a parameter which becomes independent of internal dynamics in the fast-motion limit while remaining highly sensitive to overall diffusion [42]. So the R_2/R_1 ratio can be used not only for calculating the global rotational correlation time but, by using the angular dependence of the ratio, can also be used to determine the anisotropy and rhombicity of the tensor. Once the orientational dependence of rotational diffusion tensor is determined it is possible to proceed in the dynamic analysis.

The NMR relaxation data are also dependant on the power spectral density function $J(\omega)$, since spins are relaxed by local fields fluctuating at the *Larmor* frequency ω. The probability function of finding motions at a given frequency ω can be described by the ***spectral density function*** (Fig. 2.18):

$$J(\omega) = \frac{2\tau_m}{1+(\omega\tau_m)^2} \tag{24}$$

From this equation (24) it is clear that the densities at low frequencies increase with increasing correlation times, that is to say, with larger molecules.

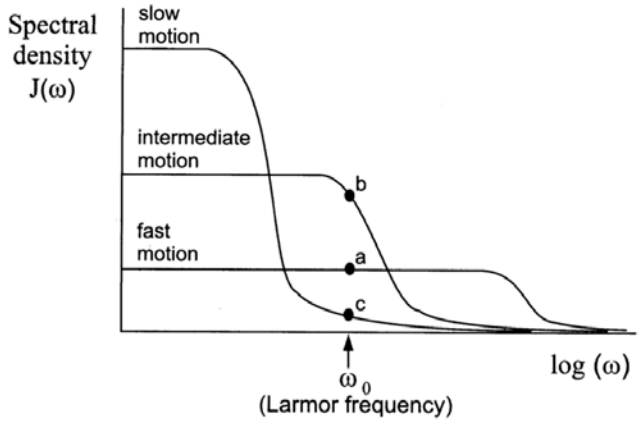


Figure 2.18: Frequency distribution of the fluctuating magnetic fields.

The technique known as ***Reduced Spectral Density Mapping*** [43]–[45] detangles the relaxation values into three spectral density values J(0), J(ω_X), and J(ω_H). The spectral density values at the three higher frequencies of $\omega_H - \omega_X$, ω_H, and $\omega_H + \omega_X$ are assumed to be approximately equal and hence J(ω_H) represents all three frequencies.

The spectral density values are a much better description of the dynamics than the relaxation data which are themselves difficult to interpret into a physical picture. Importantly in this analysis there is no separation of the Brownian rotational diffusion of the molecule from the internal mobility of individual atoms. For example if, in a protein, an α-helix relaxes slower than the core of the molecule, is this because it is more mobile than the rest of the protein or is it because the XH bond vectors of the helix are parallel with the short axis of a significantly anisotropic diffusion tensor? How does the timescale, amplitude, or other characteristics of the motion influence the spectral density values? Model-free analysis attempts to resolve these types of ambiguity.

Model-free analysis of NMR relaxation data is a technique which reveals information about internal dynamics at the atomic level and probes the chemical exchange phenomenon in a non-rigid system. This approach consists of the separation of the global Brownian rotational diffusion from internal motions relative to the diffusion frame and the description of these internal motions by amplitude

and timescale [36][46]. The internal dynamics is quantified by three types of parameter: the square of the Lipari and Szabo generalised order parameter ($\mathbf{S^2}$) which characterises the angular amplitude of the motion of a given bond vector, such as the N-H bond, with respect to the overall protein frame; the effective internal correlation time ($\boldsymbol{\tau_e}$) which links the amplitude to a timescale; and the chemical exchange relaxation parameter ($\mathbf{R_{ex}}$) which is an indicator of slower microsecond to millisecond timescale dynamics [47]. The so-called order parameters (S^2) reflect the flexibility of the polypeptide chain and range from one for complete rigidity to zero for high mobility [37] (Fig. 2.19). Timescale of these motions can range from picosecond to nanosecond.

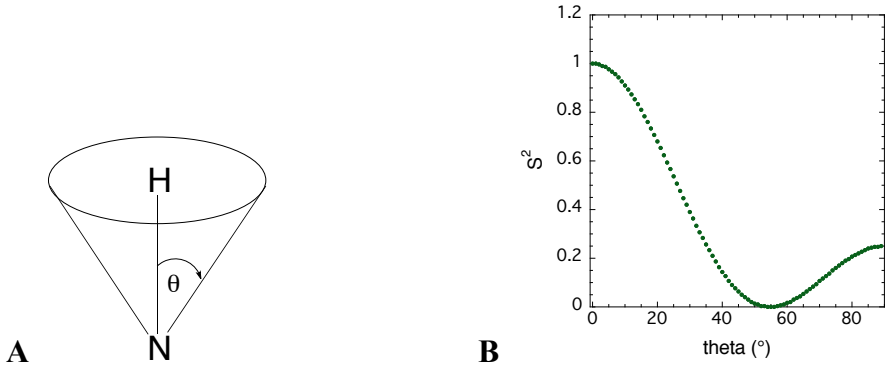


Figure 2.19: (**A**) θ is the semi-angle defining a cone of space taken by the N-H vector during internal motion; (**B**) The plot of S^2 as a function of θ.

The correlation time ($\boldsymbol{\tau}$) is given by:

$$\tau^{-1} = \tau_m^{-1} + \tau_e^{-1} \tag{25}$$

where τ_m is the overall tumbling correlation time of the macromolecule (with the assumption of an isotropic tumbling) and τ_e is an effective time constant for the internal motion of the N-H bond vector.

Chemical exchange relaxation (R_{ex}) occurs when a nucleus moves between magnetically distinct sites on the micro to millisecond timescale, either by conformational changes within the molecule or by chemical reaction; these movements causes longitudinal magnetic fluctuations, hence chemical exchange only affects the transverse relaxation rate R_2 .

Afterwards the theory was also extended to include motions on two different timescales [38] in which the *faster* (<100-200 ps) of the motions is parameterised by the amplitude $\mathbf{S_f^2}$ and correlation time $\boldsymbol{\tau_f}$ and the *slower* (>100-200 ps but still $<\tau_m$) by $\mathbf{S_s^2}$ and $\boldsymbol{\tau_s}$. The two order parameters are related by the equation:

$$S^2 = {S_f}^2 \cdot {S_s}^2 \tag{26}$$

and correlation times:

$$\tau_f^{'-1} = \tau_f^{-1} + \tau_m^{-1} \tag{27}$$

$$\tau_s^{'-1} = \tau_s^{-1} + \tau_m^{-1} \tag{28}$$

The *model-free* approach is based on a form for the total auto-correlation function that describes the overall and internal motions when it is assumed that they are independent:

$$G(t) = \frac{1}{5} G_0(t) G_i(t) \tag{29}$$

where $G_0(t)$ is the correlation function for overall motion, $G_0(t) = \mathrm{e}^{-t/\tau_m}$, and $G_i(t)$ is the simplest form for internal motions:

$$G_i(t) = \mathrm{S}^2 + (1 - \mathrm{S}^2) \cdot \mathrm{e}^{-t/\tau_e} \tag{30}$$

The correlation times τ, τ_m and τ_e are described by equation 25.

The time-dependence of the motions determines also the frequency at which the induced local fields fluctuate and therefore whether they may contribute to relaxation or not. In other words, *spectral densities*, J(ω), which are the Fourier transform of the correlation function, *G(t)*, tell us how much power is available from the motions of the molecule to cause fluctuations at the frequency. The original *model-free spectral density* function presented in Lipari and Szabo (1982a) [36] can be calculated as:

$$\mathrm{J}(\omega) = \mathrm{S}^2 \frac{\tau_m}{(1 + \omega\tau_m)^2} + \frac{(1 - \mathrm{S}^2)\tau_e}{(1 + \omega\tau_e)^2} \tag{31}$$

The internal motions in the *model-free* approach are solely characterised by the model-independent parameters S^2 and τ_e. Moreover, in this analysis, various *model-free* mathematical models are used to represent different classes of motion. By assuming certain order parameters or correlation times to be statistically negligible, either being one or zero respectively, a number of *model-free* models can be constructed. An order parameter of one means that the motion is statistically insignificant

whereas the correlation time of zero means that the motion is too fast for that parameter to be reliably extracted [48]. In combining parametric restrictions, in which statistically insignificant parameters are dropped, together with the addition of a parameter accounting for chemical exchange relaxation a number of increasingly complex models of *model-free* motions can be constructed [49]. These include models from m_0 to m_9 but commonly, for simplicity, only models m_1 to m_5 are used. As mentioned above, in certain situations, the assumption that all residues of the molecule under study will experience the same global rotational diffusion may not be the best model of the entire system. To redress this problem each residue can be assumed to tumble independently with its own global correlation time parameter called the *local* τ_m. In this case a new set of *model-free* models can be created which include an additional dimension to the above models [48], [49].

The *model-free* interpretation consists purely of data analysis methods using techniques from the mathematical fields of minimization and optimisation modelling and the statistical field of model selection [48], [50]. The quality of the fits between the experimental data and each model are calculated as χ^2 statistics. Finally the global model which best describes the entire system is selected [51] using the Akaike's Information Criteria (AIC) [52].

$$\mathrm{AIC} = \chi^2 + 2k \tag{32}$$

In which k is the number of parameters in the model. This selection criterion is based on parsimony principle of the Ockham's razor, and consists of choosing the model with the lowest number of parameters that is consistent with the data. In particular, the model with the smallest AIC value is considered as the best one, additionally taking into consideration the relative percentage errors between experimental and back-calculated values of R_1 and R_2. Over the past 15 years, many studies of backbone dynamics have been performed as extensions of NMR-based structure determinations, allowing a more complete picture of the proteins structural ensembles. In addition, there has been increasing interest in understanding the roles that the internal motions play in determining the stabilities and activities of proteins [41]. For the purpose of this thesis the dynamics analysis it has been used to better characterize the protein-protein interaction and long-range effects after binding.

3. Thermodynamics information

3. Thermodynamics information

Classical thermodynamics is a phenomenological science, concerned with precise observations on defined systems. Macromolecular solutions can be considered as thermodynamic systems. The complete set of thermodynamics equations describing a system, in a given temperature range, can be calculated from the functional dependence of enthalpy on temperature, which can be obtained, in principle, from careful calorimetric measurements. Experiments allowing the measurement of thermodynamic parameters with great precision became possible at the ending of XX century with the development of more sensitive calorimeters and microcalorimeters. Calorimetry measures heat exchange during temperature-induced changes in a system. When the system is a macromolecular solution, calorimetry experiment measures the heat capacity at constant pressure. The corresponding enthalpy and entropy functions are obtained by integrating the heat capacity and the heat capacity divided by the temperature, respectively, in the temperature range of the experiment. In Isothermal Titration Calorimetry, the amount of power required to maintain a constant temperature difference between a reaction bath and a reference cell during a titration is measured by the calorimeter; the heat absorbed or released by the chemical reaction is determined from the integral of the power curve over the appropriate time.

3.1 Isothermal Titration Calorimetry (ITC)

The Isothermal Titration Calorimetry (ITC) is considered the unique method able to provide not only the magnitude of the enthalpy changes (ΔH) upon reaction but also values for the binding affinity (K_b) and entropy changes (ΔS). Since these parameters fully define the thermodynamic of the binding process, ITC is playing an increasingly important role in the detailed study of protein-ligand interactions and the associated molecular design approaches, in particular with respect to drug design. ITC is also considered as the most quantitative technique available for measuring the thermodynamic properties of protein-protein interactions [53] and may be used as a tool for obtaining enzyme kinetic constants.

ITC analysis relies upon the accurate measurement of heat changes that follow the interaction of protein molecules in solution with a titrant, without the need to label or immobilize the binding partners, since the absorption or production of heat allows the determination of binding constants

(K_d or K_a), reaction stoichiometry (n) and all thermodynamic parameters (ΔG, ΔH, ΔS, ΔC_p) [54]. The basics that are behind this technique are reported below; the following paragraph is mostly adapted from the book "Method of molecular biophysics" [23].

Binding studies

If molecule A binds to molecule B to form the complex AB, the thermodynamic equilibrium dissociation constant, $\mathbf{K_d}$, is defined in terms of the different species concentrations, [A], [B] and [AB], at equilibrium:

$$K_d = \frac{[A][B]}{[AB]} \tag{33}$$

Where, following the binding experiment convention $K_d = 1/K_a$, the association constant **(K_a)** is a measure of the binding 'strength' or affinity of the interaction. The determination of the association/dissociation constant is probably the first aim in the detailed study of any protein-protein or protein-ligand interaction. K_d values, more often used than those of K_a, are very useful to determine at what protein concentration a complex might be formed and represent common parameters to classify the nature and the strength of a considered interaction. Moreover, the knowledge of K_d enables the calculation of the energetics involved in the interaction between a ligand and its receptor. In fact the association constant is determined by the Gibbs free energy difference (ΔG) between the bound and unbound states of the reaction at the equilibrium. At constant temperature the Gibbs free energy (ΔG) released by the reaction is given by:

$$\Delta G = \Delta H - T\Delta S \tag{34}$$

and

$$\Delta G = -RT\ln K_a \tag{35}$$

where R is the gas constant and T is the absolute temperature.

From the second law of the thermodynamic, a system that exchanges energy with its surroundings evolves in order to maximise its free energy, so, the spontaneous sense of the reaction in equations 34 and 35 is in the direction for which $\Delta G < 0$. The Gibbs free energy is a very useful parameter in the analysis of binding events. It is dependent on the changes in enthalpy (ΔH) and entropy (ΔS) of the ligands upon binding. The enthalpy term (ΔH) reflects internal energy changes associated with the intermolecular bonding interactions. The entropy term (ΔS) provides a description of the degrees of freedom of the components involved in the binding reaction. Since, whether or not the

interaction takes place depends on ΔG and not on the other terms separately, large negative (unfavourable) entropy changes can occur on binding if there are compensating enthalpy differences. These may reflect the formation of favourable atomic interactions, like salt bridges and hydrogen bonds.

Moreover, the complete set of thermodynamics equations describing a system, in a given temperature range, can be calculated from the functional dependence of enthalpy on temperature (T). Calorimetry measures heat exchange during temperature-induced changes in a system. When a change in the state of a system (e.g. protein unfolding, or ligand binding) results from a change in temperature, there is a corresponding change in enthalpy. Heat is absorbed ($\Delta H > 0$) if the change results from an increase in temperature or heat is released ($\Delta H < 0$) if the change results from a decrease in temperature. The quantity measured in a calorimetry experiment on a macromolecular solution represents the heat capacity at constant pressure **(C_p)**:

$$\Delta C_p = \frac{\partial \Delta H}{\partial T} \tag{36}$$

The calorimeter

The ITC microcalorimeter is composed of a reaction cell, which has a volume close to 1.4 ml and contains one of the reactants (sample), and an identical reference cell filled with deionized water. The other reactant is added, by injection in small volumes (close to 10 μl), and stirred in the reaction cell by means of the injector syringe. A feedback control system supplies thermal power continuously to maintain the same temperature in both reference and reaction cells. Any event taking place in the reaction cell, usually accompanied by heat, will change the temperature in that cell and the feedback control system will modulate the power supplied in order to minimize such temperature imbalance. A sequence of injections is programmed and the ligand solution is injected periodically into the reaction cell. The amount of thermal power (in mJs^{-1} or W) required to maintain a constant temperature difference between the reaction bath and a reference cell is measured by the calorimeter. The heat absorbed or released by the chemical reaction is determined from the integral of the power curve over the appropriate time.

The amount of heat $\boldsymbol{q_j}$ absorbed or released in the injection ***j*** is equal to:

$$q_j = v\Delta H\Delta\left[L_j\right] \tag{37}$$

where $\boldsymbol{v}$ is the volume of the reaction cell and $\Delta[L_j]$ is the change in concentration of bound ligand after the *j*th injection, so that *qj* is proportional to the amount of ligand, $v\ [L_j]$, that binds to the

protein. The constant of proportionality is the enthalpy change (ΔH) per mole of bound ligand. The value of the association constant (K_a) governs the equilibrium; the change in composition inside the reaction cell after each injection triggers the binding reaction and the rearrangement of populations leading to the formation of complex. The system will pass through different equilibrium states, differing in composition, as the sequence of injections proceeds. The heat associated with each injection is proportional to the increase in complex concentration and it is calculated integrating the area under the deflection of the signal measured (amount of heat per unit of time provided to maintain both cells, sample and reference, at the same temperature). At the end of the experiment, saturation of the macromolecule is reached. Applying non-linear regression using the appropriate model in the data analysis procedure, it is possible to estimate the association constant (K_a) the binding enthalpy (ΔH) and the stoichiometry (n) in a single experiment (Fig. 3.1).

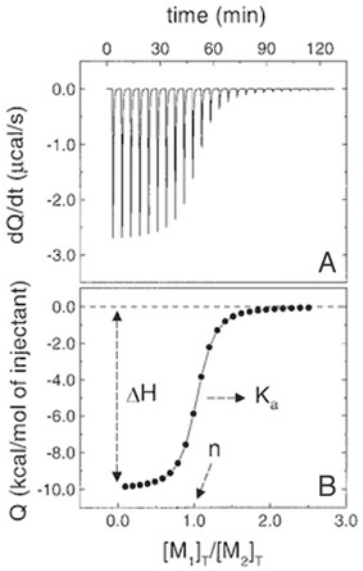


Figure 3.1: Characteristic peak sequence in the recorded signal (**A**). After saturating the macromolecule, the residual heat effects (the so-called "dilution peaks"), if any, are due to mechanical and dilution phenomena. After integration of area under each peak (and subtraction of the dilution heat effects and normalization per mol of injected ligand) the individual heats are plotted against the molar ratio (**B**) from which, through nonlinear regression, it is possible to estimate the thermodynamic parameters: n, K_a, and ΔH.

Depending on the number of sites and other aspects of the binding model, the mathematical expression relating affinity constants (K_a) to equation 37 may be quite complicated. By using the appropriate set of binding equations, the analysis can be extended to reactions with more than one binding site or involving equilibrium constants between many different states. Given the suitability of ITC to determine reaction stoichiometry (n), it is increasingly used in the analysis of systems that

involve multiple binding events, such as the formation of multiprotein complexes or the binding of multivalent ligands.

3.2 Reverse Titration of ITC

When macromolecules with multiple binding sites are titrated with ligands, the binding curve represents a description of the global energetics of the multisite system and may display multiple phases. The shape of the isotherm can differ significantly depending on whether the macromolecule or ligand is in the calorimetric cell. This is especially the case when multivalent systems display positive cooperativity and the intermediate state can be poorly populated. The isotherm is dominated by unsaturated and fully saturated states, with few singly-bound states, often giving a rather featureless binding curve. The injected reactant located in the syringe is referred to as "ligand". Usually ligand should be placed in the syringe, and the targeted protein should be placed in the cell. Sometimes reverse titrations (i.e., reversing the role of macromolecule and ligand) are conducted to check the stoichiometry or the suitability of the binding model [53] [55] as well as to fully resolve positive cooperativity [56].

Systems that involve multiple binding events that occur at two or more interacting sites often demonstrate cooperativity, a mechanism of transferring information. Cooperativity is an effective mechanism of regulation that provides a medium to transfer information, amplify or nullify a response to changes in local concentration and regulate the overall reaction pathway. Cooperative effects are either positive (synergistic) or negative (interfering), depending on whether the binding of the first ligand increases or decreases the affinity for subsequent ligands. Noncooperative (additive) binding does not affect the affinity for subsequent ligands and the interaction sites can be considered independent [56].

4. Urease system

4. Urease system

4.1. Why studying urease?

Urease is an enzyme involved in the quick hydrolysis of urea, one of the most widespread organic molecules in the biosphere. It is a nickel dependent enzyme fundamental in the biogeochemical nitrogen cycle.

Urease activity can increase the rate of urea hydrolysis of 10^{15} times [57] playing a key role in nitrogen assimilation by bacteria, plants, algae, fungi and invertebrates [58]. Nevertheless this hydrolysis results in a pH increase with negative effects in human health and agriculture. Urease is also a sentimental favourite among biochemists because it was the first enzyme ever to be crystallized [59], and the first shown to contain nickel in its catalytic site [60]. In 1926 James B. Sumner crystallized urease from jack bean (*Canavalia ensiformis*) and his work was fundamental to prove that enzymes are indeed proteins. The importance of Sumner's pioneer work was recognized two decades later and he was awarded the 1946 Nobel Prize in Chemistry for his discovery that "proteins can be crystallized" [61].

Directly or indirectly, urease plays an important role in the metabolism of all living organisms. Although it seems to be lacking only in the small *taxon* of vertebrates, even in some of them urease is fundamental for their metabolism. In particular in the rumen of herbivores bacterial urease facilitates recycling of nitrogenous wastes [62]. Some aspects about the ecological and pathological significance of ureases are highlighted below.

Urease activity in the environment

Ureases enzymatic activity in the soil plays a significant role for ecosystems and agriculture. Its primary role is to allow the organism to use external and internally generated urea as a nitrogen source [63]. Moreover in agricultural field, an effective hydrolysis of urea-based fertilizers requires controlled ureolysis to enhance efficiency and minimize ammonia toxicity or alkaline induced plant damage [62]. In fact the efficiency of soil nitrogen fertilization with urea decreases due to ammonia volatilization and root damage by soil pH increase [64].

Urease environmental activity is due to both soil microorganism and the so called "soil ureases" [64] coming from plant residues and dead bacterial cells. The last type of ureases are extra-cellular enzymes immobilized in clays and humus-derived compounds of the soil [65]. The high physicochemical stability of these ureases, due to the interaction with soil compounds, explains why the proper use of urea represents the most employed fertilizer worldwide. Furthermore, ureases may lead to the formation of calcium carbonates in soils and natural waters. The role of soil ureases in this process appear to be triple: increasing in alkalinity which favours $CaCO_3$ precipitation, increasing in the concentration of inorganic carbon dissolved (HCO_3^-) in the medium, and finally, the protein surface serves as a nucleation site for the crystals [66].

Urease in pathogenesis

Urease is a known virulence factor for a number of bacterial and fungal pathogens. The enzyme activity of ureases allows microorganisms to use urea as a nitrogen source. Urea, which is a form of nitrogen excretion for most terrestrial animals, is abundant and easily available to ureolytic bacteria. Most bacteria are urease producers, either constitutively or after induction upon specific conditions [67]. Among the urease-positive bacteria, those that cause diseases in humans or animals are the most studied and pathogenesis due to these microorganisms frequently involves urea hydrolysis that leads to increased pH and ammonia toxicity. In humans, the urinary and the digestive systems are the most common sites of infection by these bacteria. Increasing of the pH in the urinary system may cause a number of pathologies in humans, such as precipitation of ureate salts leading to the formation of infection-induced urinary and kidney stones, catheter occlusion, pyelonephritis and also hepatic encephalopathy. Important urinary pathogens include the bacteria *Proteus mirabilis* and *Ureoplasma ureoliticus* [67], [68].

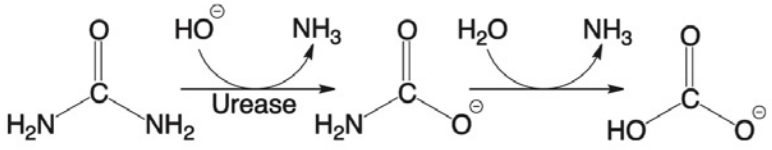
Another well known example of urease associated pathogenesis is colonization of the gastric mucosa by the highly ureolytic spirochete *Helicobacter pylori*, a causative agent of gastritis, peptic ulceration and stomach cancer [69]. The pH in the stomach is acidic (around 2-3) with respect to other organs and the colonization of the widespread microorganism depends on the enzymatic activity of its urease. The latter, breaking down urea to ammonia and carbamate, with the releasing of water, creates a favourable, less acidic, microenvironment that enables the bacterium to colonize the gastric mucosa [70][71].

Mycoplasma tuberculosis, the bacterium that causes tuberculosis, survives into phagolysosomes of infected cells thanks to its urease that, alkalinizing the acidic medium, prevents the activity of cathepsins that could destroy the microorganism [72].

Because of its role in the bacterium acid resistance, urease and all the chaperones involved in the enzyme maturation are possible targets against bacterial pathogenesis.

4.2. Urease structure

Urease (urea amidohydrolase: E.C. 3.5.1.5) is a nickel-dependent enzyme that catalyzes the hydrolysis of urea into two molecules of ammonia and one of carbon dioxide. Urea is one of the most abundant molecule in the biosphere, i.e. it is produced in large amounts by many vertebrates as the catabolic product of nitrogen-containing compounds; just consider that each human being produces ca. 10 kg of urea per year [73]. Moreover urea is very stable, with a half-time of 3.6 years at 38°C in solution [74]. Considering the long half-life for the nonenzymatic urea hydrolysis, never observed experimentally but estimated as being ca. 520 years, urease is the most efficient hydrolase known, catalyzing the hydrolysis in microseconds, enhancing of 10^{15} times the rate of the reaction [57]. The efficiency of this enzyme is ascribed to the presence of two Ni(II) in the active site of the enzyme. The reaction proceeds through a mechanism in which at first urea is decomposed to ammonia and carbamate, which then spontaneously reacts at physiological pH, to give a second molecule of ammonia and bicarbonate (Scheme 1).



Scheme 1

Ureases from different organism share 50-60% identity at the amino acid sequence level revealing common ancestral derivation. Regarding the quaternary structure, plant and fungal ureases are trimer or hexamer of single type of ~90 kDa subunit with about 840 amino acids. On the other

hand, bacterial ureases are multimers of two or three polypeptide chains that correspond to "fragments" of the single chain of the plant/fungal urease [58], [75] (Fig. 4.1).

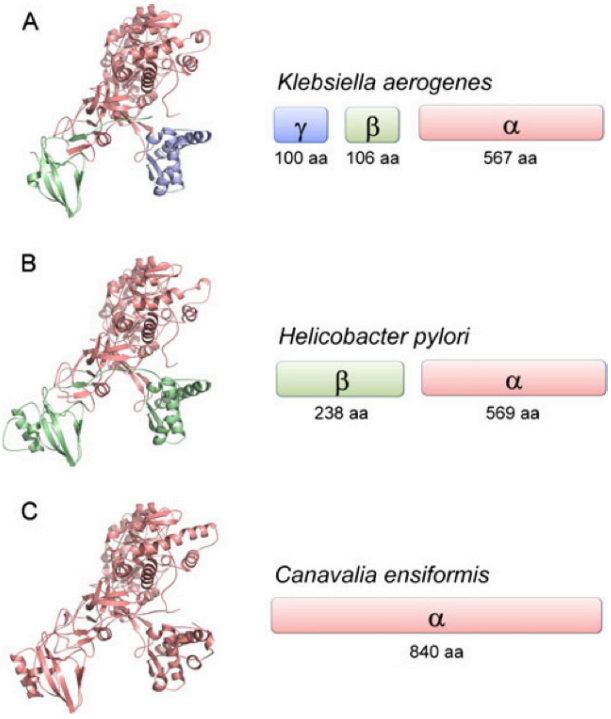


Figure 4.1: Three-dimentional and schematic representations of the subunit organization of ureases. (**A**) Typical microbial urease composed by three chains (PDB code 1FWJ). (**B**) *Helicobacter pylori* urease composed by two chains (PDB code 1E9Z). (**C**) Typical eukaryotic urease composed by a single chain (PDB code 3LA4) [75].

The quaternary structures of bacterial ureases of *Klebsiella aerogenes* [63], *Sporosarcina pasteurii* (formerly known as *Bacillus pasteurii*) [76] and *Helicobacter pylori* [77], compared with that of the urease from the seeds of the plant *Canavalia ensiformis* (jack bean) [78] revealed a marked similarity in three-dimensional shape between the single chain of the plant urease and the multimeric bacterial ones. Together, three-dimensional structure similarity and high homology of amino acid sequences suggest that all ureases are evolutionary variants of one ancestral enzyme [58].

The quaternary structure of the majority of bacterial ureases, e.g. *Klebsiella aerogenes* and *Sporosarcina pasteurii*, is composed of a trimer of trimers $(\alpha\beta\gamma)^3$, with each α subunit containing an independent active site [79] (Fig. 4.2).

In *Helicobacter pylori*, four trimers of dimers $(\alpha\beta)^3$ (with the β subunit resulting from the fusion of the corresponding β and γ subunits found in *Sporosarcina pasteurii* and *Klebsiella aerogenes*) form

a tetrahedral structure $((\alpha\beta)^3)^4$ with a total of twelve active sites per *Helicobacter pylori* functional molecule. The plant enzyme consists of a dimer of homotrimers $(\alpha^3)^2$, evolved from the fusion of the corresponding bacterial αβγ trimer, and contains six active sites.

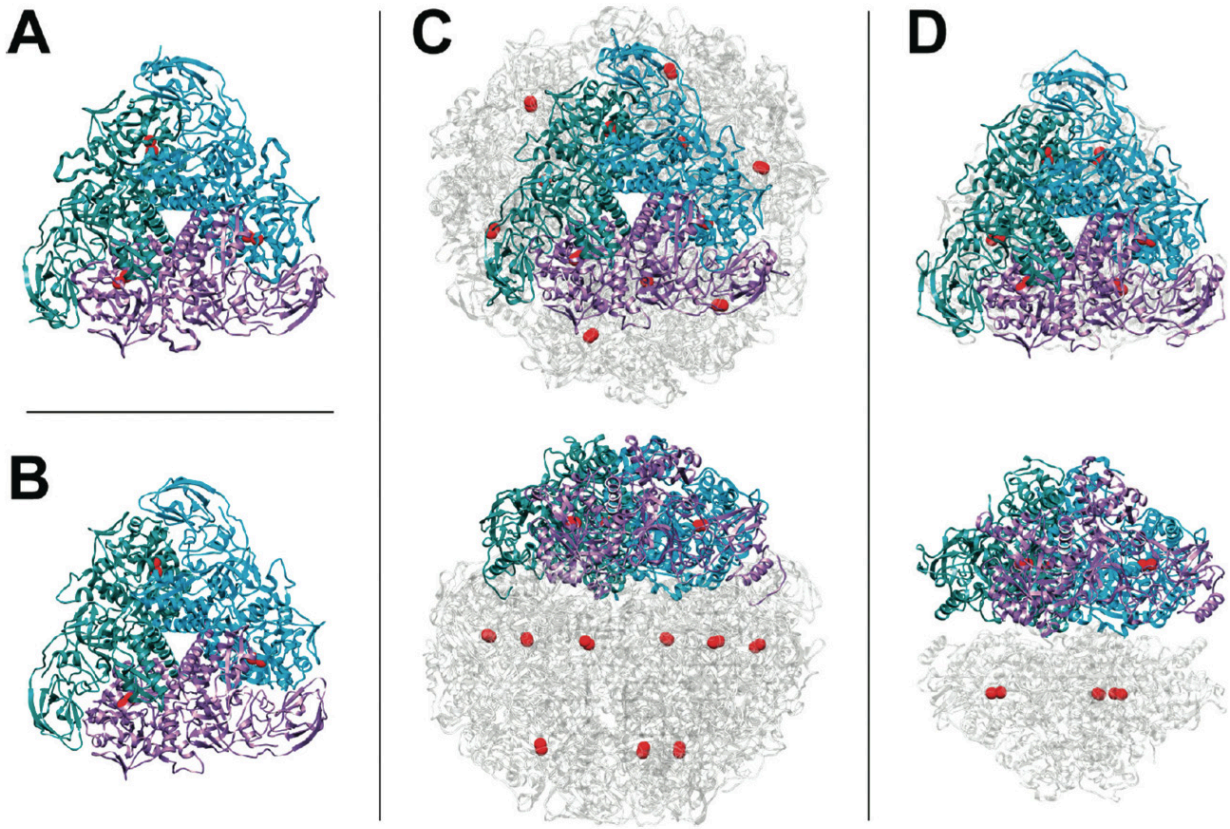


Figure 4.2: Ribbon diagram of urease from (**A**) *Sporosarcina pasteurii* (PDB code: 2UBP), (**B**) *Klebsiella aerogenes* (PDB code: 1EJZ), (**C**) *Helicobacter pylori* (PDB code: 1E9Z), and (**D**) jack bean (PDB code: 3LA4). Ribbon colors evidence the chains composing the trimer of oligomers (monomers in the case of jack bean) constituting the minimal quaternary structure of urease. Ni(II) are reported as red spheres. The bottom panels of (**C**) and (**D**) are rotated by 90° around the horizontal axis *vs* the top panels [79].

