

**Alma Mater Studiorum – Università di Bologna**

**DOTTORATO DI RICERCA IN  
BIOLOGIA CELLULARE, MOLECOLARE E INDUSTRIALE**

**Progetto n. 2: Biologia Funzionale dei Sistemi Cellulari e Molecolari**

**Ciclo XXII**

**Settore: CHIM/03**

---

**Molecular interactions:**

**metal ions and protein chaperones in the  
urease system from *Helicobacter pylori***

---

**Autore:**

Dott. Matteo Bellucci

**Coordinatore:**

Prof. Vincenzo Scarlato

**Relatore:**

Prof. Stefano Ciurli

**Esame finale anno 2010**



## Index

### A - Introduction

<b>Section A1 - Generalities of protein-protein interactions</b>	<b>7</b>
1.1. Protein diversity and evolution: from sequences to complexes	7
1.2. Classification of protein-protein interactions	9
1.3. Structural determinants of protein complexes: looking at interfaces	12
1.3.1. Interface size and shape	13
1.3.2. Interface composition: type of amino acids and interacting forces	14
1.3.3. Structural motifs and conformational changes at interface	16
1.3.4. Disorder as major component of PPIs	18
1.4. Kinetics and thermodynamics of PPIs	19
1.4.1. Determination of binding constant and thermodynamics parameters for PPIs	21
1.4.2. Calorimetrically determined thermodynamics of PPIs	23
1.5. Mutational and prediction investigation of PPIs: definition of “hotspots”	26
1.5.1. Disruption of protein-protein interfaces	26
1.5.2. Alanine scanning as a tool to discriminate hotspots: specificity and cooperativity	27
1.5.3. Computational alanine scanning to probe PPIs	29
<b>Section A2 - Metal-mediated PPIs: role of cations in protein assembly</b>	<b>30</b>
1.6. PPIs involving Ca <sup>2+</sup> -binding proteins: calmodulin and cadherin	30
1.7. Cu <sup>2+</sup> -binding proteins: PPIs involved in metal trafficking	34
1.8. Zn <sup>2+</sup> -binding proteins and PPIs: interface zinc sites and zinc finger domains	38
<b>Section A3 - Metal homeostasis is regulated by PPIs: the accessory systems of nickel enzymes urease and hydrogenase</b>	<b>50</b>
1.9 Metal trafficking: uptake, delivery and regulation	50
1.10 Urease and hydrogenase systems from <i>H. pylori</i>	55
1.10.1. Urease: structure and activity	56
1.10.2. [NiFe]-hydrogenase: structure and activity	61

1.11 Urease and [NiFe]-hydrogenase accessory systems	64
1.11.1. Urease genetic clusters in bacteria	65
1.11.2. Maturation of urease enzyme	66
1.11.3. Urease accessory proteins: structures and functions	68
1.11.4. [NiFe]-hydrogenase genetic organization in bacteria	74
1.11.5. Biosynthesis and activation of [NiFe]-hydrogenases	77
1.11.6. Accessory proteins for Ni <sup>2+</sup> -delivery and active site assembly of [NiFe]-hydrogenase	79

## **Aim of the study** **85**

## **B - Material and methods**

### **Section B1 - Experimental procedures** **89**

2.1. General molecular biology techniques	89
2.1.1. Standard DNA amplification of the genes of interest	89
2.1.2. Standard ligase reaction	90
2.1.3. Transformation of the <i>E. coli</i> cells for cloning and expression procedures	91
2.1.4. Selection of the positively transformed cells	91
2.1.5. Cloning overview	92
2.1.6. Site-directed mutagenesis protocol	94
2.2. Bacterial growth and heterologous protein expression	95
2.2.1. Culture media	95
2.2.2. <i>HpUreE</i> expression protocol	97
2.2.3. <i>HpUreG</i> expression protocol	97
2.2.4. <i>HpHypA</i> expression protocol	97
2.2.5. TEV protease expression protocol	98
2.3. Protein purification	98
2.3.1. Wild-type and mutated <i>HpUreE</i> purification protocol	98
2.3.2. Wild-type and mutated <i>HpUreG</i> purification protocol	99
2.3.3. <i>HpHypA</i> purification protocol	99
2.3.4. TEV protease purification protocol	100