4.3. Urease catalytic mechanism

The resolution of the crystal structures of *Klebsiella aerogenes* [63] and *Sporosarcina pasteurii* (Benini et al., 1999) urease, has given insights in the catalytic mechanism of the urease active site. Each α subunit of bacterial urease contain an independent binuclear nickel center, in which the Ni(II) atoms, separated by 3.5 Å (*Klebsiella aerogenes*) or 3.7 Å (*Sporosarcina pasteurii*), are bridged by the oxygen atoms of a post-translationally carbamylated lysine residue and bound to two histidines each (Fig. 4.3). One Ni(II) ($Ni_{(2)}$ in Fig. 4.3) is additionally bound to an aspartate

carboxylate oxygen. The coordination geometry of the Ni(II) ions is completed by a water molecule terminally bound to each metal ion and by a nickel-bridging hydroxide ion ($HO_{(B)}$ in Fig. 4.3). This ligand arrangement yields one penta-coordinated Ni(II) ($Ni_{(1)}$ in Fig. 4.3) with a distorted square-pyramidal geometry, and one Ni(II) hexa-coordinated with a distorted octahedral geometry ($Ni_{(2)}$ in Fig. 4.3). An additional water molecule ($H_2O_{(3)}$ in Fig. 4.3) is present in the active site, interacting with the others through hydrogen bonds, and completing a tetrahedral cluster of four water/hydroxide molecules in the close proximity of the Ni(II) atoms. Noteworthy, the residues that coordinates nickel ions and the substrate are strictly conserved among the ureases.

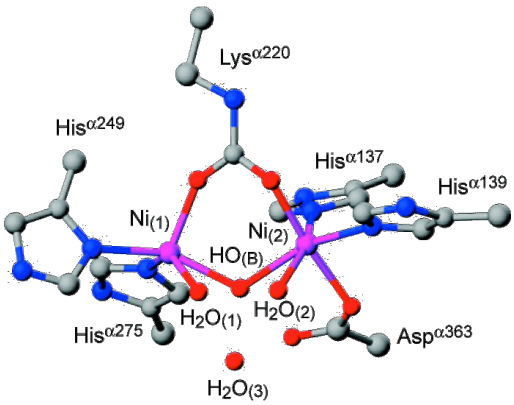


Figure 4.3: Coordination geometry of Ni(II) in native urease active site from *Sporosarcina pasteurii*. Color scheme: nickel, *purple*; carbon, *gray*; nitrogen, *blue*; oxygen, *red*.

A mobile flap modulates the entrance of the substrate to the active site cavity, with a catalytically essential histidine moving by about 5 Å between the open (Fig. 4.4A) and closed (Fig. 4.4B) conformations [79], [80].

The structure of urease active site was determined in the native hydrated form and in the presence of several inhibitors [74], [76], [80], [81], in order to elucidate both the substrate binding mode and the catalytic mechanism. The high efficiency of the urea hydrolysis catalyzed by urease makes indeed impossible to obtain structural data in the presence of this substrate.

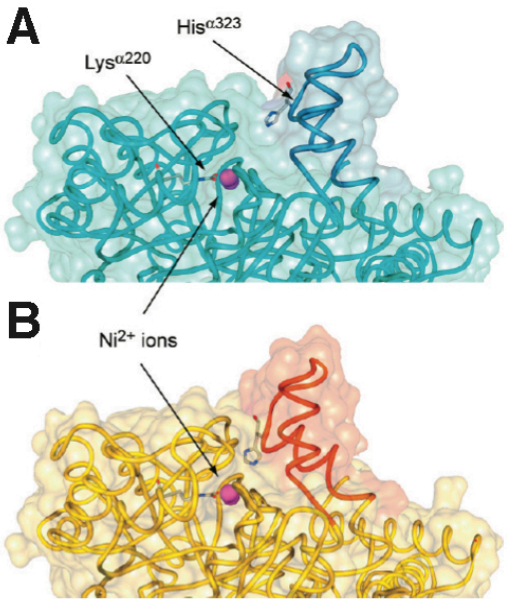
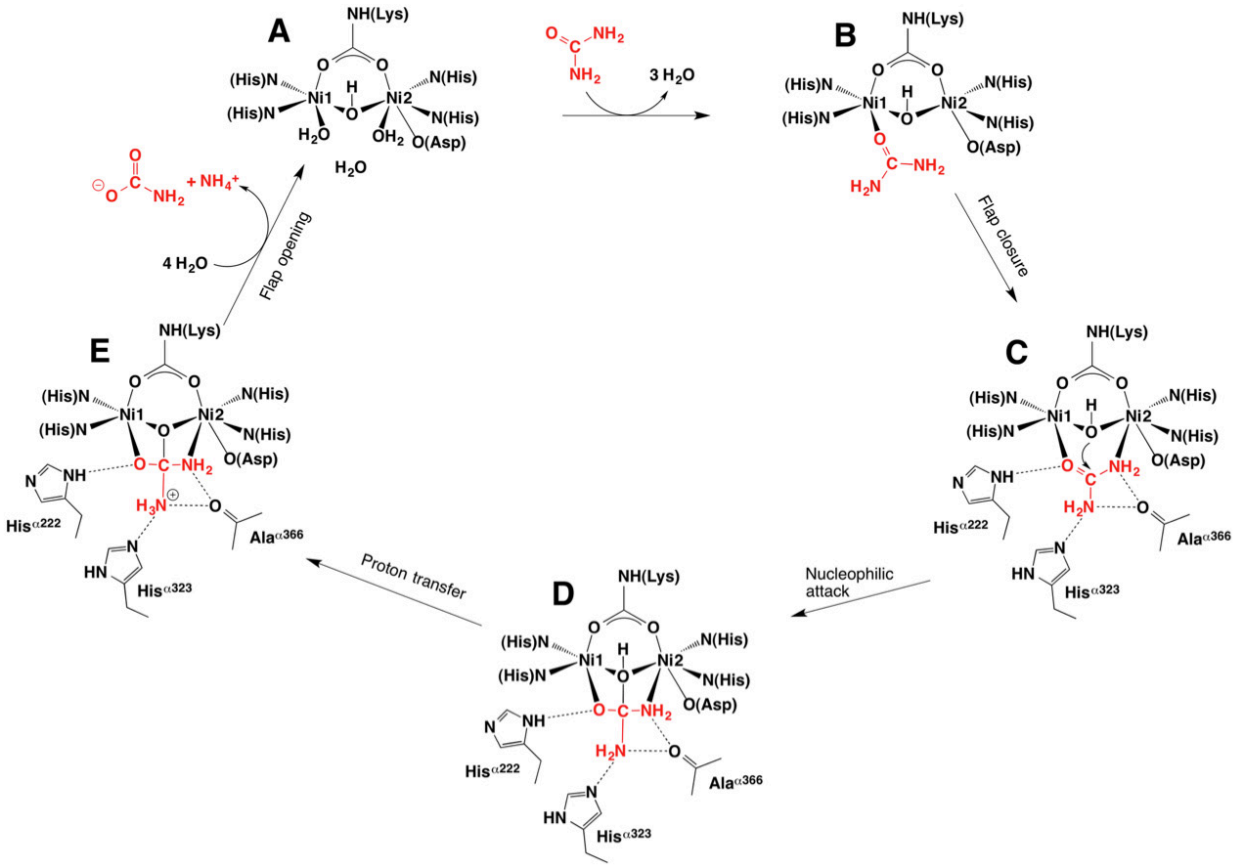


Figure 4.4: Open (**A**) and closed (**B**) conformation of the flexible flap in native urease active site from *Sporosarcina pasteurii*. Ni(II) are shown as purple spheres. Adapted from Zambelli *et al.*, 2011 [79].

The catalytic mechanism of urease (Scheme 2) assumes two different roles for the nickel ions in the active site, and the key step of the process is the nucleophilic attack of the nickel-bridging hydroxide on an urea molecule that is bound to the bimetallic nickel cluster by an oxygen atom and a nitrogen atom [80]. In this mechanism initially the enzyme is in the hydrated state and the flap is in the open conformation (A), then the carbonyl oxygen atom of urea binds the more electrophilic and coordinatively unsaturated $Ni_{(1)}$ with the concomitant displacement of water molecules from the active site due to steric hindrance (B). The correct orientation of the substrate is determined by the asymmetry of the residues binding the nickel ions, the nature and the position of these amino acids establish a specific H-bonding network, activating the substrate molecule toward the nucleophilic attack, and positioning the carbon atom of urea near the nucleophile [82]. The subsequent flap closure facilitates urea coordination to the second Ni(II) ($Ni_{(2)}$) via one of its amine group (C), and then the bridging hydroxide ($HO_{(B)}$) attacks the carbon atom of urea, forming a tetrahedral intermediate (D). The transfer of a proton to the "free" distal amino group promotes attack of water on the urea carbonyl group; the presence of the conserved histidine residue (Hα323 for *Sp* urease) stabilize the proton transfer [82]. The breakage of the distal C-N bond leads to formation of ammonia and carbamate (E), which spontaneously hydrolyzes into a second molecule of ammonia and hydrogen carbonate. The flap opening could facilitate the release of products and the rehydration of the active site enabling the return of urease in the native state. As ammonia and carbamate dissociate, water molecules bind to reconstitute the catalytic center (A) [73], [79].

The structure-based mechanism described could help to explain the requirement for Ni(II) ions instead of others metal ions as i.e. the more common and less toxic Zn(II). Moreover many studies have demonstrated that the precise position and mobility of the metal ligands, as well as those of protein residues in the active site not involved in metal binding, are important in achieving optimal urease activity.



Scheme 2: Urease catalytic mechanism of the enzymatic hydrolysis of urea. The residue-numbering scheme is relative to *Sporosarcina pasteurii* urease [73].

4.4. Urease maturation

Urease protein is assembled *in vivo* as an inactive apo-enzyme and its complete maturation, with the building of nichel-containing active site, is a multi-step process requiring a precise sequence of events and the simultaneous occurrence of different conditions [83], including the presence of metal ions. In particular the whole process, involving events such as CO_2 uptake for post-translational lysine carbamylation, hydrolysis of GTP, and Ni(II) delivery into its active site, require the involvement of specific accessory proteins [84] [79] [85].

The discovery of accessory proteins required for nickel incorporation in urease active site was reported after the sequencing of urease operon of *Klebsiella aerogenes* [84]. Homologous auxiliary genes in regions flanking urease structural genes were subsequently identified in many other ureolytic bacteria, demonstrating the presence of a unique genetic organization that possesses the characteristics of an operon. A large part of bacterial urease operons contains three structural genes: *ureA*, *ureB*, *ureC*, codifying respectively for the apo-urease subunits γ, β, α, and four accessory genes codifying urease chaperones responsible for the assembly of the catalytic site: *ureE*, *ureF*, *ureG*, *ureD*, although the gene order often changes, and *ureD* is renamed *ureH* in *Helicobacter pylori* [86] (Fig. 4.5).

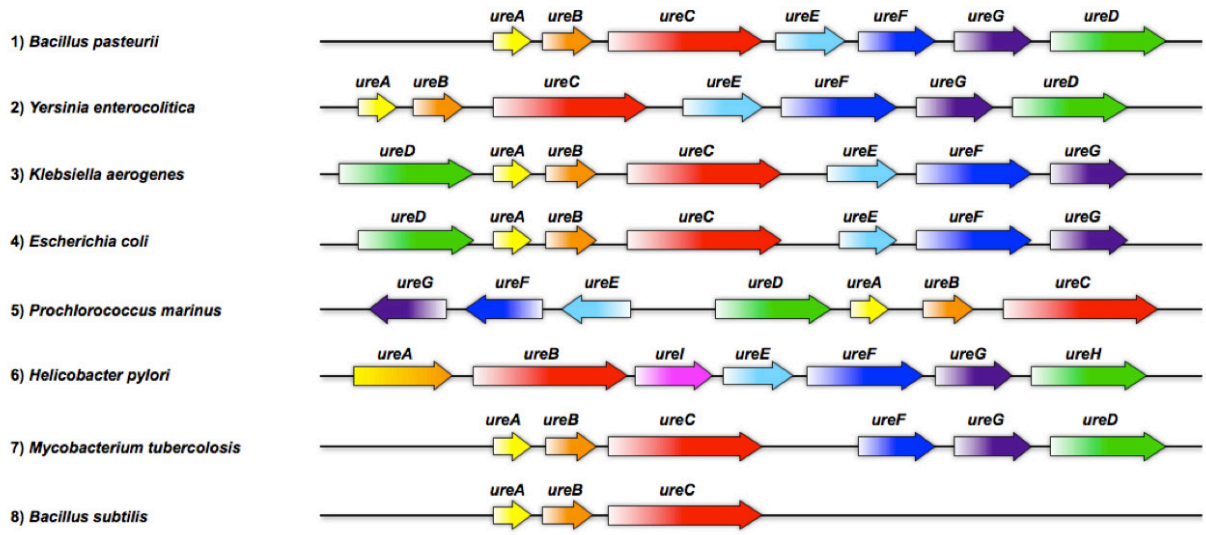


Figure 4.5. Schematic representation of urease operons of *Sporosarcina pasteurii* (formerly known as *Bacillus pasteurii* - 1) compared to some other fully sequenced genomes. Arrows indicate the direction of the gene; arrow lengths are proportional to protein lengths [25].

The generally accepted process [79], [85] (Fig. 4.6) involves the sequential binding of UreD, UreF, and UreG, or alternatively the binding of a preformed complex UreDFG in which the three proteins act together as a unique molecular chaperone, driving a protein conformational change and lysine carbamylation. The obtained pre-activation complex can bind Ni^{2+} ions deriving from UreE: the metallo-chaperone of the system, trough a route driven by GTP hydrolysis.

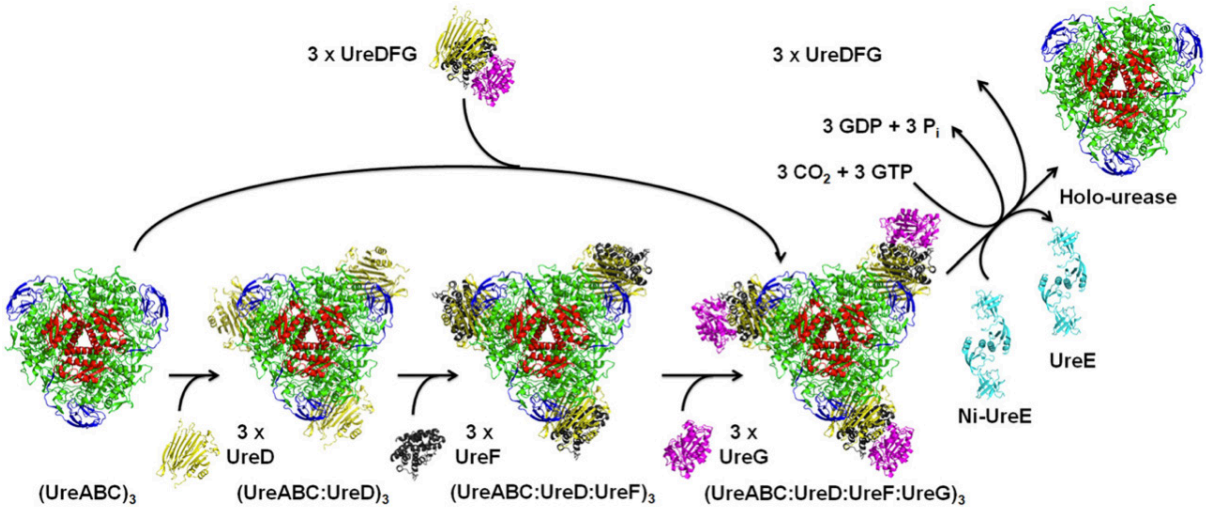


Figure 4.6: Model of *Klebsiella aerogenes* urease activation. The trimer-of-trimers urease apoprotein (UreA, *red*; UreB, *blue*; UreC, *green*) either sequentially binds UreD (*yellow*), UreF (*gray*), and UreG (*magenta*) or binds the UreDFG complex (only one protomer of each protein is shown, but the isolated complex contains two protomers of each). Formation of the active enzyme requires CO_2 to carbamylate L217 at the native active site, GTP binding to and hydrolysis by UreG, and nickel delivery by dimeric UreE (*cyan*). Taken from Farruggia *et al.*, 2013 [85].

4.5. UreD-F-G complex

UreD/UreH (in *Helicobacter pylori*) is apparently the first protein that binds apo-urease it is quite insoluble when isolated in native condition and this is the reason of the lack of information on this protein. What it is known is that UreD/UreH behaves as a scaffold for recruitment of other accessory proteins and it is also a facilitator of nickel insertion into the active site [87]. The problems of insolubility of this accessory protein were circumvented by different approaches. For *Klebsiella aerogenes* UreD, a maltose-binding protein $(MBP)^2$ fusion variant of UreD was purified and shown to be soluble and to bind Ni^{2+} and Zn^{2+} ions [88]. However, the physiological relevance of this chimerical product is uncertain, since UreD maintains its insolubility when separated from MBP. In the case of *Helicobacter pylori* UreH, solubilization is achieved by coexpression of *ureH* with *ureF*, which provides a UreH-UreF complex [87] (Fig. 4.7). The crystal structure analysis of UreF-UreH complex revealed a conformational changes of UreF upon complex formation.

UreF is proposed to be the second accessory protein that binds the apo-urease-UreD complex by interacting directly with UreD, as shown in chemical cross-linking and small-angle X-ray scattering experiments on *Klebsiella aerogenes* UreF (*Ka*UreF), and also in two-hybrid studies, light scattering experiments, pull-down assays and crystallography on the protein from *Helicobacter*

pylori (*Hp*UreF) [89]. The interaction of apourease with UreD and UreF appears to induce a conformational rearrangement of the enzyme urease, which promotes access to the active site of Ni^{2+} ions and CO_2 needed for the carbamylation of the active-site lysine [90]. Functional studies suggested that this event prevents Ni^{2+} ions from unproductively binding to the noncarbamylated active site [91]. On the basis of a protein structural model of *Sporosarcina pasteurii* UreF, obtained by fold recognition analysis [92], and on the observation that in the absence of *Ka*UreF, *Ka*UreG - the GTPase of the system - is unable to join the apo-urease-UreD complex [93], UreF was proposed to have a role in modulating the GTPase activity of UreG through a direct protein-protein interaction [92]. Moreover Pull-down assays on *Helicobacter pylori* proteins indicated that UreF-UreG interaction requires a preformed complex between UreF and UreD, suggesting that a conformational change on UreF, induced by UreD, is necessary in order to drive the formation of a UreF-UreG complex [87]. For years, structural studies on UreF were prevented by its insolubility and the crystal structure obtained for *Hp*UreF expressed and purified alone is anyhow lacking of the last 21 residues at the C-terminal region [89]. On the other hand, when the same protein is co-expressed and co-purified with *Hp*UreD, the formation of a complex is observed and in this case the C-termini of UreF is directly involved at the interface with *Hp*UreD, as revealed by the crystal structure of the complex (Fig. 4.7) [87].

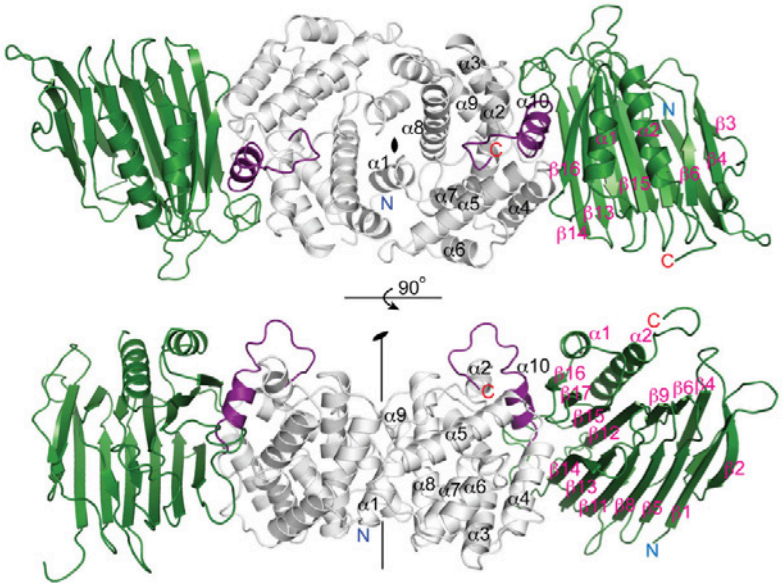


Figure 4.7: Two views of the crystal structure of *Helicobacter pylori* UreF-UreH complex (PDB code: 3SF5) differing by a 90° rotation around the horizontal axis. In each asymmetric unit, there are two copies of UreF (*white*) and UreH (*green*) related by a noncrystallographic 2-fold symmetry. Secondary structure elements of UreF (*black*), UreH (*pink*), N (*blue*), and C (*red*) termini are labeled. Residues 234 - 254 of UreF (*purple*), absent in the crystal structure of *Hp*UreF, is observed in this complex structure forming helix α10 (residue 236 - 243) and a structured loop (residue 244 - 254). Adapted from Fong *et al.*, 2011 [87].

The formation of complex thus protects the C-terminal region of UreF from degradation and indicates that this portion is essential for stabilization of the UreF-UreD interaction. After a long and detailed analysis the role of UreF appears to be that of modulating the GTPase activity of UreG so as to promote efficient coupling of GTP hydrolysis and metallocenter biosynthesis, thereby enhancing the accuracy of urease activation [85].

UreG is the chaperone responsible for coupling GTP hydrolysis to the process of urease activation. The GTPase function is encoded in the protein sequence, which features a conserved N-terminal P-loop motif. Among the four urease accessory proteins, UreG is the most conserved [94]–[96]. UreG is also proposed to catalyze, in the presence of CO_2, the formation of carboxyphosphate, an excellent carbamylation agent for the metal-binding lysine in the urease active site. The curves that correlate the urease activation to different bicarbonate concentrations indicate a higher rate and level of enzymatic activation in the presence of UreDFG-apourease complex, proving the need of a physiological bicarbonate concentration (100 μM) for urease activation only in the presence of UreG [93], [97].

UreG proteins are soluble and have been isolated and characterized from *Klebsiella aerogenes* (*Ka*UreG)[94], *Sporosarcina pasteurii* (*Sp*UreG) [95], *Mycobacterium tuberculosis* [96], and *Helicobacter pylori* (*Hp*UreG)[98]. The crystal structure of the apo-UreG was just recently obtained, although only complexed with UreH and UreF [97] (Fig. 4.8). The difficulties encountered for many years to crystallize the free form of this protein are due to its characteristic of being intrinsically disordered. UreG exists in solution as structural ensembles that present some secondary structure but they lack a well-defined tertiary structure, in a sort of ensemble of inter-converting conformations [99]. The enzyme behaviour, as a protein presenting a molten globule-like or pre-molten globule-like fold, inscribes UreG into the class of intrinsically disordered proteins [100] [99]. Enzymes are usually considered as proteins that interact with substrates using a lock-and-key or induced-fit mechanism and therefore possessing a well-determined tertiary structure. UreG is the first, and so far the only, known natural intrinsically disordered enzyme [99]. The presence of various degrees of protein flexibility in different members of this protein family suggests that a ‘disorder to order’ transition is a possible mechanism for UreG functioning *in vivo*, and that the functional activation of UreG occurs when the protein interacts with other protein partners, perhaps other urease chaperones, that trigger its active ordered form. Generally, GTP hydrolysis is regulated at different levels by the cells to avoid unnecessary consumption of GTP. Hence the native disorder could be another mechanism that cells use to regulate enzymatic activity,

allowing this protein, partially or totally inactive in the isolated state, to interact and be regulated by different protein partners [79]. In this context it is possible to conclude that the interaction with UreD, UreF, and apourease must be essential for UreG folding.

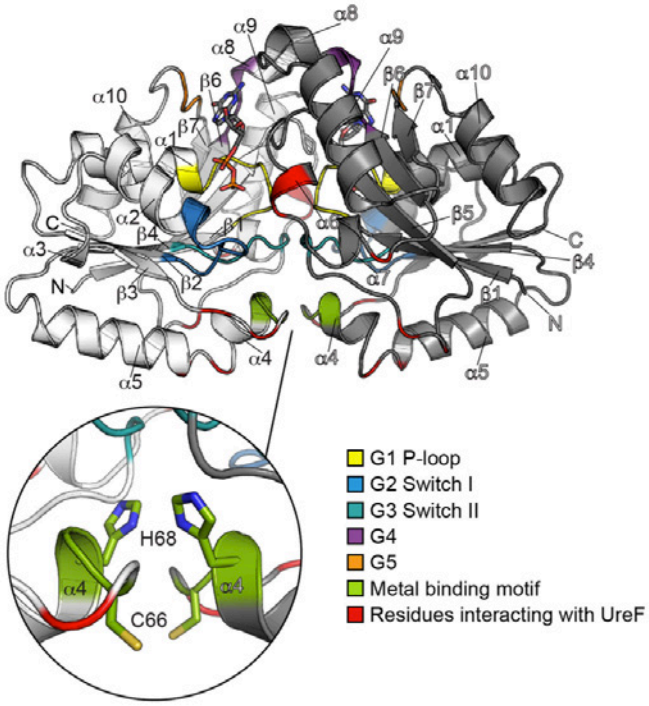


Figure 4.8: The Structure of *Hp*UreG. Two UreG protomers are colored in *white* and *grey*, respectively, with secondary structural elements labeled. GTPase structural motifs (G1 - G5), metal binding motif, and residues involved in interaction with UreF are colored as indicated. The inset shows the details of the metal binding site located at the interface between two UreG protomers, with residues involved in nickel chelation shown in sticks. Adapted from Fong *et al.*, 2013 [97].

Metal ions can have a role in the modulation of UreG activity through control of the oligomerization state. In the case of *Ka*UreG, Zn(II) binding residues are essential for the ability of this *chaperone* to activate urease *in vivo* [101], suggesting that Zn(II) binding plays a physiological role for UreG in the urease maturation pathway, driving the assembly of a functional protein complex that includes the dimeric UreG. *Sporosarcina pasteuri* UreG was demonstrated to bind two Zn(II) for dimer and up to four metal ions per dimer in the case of Ni(II), whereas the affinity is 10-fold lower for these ions. In this case Zn(II) would assume a structural role, stabilizing a certain protein conformation upon binding, rather than a role in the catalytic activity of the protein (Zambelli et al., 2005). Zn(II)-driven protein dimerization occurs *in vitro* for *Hp*UreG, with one Zn(II) binding at the protein dimerization interface using a conserved cysteine and histidine from each monomer [98]. Ni(II) binding, occurring with an affinity one order of magnitude lower as

compared to Zn(II) does not induce oligomeric changes, indicating that UreG can discriminate between different metal cofactors in the intracellular metal ion pool [98].

The UreD-F-G complex

The crystal structure of the apo UreG^2-UreF^2-UreD^2 complex from *Helicobacter pylori* [97] presents two protomers of each peptide, with the two UreG molecules in contact as expected for a protein able to dimerize (Fig. 4.9). The *Hp*UreF region observed to interact with the *Hp*UreG dimer in the crystal structure is in good agreement with that proposed on the basis of mutagenesis studies performed on *Ka*UreG [102]. On the other hand, the *Hp*UreG region that interacts with *Hp*UreF in the solid state was in fact proposed to interact with *Hp*UreE on the basis of computational studies based on calorimetric and NMR experimental data obtained in solution [103]. In the crystal structure, the conserved metal binding cysteine, histidine, proline (CHP) motif, which is observed to interact with the metal binding site of *Hp*UreE in solution, is instead found in close contact with the dimerization interface of *Hp*UreF. This difference may be reconciled by suggesting a change of partner for UreG, which could interact with UreE as calculated in the theoretical complex in a stage of the process, while forming a complex with UreF as observed in the crystal structure in another phase of the metabolic pathway of urease assembly. It should be mention that the crystal structure of the complex does not include UreE.

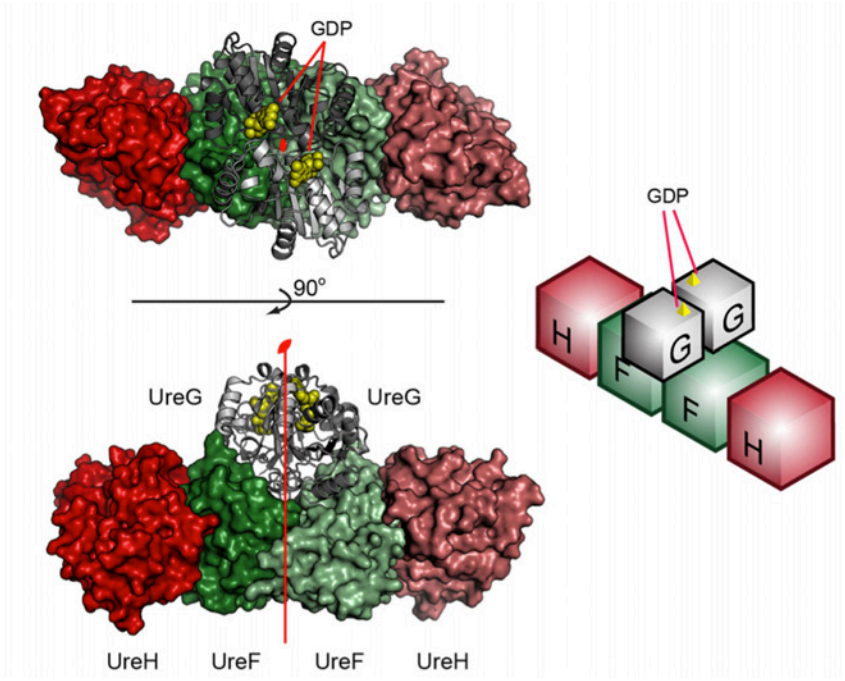


Figure 4.9: Structure of the UreG/UreF/UreH complex is shown on the left in two different orientations. UreF, UreH, and UreG are colored in shades of *green*, *red*, and *grey*, respectively. GDP ligands are shown as yellow space filling model. A schematic diagram of the relative positions of UreF, UreH, and UreG is shown on the right. The complex is a dimer of heterotrimers, with each pair of protomers related by a 2-fold symmetry. Adapted from Fong *et al.*, 2013. [97]

The calculation of a model for *Hp*UreDFGE supercomplex provides a structural rationale for the interpretation of the available experimental data on the interactions leading to the incorporation of the essential Ni(II) atoms in the active site of urease [104]. The main difference between the calculated model UreDFGE complex and the crystal structure of the UreDFG complex appears to be on the orientation of the UreG protein dimer with respect to the UreF interacting protein. The possibility exists that either UreG, being an intrinsically disordered protein, undergoes the known moonlighting behaviour [9], changing its partner as needed along a metabolic pathway, or that the crystal structure reflects one of the several possible docked complexes existing in solution, the one that actually crystallizes. In both cases, it is possible that the crystallographically determined structure represents the nonfunctional form of UreG, which needs to reorient itself on the surface of the UreDF complex and interact with UreE to accomplish its role. Alternatively, the calculated model reported here could be interpreted as one of the several possible complexes formed between these proteins. Only the experimental structure of the functional HpUreDFGE complex could solve this issue.

4.6. UreE

UreE is the best functionally and structurally characterized of the four urease accessory proteins. Thanks to its metal-binding capability, it was always proposed to function as a metallo-chaperone, in charge of delivering nickel ions to the activation complex to facilitate maturation of urease active site [105]–[108]. Structural information on UreE proteins from various bacteria has been derived from numerous crystallographic studies: UreE from *Sporosarcina pasteurii* [109], [110] (Fig. 4.10), a truncated version of UreE from *Klebsiella aerogenes* (H144**Ka*UreE) [111], and UreE from *Helicobacter pylori* (*Hp*UreE) [112], [113] display a similar fold made by a symmetric homodimer, with each protomer composed of N- and C-terminal domains connected by flexible linkers. Comparing the crystal structures in detail is possible to appreciate the strikingly large similarity of the fold despite the low sequence homology [114]. This structural comparison clarified the features of the fold that could be important for the functional role in Ni(II) trafficking.

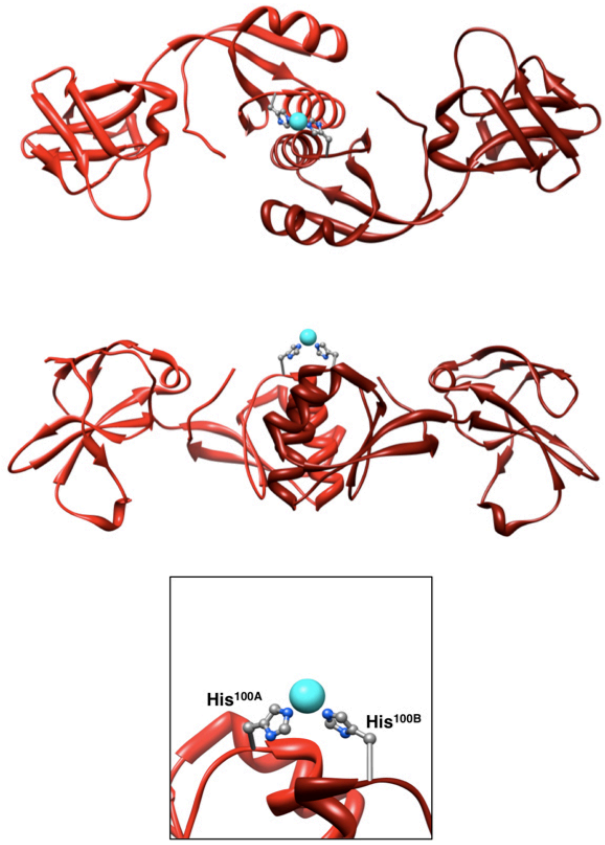


Figure 4.10: Ribbon representations of dimeric *Sp*UreE bound to Zn(II) (PDB code: 1EAR). The proteins are shown with the metal binding site toward the viewer (top panels) and rotated by 90° around the horizontal axis (bottom panels). Details of the conserved metal binding sites are also shown (square insert). Conserved His^{100} are represented as ball and stick and colored according to the following atomic scheme: carbon, *grey*; nitrogen, *blue*; zinc, *cyan* [25].

N-terminal domains of UreE, residing at the periphery of the molecule, consist of two three-stranded β-sheets stacked nearly perpendicularly upon each other, with a short helical region between the two sheets. The C-terminal domains, involved in a head-to-head dimerization, feature a four-stranded anti-parallel β-sheet and α-helices organized in a ferredoxin-like βαββαβ fold. The metal-binding site, situated at the edge of the dimerization interface on the surface of the protein, involves two conserved histidines (His^{100} for *Sp*UreE, His^{96} for *Ka*UreE, and His^{102} for *Hp*UreE). The C-terminal region of UreE, proven difficult to observe by crystallography probably because of disorder, features a variable number of histidines depending on the biological source [114] (Fig. 4.11 B).

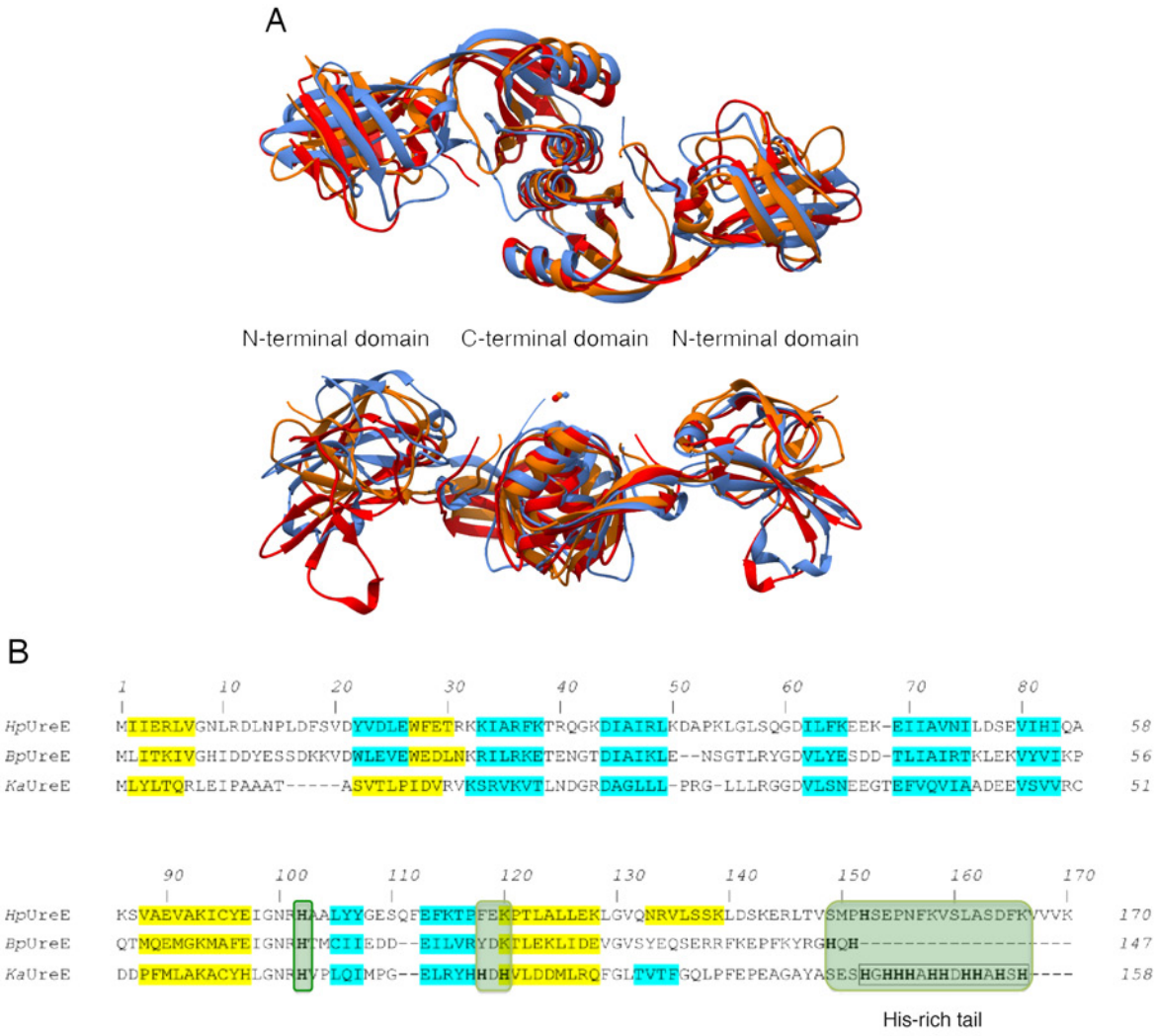
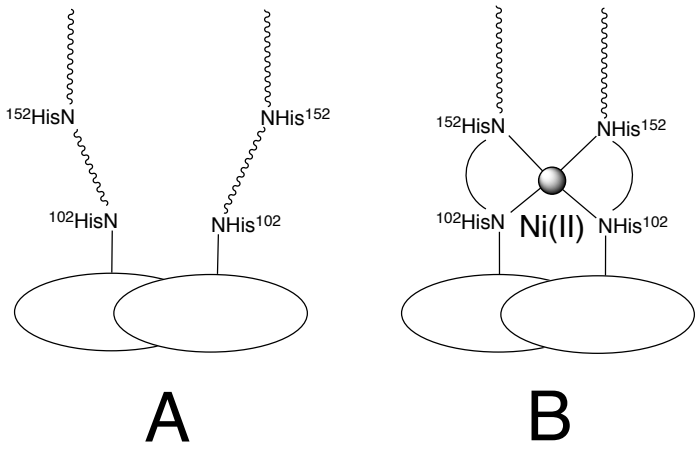


Figure 4.11: (**A**) Superimposition of the structures in ribbon representation of *Sp*UreE (in *red*, PDB code 1EAR), *Ka*UreE (in *orange*, PDB code 1GMW) and *Hp*UreE (in *light blue*, PDB code 3TJ8). The views in the top and bottom panel are rotated by 90° along the horizontal axis. Corresponding bound metal ions Zn(II), Cu(II) and Ni(II) are shown as spheres of arbitrary radius. The position of the N- and C-terminal domains is indicated. (**B**) Optimized multiple sequence alignment of UreE proteins from *Helicobacter pylori, Sporosarcina pasteurii* (formerly known as *Bacillus pasteurii* - *Bp*UreE) and *Klebsiella aerogenes*. The predicted secondary structure elements are highlighted in *yellow* (helices) and *cyan* (sheets); regions containing histidines are highlighted with *green* rectangles.

The most evident difference between the overall fold of the homologous proteins resides in the different relative orientations between the N-terminal and the C-terminal domains (Fig. 4.11 A). This movement apparently derives from a different conformation of the short linker connecting the two domains in each monomer, which results in a rotation around an ideal axis crossing the dimer from one N-terminal domain to the other [114]. This protein flexibility is likely related to induced-fit processes during formation of protein complexes involving the other urease chaperones.

Despite they share common structural features, UreE proteins display different metal binding capabilities. In particular, they exhibit a variable stoichiometry for nickel binding, that ranges from one Ni(II) per dimer in the case of *Hp*UreE [103], [113], to two Ni(II) per dimer for *Sp*UreE (Stola

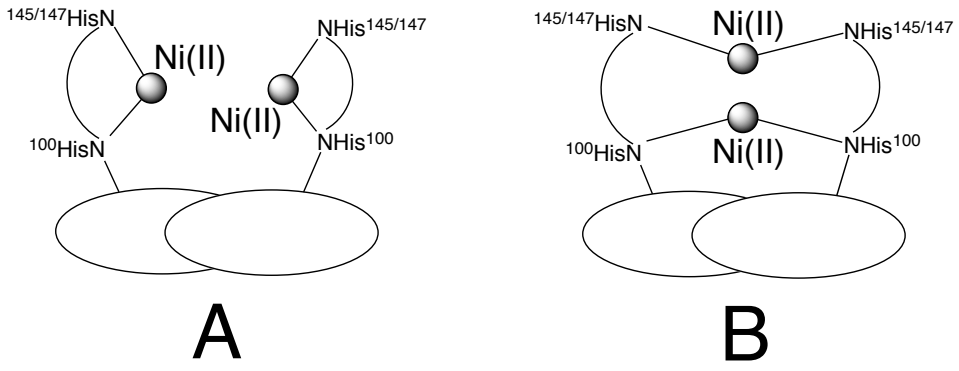
et al. 2006), up to six Ni(II) per dimer associated to *Ka*UreE [116]. This different behaviour can be correlated to the nature of the C-terminal regions of these different proteins: *Ka*UreE features a His-rich tail containing ten histidine residues among the last fifteen amino acids, *Sp*UreE displays two C-terminal histidine residues at the end of its sequence in a His-Gln-His motif, while *Hp*UreE contains a single histidine residue (His^{152}) [114] (Fig. 4.11). It is noteworthy the presence of, at least, one histidine residues that is invariably found in the C-terminal region of this class of proteins. These histidine residues in the C-terminal portion suggested a role for this region in determining the metal transport by UreE proteins, in terms of both selectivity and stoichiometry of metal binding [103], [114], [115]. Recently, the structural and functional details of these regions have been established, with the crystal structures of apo-*Hp*UreE (Scheme 3A), Ni(II)-*Hp*UreE and Zn(II)-*Hp*UreE (Scheme 3B) determined [113]. These structures, corroborated by X-ray absorption spectroscopy data, reveal that the His^{152} residue, located in the C-terminal region of *Hp*UreE, contributes to the coordination of the metal ion together with the fully conserved His^{102} residue on the protein surface (see Scheme 3) [113]. The C-terminus in *Hp*UreE proteins in the crystalline state is generally disordered (represented by the wiggle lines in Scheme 3), but becomes partially ordered upon metal ion binding and therefore can exist in an "open" (Scheme 3A) or "closed" (Scheme 3B) form. It has been suggested that a conformational change between these two forms is involved in the uptake and release of metal ions from UreE proteins [109], [113].



Scheme 3: Schematic representation of the crystal structure of apo-*Hp*UreE (**A**) and Ni-*Hp*UreE (**B**), showing the metal-binding role of the histidine residues in the C-terminal region of the protein.

Moreover it has been reported that the metal binding properties of *Sp*UreE involve the presence of two sites per protein dimer, with the Ni(II) ions in octahedral coordination geometry and an average of two histidine residues and four O/N ligands bound to each metal ion [115]. The involvement of the C-terminal portion of *Sp*UreE in the binding of Ni(II), and in particular the role of His^{145} and

His[147], was indicated [117] by the disappearance of their NMR signals due to binding of the Ni(II) ions established to be paramagnetic high-spin in octahedrally coordinated sites [107]. Two possible models were then used to fit the binding data (see Scheme 4).



Scheme 4: Schematic representation of the Ni(II) binding modes by *Sp*UreE used to fit the data by equilibrium dialysis experiments [115].

One model (A) involved the presence of two identical metal sites with negative cooperativity (K_{d1} = 1.4 μM and K_{d2} = 25 μM), while the alternative model (B) assumed the presence of two different binding sites featuring different affinity for Ni(II) (K_{d1} = 0.7 μM and K_{d2} = 50 μM) and unresolved cooperativity [115]. In the first model, each metal ion is bound to the conserved His[100] and to either one of the two histidine residues found at the C-terminal portion of each monomer, His[145] or His[147]. In the second model, one Ni(II) ion is bound to one His[100] from each monomer, and the second Ni(II) ion is bound to a pair of either His[145] or His[147] [115]. In the absence of additional structural information, no conclusion as to which of the two models is correct could then be reached.

Metal ion selectivity is a general issue concerning a large variety of processes related to metal homeostasis and cellular metal trafficking. Among the different metal ions that UreE proteins can bind *in vitro* (Ni(II), Zn(II), Cu(II), Co(II)), Zn(II) appears to have a role in the urease maturation process, through the dimerization of UreG [98]. Zn(II) and Ni(II) atoms share the same binding site at the interface of the protein dimer in the *Sp*UreE crystal structure [109]. In *Ka*UreE, the affinity for Zn(II) is 10 times lower than the affinity for Ni(II) [118], while the thermodynamics of Zn(II) binding to *Hp*UreE is very similar to that of Ni(II) [103]. Despite this similarity, the interaction of cognate Ni(II) and non-cognate Zn(II) with *Hp*UreE occurs with two different binding modes, suggesting that functional metal ion selection can be performed *in vivo* through different ligand environments and protein conformations.

The consequences of the different identity of the bound metal ion on the structural properties of the holo-protein, have been recently identified as key factors that modulate key cellular metabolic metal-dependent processes [119].

4.7. UreE-UreG interactions

In addition to its role as a Ni(II) metallo-chaperone, UreE appears to be involved in the enhancement of the GTPase activity of UreG [120]. This process involves the direct interaction of UreE with UreG observed, in the case of the proteins from *Helicobacter pylori,* both *in vivo* through two-hybrid and immunoprecipitation experiments [121] and *in vitro* using calorimetry and NMR spectroscopy [103]. A structure of the *Hp*UreE-*Hp*UreG complex was obtained by molecular modeling, suggesting that the protein-protein interface comprises the metal binding sites on both *Hp*UreE (His^{102}) and *Hp*UreG (Cys^{66} and His^{68}) [103] (Fig. 4.12). However, experimental information on the structure of UreE-UreG complexes from any biological source is still missing.

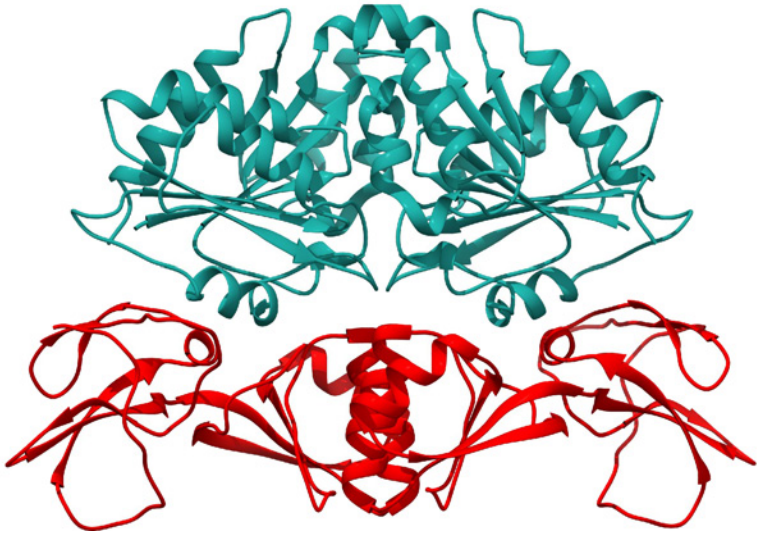


Figure 4.12: Model structure of *Hp*UreE-*Hp*UreG complex in ribbon representation [103]. *Hp*UreE (*red*) and *Hp*UreG (*sea green*) are in the apo-forms.

Consistent with this model, a novel metal site was observed in the *Hp*UreE-*Hp*UreG complex that featured a much higher binding affinity for Zn(II) as compared to the separate proteins [103]. Even a different role for *Hp*UreE bound to Zn(II) or to Ni(II) is suggested by the observation that Zn(II),

but not Ni(II), stabilizes the interaction between the two proteins [113]. In the crystal structure, the co-ordination of the Zn(II) site is tetrahedral, whereas the Ni(II) adopts an axially elongated distorted octahedral site that is approximately square bipyramidal. This reflects the intrinsic co-ordination preferences of the two metal ions, and from their electronic properties it depends the metal selectivity of UreE too. The ability of the different cations to achieve their preferred co-ordination number in the complex with UreE indicates a significant flexibility of the protein [113].

5. Aim of the study

5. Aim of the study

Urease is a nickel-dependent enzyme that catalyzes the hydrolysis of urea as the last step of organic nitrogen mineralization. Its active site contains a dinuclear center of Ni(II) ions that must be inserted in the apo-enzyme *in vivo* through the action of four accessory proteins, named UreD, UreE, UreF and UreG. This process leads to activation of the enzyme. Among these proteins, UreE appears to act as a metallo-chaperone that delivers Ni(II) to a preformed complex of apo-urease-UreDFG. Another role for UreE is the enhancement of the GTPase activity of UreG, the accessory enzyme acting as co-activator in the urease maturation process.

This study is focused on UreE from *Sporosarcina pasteurii*, a highly ureolytic organism, with the aim of deriving the structure/mobility-function relationships that control nickel binding by *Sp*UreE and its interaction with *Sp*UreG. Solution studies were planned with the purpose of obtaining structural information on the complex involving UreE and UreG from *Sporosarcina pasteurii*. In particular, the stoichiometry and the thermodynamic parameters for this interaction were investigated using isothermal titration calorimetry (ITC), and the Chemical Shift Perturbations (CSPs) induced by the protein-protein interaction were further analyzed using high resolution NMR spectroscopy, in order to map the protein surface on *Sp*UreE that is interested by the complex formation with *Sp*UreG. To this purpose, the assignment of the protein backbone resonances was required and subsequently carried out using triple resonance experiments on the labeled (^{2}H, ^{15}N, ^{13}C) protein. Moreover, in order to extend the investigation on the behaviour of the protein in solution, the backbone protein dynamics of *Sp*UreE through ^{15}N relaxation measurements was determined, providing information on the mobility of the different protein regions. The latter analysis offers the unique possibility to carry out a characterization of the C-terminal portions that were not observed in the crystal structure of UreE because of disorder, whereas they are known to contain key residues for metal ion binding.

Previous studies reported that the dimeric metallo-chaperone UreE can bind different metal ions *in vitro*, beside the cognate Ni(II). Metal ion selectivity is a general issue concerning a large variety of processes related to metal homeostasis and cellular metal trafficking. Moreover, since nickel binding properties of *Sp*UreE, previously determined, involve the presence of two sites per protein dimer, and two possible models were proposed thus far, without any conclusion about which of the two was the correct one, a combined approach was employed to discriminate between these two models. This approach coupled structural determination in the crystalline state by crystallography,

and metal binding characterization and competition in solution by calorimetry. The ITC analysis was employed for the purpose of exploring the metal selectivity and affinity of *Sp*UreE for cognate (Ni(II)) and non-cognate (Zn(II)) metal ions.

6. Materials and methods

6. Materials and methods

6.1. *Sp*UreE expression and purification

Recombinant wild-type UreE from *Sporosarcina pasteurii* (GenBank: AAD55059.1, 147 residues) was purified using a modification of the protocol employed by Stola *et al*. [115]. *Escherichia coli* BL21(DE3) (Stratagene) cells were transformed by heat shock with the pET-3d::ureE plasmid containing the wild type *ureE* gene (512 bp), as previously described [115]. The transformed cells were subjected to a double-selection process aimed at the production of high yields of the recombinant protein [122]. The selected colonies were grown aerobically under vigorous stirring in a medium supplemented with 100 mg/mL ampicillin. The M9 medium (6 g L^{-1} Na_2HPO_4, 3 g L^{-1} KH_2PO_4, 0.5 g L^{-1} NaCl, 1 mM $MgSO_4$, 4 g L^{-1} glucose, 1.25 g L^{-1} ammonium sulfate (AMS) at pH 7.5) was used for the unlabeled protein. The *E. coli* OD2-CN (^{15}N, ^{13}C > 98%, Silantes), the *E. coli* OD2-DN (^{2}H, ^{15}N > 98%, Silantes) or the *E. coli* OD2-CDN ($^{2}H \geq 95\%$, ^{13}C, ^{15}N > 98%, Silantes) rich growth media were used to obtain the correspondingly labeled *Sp*UreE. For the latter case, the cells were grown in media containing increasing amounts of D_2O (30%, 60%, 90%, 100%) to allow for adaptation [123].

Expression was obtained by incubating the cells at 37 °C for 16 hours after addition of 0.5 mM IPTG when OD_{600} reached 0.8-0.9. Cells were harvested by centrifugation at 10,000 x *g* for 30 min at 4 °C, re-suspended in 50 mM Tris-HCl buffer, pH 7.0, containing 10 mM $MgCl_2$, 2 mM dithiothreitol and 20 mg/mL DNase-I, and broken by two passages through a French Pressure cell (SLM-Aminco) operating at 20,000 psi. Cell debris was separated from the supernatant by centrifugation at 75,600 x *g* for 30 min at 4 °C, and solid AMS was added to the soluble portion of the cellular extract up to 60% saturation. The clear supernatant, obtained after centrifugation at 75,600 x *g* for 30 min at 4 °C, was loaded onto a Phenyl-Sepharose XK 26/10 column (GE Healthcare) pre-equilibrated with 50 mM Tris-HCl buffer, pH 7.0, containing 2 M AMS and 5 mM EDTA. The column was eluted using 300 mL of a linear gradient of AMS (from 2 to 0 M) in 50 mM Tris-HCl buffer, pH 7.0, 5 mM EDTA, with a flow rate of 3 mL/min. *Sp*UreE was eluted at about 1.2 M AMS and the collected fractions were dialyzed against 50 mM Tris-HCl buffer, pH 7.0, 5 mM EDTA. The resulting solution was applied onto a Q-Sepharose XK 16/10 column (GE Healthcare) pre-equilibrated with 50 mM Tris-HCl buffer, pH 7.0, 5 mM EDTA and eluted with 300 mL of a linear gradient of NaCl from 0 to 1 M, at 3 mL/min, in the same buffer. Fractions containing *Sp*UreE, eluted at about 0.2 M NaCl, were combined, concentrated to 2 mL using an

Amicon Ultra-15 centrifugal filter unit (10 kDa molecular weight cut-off, Millipore), and loaded onto a Superdex-75 16/60 size-exclusion column, eluted using 20 mM HEPES buffer, pH 7.0, 150 mM NaCl in order to exchange the buffer and remove EDTA.

6.2. *Sp*UreG expression and purification

*Sp*UreG was expressed and purified following a modification of the protocol reported in Zambelli *et al*. [95]. *E. coli* BL21(DE3) (Stratagene) cells, transformed with the pET3a::ureG plasmid [95], were grown at 37 °C in LB medium with vigorous stirring until the OD_{600} reached 0.5-0.6. Protein expression was induced with 0.5 mM IPTG at 20 °C for 16 hours. Cells were harvested by centrifugation at 10,000 x *g* for 30 min at 4 °C, re-suspended in 50 mM Tris-HCl buffer, pH 8.0, containing 10 mM $MgCl_2$, 1 mM DTT, 5 mM EDTA and 20 mg/mL DNase-I, and lysed as described above for *Sp*UreE. The obtained lysate was centrifuged and the supernatant was loaded onto a Q-sepharose XK 26/10 column (GE Healthcare), pre-equilibrated with two volumes of 20 mM Tris-HCl buffer at pH 8.0 containing 5 mM EDTA and 1 mM DTT. The unbound proteins were washed away using the equilibration buffer until the baseline was stable. *Sp*UreG was then eluted using 300 mL of a linear gradient of NaCl from 0 to 1 M, at 3 mL/min. Fractions containing *Sp*UreG were collected and solid AMS was slowly added under continuous mixing at 4 °C until 1 M concentration was reached. The resulting suspension was centrifuged at 75,600 x *g* for 30 min at 4 °C and the supernatant was loaded onto a Phenyl Sepharose 16/10 column (GE Healthcare), pre-equilibrated with 20 mM Tris-HCl buffer at pH 8.0 containing 1 M AMS and 1 mM DTT. The column was washed with the same buffer, and the protein was eluted with a linear gradient of 300 mL of AMS from 1 M to 0 M at 3 mL/min. The fractions containing *Sp*UreG were collected, concentrated using an Amicon Ultra-15 centrifugal filter unit (10 kDa molecular weight cut-off, Millipore), and then loaded onto a Superdex 75 16/60 column (GE Healthcare) pre-equilibrated with 20 mM HEPES buffer at pH 7.0, containing 150 mM NaCl and 1 mM TCEP. The fractions containing *Sp*UreG were collected and concentrated as above. TCEP was removed using Zeba Spin Desalting Columns (Thermo Scientific) just before each NMR experiment using the same buffer to maintain the protein in the reduced state as long as possible to prevent oxidation-induced dimerization.

6.3. Preliminary proteins characterization and determination of metal content in samples

Proteins purity were analyzed using SDS-PAGE and employing NuPAGE Novex Pre-Cast Gel System (Invitrogen) and NuPAGE 4-12% Bis-TRIS gels (Invitrogen). Staining was carried out using SimplyBlue Safestain (Invitrogen). Proteins quantifications were performed by absorption spectroscopy, using the theoretical value of 21,430 M^{-1} cm^{-1} for *Sp*UreE and 11,460 M^{-1} cm^{-1} for *Sp*UreG for the extinction coefficient at 280 nm, calculated using the ProtParam tool (http://us.expasy.org/cgi-bin/protparam) and the amino acid sequences of the proteins. The purified proteins were devoid of metal ions as shown by inductively coupled plasma emission spectrometry (ICP-ES), as described in protocol reported by Stola *et al*. [115]. Throughout this thesis, the provided proteins concentrations are referred to the monomer if not differently specified.

6.4. NMR spectroscopy experiments

6.4.1. Data collection and analysis for backbone assignment

NMR spectra were acquired at 25 °C on a Bruker AVANCE 900 spectrometer, operating at the proton nominal frequency of 899.2 MHz (21.1 T) on NMR samples containing 0.5 mM of labeled *Sp*UreE dimer in 20 mM HEPES buffer at pH 7, containing 150 mM NaCl, in 90% H_2O and 10% D_2O. At this protein concentration, *Sp*UreE is reported to be a dimer [107], [115], [117]. The spectrometer was equipped with a TCI 5-mm triple resonance cryo-probe with Pulsed Field Gradients along the z-axis.

2D ^{1}H-^{15}N HSQC and ^{1}H-^{15}N TROSY-HSQC, 3D TROSY-HNCO, TROSY-HNcaCO, TROSY-HNCA, TROSY-HNcoCA, TROSY-HNCACB, and TROSY-HNcoCACB experiments on the $^{2}H/^{15}N/^{13}C$-labeled *Sp*UreE were used to obtain the sequential backbone resonance assignment of ^{1}H, ^{13}C, and ^{15}N nuclei (Table 1 SI). In these pulse schemes, water suppression is achieved using selective pulse and transverse signal cancellation with pulsed field gradients associated with a flip-back pulse.

The NMR data were processed with the NMRpipe/NMRDraw software package v. 8.1 [124] using a squared cosine function to apodize the data and zero-filled once in all dimensions before Fourier transformation, using forward and backward linear prediction for the indirect dimensions to increase resolution. Spectral analysis for resonance assignment was performed using CARA 1.8.4.2 [125].

6.4.2. Data collection and analysis for backbone mobility.

The experiments for the determination of ^{15}N longitudinal ($R_1 = 1/T_1$) and transverse ($R_2 = 1/T_2$) relaxation rates, and of the ^{1}H-^{15}N cross-relaxation rate, measured via steady-state heteronuclear ^{1}H-^{15}N NOE, were acquired using phase-sensitive gradient-enhanced sequences [34], [126] at 25 °C on a Bruker AVANCE 700 spectrometer operating at the proton nominal frequencies of 700.13 MHz (16.4 T), using a TCI 5-mm triple resonance cryo-probe equipped with Pulse Field Gradients along the z-axis. The NMR samples contained 0.5 mM of $^{2}H/^{15}N$-labeled *Sp*UreE homo-dimer in 20 mM HEPES buffer at pH 7.0, containing 150 mM NaCl, in 90% H_2O and 10% D_2O. In these pulse schemes, water suppression is achieved using selective pulse and transverse signal cancellation with pulsed field gradients associated with a flip-back pulse that orients the water nuclei along the +z axis during acquisition in order to avoid saturation of water magnetization that could affect the NOE values of solvent-exposed backbone amide protons.