2.4. Preliminary protein characterization	100
2.4.1. Evaluation of protein purity and molecular mass using SDS-PAGE electrophoresis	100
2.4.2. Determination of protein identity by tryptic digestion and mass spectrometry	101
2.4.3. Determination of protein concentration and metal content of protein samples	102
2.4.4. Determination of metal content in protein samples	102
2.5. <i>HpUreE</i> biochemical and structural characterization	103
2.5.1. Circular dichroism spectroscopy	103
2.5.2. NMR spectroscopy experiments	103
2.5.3. Light scattering measurements	103
2.5.4. ITC microcalorimetry experiments	104
2.6. Characterization of the <i>HpUreE</i> - <i>HpUreG</i> protein-protein interaction	105
2.6.1. Native PAGE electrophoresis	105
2.6.2. Light scattering measurements	105
2.6.3. Thermal denaturation scanning	105
2.6.4. NMR spectroscopy experiments	106
2.6.5. ITC microcalorimetry experiments	107
2.6.6. Measurement of GTPase activity for <i>HpUreG</i> in complex with <i>HpUreE</i>	107
2.7. Characterization of the <i>HpUreE</i> - <i>HpHypA</i> protein-protein interaction	107
2.7.1. ITC microcalorimetry experiments	107
<b>Section B2 - Biocomputing procedures</b>	<b>108</b>
2.8. Structural modeling of the <i>HpUreE</i> - <i>HpUreG</i> complex	108
2.8.1. Sequence search and alignment of UreE proteins	108
2.8.2. <i>HpUreE</i> structure prediction and homology modeling	108
2.8.3. <i>HpUreG</i> homology modeling	108
2.8.4. <i>HpUreE</i> - <i>HpUreG</i> molecular docking	108
2.8.5. Calculation of electrostatic potential for the he <i>HpUreE</i> - <i>HpUreG</i> structural model	109
2.9. Structural modeling of the <i>HpUreE</i> - <i>HpHypA</i> complex	109
2.9.1. Sequence search and alignment of HypA proteins	109
2.9.2. <i>HpHypA</i> dimeric structure prediction and homology modeling	109
2.9.3. <i>HpUreE</i> - <i>HpHypA</i> molecular docking	110
2.9.4. Calculation of electrostatic potential for the he <i>HpUreE</i> - <i>HpUreG</i> structural model	110

## C - Results and discussion

<b>Section C1 - Experimental results</b>	<b>113</b>
3.1. Protein cloning, expression and purification	113
3.1.1. Cloning and site-directed mutagenesis of <i>HpUreE</i>	113
3.1.2. <i>HpUreE</i> heterologous expression and purification	114
3.1.3. Heterologous expression and purification of <i>HpUreG</i>	117
3.1.4. Cloning of native, His <sub>6</sub> -tagged and His <sub>6</sub> -GB1-tagged <i>HpHypA</i>	118
3.1.5. Heterologous expression and purification of <i>HpHypA</i>	119
3.2. Biochemical and structural characterization of <i>HpUreE</i>	121
3.2.1. Determination of <i>HpUreE</i> structural properties	121
3.2.2. <i>HpUreE</i> metal binding properties	124
3.2.3. Influence of metal binding on <i>HpUreE</i> structure	127
3.3. Characterization of the <i>HpUreE</i> - <i>HpUreG</i> interaction	127
3.3.1. Preliminary characterization of the <i>HpUreE</i> - <i>HpUreG</i> interaction	127
3.3.2. Elucidation of the <i>HpUreE</i> - <i>HpUreG</i> complex formation	128
3.3.3. The role of Zn <sup>2+</sup> in the stabilization of the <i>HpUreE</i> - <i>HpUreG</i> complex	130
3.4. Preliminary characterization of the <i>HpUreE</i> - <i>HpHypA</i> interaction	134
<b>Section C2 - Biocomputing results</b>	<b>136</b>
3.5. Molecular modeling of the <i>HpUreE</i> - <i>HpUreG</i> complex	136
3.6. Molecular modeling of the <i>HpUreE</i> - <i>HpHypA</i> complex	138
3.6.1. HypA multiple sequence alignment construction and analysis	142
3.6.2. Structural model of dimeric <i>HpHypA</i>	142
3.6.3. Molecular models of the <i>HpUreE</i> - <i>HpHypA</i> interaction	149
<b>D - Conclusions</b>	
<b>Section D -Conclusions</b>	<b>149</b>
<b>Bibliography</b>	<b>151</b>

---

## **A - Introduction**

---



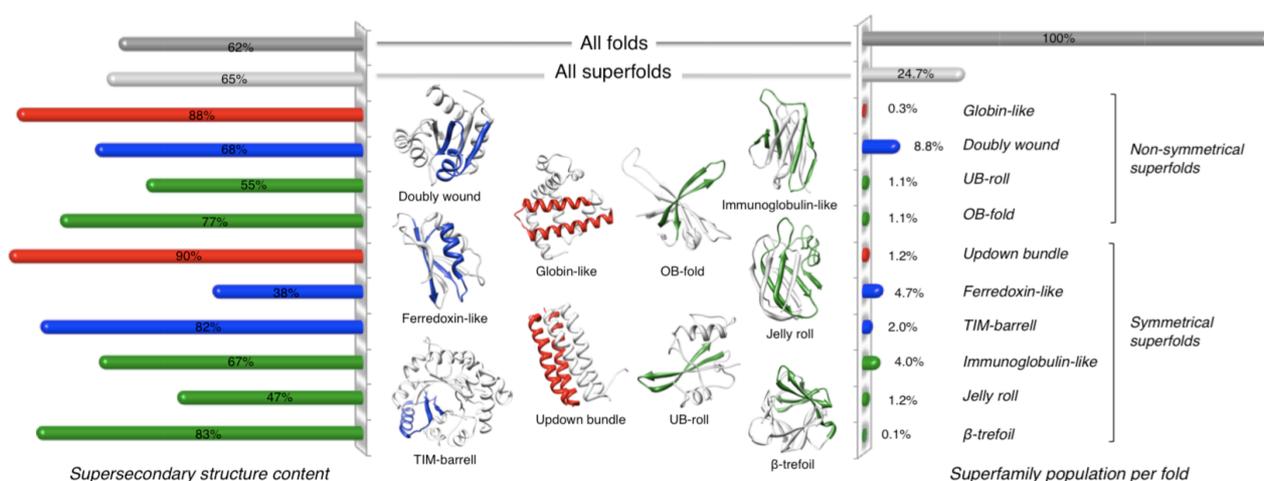
## Section A1 - Generalities of protein-protein interactions