The NMR experiments used to measure relaxation rates consist of a series of ^{1}H-^{15}N HSQC experiments in which spectra are acquired at different time intervals after the radiofrequency pulse. T_1 measurements were based on inversion-recovery type of experiments recorded using eight different delays: 10, 70, 150, 250, 380, 540, 740, and 1350 ms. T_2 measurements were carried out using a Carr-Purcell-Meiboom-Gill (CPMG) spin-echo pulse sequences recorded with ten different delays prior to the 180° refocusing pulse: 8.48, 16.96, 25.44, 33.92, 42.40, 50.88, 59.36, 67.84, 76.32, and 84.8 ms. Recycle delays of 4 s were used in both experiments. In both T_1 and T_2 measurements, the large delay was duplicated. ^{1}H-^{15}N NOE values were obtained recording two sets of spectra in the presence and absence of a 6 s proton saturation period. The two spectra were recorded in an interleaved manner to ensure identical conditions in the two experiments. Spectra were processed using TopSpin 3.2 (Bruker BioSpin) and peak intensities were analyzed using Dynamics Center 2.1.5 (Bruker BioSpin). The values of R_1 and R_2 were calculated by fitting the peak intensities in ^{15}N-^{1}H HSQC spectra acquired at different relaxation delays to a two-parameters exponential decay function, using a non-linear Marquard algorithm. The errors on the parameters derived from the fit were estimated from the inverse of the weighted curvature matrix, using a confidence level of 95%. The heteronuclear NOE values were estimated as the ratio of the intensities in the saturated *vs* non-saturated spectra. In all cases, the uncertainty on the intensities was derived from the standard deviation of the noise in each spectrum and differences in intensity values for repetition experiments. The Dynamics Center 2.1.5 (Bruker BioSpin) software was used to carry out the quantitative analysis of the relaxation data, assuming a chemical shift anisotropy of the ^{15}N spin of -160 ppm and 1.02 Å for the bond length of the amide NH vector.

6.4.3. *Sp*UreE-*Sp*UreG interaction followed by NMR

$^{2}H/^{15}N$-labeled 0.5 mM *Sp*UreE dimer was titrated with unlabeled *Sp*UreG monomer at ratios of 1:1, 1:2, 1:3, and 1:4 in 20 mM HEPES buffer at pH 7.0, containing 150 mM NaCl, in 90% H_2O and 10% D_2O, at 25 °C using a Bruker AVANCE 950 spectrometer, operating at the proton nominal frequency of 950.2 MHz (22.3 T). The spectrometer was equipped with a TCI 5-mm triple resonance cryo-probe with Pulsed Field Gradients along the z-axis. Corresponding ^{1}H-^{15}N TROSY-HSQC spectra were recorded and compared to the ^{1}H-^{15}N TROSY-HSQC spectrum of free *Sp*UreE. The original data were zero-filled four times in F_1 and eight times in F_2 prior to Fourier transformation, and mild resolution enhancement was achieved by applying a p/3-shifted sine-squared apodization function in both dimensions, using the iNMR software (www.inmr.net). The chemical shifts were determined using the spectrum of *Sp*UreE in the absence and in the presence of four equivalents of *Sp*UreG monomer, using the peak-picker tool in iNMR. Chemical shift perturbations (CSP) were calculated using the formula: $\Delta\delta = \Delta H_N + \Delta N/7$ (or $\Delta\delta = \Delta H_N + \Delta N/5$ for glycines), where ΔH_N and ΔN are the absolute values of the chemical shift differences (ppm) of the amide proton (H_N) and nitrogen (N) resonances, respectively [33], [127].

6.5. ITC experiments

Isothermal Titration Calorimetry experiments were performed at 25 °C using a high-sensitivity VP-ITC microcalorimeter (MicroCal LLC, Northampton, MA, U.S.A.).

The integrated heat data were analyzed using the Origin software package (MicroCal), and fitted using a non-linear least-squares minimization algorithm to theoretical titration curves that involved different binding models. The reduced parameter χ_v^2 ($\chi_v^2 = \chi^2/N$, where N is the degrees of freedom, $N = N_{idp} - N_{par}$, N_{idp} = number of points, and N_{par} = number of parameters floating in the fit) was used to establish the best fit among the tested models. Values for the enthalpy change of reaction (ΔH), the binding affinity constant (K_b) and the number of sites (n) were the parameters of the fit. The reaction entropy was calculated using the equations $\Delta G = - RT\ln K_b$ (R = 1.9872 cal mol^{-1} K^{-1}, T = 298 K) and $\Delta G = \Delta H - T\Delta S$ (see Chapter 3).

6.5.1. Characterization of metal binding selectivity and affinity

Direct and reverse titration experiments with metals were performed in order to investigate the metal binding selectivity and affinity of *Sp*UreE.

*Sp*UreE, purified as described above, was eluted from a Superdex-75 size-exclusion column, using 20 mM TRIS-HCl buffer, 150 mM NaCl, pH 7.0, immediately before the ITC measurement to exchange the buffer and to freshly purify the protein. *Sp*UreE and the metal ions (from 100 mM stock solutions) were diluted using the same buffer. The measuring cell contained 1.4093 mL of either *Sp*UreE or metal ion solution for direct or reverse titrations, respectively, and the reference cell was filled with deionized water. Before starting the experiments, the baseline stability was verified. A spacing of 300-360 s between injections was applied in order to allow the system to reach thermal equilibrium after each addition. For each titration, a control experiment was carried out by adding the titrating protein or metal ions solution into the buffer alone, under identical conditions. Heats of dilution were negligible. In the case of direct titrations, a *Sp*UreE solution (15 mM dimer) in the cell was titrated with 55 injections (5 mL each) of $NiSO_4$ or $ZnSO_4$ solutions (0.5 mM). In the case of reverse titrations, the $NiSO_4$ or $ZnSO_4$ solutions (20 μM) were titrated with 25 (for Ni(II)) or 40 (for Zn(II)) injections (10 μL each) of a solution containing 70-80 μM *Sp*UreE dimer. In the case of Zn(II) titration, three subsequent experiments were performed to reach binding saturation, refilling the syringe with the metal ion solution and continuing the titration. The three binding curves were then concatenated in a single curve using the program ConCat32 (MicroCal). For competition experiments, *Sp*UreE (19 mM dimer) was incubated in the cell with equimolar concentration of Ni(II) (or Zn(II)) and titrated with 55 injections (5 mL each) of 0.5 mM Zn(II) (or Ni(II)) solutions (in the syringe).

The dissociation constants and thermodynamic parameters obtained from this analysis do not take into account possible events of proton transfer linked to metal binding. However, the values of the measured equilibrium constants compare well with those reported in the literature and determined using equilibrium dialysis coupled to metal analysis, which, in principle, should also take into account similar effects [115].

6.5.2. *Sp*UreE-*Sp*UreG interaction followed by ITC

*Sp*UreE and *Sp*UreG samples were freshly eluted from a Superdex-75 size-exclusion column, using 20 mM Tris-HCl buffer at pH 7.0 containing 150 mM NaCl, immediately before the ITC

experiments to exchange the buffer that has to be the same for both proteins. A solution of *Sp*UreG (50 μM monomer) was degassed and loaded into the sample cell (V = 1.4093 mL), while the reference cell was filled with deionized water. The syringe was filled with a 150 μM solution of *Sp*UreE dimer and the temperature of the two cells was set and stabilized at 25 °C. Stirring speed was 300 rpm, and thermal power was monitored every 2 s using high instrumental feedback. A series of 28 injections (10 μL each) of *Sp*UreE solution was carried out at intervals of 300 s, a time necessary to allow the system to reach thermal equilibrium after each injection. A control experiment was carried out by adding the titrating solution into the buffer alone, under identical conditions. Heats of dilution were negligible.

6.6. Bioinformatics analysis

6.6.1. NMR data processing and analysis

NMRPipe

NMR data were processed with the NMRPipe/NMRDraw software package v. 8.1 [124]. The NMRPipe suite of programs is a UNIX/linux software developed for processing, graphics, and analysis of multidimensional NMR data. The software is based on UNIX pipes and allows a spectral data flows through a pipeline of processing functions, such as Fourier transformation or linear prediction. Complete multidimensional processing scheme are constructed as simple UNIX shell scripts. The NMRPipe processing program is the central part of a system of tools for multi-dimensional spectral processing and analysis carried out using UNIX pipes. It applies a succession of spectral processing functions, i.e. a one-dimensional Fourier Transform, to a series of spectral data vectors. When used in conjunction with data input and output programs NMRPipe can be used to construct complete multi-dimensional processing schemes.

iNMR

The processing using iNMR (http://www.inmr.net/) was performed using a preformed sequence of steps for each dimension of the spectrum: zero-filling, apodization function, Fourier transformation and phase correction. Before starting the processing flow, the solvent peak was digitally removed.

After the processing was carried out, many frequency-domain algorithms were applied, such as baseline correction, symmetrization, noise reduction, *etc*.

CARA

The analysis of NMR spectra for resonance assignment was performed using the software application CARA 1.8.4.2 (computer aided resonance assignment) http://cara.nmr.ch/doku.php [125]. CARA is particularly suited for biomacromolecules structure determination since it includes a number of dedicated tools for backbone assignment, side chain assignment, and peak integration that support the entire process.

Dynamics Center

Dynamics Center 2.1.5 (Bruker BioSpin) is an integrated platform used for analysing NMR dynamic data. The basic dynamics analysis includes hetero nuclear NOE, T_1, T_2, and R_{ex} data, and yields relaxation parameters. The modeling of dynamical parameters of protein backbones such as correlation times and order parameters is performed using reduced spectral density analysis and model free modeling. If the molecular structure is available, determination of the diffusion tensor and anisotropic modeling can be done.

6.6.2 Structural parameters predictions

SPHERE

SPHERE is a server program (http://www.fccc.edu/research/labs/roder/sphere/sphere.html) which is able to predict amide hydrogen exchange rates [128]. It was applied to the crystal structure of dimeric *Sp*UreE (PDB code 1EAR) to calculate the intrinsic exchange rates of every single unassigned amino acid.

PONDR

The Predictors of Natural Disordered Regions (PONDR) is developed as a series of neural network predictors (NNPs) that use amino acid sequence data to predict disorder in a given region

(http://www.pondr.com). The predictors are feed-forward neural networks that use sequence information from windows of generally 21 amino acids. Attributes, such as the fractional composition of particular amino acids or hydropathy, are calculated over this window, and these values are used as inputs for the predictor. The neural network, which has been trained on a specific set of ordered and disordered sequences, then outputs a value for the central amino acid in the window. The predictions are then smoothed over a sliding window of 9 amino acids. If a residue value exceeds a threshold of 0.5 (the threshold used for training) the residue is considered disordered [129]. In particular PONDR **VL-XT** was employed as it refers to the merger of three predictors, one trained on **V**ariously characterized **L**ong disordered regions and two trained on **X**-ray characterized **T**erminal disordered regions [129], [130]. VL-XT outputs are real numbers between 1 and 0, where 1 is the ideal prediction of disorder and 0 is the ideal prediction of order. VL-XT outputs are typically not ideal and a threshold is applied with disorder assigned to values greater than or equal to 0.5. The graph shows the residue by residue output of the neural network. Any region that exceeds 0.5 on the Y-axis is considered disordered (see Fig. 7.4 A). Note that as the length of the predicted disordered region increases, the accuracy of the prediction increases as well. Extremely long predictions of disorder have a very high level of confidence. Based on these outputs the position of *Sp*UreE in the Uversky plot [131] was also obtained (see Fig. 7.3). This plot, taking in consideration the mean scaled of hydropaty and the absolute mean of net charge of proteins, allows to reliably separate the "natively unfolded" proteins from native ones.

SSP

The Secondary Structure Propensity (SSP) method provides a quantitative estimation of secondary structure propensity. SSP is an algorithm that combines chemical shifts from different nuclei into a score, which reflects the expected fraction of α- or β-structure. The contributions of different chemical shifts are weighted by their sensitivity to α- or β-structure. A SSP score at a given residue of 1 or -1 reflects fully formed α- or β-structure respectively, while a score of 0.5 indicates that 50% of the conformers in the disordered state ensemble are helical at that position [132]. (http://pound.med.utoronto.ca/~JFKlab)

TALOS+

TALOS+ (http://spin.niddk.nih.gov/NMRPipe/talos/) is a hybrid system for the empirical prediction of protein backbone torsion angles (Φ and Ψ) values and their uncertainty. The program uses a

combination of some nuclei (HN, HA, CA, CB, CO and N) chemical shift assignment for a given amino acid sequence of the polypeptide. The original TALOS approach is based on the well-known observation that many kinds of secondary chemical shifts (i.e. differences between chemical shifts and their corresponding random coil values) are highly correlated with aspects of protein secondary structure. Original TALOS approach, in a high-resolution structural database, searched the 10 best matches to the secondary chemical shifts of given residue in a target protein along with its two flanking neighbours. If there was a consensus of Φ and Ψ angles among the 10 best database matches, then these database triplet structures was used to form a prediction for the backbone angles of the target residue. TALOS+ implements the database searching method of the previous version with a new classification scheme based on artificial neural networks, as well as a larger database of 200 proteins. This improved approach allows TALOS+ to make a larger number of useful backbone angle predictions, 88% of residues in a given protein on average. The goal of TALOS+ is to use secondary shift and sequence information in order to make quantitative predictions for the protein backbone angles Φ and Ψ, and to provide a measure of the uncertainties in these predictions. The neural network analyzes the chemical shifts and sequence to estimate the likelihood of a given residue being in a sheet, helix, or loop conformation. The results from the calculation of each residue associated with those deriving from research in databases increase the chance of getting a correct prediction of the angles of the backbone [133].

7. Results and discussion

7. Results and discussion

7.1. Characterization of *Sp*UreE-*Sp*UreG complex

The study of the detailed structural properties of protein-protein complexes must rely on X-ray diffraction crystallography, nuclear magnetic resonance spectroscopy, or computational studies [11]. Whereas the interaction between UreE and UreG proteins by computational approaches have been previously reported [103], and attempts to co-crystallize UreE and UreG are currently in progress, here is reported an NMR investigation of structural properties of the complex formation in solution. Initial trials focused on the proteins from *Helicobacter pylori* were unsuccessful: the deuterated form of *Hp*UreE, necessary to improve the spectral resolution for this large 35-kDa protein and its complexes, was unstable, while UreG proteins are intrinsically disordered in solution independently of the biological source [95], [100], [134], with NMR spectra characterized by large signal broadening and overlap [95], [96], [98], [135], preventing an extended assignment of their NMR spectra. In particular, *Hp*UreG is the most folded UreG so far investigated [98], but it was unstable in solution over the course of the several days necessary to carry out the full backbone NMR signal assignment. Therefore, the attention was turned onto the two proteins from *Sporosarcina pasteurii*, which have proved to be much more resilient to denaturation and aggregation.

7.1.1. *Sp*UreE-*Sp*UreG by ITC

The binding stoichiometry and thermodynamic parameters for this protein-protein interaction were investigated using isothermal calorimetric measurements carried out by titrating a diluted *Sp*UreG solution with a concentrated *Sp*UreE solution. The occurrence of a binding event was revealed by the presence of exothermic peaks that followed each addition (Fig. 7.1 A). Integration of the peak areas revealed a binding isotherm (Fig. 7.1 B) that was initially fit considering a direct titration of *Sp*UreE dimer into *Sp*UreG monomer and a single set of identical binding sites (Table 1). The obtained fitting parameters (χ_v^2 = 6,765, n= 0.37 ± 0.01, K_b = (8.3 ± 0.5) x 10^5 M^{-1} (K_{d1} = 1.20 ± 0.07 μM), ΔH = -5.75 ± 0.05 kcal mol^{-1} and ΔS = +7.80 cal mol^{-1} K^{-1}) were statistically consistent but showed an unrealistic stoichiometry. An alternative model involving two sets of independent binding sites, even though resulting in a better fit of the data (χ_v^2 = 3,050) also showed unrealistic stoichiometries, as well as an unacceptable error in one of the affinity constants, suggesting an over-

parametrization of the fitting procedure (n_1= 0.30 ± 0.03, K_{b1} = (7.7 ± 0.4) x 10^5 M^{-1} (K_{d1} = 1.30 ± 0.07 μM), ΔH_1 = -5.49 ± 0.06 kcal mol^{-1} and ΔS_1 = +8.52 cal mol^{-1} K^{-1}; n_2= 0.08 ± 0.01, K_{b2} = (1.8 ± 2.7) x 10^8 M^{-1} (K_{d1} = 6 ± 9 nM), ΔH_2 = -5.48 ± 0.05 kcal mol^{-1} and ΔS_2 = +19.4 cal mol^{-1} K^{-1}). The fitting model that considers the sequential binding of *Sp*UreE dimers to *Sp*UreG monomers was not applicable in this case, because of the half-integer stoichiometry evident from the binding isotherm. An alternative fitting approach involved a reverse titration model [55], [56] that considers the titration of *Sp*UreG monomers onto *Sp*UreE dimers. The binding scheme involving a single set of sites gave, as before, an unreliable stoichiometry, considering the homo-dimeric nature of *Sp*UreE and the dimerization equilibrium involving *Sp*UreG: (χ_v^2 = 6,765, n= 2.73 ± 0.02, K_b = (3.0 ± 0.5) x 10^5 M^{-1} (K_d = 3.3 ± 0.5 μM), ΔH = -2.10 ± 0.02 kcal mol^{-1} and ΔS = +18.0 cal mol^{-1} K^{-1}). A second model considering two independent sets of binding sites did not go to convergence and was therefore discarded. Finally, a third model involving a sequential binding scheme gave realistic and statistically consistent fitting parameters (χ_v^2 = 5,612, n_1= 1.0, K_{b1} = (2.4 ± 0.5) x 10^4 M^{-1} (K_{d1} = 42 ± 9 μM), ΔH_1 = +2.7 ± 0.9 kcal mol^{-1} and ΔS_1 = +29.2 cal mol^{-1} K^{-1}, n_2= 1.0, K_{b2} = (6 ± 1) x 10^5 M^{-1} (K_d = 1.7 ± 0.3 μM), ΔH_2 = -8.5 ± 0.9 kcal mol^{-1} and ΔS_2 = -2.18 cal mol^{-1} K^{-1}). This model indicate that two *Sp*UreG monomers sequentially interact with a *Sp*UreE dimer showing positive cooperativity. This could be explained considering that binding of the second *Sp*UreG monomer onto *Sp*UreE, in a way analogous to the model of the UreE-UreG complex in *Helicobacter pylori* (Fig. 4.12), would induce *Sp*UreG dimerization, thus contributing to the binding affinity of the second *Sp*UreG monomer.

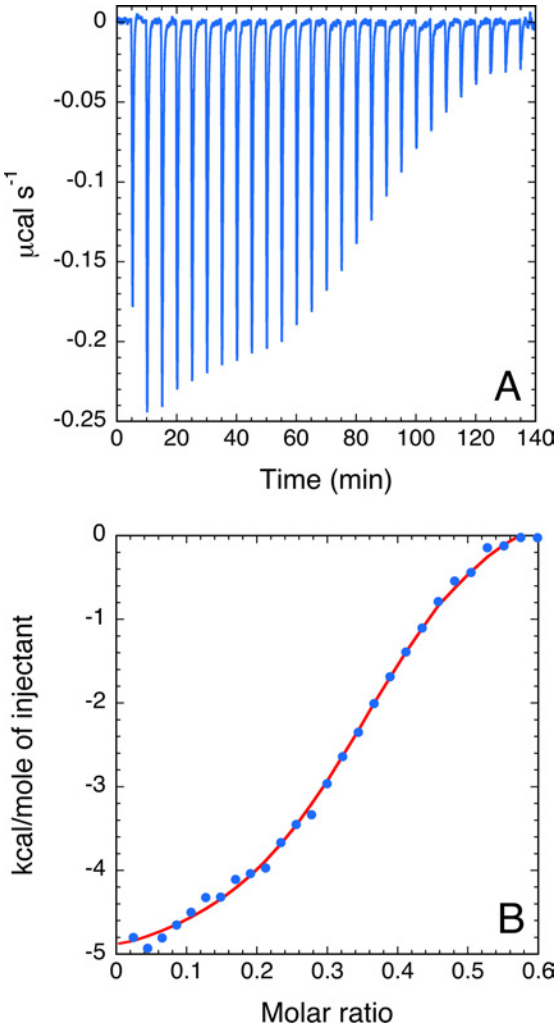


Figure 7.1: Analysis of *Sp*UreE binding to *Sp*UreG performed using isothermal titration calorimetry. (**A**) Representative plot of raw titration data of *Sp*UreE dimer (0.15 µM) onto *Sp*UreG monomer (0.05 µM). (**B**) Integrated heat data as a function of *Sp*UreE/*Sp*UreG molar ratio. The solid line represents the best fit obtained using a model that involves the sequential binding of two monomers of *Sp*UreG onto one *Sp*UreE dimer.

Table 1

Direct titration analysis - *Sp*UreE titrated onto *Sp*UreG

Binding scheme	χ_v^2	*n*	k_b (M^{-1})	ΔH (kcal mol^{-1})	ΔS (cal mol^{-1} K^{-1})
One set of sites	6,765	0.37 ± 0.01	(8.3 ± 0.5) x 10^5	-5.75 ± 0.05	+7.80
Two sets of sites	3,050	n_1 =0.30 ± 0.03	(7.7 ± 0.4) x 10^5	-5.49 ± 0.06	+8.52
		n_2 = 0.08 ± 0.01	(1.8 ± 2.7) x 10^8	-5.48 ± 0.05	+19.4
Sequential binding sites	NA	-	-	-	-

Reverse titration analysis – *Sp*UreG titrated onto *Sp*UreE

Binding scheme	χ_v^2	*n*	k_b (M^{-1})	ΔH (kcal mol^{-1})	ΔS (cal mol^{-1} K^{-1})
One set of sites	6,765	*2.73* ±0.02	(3.0 ± 0.2) x 10^5	-2.10 ± 0.02	+ 18.0
Two sets of sites	NA	-	-	-	-
Sequential binding events	5,612	n_1 = 1.0	(2.4 ± 0.5) x 10^4	+2.7 ± 0.9	+29.2
		n_2 = 1.0	(6.2 ± 1.4) x 10^5	-8.5 ± 0.9	-2.18

7.1.2. *Sp*UreE-*Sp*UreG by NMR

In order to derive structural information for this complex, it is necessary to determine the residues on *Sp*UreE involved in the interaction with *Sp*UreG and vice versa. Initial attempts carried out by NMR indicated that the ^{1}H-^{15}N HSQC spectrum of a ^{15}N-labeled *Sp*UreG sample obtained in the presence of equimolar amounts of unlabeled ^{14}N-*Sp*UreE maintained the characteristics of a disordered protein, analogously to what was observed for the same proteins from *Helicobacter pylori* [103], preventing the realization of this approach. The alternative strategy of studying ^{15}N-labeled *Sp*UreE in the presence of unlabeled ^{14}N-*Sp*UreG was instead achievable provided that no metal ions (Ni(II) or Zn(II)) were present in solution, a condition that lead to protein precipitation. The NMR spectra of *Sp*UreE have been previously studied, and the backbone amide NH signals have been assigned at pH 6.5, 0.5 M NaCl, and 35 °C [136]. However, since *Sp*UreG is less soluble and prone to aggregation at pH lower than 7.0, the new backbone assignment of *Sp*UreE was carried out at the same pH 7.0 utilized for the determination of the binding parameters for the *Sp*UreE-*Sp*UreG complex in solution by ITC, and at ionic strength similar to physiologic conditions (150 mM NaCl). Moreover it has been previously reported that *Sp*UreE slowly precipitates at 35 °C at NaCl concentrations lower than 0.5 M [137] and this fact, on one hand prevented an extensive NMR study at high temperature and low ionic strength, and, on the other hand prompted to use a temperature of 25 °C, as in the case of ITC.

Assignment of the NMR spectra of *Sp*UreE and intrinsic exchange rates

Heteronuclear 2D and 3D NMR spectra of *Sp*UreE were recorded and analyzed, and backbone assignments were obtained using the scalar connectivity provided by the triple resonance experiments, analyzed using the computer-assisted resonance assignment (CARA) software [125]. As previously reported [136], deuterium random labeling of the protein and TROSY-based pulse sequences were necessary to improve the resolution and sensitivity of the NMR spectra for this 34.8 kDa (147 residues per monomer) homo-dimeric protein [30]. These would otherwise be affected by rapid relaxation rates of the nuclei (especially ^{13}C) that participate in the magnetization transfer steps occurring during the triple resonance sequences, as well as by lack of resolution (especially in the ^{13}C dimension) due to the short acquisition times imposed by the fast transverse relaxation rates. Deuteration was accomplished using 98% ^{2}H-labeled rich growth media in 100% deuterium oxide. The large spectral dispersion of proton signals unambiguously indicates that the protein is generally well folded, while the number of cross-peaks is consistent with the number of residues in one

subunit, coherent with symmetrical orientation of the two monomers. The assigned 2D ^{1}H-^{15}N TROSY-HSQC spectrum and assignments of the amide resonances are shown in Figure 7.2, while Table 1-SI reports the chemical shifts of all assigned nuclei. The identification of backbone ^{1}HN and ^{15}N amide peaks was obtained for 125 out of the expected 144 residues (not counting M1, P83 and P139) following a standard sequential assignment procedure. Nearly complete resonance assignments were achieved for the other backbone nuclei: 85.0% for $^{13}C\alpha$, 84.3% for $^{13}C\beta$, and 77.6% for ^{13}CO. The ^{1}HN, ^{15}N, $^{13}C\alpha$, $^{13}C\beta$, and ^{13}CO chemical shifts have been deposited in the BioMagResBank (http://www.bmrb.wisc.edu) under accession number 19820. A comparison with the previous assignment carried out at 35 °C in a different buffer at pH 6.5 (BMRB code 5484) reveals an overall good agreement, except for the first few residues at the N-terminus, probably as a consequence of the different conditions of the protein solution. The 19 unassigned residues in the TROSY-HSQC spectrum (E28, D29, N31, N51, G58, D65, C103, I110, R113, K116, V125, R135-R143) were not observable in the spectrum at pH 7.0. To understand whether this was due to fast exchange with the solvent, intrinsic exchange rates were calculated [128] using the program SPHERE (http://www.fccc.edu/research/labs/roder/sphere/sphere.html) applied to the crystal structure of dimeric *Sp*UreE (PDB code 1EAR). The data indicate that residues N31, N51, G58, D65, C103, R113, K116, R135, F136, K137, K141, Y142 and R143 feature exchange rates ≥ 12 s^{-1}, consistently with their position in exposed loops connecting strands in the N-terminal domains, or in the C-terminal region. On the other hand, amide signals for residues E28, D29, I110, V125, E138 and F140 appear not to be involved in exchange with the solvent, suggesting that they are broadened beyond detection because of conformational exchange phenomena involving mobility with rates in the 10^{2} - 10^{3} s^{-1} range. These rates are intermediate between those related to chemical or solvent exchange phenomena (1 - 10^{3} s^{-1}) and those due to conformational changes involving small protein segments or slow loop reorientations (10^{6} - 10^{9} s^{-1}) or local motions involving few atoms (bond librations, 10^{9} - 10^{12} s^{-1}), and are typical for large inter-domain motions. Indeed, all these residues, with the exception of V125 located at the end of helix α4 in the C-terminal domain and possibly subject to helix fluctuations, are located at the interface between the N- and C-terminal domains, supporting the existence of inter-domain flexibility in UreE proteins, as suggested by X-ray crystallography [109] and molecular modeling studies [114]. Increasing the temperature from 25 °C to 35 °C did not lead to the appearance of additional signals, and the lower temperature was thus chosen to avoid sample degradation.

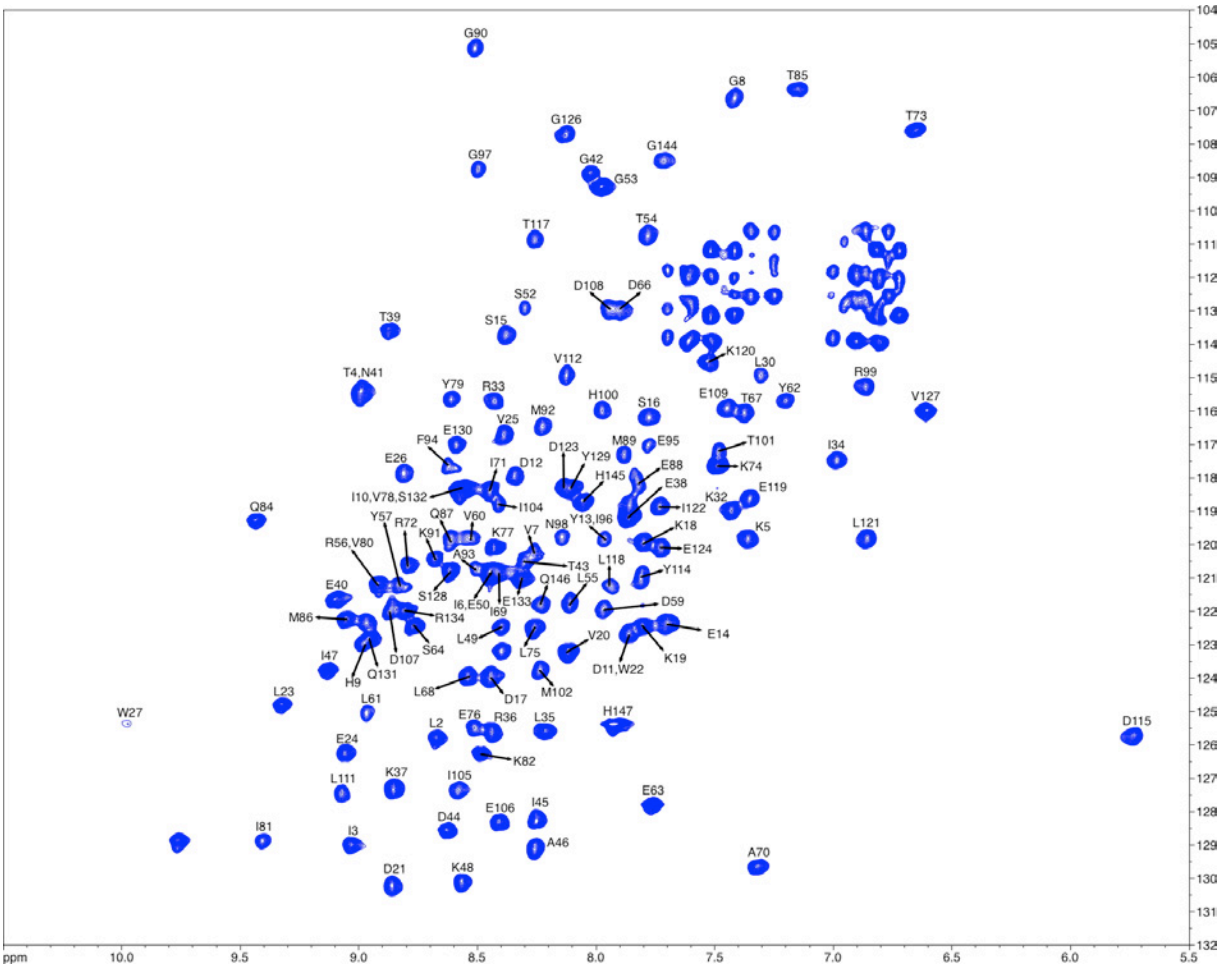


Figure 7.2: 900 MHz ^{1}H-^{15}N TROSY-HSQC spectrum of *Sp*UreE in 90% H_2O, 10% D_2O, pH 7.0, 25 °C. Assigned cross-peaks are labeled with one-letter amino acid type and sequence number.

Secondary structure propensity of *Sp*UreE from NMR chemical shifts

Considering charge and hydrophobicity in the Uversky plot [131] obtained using the disorder predictor PONDR [129] (http://www.pondr.com), *Sp*UreE is found to lie close to the boundary between proteins predicted to be folded and unfolded (Fig. 7.3). In particular, PONDR VL-XT predicts the presence of three disordered regions covering 43 residues in the segments 25-49, 96-100, and 127-139 (Fig. 7.4 A). The first two regions belong to structural fragments (Fig. 7.4 B) that were predicted to be involved in forming a complex (Fig. 4.12) between UreE and UreG [103], while the third segment covers a long final stretch of the protein that is inserted between the N- and C-terminal domains, possibly modulating the inter-domain flexibility of the protein.

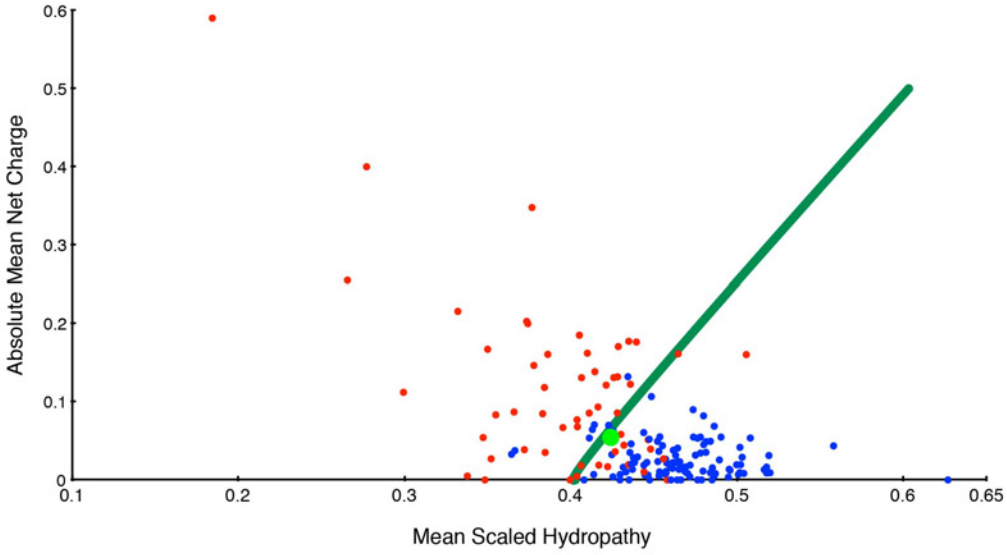


Figure 7.3: Charge-hydropathy plot for *Sp*UreE calculated using PONDR; data for ordered and disordered proteins are *blue* and *red*, respectively, while the position of *Sp*UreE is shown as a *green* dot; the *green* line represents the border between "natively unfolded" and native proteins.

This picture of protein disorder fits well with the idea that protein-protein interactions depend on mobile regions at the molecular interface, which is also consistent with molecular dynamics calculations on UreG, which indicated the presence of disorder in regions predicted to be interacting with UreE [134], and also found by crystallography to be involved in the interaction between UreG and the UreDF complex in *Helicobacter pylori* [97]. This should not be surprising, because the phenomenon of moonlighting, by means of which a given protein fulfills more than one function by interacting with different partners while adopting different conformations, depending on the metabolic process, is well-known for intrinsically disordered proteins [9].

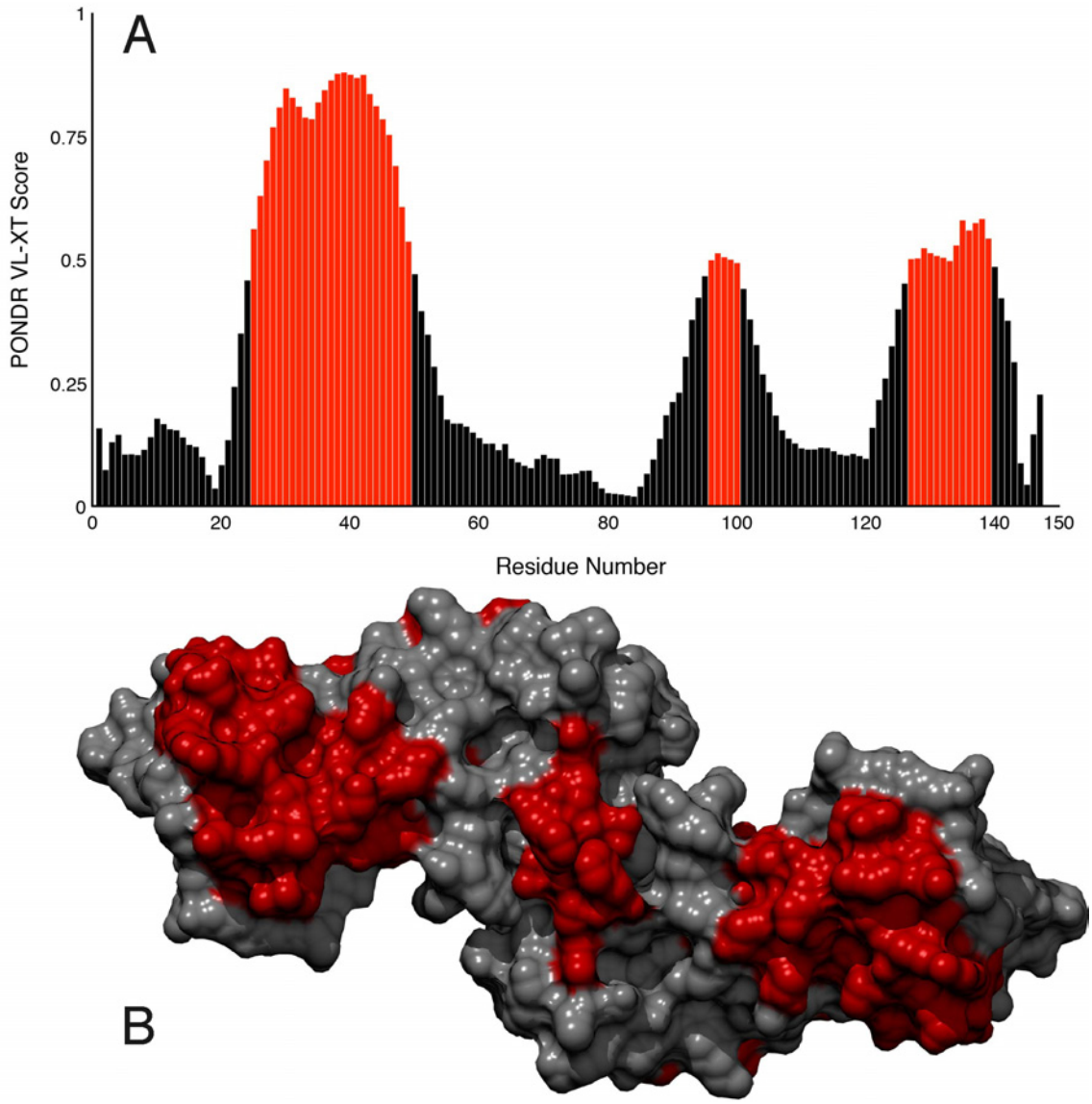


Figure 7.4: Predicted regions of order and disorder using PONDR VL-XT; (**A**) sections with scores higher than 0.5, shown in red, are predicted to be disordered; (**B**) surface of *Sp*UreE (PDB code 1EAR) showing in *red* the regions predicted as disordered by PONDR VL-XT.

In order to validate the disorder prediction, the experimental NMR chemical shifts were used to predict the secondary structure propensity for *Sp*UreE in solution using the programs SSP [132] (Fig. 7.5 A) and TALOS+ [133] (Fig. 7.5 B). The secondary structure elements estimated using these calculations are in good agreement with those determined by X-ray crystallography in the solid state. In addition, the regions predicted as disordered by PONDR indeed show small or negative SSP scores, suggesting the presence of mobility in these protein segments. The results of the analyses of chemical shifts described above prompted to investigate more directly the dynamics of *Sp*UreE using ^{15}N relaxation measurements.

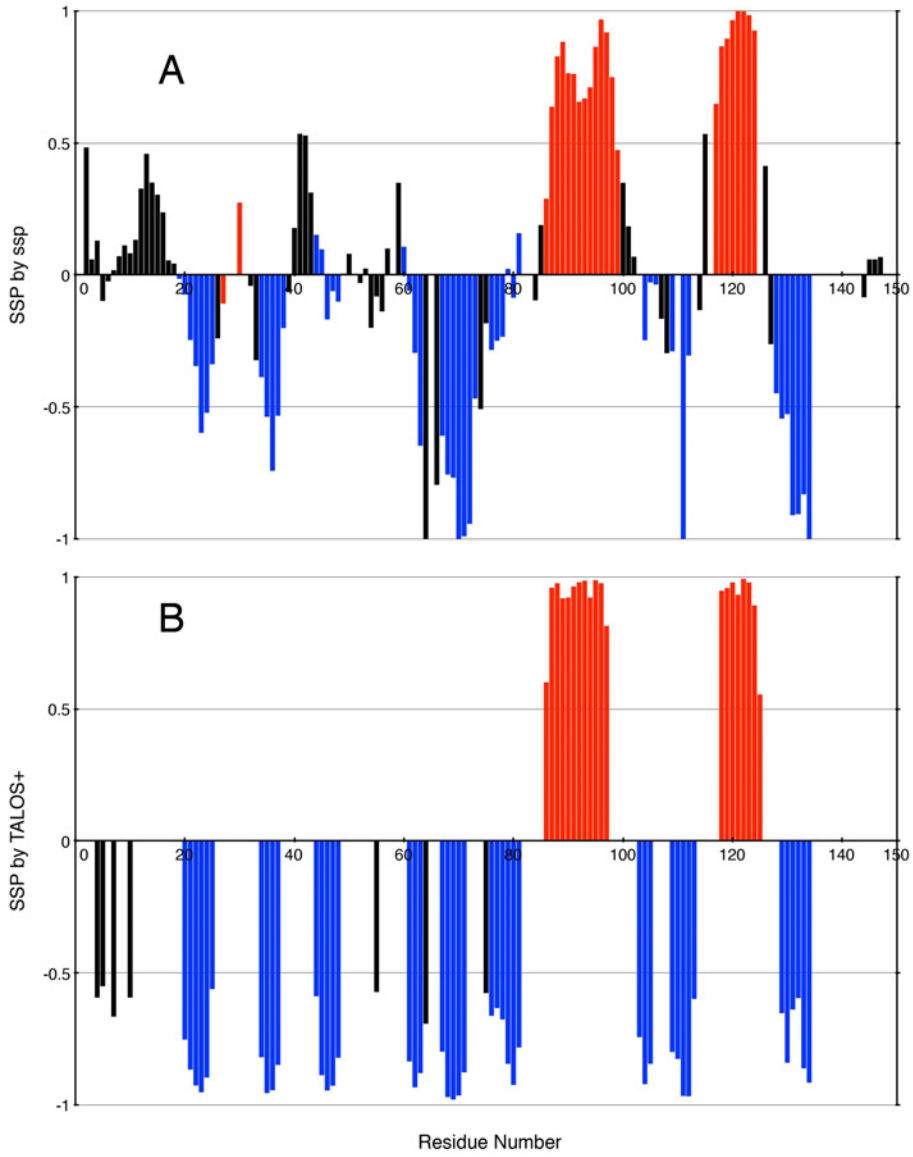


Figure 7.5: Secondary structure propensity by NMR. SSP (**A**) and TALOS+ (**B**) scores calculated using the Cα, Cβ and Hα chemical shifts for *Sp*UreE. A positive score indicates a propensity for α-structure, while a negative score indicates a propensity for β-structure or extended loops, with residues predicted in fully formed α-helices and β-strands given scores of +1 and -1, respectively. The α-helices and β-strands established by UCSF-Chimera [138] using the PDB code 1EAR are shown as *red* and *blue* bars, respectively.

NMR studies of *Sp*UreE dynamics

The ^{15}N relaxation rates R_1 and R_2, and the ^{1}H-^{15}N NOE values of 123 out of 144 (85.4%, not including M1, P83 and P139) backbone amide groups are shown in Fig. 7.6 A, 7.6 B and 7.6 C, respectively. The missing residues were either too weak to quantify accurately or were highly overlapped. The presence of local internal motions in the picosecond-nanosecond time scales contribute to the R_1, R_2 and NOE values, while conformational exchange processes occurring on the microsecond-millisecond time scale additionally contribute to increase the R_2 rates [139].

Therefore, the analysis of these parameters can provide information on local backbone mobility of *Sp*UreE at different time scales. NOEs are especially more sensitive to fast internal dynamics than R_1 and R_2 [39].

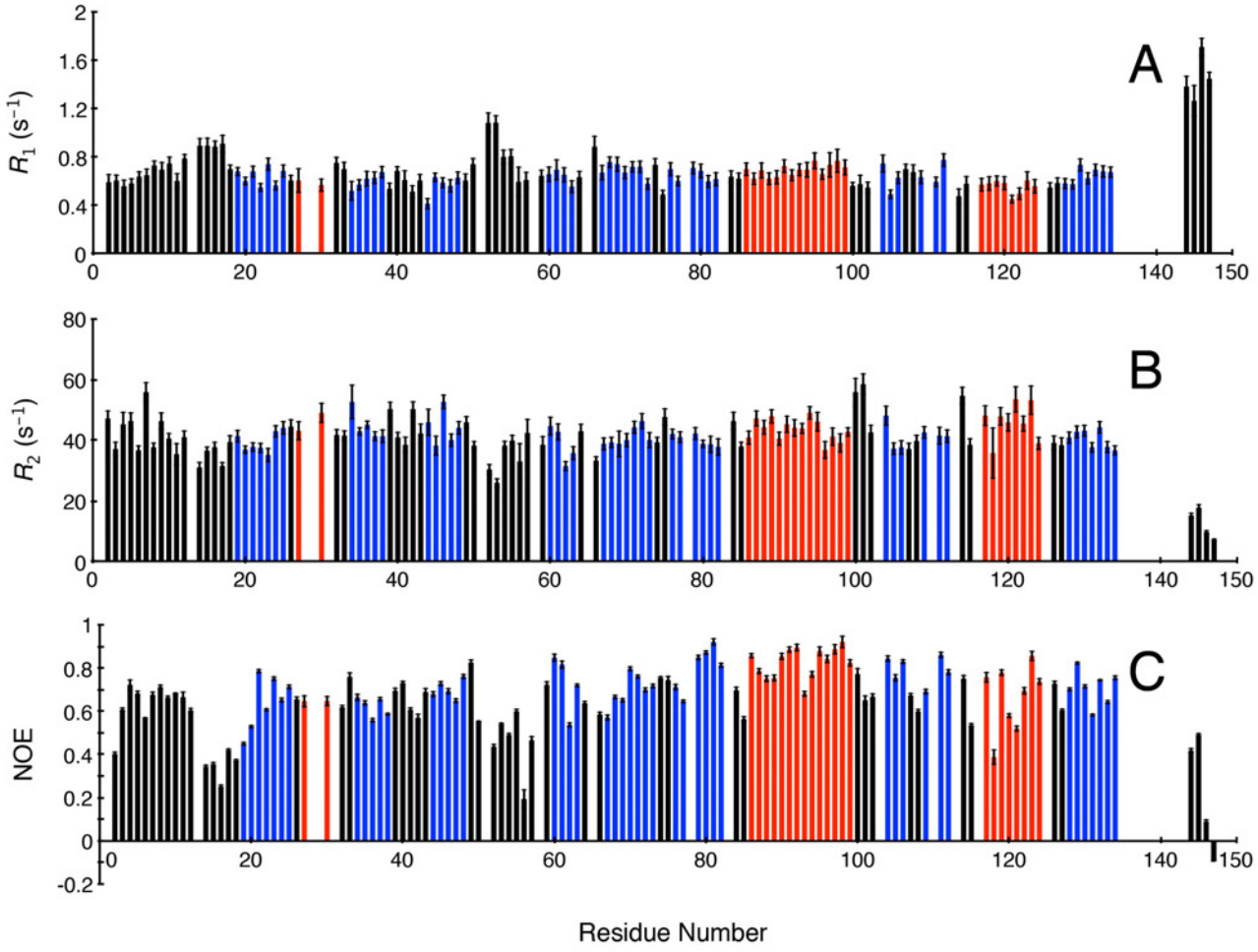


Figure 7.6: Backbone amide ^{15}N relaxation properties recorded at 25 °C and 16.4 T on $^{2}H/^{15}N$-labeled *Sp*UreE. (**A**) longitudinal R_1 relaxation rate; (**B**) transverse R_2 relaxation rates; (**C**) steady-state heteronuclear ^{1}H-^{15}N NOE. The α-helices and β-strands established by UCSF-Chimera [138] using the PDB code 1EAR are shown as *red* and *blue* bars, respectively.

A qualitative analysis of the relaxation data for *Sp*UreE suggests that the regions with defined α-helices or β-strands are characterized by large NOE values, indicative of low mobility, whereas small or negative NOE values are found on the initial extended loop, on the first β-strand at the N-terminus, and in the last four residues at the C-terminus. The values of R_1 and R_2 are relatively constant for structured regions of the protein, while the mobile portion detected in the NOE analysis at the C-terminus are characterized by larger R_1 and smaller R_2 values, suggesting the occurrence of

faster dynamics at the C-terminus. This directly explains the small or absent electron density in the solid-state structures of *Sp*UreE, an observation initially interpreted as a consequence of disorder [109], [110] and now fully supported by NMR relaxation data. A progressive increase of backbone mobility upon moving towards the C-terminus would also explain the lack of signals for residues in the region 135-143, expected to experience exchange at intermediate time scales, probably arising from conformational fluctuations, which would lead to broadening beyond detection.

The NMR relaxation data were used to quantify the ps-ns dynamics of *Sp*UreE using the model-free formalism by Lipari and Szabo [36], [46], further extended by Clore *et al*. [140] and D'Auvergne and Gooley [50]. This approach assumes independence between local internal dynamics and molecular tumbling and employs four types of parameters to describe the molecular motion: a) the overall protein rotational diffusion tensor to quantify the rate of molecular rotation about each of the three axes, b) the site-specific correlation time (τ_e), reflecting the time scale of bond vector reorientation, c) the site-specific order parameter (S^2), quantifying the amplitude of internal motions of the bond vectors [37], and d) the site-specific exchange broadening (R_{ex}) that contributes to R_2 through μs-ms chemical exchange [47], [141] (see Chapter 2).

In order to determine the diffusion tensor, a preliminary estimate of the rotational correlation time (τ_m = 22.3 ± 0.2 ns) was obtained according to $\tau_m = \frac{1}{2\omega_N}\sqrt{6\frac{R_2}{R_1} - 7}$ [40] using 49 experimental R_1 and R_2 relaxation rates, selected by excluding residues characterized by significant internal mobility as shown by their small R_2 ($R_2 < (\overline{R}_2 - \sigma)$ and $(\overline{R}_2 - R_2)/ R_2 > (\overline{R}_2 - R_1)$) and small NOE (NOE < 0.65). This value of τ_m was used to obtain a more accurate value by fitting the same 49 R_2/R_1 ratios to the general equation that correlates this ratio to the spectral densities assuming isotropic tumbling [39], an approach that yielded τ_m = 22.5 ± 2.0 ns. This value corresponds to a relative molecular mass M_r of 37.5 ± 3.3 kDa estimated using the empirical relationship τ_m (ns) ≈ 0.6 · M_r (kDa) [142], supporting the presence of the *Sp*UreE dimer (34.8 kDa) in solution under the experimental conditions used. The high-frequency corrected ^{15}N spin relaxation rates [42], [143] were then used to calculate the rotational diffusion tensor of *Sp*UreE as a nearly axial spheroid described by the components D_{xx} = 6.03 ± 0.05 x 10^6 s^{-1}, D_{yy} = 5.80 ± 0.05 x 10^6 s^{-1}, and D_{zz} = 9.67 ± 0.07 x 10^6 s^{-1} and the polar angles α = 72°, β = 33°, and γ = 47°. The resulting anisotropy of 1.64 and rhombicity of -0.093, consistent with the protein shape derived from crystallography, imposed the use of anisotropic modeling for five possible mathematical models [144] that consider the following variables as parameters for the fit:

TM1 - S^2 (in this model S^2 is the square of the generalized order parameter reflecting the amplitude of the motion of the NH bond vector relative to the rotational diffusion frame);

TM2 - S^2 and τ_e (in this model τ_e is the effective correlation time reflecting the ps-ns timescale of the motion);

TM3 - S^2 and R_{ex} (model in which R_{ex} accounts for the relaxation due to chemical exchange and is an indicator of motions in the μs-ms timescale);

TM4 - S^2, τ_e and R_{ex} (all the three parameters are considered in this model);

TM5 - S^2_f, S^2_s, and τ_s (model in which the motions are further divided in two timescales, the faster of which described by S^2_f and the slower by S^2_s and τ_s).

This approach was applied to each residue of the protein, in order to characterize the local dynamics of the NH bonds, and in each case the best model was selected using the Akaike's Information Criteria (AIC) [52] as previously described [51]. This selection criterion is based on the Occam razor principle, choosing the model with the lowest number of parameters that is consistent with the data. In particular, the model with the smallest AIC is considered as the best choice, additionally taking into consideration the relative percentage errors between experimental and back-calculated values of R_1 and R_2. As a result, the TM4 model with parameters S^2, τ_e, and R_{ex} was chosen to describe all the assigned residues with only few exceptions. As can be seen in Figure 7.7 A, almost all residues have high order parameters, indicating small amplitudes of the dynamics, while lower S^2 values are observed for the residues belonging to two loops covering the range D12-K18 and E50-R56, in addition to the C-terminal region G144-H147. For the last two residues, the best model-free model was TM5, necessary to take into account the faster local motions occurring in this region. The values of the local correlation time τ_e (Fig. 7.7 B) are of the order of 1 ns for most of the residues, with larger values detected in the helical region covering residues M86-R99, consistently with their location at the monomer-monomer interface, which may cause their dynamics to slow down. Analysis of the chemical exchange contribution R_{ex} (Fig. 7.7 C) revealed that conformational fluctuations within the micro-millisecond time range are widespread throughout the protein backbone, with higher values remarkably observed in the region involving the conserved metal binding residue H100.

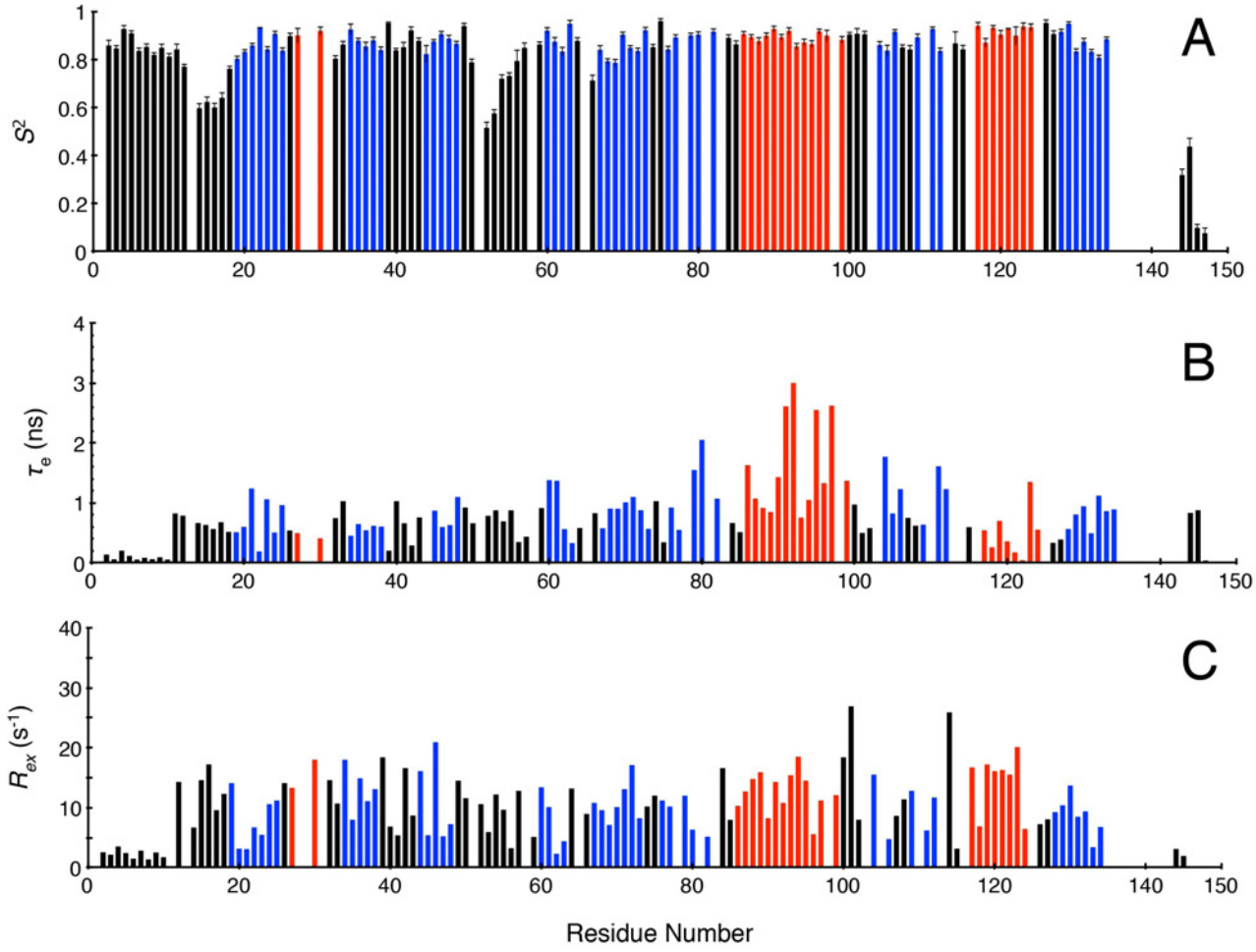


Figure 7.7: Model-free characteristics of the backbone dynamics of *Sp*UreE. (**A**) Order parameter values (S^2) of the backbone of *Sp*UreE; (**B**) local correlation time (τ_e); (**C**) conformational exchange contribution to R_2 (R_{ex}) as a function of residue number. The α-helices and β-strands established by UCSF-Chimera [138] using the PDB code 1EAR are shown as *red* and *blue* bars, respectively.

The NMR relaxation data were also analyzed using a reduced spectra density mapping approach [34], [44], [45]. This represents an alternative to the model-free approach because it does make any assumption about the nature of the correlation function that describes the overall rotational diffusion, nor about the value of ^{15}N chemical shift anisotropy. The reduced spectral density functions at three frequencies ($J(0)$, $J(\omega_N)$ and $J(0.87\omega_H)$) are shown in Figure 7.8 A, 7.8 B and 7.8 C, respectively. Their trend is as expected for a protein in which significantly higher internal mobility on the fast time scale dominates the relaxation of the C-terminal tail ($J(0)$ = 3.1 ± 1.3 ns rad^{-1}; $J(\omega_N)$ = 0.230 ± 0.025 ns rad^{-1}; $J(0.87\omega_H)$ = 0.0179 ± 0.008 ns rad^{-1}) as compared to the bulk of the protein ($J(0)$ = 11.1 ± 1.5 ns rad^{-1}; $J(\omega_N)$ = 0.112 ± 0.018 ns rad^{-1}; $J(0.87\ \omega_H)$ = 0.0034 ± 0.018 ns rad^{-1}). Moreover, the region around the conserved Ni(II) binding residue H100, which

shows high R_{ex} values indicative of enhanced mobility in the micro-millisecond time scale, consistently shows high values of $J(0)$.

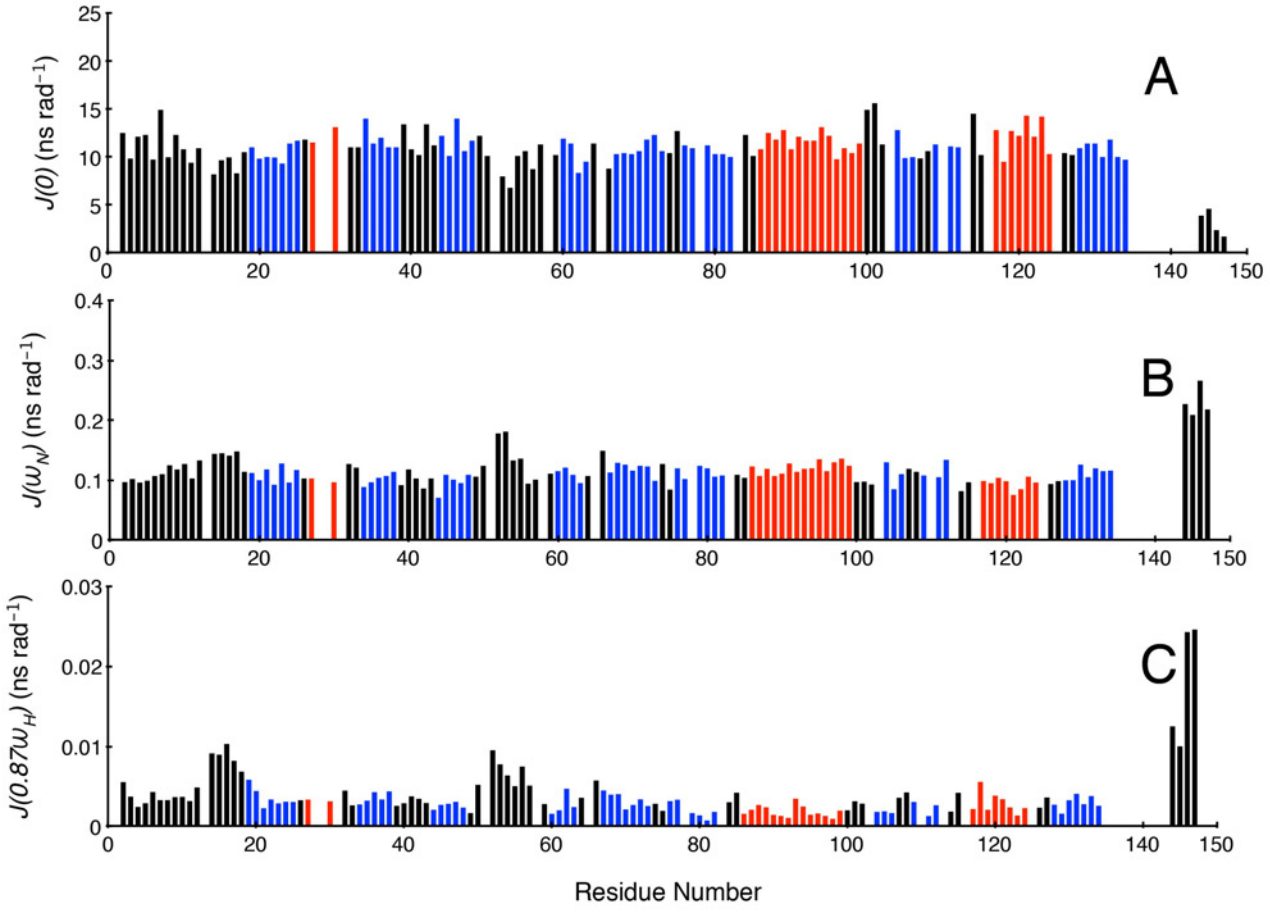


Figure 7.8: Values of $J(0)$ (**A**), $J(\omega_N)$ (**B**), and $J(0.87\omega_H)$ (**C**) resulting from a reduced spectral density analysis of the relaxation data for *Sp*UreE. The α-helices and β-strands established by UCSF-Chimera [138] using the PDB code 1EAR are shown as *red* and *blue* bars, respectively.

This quantitative analysis of the protein backbone dynamics confirms the qualitative indications that the C-terminal portion of *Sp*UreE experiences large amplitude movements in the nanosecond time scale, while the regions predicted to be involved in protein-protein interactions between *Sp*UreE and *Sp*UreG are characterized by significant motions in the micro-millisecond time scale. This is not surprising, considering that conformational *equilibria* pre-existing in the unbound state of proteins are known to sample the most suitable conformers to optimize binding to the partner and to stabilize complex formation [145], [146].

CSP analysis of *Sp*UreE-*Sp*UreG interaction

With the purpose of further supporting the predicted model of interaction between UreE and UreG with direct observations, and to map the residues involved in this interaction, an analysis of the perturbations of the amide ^{1}H and ^{15}N chemical shifts upon protein-protein complex formation was carried out. A series of ^{1}H-^{15}N TROSY HSQC experiments were therefore acquired for $^{2}H/^{15}N$ -labeled *Sp*UreE alone and in the presence of unlabeled *Sp*UreG at different stoichiometric ratios up to 1 dimer *Sp*UreE : 4 monomer *Sp*UreG. The data were analyzed with respect to changes in chemical shift (Δδ) (Figure 7.9 A). Addition of *Sp*UreG gave rise to small CSPs, with maximum observed shifts of about 0.05 ppm. Single, averaged resonances were observed, indicating fast exchange between the free and bound *Sp*UreE on the NMR time scale, consistently with the value of the dissociation constant measured by ITC [127]. The average magnitude of the observed CSPs is too small to determine the dissociation constant, whose value was anyway determined by ITC. The amino acids showing the largest perturbations of chemical shift belong to the unstructured regions L2 - S16 and G144 - H147, while some smaller changes are observed throughout the sequence. Similarly small CSPs are not unprecedented [147], [148], and have been ascribed to either weak complexes involving a well-defined patch, or to multiple binding models [127]. The surface map of the residues affected by CSP is shown in Figure 7.9 B, reveals a consistent distribution on the protein surface, a result that can be interpreted as indicating that the small CSPs observed are due to the formation of a weak complex involving a single protein-protein orientation, or multiple complexes with small structural deviations from each other. The protein patch identified by NMR as showing the largest and most consistent alterations of the backbone amide chemical shifts is consistent with the surface region on UreE suggested by molecular modeling to be involved in the formation of the UreE-UreG complex in *Helicobacter pylori* [103]. Another region of perturbed residues (E88- M102) constitutes the interface between the two monomers at the core of the *Sp*UreE dimer, and is thus hidden from direct contact with the partner protein. Its perturbation, however, could be induced by *Sp*UreG binding to the *Sp*UreE homo-dimer, which leads to a small rearrangement of the two monomers by perturbing the hydrophobic interactions between the α-helices through an allosteric mechanism. The other region shown to be perturbed by the interaction with UreG covers residues G144 - H147 and represents the disordered C-terminus of *Sp*UreE, which along with H100, have been shown to be involved in Ni(II) coordination. Their large CSP values suggest that they are also involved in the interaction of UreE with UreG.

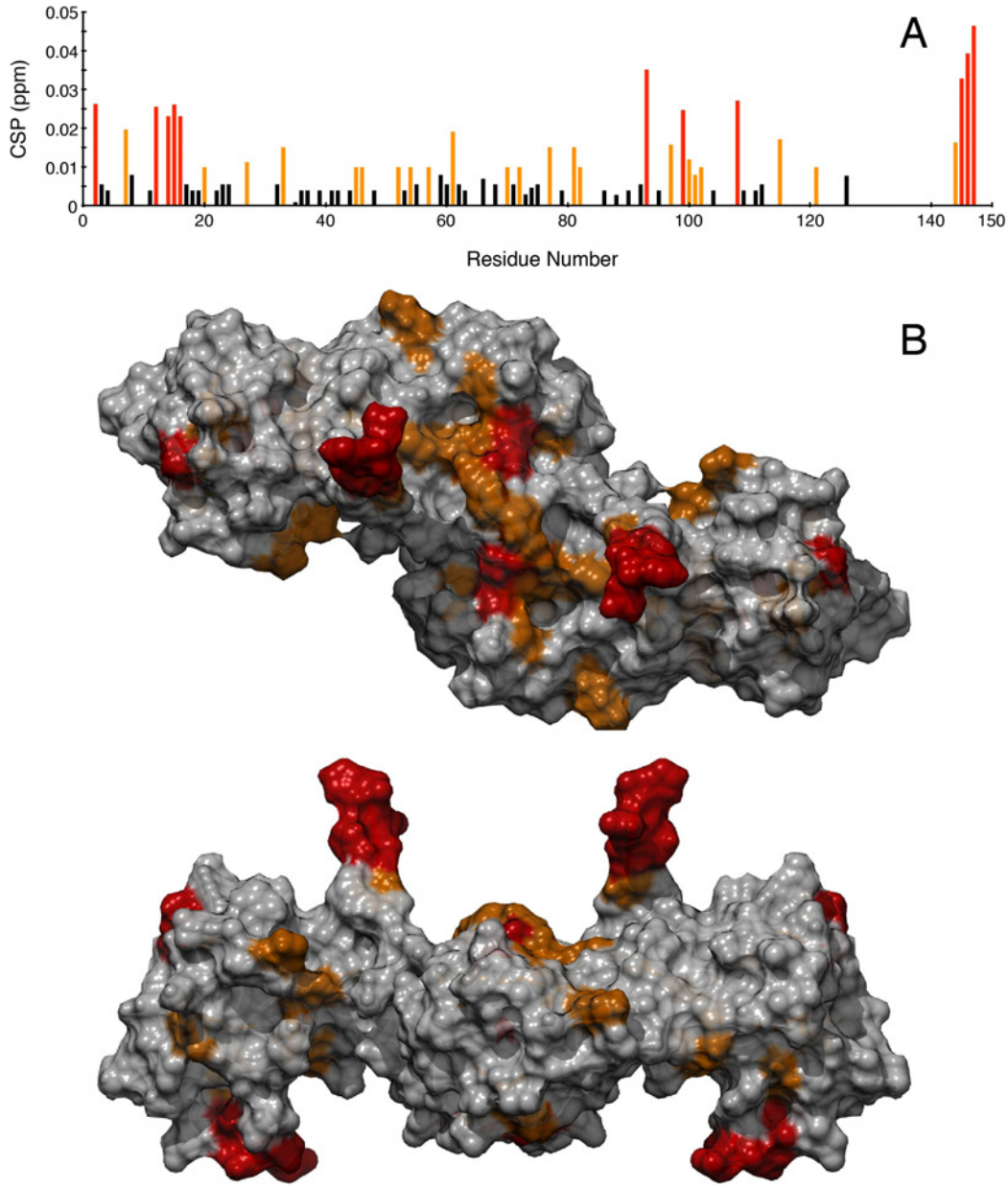


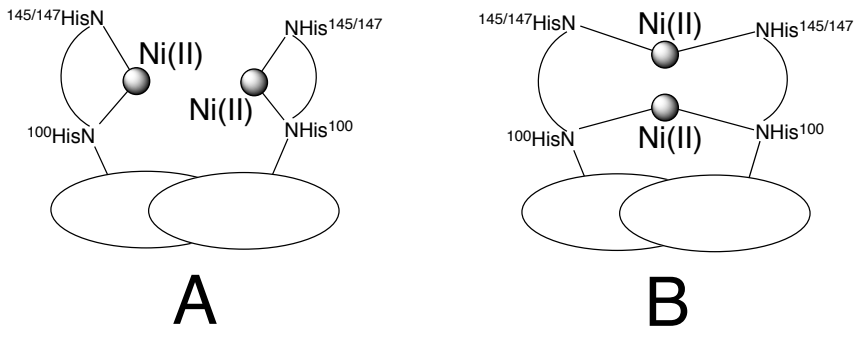
Figure 7.9: *Sp*UreE-*Sp*UreG complex interaction site identified by mean of chemical shift mapping (**A**) Chemical shift perturbations (CSP) measured for all *Sp*UreE assigned residues. (**B**) Surface representation of the structure of *Sp*UreE, highlighting regions experiencing CSP between 0.01 and 0.02 mapped in *orange* and larger than 0.02 colored in *red*, in agreement with panel **A**; top and bottom panels are rotated by 90° along the horizontal axis.

Interestingly, mutation of the C-terminal histidine (H152) in *Hp*UreE, involved in Ni(II) coordination and corresponding to either H145 or H147 in *Sp*UreE, abolished the high affinity metal binding site built by the UreE-UreG complex [103], supporting the idea that the position of the C-terminal arm of UreE is influenced by it interaction with UreG. Even though the residues preceding the last portion of the protein are not assigned because unobserved (R135 - R143), protein docking [103] predicted that these residues assist in binding UreG in the case of the proteins

from *Helicobacter pylori*. Supporting this idea, the cross-peak corresponding to R134 experiences a large intensity loss upon addition of UreG, suggesting that the whole C-terminal tail (R135 - H147) is involved in the formation of this protein-protein complex. The residues L2 - S16 and R56 - L61, also undergoing small but significant CSPs, are positioned on the opposite side of the proposed *Sp*UreE-*Sp*UreG interaction surface, which suggests a long-range allosteric structural rearrangement. If the perturbations were the result of direct interaction, a more consistent set of CSPs modifications in this region would be expected, considering the size of the UreG dimer. A larger sequence conservation is observed in the proposed interaction site as compared to this latter loop [103], further supporting the hypothesis that the UreE-UreG protein complex occurs at the K32 - R49, H100 - M102, and G144 - H147 interface.

7.2. Characterization of metal binding properties of *Sp*UreE

The nickel binding properties of *Sp*UreE, previously determined through the direct measurement of free and bound metal ion concentrations at dialysis equilibrium using inductively coupled plasma emission spectrometry, and supported by X-ray absorption spectroscopy, involve the presence of two sites per protein dimer, with the Ni(II) ions in octahedral coordination geometry and an average of two histidine residues and four O/N ligands bound to each metal ion [115]. Two possible models were proposed (Scheme 1) since, without any additional structural information, it was impossible to draw a conclusion about which of the two is correct. The analysis carried out by mean of isothermal titration calorimetry and X-ray crystallographic data had the purpose of reevaluating the thermodynamics parameters and the molecular details of Ni(II) binding to *Sp*UreE. Concomitantly, a comparison with Zn(II) binding was used to sort out the possible determinants of metal ion specificity featured by this metallo-chaperone.



Scheme 1 (formerly Scheme 4 in Chapter 4)**:** Representation of the Ni(II) binding modes by *Sp*UreE used to fit the data by equilibrium dialysis experiments [115].

7.2.1. Nickel binding by ITC

The ITC measurements were initially carried out by performing a direct titration, namely by adding Ni(II) to the apo-protein, and the occurrence of a binding event was revealed by the presence of exothermic peaks that followed each addition (Fig. 7.10 A). At the protein concentrations utilized for this type of experiment, *Sp*UreE is reported to be a dimer in the presence of Ni(II) [107], [117]. Fits of the integrated heat data (Fig. 7.10 B), attempted assuming a single set of identical sites, multiple sets of non-interacting sites, or a stepwise binding model, consistently yielded poor results characterized by an unrealistically low stoichiometry or negative binding constants.

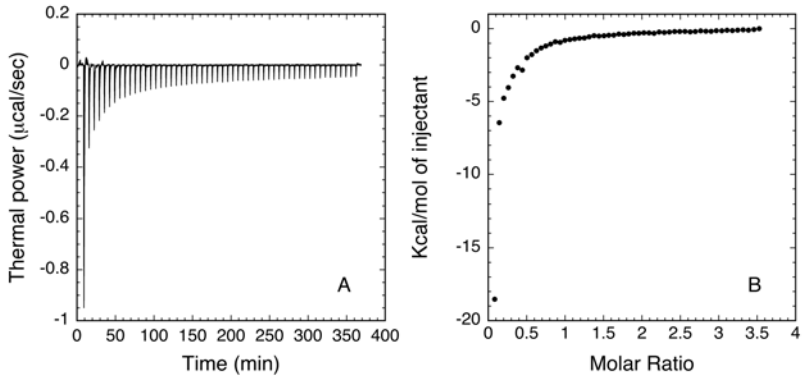


Figure 7.10: Nickel binding properties of *Sp*UreE determined using isothermal titration calorimetry. (**A**) Representative plot of raw direct titration data of Ni(II) (500 mM, in the syringe) into *Sp*UreE (15 mM, in the sample cell). (**B**) Integrated heat data as a function of metal/protein molar ratio.

Considering the previously determined 2:1 stoichiometry of the Ni(II):dimer for *Sp*UreE, together with the observation of a rather featureless binding curve, as well as saturation of binding that occurs at very low metal binding stoichiometry, the presence of positive cooperativity was suspected. The positive cooperativity observed here could be resolved using a reverse titration method [55], [56]. Figure 7.11 A shows a representative plot of raw titration data reporting the thermal effect of injecting *Sp*UreE onto a solution of $NiSO_4$, while Table 2 summarizes the results of the analysis. The integrated heat data, shown in Figure 7.11 B, were initially fitted using a model involving a single set of identical and independent Ni(II) binding sites. This fit yielded $\chi_v^2 = 6.9 \times 10^4$, a stoichiometry $n = 2.1 \pm 0.1$ Ni(II) per dimer, $K_b = 1.8 \pm 0.2 \times 10^7$ M^{-1} ($K_d = 56 \pm 6$ nM), $\Delta H = -6.0 \pm 0.1$ kcal mol^{-1} and $\Delta S = +13.0$ cal mol^{-1} K^{-1}. The fit using an alternative model involving two independent and different binding sites produced a similar value for $\chi_v^2 = 7.8 \times 10^4$, but yielded two very similar binding constants ($1.4 \pm 15 \times 10^7$ M^{-1} and $2.6 \pm 60 \times 10^7$ M^{-1}) with unacceptably large errors in the stoichiometry (1.1 ± 1.2 and 1.0 ± 1.2) and thermodynamic parameters, indicating an

over-parameterization of the model, and was thus discarded. A third model, involving two sequential steps of single ion binding (n = 1) was also considered, which resulted in a χ_v^2 = 6.4 x 10^4, K_{b1} = 6.0 ± 2.6 x 10^5 M^{-1} (K_d = 1.7 ± 0.7 μM), ΔH_1 = +3.2 ± 3.1 kcal mol^{-1} and ΔS_1 = +37.3 cal mol^{-1} K^{-1}, K_{b2} = 2.3 ± 0.9 x 10^7 M^{-1} (K_d = 43 ± 17 nM), ΔH_2 = -15.0 ± 0.3 kcal mol^{-1} and ΔS_2 = -16.7 cal mol^{-1} K^{-1}. This model implies large positive cooperativity, with the second binding constant increasing ca. 40-fold as compared with the first binding constant.

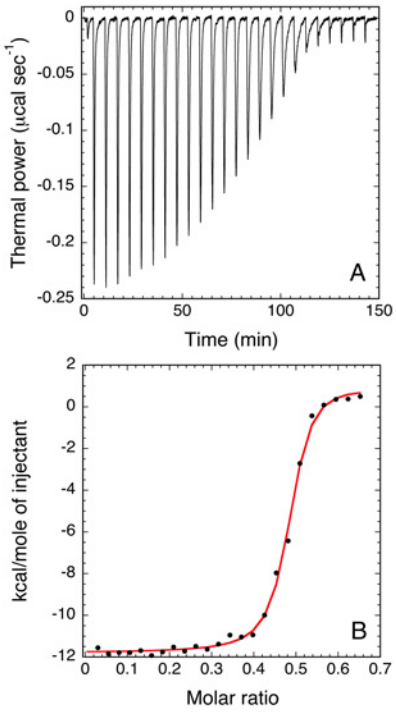


Figure 7.11: Nickel binding properties of *Sp*UreE by ITC (**A**) Representative plot of raw inverse titration data of Ni(II) (20 mM) with *Sp*UreE (70 mM). (**B**) Integrated heat data as a function of metal/protein molar ratio. The continuous line represents the best fit obtained using a model involving sequential binding of two Ni(II) ions.

These parameters are characterized by relatively larger errors in the thermodynamics of the first binding site, but feature similar statistics as compared to the first model. In addition, the presence of such a large positive cooperativity would explain the peculiar shape and the difficult interpretation of the direct titration experiment, making this latter fitting model more consistent, overall, with the data obtained. The difficulty of analyzing the data using direct calorimetric titrations can explain the different conclusions that can be drawn here as compared to those obtained using the same direct titrations with dialysis and metal analysis, which were interpreted using negative cooperativity [115]. In the case of similar calorimetric titrations carried out on *Hp*UreE, a single binding site for Ni(II) was observed, with K_d in the micromolar range [103]. A notable difference between *Hp*UreE and *Sp*UreE, which could explain the different stoichiometry and thermodynamics of Ni(II) binding, is the presence of a single histidine residue in the C-terminal arm of *Hp*UreE (H152), while

two histidines are found in *Sp*UreE (H145 and H147) in addition to the conserved H102 (for *Hp*UreE) and H100 (in *Sp*UreE) [114]: in the case of *Hp*UreE, Ni(II) binds to a pair of H102 and a pair of H152, as observed by crystallography and X-ray absorption spectroscopy (Scheme 3 B in Chapter 4) [113]. In the case of *Sp*UreE, the first micromolar binding event could involve H145 and H147 (low affinity, LA site) causing a reorganization of the disordered C-terminal arms of the protein, moving the two H145 residues on the two arms closer to each other and to the two H100 residues, bringing about a nanomolar binding event involving H100 and H145 (high affinity, HA site). In this way, the two distal histidine residues would act as a molecular funnel to collect and direct Ni(II) ions onto the putatively functional binding HA site involving the pair of H100. Alternatively, the low affinity event could involve H100 and H145, bringing closer to each other the two H147 residues and causing the formation of the second, high-affinity binding site. To resolve this issue further titrations of apo-*Sp*UreE and Ni-loaded *Sp*UreE with Zn(II), were performed as described below.

7.2.2. Zinc binding by ITC

In order to compare the different affinity of *Sp*UreE for Ni(II) and Zn(II), the thermodynamics of zinc binding were also investigated using the same methodology. Similarly to what observed in the case of the titration with Ni(II), the direct titration of Zn(II) onto the apo-protein showed the occurrence of a binding event, revealed by exothermic peaks that followed each injection (Fig. 7.12 A), but the integrated heat data (Fig. 7.12 B) could not be reliably fit using any of the binding models, consistently yielding poor results characterized by an unrealistically low stoichiometry.

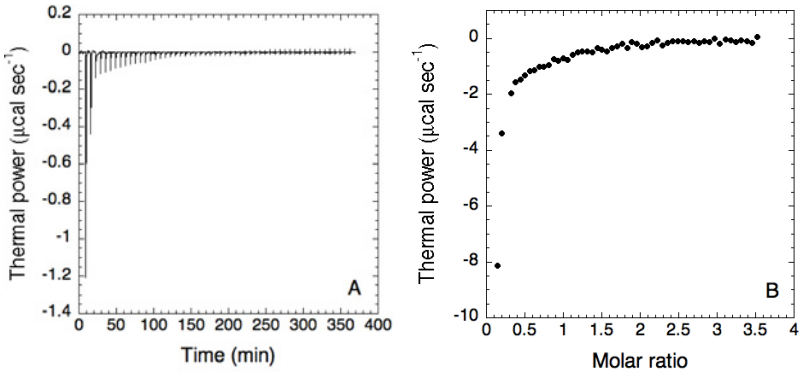


Figure 7.12: Zinc binding properties of *Sp*UreE determined using isothermal titration calorimetry. (**A**) Representative plot of raw direct titration data of Zn(II) (500 mM, in the syringe) into *Sp*UreE (15 mM, in the sample cell). (**B**) Integrated heat data as a function of metal/protein molar ratio.

Therefore, as in the case of nickel binding, the reverse titration approach was carried out, with data shown in Figure 7.13 A. The occurrence of an exothermic effect following each addition of *Sp*UreE to the metal solution was observed, with the resulting integrated heat data (Fig. 7.13 B) indicating the presence of at least two different events of binding and thus ruling out the single set of sites model.

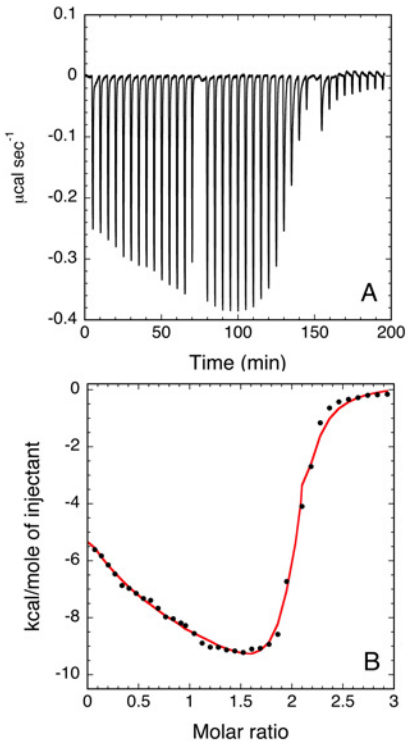
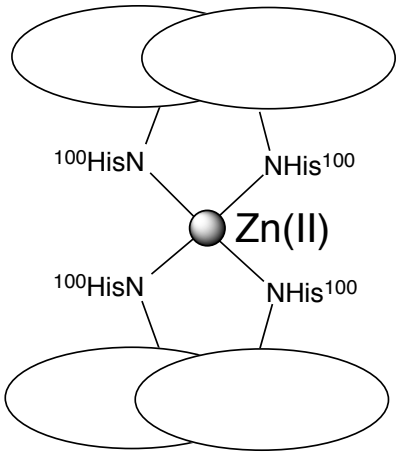


Figure 7.13: Zinc binding properties of *Sp*UreE determined using isothermal titration calorimetry. (**A**) Representative plot of raw inverse titration data of Zn(II) (20 mM) with *Sp*UreE (80 mM). (**B**) Integrated heat data as a function of metal/protein molar ratio. The continuous line represents the best fit obtained using a model involving two independent sets of binding sites for Zn(II).

Modeling the curve using a sequential set of two binding events with a single Zn(II) ion bound to the *Sp*UreE dimer in each event, did not lead to an acceptable fit. Attempts to fit the data using two sets of sites with different stoichiometry of Zn(II) binding to *Sp*UreE yielded the following results, characterized by large uncertainties in the fitted parameters: $\chi_v^2 = 7.9 \times 10^4$, $n_1 = 0.6 \pm 0.3$, $K_{b1} = 5.8 \pm 6.2 \times 10^4$ M^{-1} ($K_d = 17 \pm 18$ µM), $\Delta H_1 = +10.2 \pm 10.1$ kcal mol^{-1} and $\Delta S_1 = +55.9$ cal $mol^{-1}K^{-1}$; $n_2 = 0.54 \pm 0.01$ $K_{b2} = 3.9 \pm 2.9 \times 10^7$ M^{-1} ($K_d = 26 \pm 19$ nM), $\Delta H_2 = -17.3 \pm 0.1$ kcal mol^{-1} and $\Delta S_2 = -23.4$ cal mol^{-1} K^{-1}. Regardless of the precision of the fit, the two binding events observed for Zn(II) show a half-integer stoichiometry, suggesting that, in the experimental conditions used, the dimeric *Sp*UreE forms a dimer of dimers in the presence of Zn(II). Therefore, the data were fitted using a model that considered the sequential binding of two *Sp*UreE dimers to a single Zn(II) ion. The obtained fit (Fig. 7.13 B) resulted in improved statistics and smaller errors on the thermodynamics parameters ($\chi_v^2 = 4.5 \times 10^4$, $K_{b1} = 1.1 \pm 0.2 \times 10^7$ M^{-1} ($K_{d1} = 91 \pm 16$ nM), $\Delta H_1 =$ -

5.4 ± 0.2 kcal mol^{-1} and ΔS_1 = +14.3 cal mol^{-1} K^{-1}; K_{b2} = 4.6 ± 0.4 x 10^6 M^{-1} (K_{d2} = 217 ± 19 nM), ΔH_2 = -12.4 ± 0.2 kcal mol^{-1} and ΔS_2 = -11.0 cal mol^{-1} K^{-1}). This model (Scheme 2) is consistent with the crystallographic observation of a dimer-of-dimers of *Sp*UreE bridged by a single Zn(II) ion bound to four H100 residues and two water molecules [109], further supporting its reliability.

The slightly negative cooperativity deduced from the fit could be explained by steric hindrance between the two protein dimers bound to the same Zn(II) ion, as also revealed by the crystal structure [109]. This observation marks a distinctly different behaviour of *Sp*UreE in the presence of Ni(II) and Zn(II), with the two metal ions possibly modulating in different ways the chaperone function of the protein.



Scheme 2: Representation of the dimers of *Sp*UreE binding Zn(II) ion, as derived from ITC [110] and crystallographic studies [109].

7.2.3. Metal binding by X-ray crystallography

In order to provide structural support to the calorimetric metal binding data, a concomitant crystallographic analysis was carried out by collaborators [110] that used samples of *Sp*UreE purified as described in Chapter 6. In this case, *Sp*UreE was eluted from a Superdex-75 size-exclusion column using 20 mM TRIS-HCl buffer, 150 mM NaCl, at pH 8.0, and concentrated to 5.3 mg/mL. The co-crystallization of *Sp*UreE was carried out in the presence of both Ni(II) and Zn(II) added in equimolar amounts. This experiment allowed to obtain diffracting crystals, which were used to provide structural support to the models drawn on the basis of the calorimetric metal competition data. Crystallographic data analysis resulted in the modeling of two protein chains in the asymmetric unit. The overall polypeptide fold is similar to that previously reported [109]: each

monomer is made of two domains, N- and C-terminal, connected by a flexible linker (K74-K77), with the N-terminal domain constituted by two three-stranded mixed β-sheets that stack upon each other in a nearly perpendicular orientation, and the C-terminal domain made of a four-stranded anti-parallel β-sheet and two α-helices organized in a ferredoxin-like βαββαβ-fold (Fig. 7.14 A). The C-terminal domains are responsible for head-to-head dimerization contacts stabilized by hydrophobic interactions formed between α-helices. However, in contrast to the previously observed dimer-of-dimer arrangement in the crystal [109], the protein chains do not form a dimer of dimers in the crystal lattice, but simply dimers arranged around the 6_3 axis, forming a large solvent channel (Fig. 7.14 B).

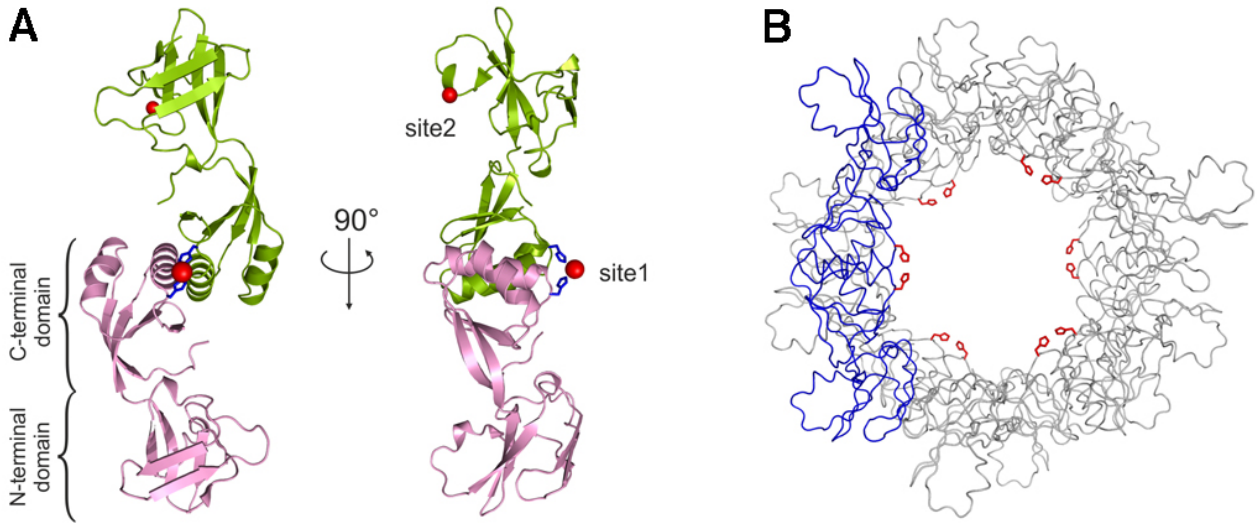


Figure 7.14: (**A**) Ribbon representation of crystallographic structural model of *Sp*UreE dimer shown in two perspectives (PDB code 4L3K), polypeptide chain A and B are shown in *green* and *pink* respectively, metal ions shown as a *red* ball, two H100 shown as *blue* sticks; (**B**) Packing of *Sp*UreE dimers in the crystal lattice, pairs of H100 shown in *red* [110].

Two metal binding sites were detected in the structure (Fig. 7.14 A). One is located on the surface of the protein, in a central position between the two protomers of the dimer (site 1), on the peripheral surface of the *Sp*UreE dimer. The coordination environment of the bound metal ion is consistent with octahedral geometry, but only two of its ligands can be identified unambiguously. These are the two conserved and adjacent H100 Nε, one from each chain, which form an angle of approximately 90º with the metal ion. Each H100 is stabilized by an H-bond formed between H100 Nδ and N98 Oδ from the neighboring chain, stabilizing the homodimer (Fig. 7.15). The electron density for the residues Y142-R143-G144 is weak, because of disorder, in both chains A and B, and

only the main chain can be tentatively traced. Residual electron density is most likely to correspond to the two H145 residues, one from each chains of the monomer, bound to the metal ion (Fig. 7.15).

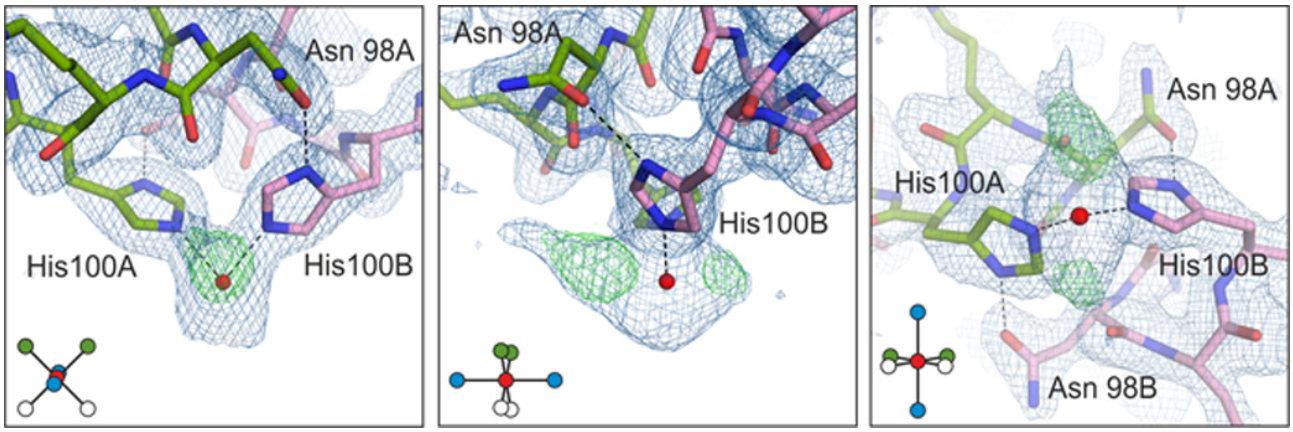


Figure 7.15: Details of the nickel-binding site in the center of *Sp*UreE dimer (site 1). Three views are presented with insets in the lower left corner of each figure, illustrating the indicated octahedral geometry around the metal ion: the latter is represented by a *red* sphere, the two H100 by *green* spheres and the presumed positions of the two H145 as *blue* spheres; *white* spheres illustrate missing interactions in octahedral geometry. $2F_o$-F_c electron density map is contoured at the 1σ level (*blue*), while the F_o-F_c electron density map is contoured at 3σ (*green*) [110].

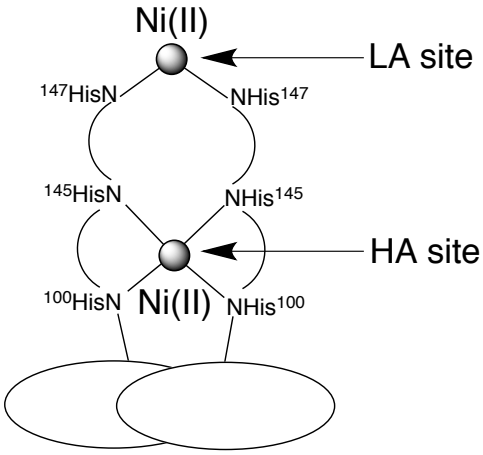
This is consistent with what was observed in the case of *Hp*UreE, in which a pair of histidine residues (H152) from the two C-terminal portions of the protein were found coordinated to Ni(II) in addition to H102, corresponding to H100 in *Sp*UreE [113]. The two remaining ligands required to complete the octahedral geometry, on the side exposed to the wide solvent channel in the crystal lattice, are not visible in the electron density. These ligands were suggested to be water molecules in the case of *Hp*UreE on the basis of X-ray absorption spectroscopic data [113]. The last two residues of the polypeptide chain, Q146-H147, were not observed, due to disorder. This is consistent with the observation that, also in *Hp*UreE, the last 18 residues that follow the Ni(II)-bound H152 are disordered [113]. The results obtained using isothermal titration calorimetry in solution, indicate that H147 could be involved in metal binding of a second Ni(II), which however is not visible in the crystal structure. An explanation for this could be related to the lower affinity of this distal site for Ni(II). Alternatively, the protein disorder involving the protein C-terminal arm observed in the crystalline state could prevent the observation of the H147 electron density, thus precluding a metal ion bound to this residue to be visible in the diffraction map.

The second metal ion (site 2) was found in the N-terminal domain, linking symmetry-related dimers, and coordinated with a pseudo-tetrahedral geometry. Anyway, it is unlikely that the *Sp*UreE molecules in diluted solution form such stable contacts between two dimers, and therefore this

interaction is likely to be an artifact of crystallization [110]. Moreover, calorimetric measurements indicate that, in the case of Ni(II), a 1+1 stoichiometry per protein dimer is observed, while Ni(II) binding to sites 2 in the dimeric protein, in addition to site 1, would create a 1+2 stoichiometry. In the case of Zn(II), the 2:1 stoichiometry between the *Sp*UreE dimer and the metal ion determined by calorimetry is not compatible with Zn(II) binding to site 2 in the tetrameric quaternary structure observed in crystallization experiments.

7.2.4. Ni(II) and Zn(II) competition for *Sp*UreE

A comparison of the occupancy of the two metal binding sites in the presence of an equimolar Ni(II)/Zn(II) crystals mixture was used to assess their relative affinities for the two metal ions. What has emerged from this analysis is that the conserved binding site in *Sp*UreE is highly selective for Ni(II) and the other site is not selective, with Ni(II) and Zn(II) binding with similar affinity. The coordination spheres around the two sites also provide some evidence as to the identity of the bound metal ion. Site 2 is clearly tetrahedral, which is the preferred coordination of Zn(II), while the coordination sphere around site 1, although partly disordered, is indicative of an octahedral coordination, preferred by Ni(II). As described above, calorimetric experiments show that a dimer of *Sp*UreE binds Ni(II) with a 1+1 stoichiometry and positive cooperativity, in a low affinity (LA) site and a high affinity (HA) site that involve either the H100-H145 or the H145-H147 pair. In particular, the first binding event has K_{d1} = 1.7 μM (LA site) and the second event K_{d2} = 43 nM (HA site). The same approach was used to establish that two dimers of *Sp*UreE bind Zn(II) with K_d = 91 nM, most likely in the crystallographically established site that comprises H100 [109]. These values suggest that the HA site for Ni(II) is the crystallographically observed site 1, involving H100 and H145, consistently with the anomalous scattering analysis that indicates that Ni(II) binds to site 1 better than Zn(II). Consequently, the Ni(II) binding event at the LA site involves H145 and H147, confirming a role for storage and molecular funnel for the C-terminal portions of UreE proteins [114] (Scheme 3).



Scheme 3: Representation of Ni(II) binding to the high affinity (HA) and low affinity (LA) sites of the *Sp*UreE dimer, as derived from ITC and crystallographic studies.

In order to confirm the metal selectivity of the HA site in solution, we titrated Ni(II) onto a solution containing *Sp*UreE pre-incubated with one equivalent of Zn(II) per protein dimer and vice versa, and monitored the reaction using ITC. The absence of significant heat of reaction in the Zn(II) titration onto Ni(II)-bound *Sp*UreE experiment (Fig. 7.16) confirmed that the Ni(II) bound to the HA site of *Sp*UreE prevents Zn(II) from binding to the protein in the same site, highlighting the metal ion specificity of the HA metal site for Ni(II) as compared to Zn(II). Moreover, this observation demonstrates that, in the absence of Ni(II), Zn(II) binding occurs in the HA site.

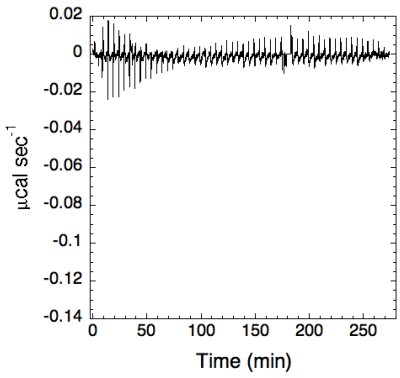


Figure 7.16: Representative plot of raw direct titration data of Zn(II) (500 mM, in the syringe) into *Sp*UreE (19 mM, in the sample cell) preincubated with an equimolar amount of Ni(II).

On the other hand, when Ni(II) is added to the *Sp*UreE containing Zn(II) in the HA site, exothermic peaks are observed following each injection (Fig. 7.17 A), indicating that Ni(II) can outcompete

Zn(II) for binding to the protein in the HA site. Considering the sequential mode of binding Ni(II) to apo-*Sp*UreE, established by ITC, we used the same model to fit the data in the competition experiment. This model produced a statistically consistent fit: $\chi_v^2 = 3.4 \times 10^3$, $K_{b1} = 2.4 \pm 0.3 \times 10^5$ M^{-1} ($K_{d1} = 4.2 \pm 0.5$ μM), $\Delta H_1 = -2.9 \pm 0.1$ kcal mol^{-1} and $\Delta S_1 = +15.0$ cal mol^{-1} K^{-1}, $K_{b2} = 1.1 \pm 0.3 \times 10^3$ M^{-1} ($K_{d2} = 0.9 \pm 0.2$ mM), $\Delta H_2 = -15.8 \pm 0.4$ kcal mol^{-1} and $\Delta S_2 = -39.3$ cal mol^{-1} K^{-1}. This fit features a first event with a binding constant similar to the low affinity event observed upon titrating Ni(II) on apo-*Sp*UreE, suggesting that, in the presence of Zn(II) in the HA site, Ni(II) binds to the LA site. The second Ni(II) binding event presents a significantly lower binding constant as compared to the apo-protein, and a consequent absence of positive cooperativity, further supporting the hypothesis that Zn(II) binds to the physiological HA site, but it is outcompeted by Ni(II).

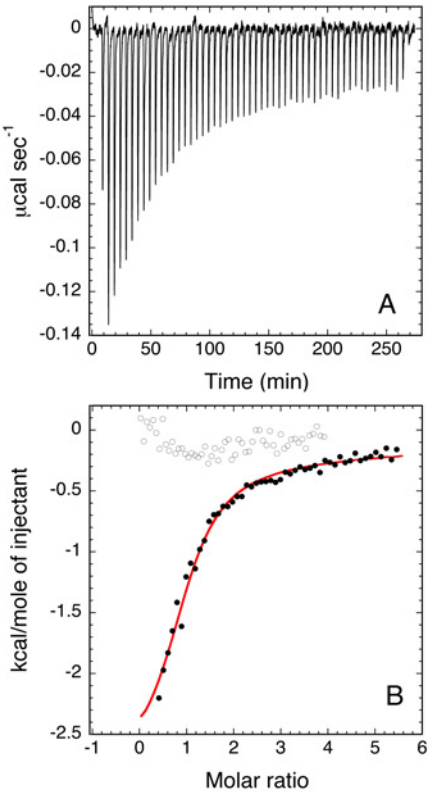


Figure 7.17: Ni(II) titration onto Zn(II)-bound binding *Sp*UreE determined using isothermal titration calorimetry. (**A**) Representative plot of raw direct titration data of Ni(II) (0.5 mM) onto *Sp*UreE (19 mM) incubated with one equivalent of Zn(II). (**B**) Integrated heat data as a function of metal/protein molar ratio. Filled circles represent the integrated titration data of Ni(II) onto Zn(II)-bound *Sp*UreE. The continuous line represents the best fit obtained using a model involving sequential binding of two Ni(II) ions. The hollow circles represent the integrated data of Zn(II) titrated onto Ni-bound *Sp*UreE, whose raw data is shown in Figure 7.16.

Table 2: Results of the fitting analysis of ITC experiments on Ni(II) and Zn(II) binding to *Sp*UreE.

Ni(II) binding to homodimeric apo-*Sp*UreE

Binding scheme	χ_v^2	*n*	k_b (M^{-1})	ΔH (kcal mol^{-1})	ΔS (cal mol^{-1} K^{-1})
One set of sites	6.9 x 10^4	2.1 ± 0.1	(1.8 ± 0.2) x 10^7	-6.0 ± 0.1	+13
Two sets of Ni(II) sites	7.8 x 10^4	n_1 = 1.1 ± 1.2	(1.4 ± 15) x 10^7	-23 ± 8,400	-45.1
		n_2 = 1.0 ± 1.2	(2.6 ± 60) x 10^7	15 ± 9,800	+82.4
Sequential binding of two Ni(II)	6.4 x 10^4	n_1 = 1.0	(6.0 ± 2.6) x 10^5	+3.2 ± 3.1	+37.3
		n_2 = 1.0	(2.3 ± 0.9) x 10^7	-15.0 ± 0.3	-16.7

Zn(II) binding to homodimeric apo-*Sp*UreE

Binding scheme	χ_v^2	*n*	k_b (M^{-1})	ΔH (kcal mol^{-1})	ΔS (cal mol^{-1} K^{-1})
One set of sites	n.a.*	n.a.	n.a.	n.a.	n.a.
Two sets of Zn(II) sites	7.9 x 10^4	n_1 = 0.6 ± 0.3	(5.8 ± 6.2) x 10^4	+10.2 ± 10.1	+55.9
		n_2 = 0.54 ± 0.01	(3.9 ± 2.9) x 10^7	-17.3 ± 0.1	-23.4
Sequential binding of two Zn(II)	n.a.	n.a.	n.a.	n.a.	n.a.
Sequential binding of two *Sp*UreE dimers to one Zn(II)	4.5 x 10^4	n_1 = 1.0	(1.1 ± 0.2) x 10^7	-5.4 ± 0.2	+14.3
		n_2 = 1.0	(4.6 ± 0.4) x 10^6	-12.4 ± 0.2	-11.0

Ni(II) binding to Zn(II)-*Sp*UreE

Binding scheme	χ_v^2	*n*	k_b (M^{-1})	ΔH (kcal mol^{-1})	ΔS (cal mol^{-1} K^{-1})
Sequential binding of two Ni(II) to one Zn-*Sp*UreE dimer	3.4 x 10^3	n_1 = 1.0	(2.4 ± 0.3) x 10^5	-2.9 ± 0.1	+15.0
		n_2 = 1.0	(1.1± 0.3) x 10^3	-15.8 ± 0.4	-39.3

* Not acceptable statistics

8. Conclusions

8. Conclusions

The present study is focused on urease system and in particular on the required protein-protein interaction between UreE and UreG accessory proteins, which is responsible in Ni(II) trafficking and delivery into the correct position in the enzyme active site. The structural basis of this interaction, analyzed at the molecular level, provided a contribution for the overall description of the protein interaction network essential for urease activation. The results presented here are consistent with two monomers of *Sp*UreG cooperatively binding at the *Sp*UreE interface forming a hetero-dimer of dimers $(UreE)_2$-$(UreG)_2$, previously observed for the same proteins from *Helicobacter pylori* [103]. The regions involved in protein-protein interaction were mapped by chemical shift perturbation (CSP) analysis, and correspond to those calculated for the *Hp*UreE-*Hp*UreG interaction. NMR analysis of the mobility of *Sp*UreE, as well as disorder prediction studies, reveal that the regions involved in this interaction experience significant mobility in the free state of the protein. Protein flexibility has been proposed to be important also for the regions involved in the interaction of the counterpart UreG, on the basis of molecular dynamics calculations [134], confirming a role of disorder for molecular recognition in the urease interaction network. The NMR dynamic analysis of *Sp*UreE demonstrates that the C-terminal portion, containing two histidine residues (H145 and H147) responsible together with H100 for metal ion binding, is the most mobile region of the protein. The residues belonging to this part of *Sp*UreE feature a large CSP upon addition of *Sp*UreG, showing that their chemical environment is directly affected by protein-protein interaction. This observation, together with the observation that the *Sp*UreE surface involved in interaction contains H100, responsible for metal binding, strongly suggests that the interaction between the two proteins directly dictates the metal binding properties of the complex and is thus essential for metal ion delivery. This is consistent with the previous observation that a new metal binding site is formed at the dimer interface upon complex formation for the *Helicobacter pylori* orthologues [103]. The present study implies therefore that the UreE-UreG interaction is important not only for facilitating GTP hydrolysis by UreG, as previously observed [120], but also for directly aiding metal ion delivery by UreE into the urease activation complex.

To completely characterize UreE, the nickel chaperone of urease system, the structural basis of metal ion selection and the preference of Ni(II) over Zn(II) in *Sp*UreE have been investigated, using an approach that coupled structural determination in the crystalline state by crystallography, and metal binding characterization and competition in solution by calorimetry. The obtained data are

consistent with the presence of a high affinity Ni(II) binding site at the dimer interface involving the two conserved H100 and the C-terminal H145 residues, in addition to a low affinity site found in the C-terminal disordered arms. The latter site, comprising H145 and H147 which serve as nickel storage, apt to funnel the Ni(II) ions to the physiological site for metal transport. The calorimetric data, showing a strong positive cooperativity between the two metal binding events, allowed to discriminate between two models, previously proposed on the basis of X-ray absorption spectroscopy experiments [115]. The crystal structure of the protein obtained in the presence of equimolar amounts of both metal ions indicates that the high affinity metal binding site preferentially binds Ni(II) over Zn(II). The selectivity of this metal binding site for Ni(II) over Zn(II) has been confirmed in solution by competition experiments using calorimetry. The selective affinity of UreE dimer for Ni(II) is consistent with its proposed role in transporting Ni(II) cations.

9. Supplementary information

Table 1 SI - *Chemical shifts* of all assigned nuclei of $^{2}H/^{15}N/^{13}C$ *Sp*UreE.

Residue	*Chemical shifts* (ppm)				
	^{15}N	^{1}HN	$^{13}C\alpha$	$^{13}C\beta$	^{13}CO
1 MET					
2 LEU	126,34	8,70	60,88	31,95	175,07
3 ILE	129,50	9,06	57,13	37,38	175,23
4 THR	115,78	9,02	60,46	69,69	
5 LYS	120,40	7,39	54,72	34,14	173,79
6 ILE	121,54	8,48	60,09	36,92	
7 VAL	120,72	8,29	62,38	31,49	176,17
8 GLY	107,10	7,44	45,03		170,30
9 HIS	123,52	9,01	56,68	32,96	
10 ILE	118,71	8,60	55,25	39,67	
11 ASP	123,13	7,95	56,25	40,89	177,43
12 ASP	118,47	8,37	54,87	40,36	176,33
13 TYR	120,50	8,03	57,89	37,45	
14 GLU	122,89	7,73	57,90	28,84	177,00
15 SER	114,26	8,41	57,70	62,67	174,13
16 SER	116,65	7,80	57,87	63,73	174,84
17 ASP	124,46	8,47	53,69	40,10	176,02
18 LYS	120,42	7,83	56,35	32,15	175,67
19 LYS	122,91	7,83	55,24	31,69	175,05
20 VAL	123,72	8,14	60,94	31,42	175,09
21 ASP	130,73	8,89	51,42	40,89	174,68
22 TRP	123,15	7,89	56,87	29,44	175,28
23 LEU	125,32	9,36	52,98	42,15	174,30
24 GLU	126,78	9,09	54,79	29,97	176,02
25 VAL	117,27	8,41	57,83	32,22	175,38
26 GLU	118,37	8,84	54,31	28,97	178,85
27 TRP	125,88	10,0	58,99	28,78	179,16
28 GLU					
29 ASP					
30 LEU	115,41	7,33	55,63	39,90	177,10
31 ASN					
32 LYS	119,46	7,45	56,31	33,67	175,38
33 ARG	116,20	8,45	55,87	30,83	175,05
34 ILE	117,97	7,01	58,64	39,70	175,07
35 LEU	126,10	8,24	54,54	44,93	173,82
36 ARG	126,17	8,47	54,50	30,83	174,45
37 LYS	127,83	8,88	54,88	35,73	172,83
38 GLU	119,64	7,89	54,13	31,16	178,22
39 THR	114,11	8,90	60,90	69,29	176,85
40 GLU	122,14	9,13	58,98	28,78	178,22
41 ASN	115,96	9,02	52,63	37,18	
42 GLY	109,44	8,05	45,10		173,99

43 THR	121,03	8,33	62,60	67,50	175,69
44 ASP	129,08	8,65	54,46	40,89	173,66
45 ILE	128,77	8,28	57,50	36,72	172,85
46 ALA	129,61	8,29	49,58	18,65	175,42
47 ILE	124,31	9,16	60,53	38,18	175,71
48 LYS	130,58	8,59	54,84	32,08	175,11
49 LEU	123,05	8,43	54,65	40,69	178,22
50 GLU	121,34	8,48	56,02	30,17	176,85
51 ASN					
52 SER	113,45	8,33	58,05	62,67	
53 GLY	109,76	8,01	45,03		172,54
54 THR	111,25	7,81	59,94	71,14	
55 LEU	122,27	8,14	54,87	42,88	175,24
56 ARG	121,81	8,95	53,06	31,75	
57 TYR	121,79	8,86	60,36	38,18	177,25
58 GLY					
59 ASP	122,43	8,00	56,05	40,89	175,71
60 VAL	120,39	8,55	61,34	30,89	175,36
61 LEU	125,57	9,00	54,09	41,62	175,07
62 TYR	116,18	7,23	57,09	41,35	171,27
63 GLU	128,34	7,79	54,13	31,36	173,18
64 SER	122,91	8,80	55,20	64,86	173,40
65 ASP					
66 ASP	113,53	7,93	52,98	42,48	176,02
67 THR	116,58	7,40	61,60	72,27	171,11
68 LEU	124,47	8,57	52,61	44,20	175,09
69 ILE	121,34	8,44	60,01	38,51	174,63
70 ALA	130,10	7,34	48,80	22,35	174,09
71 ILE	118,90	8,47	58,38	40,10	175,05
72 ARG	121,16	8,82	53,72	31,42	173,18
73 THR	108,08	6,67	58,27	69,56	174,99
74 LYS	118,20	7,51	55,24	32,68	174,41
75 LEU	123,05	8,29	54,87	41,29	176,02
76 GLU	126,04	8,54	54,24	33,14	174,20
77 LYS	120,57	8,46	56,61	31,56	177,43
78 VAL	118,84	8,59	59,00	34,83	177,14
79 TYR	116,17	8,64	57,09	38,90	174,45
80 VAL	121,71	8,95	60,86	30,89	175,40
81 ILE	129,38	9,44	60,89	37,78	175,69
82 LYS	126,80	8,52	53,39	31,62	173,22
83 PRO					
84 GLN	119,77	9,47	55,46	31,36	176,00
85 THR	106,88	7,18	57,57	71,81	174,41
86 MET	122,75	9,08	57,87	32,68	177,93
87 GLN	120,41	8,64	58,83	28,44	177,79

88 GLU	118,68	7,85	59,24	30,50	177,99
89 MET	117,82	7,91	57,51	31,09	177,29
90 GLY	105,69	8,54	46,99		174,74
91 LYS	120,93	8,70	59,86	31,62	179,15
92 MET	116,96	8,26	56,42	34,20	178,06
93 ALA	121,27	8,54	55,57	19,24	179,73
94 PHE	118,19	8,65	60,90	38,84	177,77
95 GLU	117,59	7,81	57,87	28,11	180,10
96 ILE	120,38	7,99	64,67	36,79	178,18
97 GLY	109,29	8,52	47,36		176,19
98 ASN	120,31	8,17	54,87	38,77	175,19
99 ARG	115,76	6,89	54,54	29,37	175,36
100 HIS	116,49	8,00	55,24	25,47	173,93
101 THR	117,71	7,51	62,34	69,69	172,64
102 MET	124,30	8,26	55,87	32,22	175,38
103 CYS					
104 ILE	119,31	8,44	59,57	39,70	174,24
105 ILE	127,91	8,61	59,76	36,92	174,74
106 GLU	128,84	8,44	54,46	30,96	174,74
107 ASP	122,47	8,89	56,35	39,37	173,82
108 ASP	113,49	7,98	52,43	39,76	174,44
109 GLU	116,43	7,48	54,09	32,75	174,28
110 ILE					
111 LEU	127,95	9,10	53,07	43,47	176,04
112 VAL	115,40	8,15	58,20	34,87	174,93
113 ARG					
114 TYR	121,48	7,84	59,38	37,51	174,39
115 ASP	126,20	5,76	52,76	40,69	175,88
116 LYS					
117 THR	111,33	8,29	64,23	68,43	178,78
118 LEU	121,75	7,97	55,13	41,09	175,98
119 GLU	119,15	7,38	60,09	28,38	176,98
120 LYS	115,03	7,56	58,12	30,76	178,53
121 LEU	120,38	6,88	56,72	40,82	177,89
122 ILE	119,32	7,76	64,97	36,79	179,14
123 ASP	118,61	8,16	56,57	40,16	
124 GLU	120,56	7,75	58,61	28,64	178,84
125 VAL					
126 GLY	108,19	8,15	45,47		174,12
127 VAL	116,49	6,63	59,35	32,75	174,12
128 SER	121,34	8,65	58,90	63,00	172,87
129 TYR	118,83	8,14	55,31	40,82	172,98
130 GLU	117,58	8,62	53,61	33,01	174,76
131 GLN	123,33	8,99	55,09	28,97	175,15
132 SER	118,53	8,59	55,24	63,93	

133 GLU	121,51	8,34	54,50	31,36	175,71
134 ARG	122,47	8,83	52,61	34,47	173,52
135 ARG					
136 PHE					
137 LYS					
138 GLU					
139 PRO					
140 PHE					
141 LYS					
142 TYR					
143 ARG					
144 GLY	108,98	7,74	44,58		176,31
145 HIS	119,23	8,09	55,52	29,77	174,90
146 GLN	122,29	8,26	55,24	28,97	175,94
147 HIS	125,88	7,95	56,76	29,44	179,16

Bibliography

[1] J. De Las Rivas and C. Fontanillo, “Protein-protein interactions essentials: key concepts to building and analyzing interactome networks.,” *PLoS Comput. Biol.*, vol. 6, no. 6, Jun. 2010.

[2] M. Baker, “The interaction map,” *Nature*, vol. 484, pp. 271–275, 2012.

[3] L. Bonetta, “Interactome under construction,” *Nature*, vol. 468, pp. 851–854, 2010.

[4] V. N. Uversky, “The triple power of D3: Protein intrinsic disorder in degenerative diseases,” *Front. Biosci.*, vol. 19, pp. 181–258, 2014.

[5] C. Royer, *Proteins - Chapter 4: Protein-Protein Interactions*. Biophysics Textbook On-Line, 2000.

[6] D. Chandler, “Interfaces and the driving force of hydrophobic assembly.,” *Nature*, vol. 437, no. 7059, pp. 640–7, Sep. 2005.

[7] P. E. Wright and H. J. Dyson, “Intrinsically unstructured proteins: re-assessing the protein structure-function paradigm.,” *J. Mol. Biol.*, vol. 293, no. 2, pp. 321–31, Oct. 1999.

[8] H. J. Dyson and P. E. Wright, “Intrinsically unstructured proteins and their functions.,” *Nat. Rev. Mol. Cell Biol.*, vol. 6, no. 3, pp. 197–208, Mar. 2005.

[9] P. Tompa, C. Szász, and L. Buday, “Structural disorder throws new light on moonlighting.,” *Trends Biochem. Sci.*, vol. 30, no. 9, pp. 484–9, Sep. 2005.

[10] A. Stein, R. Pache, P. Bernadó, M. Pons, and P. Aloy, “Dynamic interactions of proteins in complex networks: a more structured view.,” *FEBS J.*, vol. 276, no. 19, pp. 5390–405, Oct. 2009.

[11] J. Janin and A. M. J. J. Bonvin, “Protein-protein interactions.,” *Curr. Opin. Struct. Biol.*, vol. 23, no. 6, pp. 859–61, Dec. 2013.

[12] X. Zhang, T. Perica, and S. Teichmann, “Evolution of protein structures and interactions from the perspective of residue contact networks.,” *Curr. Opin. Struct. Biol.*, vol. 23, no. 6, pp. 954–63, Dec. 2013.

[13] T. Perica, J. A. Marsh, F. L. Sousa, E. Natan, L. J. Colwell, S. E. Ahnert, and S. A. Teichmann, “The emergence of protein complexes: quaternary structure, dynamics and allostery,” *Biochem. Soc. Trans.*, vol. 40, no. 3, pp. 475–491, Jun. 2012.

[14] B. Zambelli, F. Musiani, and S. Ciurli, *Interplay between Metal Ions and Nucleic Acids*, vol. 10. Dordrecht: Springer Netherlands, 2012, pp. 135–170.

[15] L. Finney and T. V O’Halloran, “Transition metal speciation in the cell: insights from the chemistry of metal ion receptors.,” *Science*, vol. 300, no. 5621, pp. 931–6, May 2003.

[16] C. Andreini, I. Bertini, G. Cavallaro, G. L. Holliday, and J. M. Thornton, “Metal ions in biological catalysis: from enzyme databases to general principles.,” *J. Biol. Inorg. Chem.*, vol. 13, no. 8, pp. 1205–1218, Nov. 2008.

[17] K. J. Waldron, J. C. Rutherford, D. Ford, and N. J. Robinson, “Metalloproteins and metal sensing.,” *Nature*, vol. 460, no. 7257, pp. 823–30, Aug. 2009.

[18] H. Irving and R. J. P. Williams, “Order of Stability of Metal Complexes,” *Nature*, vol. 162, no. 4123, pp. 746–747, 1948.

[19] K. J. Waldron and N. J. Robinson, “How do bacterial cells ensure that metalloproteins get the correct metal?,” *Nat. Rev. Microbiol.*, vol. 7, no. 1, pp. 25–35, Jan. 2009.

[20] L. Rulisek and Vondrasek, “Coordination geometries of selected transition metal ions (Co2+, Ni2+, Cu2+, Zn2+, Cd2+, and Hg2+) in metalloproteins,” *J. Inorg. Biochem.*, vol. 71, pp. 115–127, 1998.

[21] S. Tottey, K. J. Waldron, S. J. Firbank, B. Reale, C. Bessant, K. Sato, T. R. Cheek, J. Gray, M. J. Banfield, C. Dennison, and N. J. Robinson, “Protein-folding location can regulate manganese-binding versus copper- or zinc-binding.,” *Nature*, vol. 455, no. 7216, pp. 1138–42, Oct. 2008.

[22] L. H. Hartwell, J. J. Hopfield, S. Leibler, and W. Murray, “From molecular to modular cell biology.,” *Nature*, vol. 402, no. 6761 Suppl, pp. C47–52, Dec. 1999.

[23] I. N. Serdyuk, N. R. Zaccai, and J. Zaccai, *Methods in Molecular Biophysics Structure, Dynamics, Function*. 2012.

[24] K. Takeuchi and G. Wagner, “NMR studies of protein interactions.,” *Curr. Opin. Struct. Biol.*, vol. 16, no. 1, pp. 109–17, Feb. 2006.

[25] F. Agostini, “Spettroscopia NMR di UreE da Bacillus pasteurii: struttura e dinamica di una nichel-chaperonina,” 2009.
[26] R. R. Ernst, G. Bodenhausen, and A. Wokaun, *Principles of Nuclear Magnetic Resonance in One and Two Dimensions*. Oxford University Press., 1987.
[27] G. Otting and K. Wuthrich, “Heteronuclear filters in two-dimensional [1H, 1H]-NMR spectroscopy: combined use with isotope labelling for studies of macromolecular conformation and intermolecular interactions.,” *Q. Rev. Biophys.*, vol. 23, pp. 39–56, 1990.
[28] K. Pervushin, R. Riek, G. Wider, and K. Wüthrich, “Attenuated T2 relaxation by mutual cancellation of dipole-dipole coupling and chemical shift anisotropy indicates an avenue to NMR structures of very large biological macromolecules in solution.,” *Proc. Natl. Acad. Sci. U. S. A.*, vol. 94, no. 23, pp. 12366–71, Nov. 1997.
[29] C. Fernandez, “TROSY in NMR studies of the structure and function of large biological macromolecules,” *Curr. Opin. Struct. Biol.*, vol. 13, no. 5, pp. 570–580, Oct. 2003.
[30] K. H. Gardner and L. E. Kay, “The use of 2H, 13C, 15N multidimensional NMR to study the structure and dynamics of proteins.,” *Annu. Rev. Biophys. Biomol. Struct.*, vol. 27, pp. 357–406, Jan. 1998.
[31] E. R. P. Zuiderweg, “Mapping Protein-Protein Interactions in Solution by NMR Spectroscopy †,” *Biochemistry*, vol. 41, no. 1, 2002.
[32] Y. Chen, J. Reizer, M. H. Saier, W. J. Fairbrother, and P. E. Wright, “Mapping of the binding interfaces of the proteins of the bacterial phosphotransferase system, HPr and IIAglc.,” *Biochemistry*, vol. 32, no. 1, pp. 32–7, Jan. 1993.
[33] R. Williamson, M. D. Carr, T. a Frenkiel, J. Feeney, and R. B. Freedman, “Mapping the binding site for matrix metalloproteinase on the N-terminal domain of the tissue inhibitor of metalloproteinases-2 by NMR chemical shift perturbation.,” *Biochemistry*, vol. 36, no. 45, pp. 13882–9, Nov. 1997.
[34] N. A. Farrow, J. R. Muhandham, A. U. Singer, S. M. Pascal, C. M. Kay, G. Gish, S. E. Shoelson, T. Pawson, J. D. Forman-kay, and L. E. Kay, “Backbone Dynamics of a Free and a Phosphopeptide-Complexed Src Homology 2 Domain Studied by 1 5 N NMR Relaxation,” pp. 5984–6003, 1994.
[35] B. K. Vu, J. D. Walsh, D. S. Dimitrov, and R. Ishima, “Dynamics of Antibody Domains Studied by Solution NMR,” *Methods Mol Biol.*, pp. 525–533, 2009.
[36] G. Lipari and A. Szabo, “Model-Free Approach to the Interpretation of Nuclear Magnetic Resonance Relaxation in Macromolecules . 1 . Theory and Range of Validity,” *J. Am. Chem. Soc.*, vol. I, no. 1, pp. 4546–4559, 1982.
[37] R. Ishima and D. A. Torchia, “Protein dynamics from NMR,” vol. 7, no. 9, pp. 7–10, 2000.
[38] G. M. Clore, A. Szabo, A. Bax, L. E. Kay, P. C. Driscoll, and A. M. Gronenborn, “Deviations from the Simple Two-Parameter Model-Free Approach to the Interpretation of Nitrogen-15 Nuclear Magnetic Relaxation of Proteins,” vol. 1, no. 2, pp. 4989–4991, 1990.
[39] L. E. Kay, D. A. Torchia, and A. Bax, “Backbone Dynamics of Proteins As Studied by 15N Inverse Detected Heteronuclear NMR Spectroscopy : Application to Staphylococcal Nuclease,” *Biochemistry*, vol. 28, pp. 8972–8979, 1989.
[40] D. Fushman, R. Weisemann, H. Thüring, and H. Rüterjans, “Backbone dynamics of ribonuclease T1 and its complex with 2’GMP studied by two-dimensional heteronuclear NMR spectroscopy.,” *J. Biomol. NMR*, vol. 4, no. 1, pp. 61–78, Jan. 1994.
[41] V. A. Jarymowycz and M. J. Stone, “Fast Time Scale Dynamics of Protein Backbones : NMR Relaxation Methods , Applications , and Functional Consequences,” *Chem. Rev.*, no. 812, pp. 1624–1671, 2006.
[42] D. Fushman, R. Xu, and D. Cowburn, “Direct determination of changes of interdomain orientation on ligation: use of the orientational dependence of 15N NMR relaxation in Abl SH(32).,” *Biochemistry*, vol. 38, no. 32, pp. 10225–30, Aug. 1999.
[43] J. F. Lefevre, K. T. Dayie, J. W. Peng, and G. Wagner, “Internal mobility in the partially folded DNA binding and dimerization domains of GAL4: NMR analysis of the N-H spectral density functions.,” *Biochemistry*, vol. 35, no. 8, pp. 2674–86, Feb. 1996.
[44] J. W. Peng and G. Wagner, “Mapping of Spectral Density Functions Using Heteronuclear NMR Relaxation Measurements,” *J. Magn. Reson.*, vol. 98, pp. 308–332, 1992.
[45] N. Farrow, O. Zhang, Szabo, D. Torchia, and L. E. Kay, “Spectral density function mapping using 15N relaxation data exclusively.,” *J. Biomol. NMR*, vol. 6, no. 2, pp. 153–62, Sep. 1995.

[46] G. Lipari and A. Szabo, “Model-Free Approach to the Interpretation of Nuclear Magnetic Resonance Relaxation in Macromolecules. 2. Analysis of Experimental Results,” *J. Am. Chem. Soc.*, vol. 2, no. 1, 1982.
[47] P. J. Sapienza and A. L. Lee, “Using NMR to study fast dynamics in proteins: methods and applications.,” *Curr. Opin. Pharmacol.*, vol. 10, no. 6, pp. 723–30, Dec. 2010.
[48] E. J. D’Auvergne and P. R. Gooley, “Model-free model elimination : A new step in the model-free dynamic analysis of NMR relaxation data,” *J. Biomol. NMR*, pp. 117–135, 2006.
[49] D. Fushman, S. Cahill, and D. Cowburn, “Homology (PH) Domain in Solution : Analysis of 15 N Relaxation with Monomer / Dimer Equilibration,” 1997.
[50] E. J. D’Auvergne and P. R. Gooley, “Optimisation of NMR dynamic models I . Minimisation algorithms and their performance within the model-free and Brownian rotational diffusion spaces,” *J. Biomol. NMR*, pp. 107–119, 2008.
[51] E. J. d’Auvergne and P. R. Gooley, “The use of model selection in the model-free analysis of protein dynamics.,” *J. Biomol. NMR*, vol. 25, no. 1, pp. 25–39, Jan. 2003.
[52] H. Akaike, “Factor Analysis and AIC,” *Psychometrika*, vol. 52, no. 3, pp. 317–332, 1987.
[53] Y. Liang, “Applications of isothermal titration calorimetry in protein science,” *Acta Biochim. Biophys. Sin. (Shanghai).*, vol. 40, no. 7, pp. 565–576, Jul. 2008.
[54] A. Velázquez Campoy and E. Freire, “ITC in the post-genomic era...? Priceless.,” *Biophys. Chem.*, vol. 115, no. 2–3, pp. 115–24, Apr. 2005.
[55] A. Velazquez-Campoy and E. Freire, “Isothermal titration calorimetry to determine association constants for high affinity ligands,” *Nat. Publ. Gr.*, pp. 4–9, 2006.
[56] A. Brown, “Analysis of cooperativity by isothermal titration calorimetry.,” *Int. J. Mol. Sci.*, vol. 10, no. 8, pp. 3457–77, Oct. 2009.
[57] B. P. Callahan, Y. Yuan, and R. Wolfenden, “The Burden Borne by Urease,” *J. Am. Chem. Soc.*, vol. 127, no. 31, pp. 10828–10829, 2005.
[58] B. Krajewska, “Ureases I. Functional, catalytic and kinetic properties: A review,” *J. Mol. Catal. B Enzym.*, vol. 59, no. 1–3, pp. 9–21, Jul. 2009.
[59] J. B. Sumner, “The Isolation and Crystallization of the Enzyme Urease,” pp. 435–441, 1926.
[60] N. E. Dixon, C. Gazzola, J. J. Watters, R. L. Blakeley, and B. Zerner, “Jack Bean Urease (EC 3.5.1.5). A Metalloenzyme. A Simple Biological Role for Nickel?,” *J. Am. Chem. Soc.*, pp. 4131–4133, 1975.
[61] J. B. Sumner, “The story of urease,” *J. Chem. Educ.*, 1937.
[62] H. L. Mobley and R. P. Hausinger, “Microbial ureases: significance, regulation, and molecular characterization.,” *Microbiol. Rev.*, vol. 53, no. 1, pp. 85–108, Mar. 1989.
[63] E. Jabri, M. B. Carr, R. P. Hausinger, and P. A. Karplus, “The Crystal Structure of Urease from Kiebsiella aerogenes,” *Science (80-.).*, vol. 268, no. May, 1995.
[64] J. M. Bremner and M. J. Krogmeier, “Evidence that the adverse effect of urea fertilizer on seed germination in soil is due to ammonia formed through hydrolysis of urea by soil urease.,” *Proc. Natl. Acad. Sci. U. S. A.*, vol. 86, no. 21, pp. 8185–8, Nov. 1989.
[65] S. Ciurli, C. Marzadori, S. Benini, S. Deiana, and C. Gessa, “Urease from the soil bacterium Bacillus pasteurii: Immobilization on Ca-polygalacturonate,” *Soil Biol. Biochem.*, vol. 28, no. 6, pp. 811–817, Jun. 1996.
[66] I. Sondi and B. Salopek-sondi, “Influence of the Primary Structure of Enzymes on the Formation of CaCO 3 Polymorphs : A Comparison of Plant (Canavalia ensiformis) and Bacterial (Bacillus pasteurii) Ureases,” no. 18, pp. 8876–8882, 2005.
[67] R. Burne and Y. Y. Chen, “Bacterial ureases in infectious diseases.,” *Microbes Infect.*, vol. 2, no. 5, pp. 533–42, Apr. 2000.
[68] H. L. Mobley, M. D. Island, and R. P. Hausinger, “Molecular biology of microbial ureases.,” *Microbiol. Rev.*, vol. 59, no. 3, pp. 451–80, Sep. 1995.
[69] M. G. Bruce and H. I. Maaroos, “Epidemiology of Helicobacter pylori infection.,” *Helicobacter*, vol. 13 Suppl 1, pp. 1–6, Oct. 2008.
[70] H. Yoshiyama and T. Nakazawa, “Unique mechanism of Helicobacter pylori for colonizing the gastric mucus.,” *Microbes Infect.*, vol. 2, no. 1, pp. 55–60, Jan. 2000.
[71] J. R. Warren, “Helicobacter – the ease and difficulty of a new discovery,” *Nobel Lect.*, pp. 292–305, 2005.

[72] K. Sendide, A. Deghmane, J. Reyrat, A. Talal, Z. Hmama, and B. Abdallah, “Mycobacterium bovis BCG Urease Attenuates Major Histocompatibility Complex Class II Trafficking to the Macrophage Cell Surface,” vol. 72, no. 7, pp. 4200–4209, 2004.
[73] M. J. Maroney and S. Ciurli, “Nonredox Nickel Enzymes.,” *Chem. Rev.*, Dec. 2013.
[74] S. Benini, W. R. Rypniewski, K. S. Wilson, S. Mangani, and S. Ciurli, “Molecular Details of Urease Inhibition by Boric Acid : Insights into the Catalytic Mechanism,” pp. 3714–3715, 2004.
[75] R. Ligabue-Braun, F. C. Andreis, H. Verli, and C. R. Carlini, “3-To-1: Unraveling Structural Transitions in Ureases.,” *Naturwissenschaften*, vol. 100, no. 5, pp. 459–67, May 2013.
[76] 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 structures of the native and inhibited enzyme from Bacillus pasteurii: why urea hydrolysis costs two nickels.,” *Structure*, vol. 7, no. 2, pp. 205–16, Feb. 1999.
[77] N. Ha, S. Oh, J. Y. Sung, K. A. Cha, M. H. Lee, and B. Oh, “Supramolecular assembly and acid resistance of Helicobacter pylori urease,” *Nat. Struct. Biol.*, vol. 8, no. 6, pp. 505–509, 2001.
[78] A. Balasubramanian and K. Ponnuraj, “Crystal structure of the first plant urease from jack bean: 83 years of journey from its first crystal to molecular structure.,” *J. Mol. Biol.*, vol. 400, no. 3, pp. 274–83, Jul. 2010.
[79] B. Zambelli, F. Musiani, S. Benini, and S. Ciurli, “Chemistry of Ni in Urease: Sensing, Trafficking, and Catalysis,” *Acc. Chem. Res.*, vol. 44, no. 7, 2011.
[80] 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 mechanisms,” *Coord. Chem. Rev.*, vol. 190–192, pp. 331–355, Sep. 1999.
[81] S. Benini, P. Kosikowska, M. Cianci, L. Mazzei, A. G. Vara, Ł. Berlicki, and S. Ciurli, “The crystal structure of Sporosarcina pasteurii urease in a complex with citrate provides new hints for inhibitor design.,” *J. Biol. Inorg. Chem.*, vol. 18, no. 3, pp. 391–9, Mar. 2013.
[82] F. Musiani, E. Arnofi, R. Casadio, and S. Ciurli, “Structure-based computational study of the catalytic and inhibition mechanisms of urease,” *J. Biol. Inorg. Chem.*, vol. 6, no. 3, pp. 300–314, Mar. 2001.
[83] I. S. Park and R. P. Hausinger, “Requirement of carbon dioxide for in vitro assembly of the urease nickel metallocenter.,” *Science*, vol. 267, no. 5201, pp. 1156–8, Feb. 1995.
[84] S. B. Mulrooney and R. P. Hausinger, “Sequence of the Klebsiella aerogenes urease genes and evidence for accessory proteins facilitating nickel incorporation.,” *J. Bacteriol.*, vol. 172, no. 10, pp. 5837–43, Oct. 1990.
[85] M. Farrugia, L. Macomber, and R. P. Hausinger, “Biosynthesis of the urease metallocenter.,” *J. Biol. Chem.*, vol. 288, no. 19, pp. 13178–85, May 2013.
[86] V. Cussac, R. L. Ferrero, and A. Labigne, “Expression of Helicobacterpylori Urease Genes in Escherichia coli Grown under Nitrogen-Limiting Conditions,” *J. Bacteriol.*, vol. 174, no. 8, 1992.
[87] Y. H. Fong, H. C. Wong, C. P. Chuck, Y. W. Chen, H. Sun, and K.-B. Wong, “Assembly of preactivation complex for urease maturation in Helicobacter pylori: crystal structure of UreF-UreH protein complex.,” *J. Biol. Chem.*, vol. 286, no. 50, pp. 43241–9, Dec. 2011.
[88] E. L. Carter and R. P. Hausinger, “Characterization of the Klebsiella aerogenes urease accessory protein UreD in fusion with the maltose binding protein.,” *J. Bacteriol.*, vol. 192, no. 9, pp. 2294–304, May 2010.
[89] B. Zambelli, A. Berardi, V. Martin-Diaconescu, L. Mazzei, F. Musiani, M. J. Maroney, and S. Ciurli, “Nickel binding properties of Helicobacter pylori UreF, an accessory protein in the nickel-based activation of urease.,” *J. Biol. Inorg. Chem.*, vol. 19, pp. 319–334, 2014.
[90] Z. Chang, J. Kuchar, and R. P. Hausinger, “Chemical cross-linking and mass spectrometric identification of sites of interaction for UreD, UreF, and urease.,” *J. Biol. Chem.*, vol. 279, no. 15, pp. 15305–13, Apr. 2004.
[91] M. B. C. Moncrief and R. P. Hausinger, “Purification and Activation Properties of UreD-UreF-Urease Apoprotein Complexes,” *J. Bacteriol.*, vol. 178, no. 18, pp. 5417–5421, 1996.
[92] M. Salomone-stagni, B. Zambelli, F. Musiani, and S. Ciurli, “A model-based proposal for the role of UreF as a GTPase-activating protein in the urease active site biosynthesis,” *Proteins*, 2007.
[93] M. Soriano and R. P. Hausinger, “GTP-dependent activation of urease apoprotein in complex with the UreD, UreF, and UreG accessory proteins.,” *Proc. Natl. Acad. Sci. U. S. A.*, vol. 96, no. 20, pp. 11140–4, Sep. 1999.

[94] M. B. C. Moncrief and R. P. Hausinger, “Characterization 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.*, vol. 179, no. 13, pp. 4081–4086, 1997.

[95] B. Zambelli, M. Stola, F. Musiani, K. De Vriendt, B. Samyn, B. Devreese, J. Van Beeumen, P. Turano, A. Dikiy, D. Bryant, and S. Ciurli, “UreG, a chaperone in the urease assembly process, is an intrinsically unstructured GTPase that specifically binds Zn2+.,” *J. Biol. Chem.*, vol. 280, no. 6, pp. 4684–95, Feb. 2005.

[96] B. Zambelli, F. Musiani, M. Savini, P. Tucker, and S. Ciurli, “Biochemical studies on Mycobacterium tuberculosis UreG and comparative modeling reveal structural and functional conservation among the bacterial UreG family.,” *Biochemistry*, vol. 46, no. 11, pp. 3171–82, Mar. 2007.

[97] Y. H. Fong, H. C. Wong, M. H. Yuen, P. H. Lau, Y. W. Chen, and K.-B. Wong, “Structure of UreG/UreF/UreH complex reveals how urease accessory proteins facilitate maturation of Helicobacter pylori urease.,” *PLoS Biol.*, vol. 11, no. 10, p. e1001678, Oct. 2013.

[98] B. Zambelli, P. Turano, F. Musiani, P. Neyroz, and S. Ciurli, “Zn2+-linked dimerization of UreG from Helicobacter pylori, a chaperone involved in nickel trafficking and urease activation.,” *Proteins*, vol. 74, no. 1, pp. 222–39, Jan. 2009.

[99] B. Zambelli, N. Cremades, P. Neyroz, P. Turano, V. N. Uversky, and S. Ciurli, “Insights in the (un)structural organization of Bacillus pasteurii UreG, an intrinsically disordered GTPase enzyme.,” *Mol. Biosyst.*, vol. 8, no. 1, pp. 220–8, Jan. 2012.

[100] P. Neyroz, B. Zambelli, and S. Ciurli, “Intrinsically disordered structure of Bacillus pasteurii UreG as revealed by steady-state and time-resolved fluorescence spectroscopy.,” *Biochemistry*, vol. 45, no. 29, pp. 8918–30, Jul. 2006.

[101] J. L. Boer, S. Quiroz-valenzuela, K. L. Anderson, and R. P. Hausinger, “Mutagenesis of Klebsiella aerogenes UreG To Probe Nickel Binding and Interactions with Other Urease-Related Proteins,” *Biochemistry*, vol. 49, pp. 5859–5869, 2010.

[102] J. L. Boer and R. P. Hausinger, “Klebsiella aerogenes UreF: Identification of the UreG Binding Site and Role in Enhancing the Fidelity of Urease Activation.,” *Biochemistry*, vol. 51, no. 11, pp. 2298–308, Mar. 2012.

[103] M. Bellucci, B. Zambelli, F. Musiani, P. Turano, and S. Ciurli, “Helicobacter pylori UreE, a urease accessory protein: specific Ni(2+)- and Zn(2+)-binding properties and interaction with its cognate UreG.,” *Biochem. J.*, vol. 422, no. 1, pp. 91–100, Aug. 2009.

[104] F. Biagi, F. Musiani, and S. Ciurli, “Structure of the UreD – UreF – UreG – UreE complex in Helicobacter pylori : a model study,” 2013.

[105] 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*, vol. 38, pp. 4078–4088, 1999.

[106] G. J. Colpas and R. P. Hausinger, “In Vivo and in Vitro Kinetics of Metal Transfer by the Klebsiella aerogenes Urease Nickel Metallochaperone, UreE,” *J. Biol. Chem.*, vol. 275, no. 15, pp. 10731–10737, Apr. 2000.

[107] S. Ciurli, N. Safarov, S. Miletti, A. Dikiy, S. K. Christensen, K. Kornetzky, D. A. Bryant, I. Vandenberghe, B. Devreese, B. Samyn, H. Remaut, and J. Van Beeumen, “Molecular characterization of Bacillus pasteurii UreE , a metal-binding chaperone for the assembly of the urease active site,” *J. Biol. Inorg. Chem.*, vol. 7, pp. 623–631, 2002.

[108] E. L. Carter, N. Flugga, J. L. Boer, S. B. Mulrooney, and R. P. Hausinger, “Interplay of metal ions and urease.,” *Metallomics*, vol. 1, no. 3, pp. 207–21, Jan. 2009.

[109] H. Remaut, N. Safarov, S. Ciurli, and J. Van Beeumen, “Structural basis for Ni2+ transport and assembly of the urease active site by the metallochaperone UreE from Bacillus pasteurii.,” *J. Biol. Chem.*, vol. 276, no. 52, pp. 49365–70, Dec. 2001.

[110] B. Zambelli, K. Banaszak, A. Merloni, A. Kiliszek, W. Rypniewski, and S. Ciurli, “Selectivity of Ni(II) and Zn(II) binding to Sporosarcina pasteurii UreE, a metallochaperone in the urease assembly: a calorimetric and crystallographic study.,” *J. Biol. Inorg. Chem.*, vol. 18, no. 8, pp. 1005–17, Dec. 2013.

[111] H. K. Song, S. B. Mulrooney, R. Huber, and R. P. Hausinger, "Crystal structure of Klebsiella aerogenes UreE, a nickel-binding metallochaperone for urease activation.," *J. Biol. Chem.*, vol. 276, no. 52, pp. 49359–64, Dec. 2001.
[112] R. Shi, C. Munger, A. Asinas, S. L. Benoit, E. Miller, A. Matte, R. J. Maier, and M. Cygler, "Crystal structures of apo and metal-bound forms of the UreE protein from Helicobacter pylori: role of multiple metal binding sites.," *Biochemistry*, vol. 49, no. 33, pp. 7080–8, Aug. 2010.
[113] K. Banaszak, V. Martin-Diaconescu, M. Bellucci, B. Zambelli, W. Rypniewski, M. J. Maroney, and S. Ciurli, "Crystallographic and X-ray absorption spectroscopic characterization of Helicobacter pylori UreE bound to Ni^{2+} and Zn^{2+} reveals a role for the disordered C-terminal arm in metal trafficking.," *Biochem. J.*, vol. 441, no. 3, pp. 1017–26, Feb. 2012.
[114] F. Musiani, B. Zambelli, M. Stola, and S. Ciurli, "Nickel trafficking: insights into the fold and function of UreE, a urease metallochaperone.," *J. Inorg. Biochem.*, vol. 98, no. 5, pp. 803–13, May 2004.
[115] M. Stola, F. Musiani, S. Mangani, P. Turano, N. Safarov, B. Zambelli, and S. Ciurli, "The nickel site of Bacillus pasteurii UreE, a urease metallo-chaperone, as revealed by metal-binding studies and X-ray absorption spectroscopy.," *Biochemistry*, vol. 45, no. 20, pp. 6495–509, May 2006.
[116] M. H. Lee, H. S. Pankratz, S. Wang, R. A. Scott, M. G. Finnegan, M. K. Johnson, J. A. Ippolit, D. W. Christianson, and R. P. Hausinger, "Purification and characterization of Klebsiella aerogenes UreE protein : A nickel-binding protein that functions in urease metallocenter assembly," pp. 1042–1052, 1993.
[117] H.-S. Won, Y.-H. Lee, J.-H. Kim, I. S. Shin, M. H. Lee, and B.-J. Lee, "Structural characterization of the nickel-binding properties of Bacillus pasteurii urease accessory protein (Ure)E in solution.," *J. Biol. Chem.*, vol. 279, no. 17, pp. 17466–72, Apr. 2004.
[118] N. E. Grossoehme, S. B. Mulrooney, R. P. Hausinger, and D. E. Wilcox, "Thermodynamics of Ni2+, Cu2+, and Zn2+ binding to the urease metallochaperone UreE.," *Biochemistry*, vol. 46, no. 37, pp. 10506–16, Sep. 2007.
[119] K. Higgins, C. E. Carr, and M. J. Maroney, "Specific Metal Recognition in Nickel Trafficking.," *Biochemistry*, Sep. 2012.
[120] A. Soriano, G. J. Colpas, and R. P. Hausinger, "UreE Stimulation of GTP-Dependent Urease Activation in the UreD-UreF-UreG-urease Apoprotein Complex," *Biochemistry*, vol. 39, pp. 12435–12440, 2000.
[121] P. Voland, D. L. Weeks, E. a Marcus, C. Prinz, G. Sachs, and D. Scott, "Interactions among the seven Helicobacter pylori proteins encoded by the urease gene cluster.," *Am. J. Physiol. Gastrointest. Liver Physiol.*, vol. 284, no. 1, pp. G96–G106, Jan. 2003.
[122] A. Sivashanmugam, V. Murray, C. Cui, Y. Zhang, J. Wang, and Q. Li, "Practical protocols for production of very high yields of recombinant proteins using Escherichia coli.," *Protein Sci.*, vol. 18, no. 5, pp. 936–48, May 2009.
[123] O. Paliy, D. Bloor, D. Brockwell, P. Gilbert, and J. Barber, "Improved methods of cultivation and production of deuteriated proteins from E. coli strains grown on fully deuteriated minimal medium.," *J. Appl. Microbiol.*, vol. 94, no. 4, pp. 580–6, Jan. 2003.
[124] F. Delaglio, S. Grzesiek, G. W. Vuister, G. Zhu, J. Pfeifer, and a Bax, "NMRPipe: a multidimensional spectral processing system based on UNIX pipes.," *J. Biomol. NMR*, vol. 6, no. 3, pp. 277–93, Nov. 1995.
[125] R. L. J. Keller, "Optimizing the process of nuclear magnetic resonance spectrum analysis and computer aided resonance assignment," 2004.
[126] L. E. Kay, D. A. Torchia, and A. B. J, "Backbone Dynamics of Proteins AS Studied by 15N Inverse Detected Heteranuckar," pp. 8972–8979, 1989.
[127] M. P. Williamson, "Using chemical shift perturbation to characterise ligand binding," *Prog. Nucl. Magn. Reson. Spectrosc.*, vol. 73, pp. 1–16, 2013.
[128] Y. Bai, J. S. Milne, L. Mayne, and S. W. Englander, "Primary Structure Effects on Peptide Group Hydrogen Exchange," *Proteins*, vol. 17, no. 1, pp. 75–86, 1993.
[129] P. Romero, Z. Obradovic, X. Li, E. C. Garner, C. J. Brown, and a K. Dunker, "Sequence complexity of disordered protein.," *Proteins*, vol. 42, no. 1, pp. 38–48, Jan. 2001.
[130] X. Li, P. Romero, M. Rani, A. Dunker, and Z. Obradovic, "Predicting Protein Disorder for N-, C-, and Internal Regions.," *Genome Inform. Ser. Workshop Genome Inform.*, vol. 10, no. I, pp. 30–40, Jan. 1999.

[131] V. N. Uversky, J. R. Gillespie, and a L. Fink, “Why are ‘natively unfolded’ proteins unstructured under physiologic conditions?,” *Proteins*, vol. 41, no. 3, pp. 415–27, Nov. 2000.
[132] J. A. Marsh, V. K. Singh, Z. Jia, and J. D. Forman-kay, “Sensitivity of secondary structure propensities to sequence differences between a- and g-synuclein : Implications for fibrillation,” *Protein Sci.*, vol. 15, pp. 2795–2804, 2006.
[133] Y. Shen, F. Delaglio, G. Cornilescu, and A. Bax, “TALOS+: a hybrid method for predicting protein backbone torsion angles from NMR chemical shifts.,” *J. Biomol. NMR*, vol. 44, no. 4, pp. 213–23, Aug. 2009.
[134] F. Musiani, E. Ippoliti, C. Micheletti, P. Carloni, and S. Ciurli, “Conformational Fluctuations of UreG, an Intrinsically Disordered Enzyme,” *Biochemistry*, vol. 52, pp. 2949–2954, 2013.
[135] B. Zambelli, N. Cremades, P. Neyroz, P. Turano, V. N. Uversky, and S. Ciurli, “Insights in the (un)structural organization of Bacillus pasteurii UreG, an intrinsically disordered GTPase enzyme.,” *Mol. Biosyst.*, vol. 8, no. 1, pp. 220–8, Jan. 2012.
[136] Y. Lee, H. Won, H. Ahn, S. Park, H. Yagi, H. Akutsu, and B. Lee, “Letter to the Editor : Backbone NMR assignments of the metal-free UreE from Bacillus pasteurii,” pp. 361–362, 2002.
[137] Y. Lee, H. Won, M. Lee, and B. Lee, “Effects of salt and nickel ion on the conformational stability of Bacillus pasteurii UreE,” *Fed. Eur. Biochem. Soc.*, vol. 522, pp. 135–140, 2002.
[138] E. F. Pettersen, T. D. Goddard, C. C. Huang, G. S. Couch, D. M. Greenblatt, E. C. Meng, and T. E. Ferrin, “UCSF Chimera--a visualization system for exploratory research and analysis.,” *J. Comput. Chem.*, vol. 25, no. 13, pp. 1605–12, Oct. 2004.
[139] A. G. Palmer, “Probing molecular motion by NMR.,” *Curr. Opin. Struct. Biol.*, vol. 7, no. 5, pp. 732–7, Oct. 1997.
[140] G. M. Clore and A. M. Gronenborn, “Structures of Larger Proteins in Solution : Spectroscopy,” *Science*, vol. 252, pp. 1390–1399, 1991.
[141] J. M. Kneller, M. Lu, and C. Bracken, “An effective method for the discrimination of motional anisotropy and chemical exchange.,” *J. Am. Chem. Soc.*, vol. 124, no. 9, pp. 1852–3, Mar. 2002.
[142] P. Rossi, G. V. T. Swapna, Y. J. Huang, J. M. Aramini, C. Anklin, K. Conover, K. Hamilton, R. Xiao, T. B. Acton, A. Ertekin, J. K. Everett, and G. T. Montelione, “A microscale protein NMR sample screening pipeline.,” *J. Biomol. NMR*, vol. 46, no. 1, pp. 11–22, Jan. 2010.
[143] D. Fushman, N. Tjandra, and D. Cowburn, “An Approach to Direct Determination of Protein Dynamics from 15N NMR Relaxation at Multiple Fields, Independent of Variable 15N Chemical Shift Anisotropy and Chemical Exchange Contributions,” *J. Am. Chem. Soc.*, vol. 121, pp. 8577–8582, 1999.
[144] A. M. Mandel, M. Akke, and A. G. Palmer, “Dynamics of Ribonuclease H : Temperature Dependence of Motions on Multiple Time Scales,” *Byochemistry*, vol. 35, no. 50, pp. 16009–16023, 1996.
[145] D. Tobi and I. Bahar, “Structural changes involved in protein binding correlate with intrinsic motions of proteins in the unbound state,” *Proc. Natl. Acad. Sci. U.S.A.*, vol. 102, no. 52, pp. 18908–18913, 2005.
[146] D. D. Boehr, R. Nussinov, and P. E. Wright, “The role of dynamic conformational ensembles in biomolecular recognition,” *Nat Chem Biol*, vol. 5, no. 11, pp. 789–796, 2009.
[147] J. Morrison, J.-C. Yang, M. Stewart, and D. Neuhaus, “Solution NMR Study of the Interaction Between NTF2 and Nucleoporin FxFG Repeats,” *J. Mol. Biol.*, vol. 333, pp. 587–603, Oct. 2003.
[148] J. A. R. Worrall, W. Reinle, R. Bernhardt, and M. Ubbink, “Transient Protein Interactions Studied by NMR Spectroscopy : The Case of Cytochrome c and Adenodoxin,” *Biochemistry*, vol. 42, no. 23, pp. 7068–7076, 2003.

Acknowledgements

I would like to thank my PhD supervisor Prof. Stefano Ciurli for giving me the opportunity to work on this project and for his support that allowed me to grow as a scientist and as a person too.

I also want to express my thankfulness to all the members of the research group that always provided a stimulating environment and insightful comments. In particular a sincere thanks goes to Olena Dobrovolska, Luca Mazzei, Francesco Musiani and Barbara Zambelli, since it was a real pleasure to work with all of them.

I would also like to thank Daniela and Silvia, my dear friends and biologist fellow colleagues, who shared with me the difficulties during the PhD route, it would not have been the same without their encouragements and their fruitful exchange on methods and topics.

A special thanks goes to my family. Words cannot express how grateful I am to my father and my mother for all of the support that they give me every day and especially in the moments when there is no one else to answer my questions. The fact that they so much believe in me is what motivates me to never give up. In addition a heartfelt thank goes to the brilliant insights of my father, who often have unravelled the most difficult knots that occurred during the writing of this thesis.

Finally, I would like to express my deep gratitude to my huckleberry friend Enrica, who did a real contribution to the corrections of this thesis. My English could have been even worse without her.

I would also like to thank all of my friends who supported me in writing, and incited me to strive towards my goal, especially to Daniela, Matteo, Luca, Lucia, Trula, Mattia, Andrenzo, Michela and Federica.