### 1.1. Protein diversity and evolution: from sequences to complexes

In the highly crowded environment of a living cell, biological macromolecules occur at a concentration of 300-400 g l<sup>-1</sup> and they physically occupy a significant fraction (typically 20-30%) of the total volume [1]. Proteins constitute the large majority of these molecules and are actively involved in virtually every cellular process, such as biochemical catalysis, structural or mechanical functions, regulation of gene expression, signal transduction, molecules' uptake and transport and biosynthesis reactions. This massive variety of functions probably relies on the fact that proteins are extremely diverse. But how much proteins differ? It is possible to estimate that the number of unique protein sequences could round on 10<sup>10</sup>-10<sup>13</sup>, simply considering the number of species on Earth (10<sup>7</sup>-10<sup>8</sup>), most of which are microbial [2], and the amount of protein-coding genes per genome (10<sup>3</sup>-10<sup>5</sup>) [3]. But this huge range of protein variants represents just a little fraction of the total possible amino acid permutations (10<sup>321</sup> - 10<sup>469</sup> arrangements), considering average protein length in genomes [4]. Moreover, these calculations don't take into account intra-species variations. If we consider that 5 · 10<sup>30</sup> microbial cells in our planet (representing ~70% of life in certain habitats) have a turnover rate of 8 · 10<sup>29</sup> cells per year [2] and that mutations occur at rate of 4 · 10<sup>-7</sup> per microbial cell per generation [5], we would expect up to 2 · 10<sup>32</sup> total amino acid changes in microbial proteins in the 4 billion-year-long history of life. This represents a still minute contribute to sequence variability, even in the case that all mutations would occur in protein sequences that fold successfully. Accordingly, the majority of changes in proteins result from point mutations, which very rarely affect the overall structure significantly. In fact, we have also to remind that proteins organize into energetically favored three-dimensional shapes, based on their amino acid composition: these well-packed highly ordered architectures embed protein function.

Since the 1950s [6] it has been known that protein structure is hierarchical and composed by four level of complexity: *i*) primary structure, the linear sequence of amino acid linked by peptide bonds; *ii*) secondary structure, the local spatial conformation of the polypeptide backbone that originate helix, sheet and turn elements; *iii*) tertiary structure, the actual arrangement of the protein molecule stabilized by side-chain interactions between the secondary structure elements; *iv*) quaternary structure, the assembly of several polypeptide chains into an ordered supramolecular unit. Moreover, the observation of redundancy and modularity in protein architectures leads to the addition of new levels in terms of structural organization, such as supersecondary structures, recurrent structural motifs ( $\beta\beta$ -hairpins,  $\alpha\alpha$ -hairpins,  $\beta\alpha\beta$  elements, see Figure 1) sometimes

repeated in tandem to organize folds and protein domains, distinct compact folding units acting as structural modules that appear singly or in combination with other domains in multi-domain proteins [7, 8]. The structures of domains and supersecondary architectures reflect the evolutionary process that shaped them and retain traces of common ancestry; this criteria was used to arrange structures into families, superfamilies and folds, the latter grouping analogous convergent evolved superfamilies [9]. This classification showed that ~25% of all domains with known structure assume one of ten folds, named superfolds, and that the most frequent supersecondary structures are  $\beta\beta$ -hairpins,  $\alpha\alpha$ -hairpins and  $\beta\alpha\beta$  elements, which constitute more than 60% of the average protein structure (Figure 1).



**Figure 1 - Supersecondary structural elements and common superfolds**

The central panel shows the most common superfolds; a representative structure for each superfold is shown as an example:  $\beta$ -trefoil (PDB entry: 4fgf); jelly-roll (PDB code: 1goh); immunoglobulin-like (PDB entry: 1jp5); TIM-barrel (PDB entry: 1hti); ferredoxin-like (PDB entry: 1aps); updown bundle (PDB entry: 1rpr); OB-fold (PDB entry: 1qvc); UB-roll (PDB entry: 1lkk); globin (PDB entry: 1ebc); doubly wound (PDB entry: 5chy). The structural elements are colored as follow:  $\alpha\alpha$ -hairpins in red,  $\beta\alpha\beta$  elements in blue,  $\beta\beta$ -hairpins in green. Side panels show the fraction of residues contained in the supersecondary structural elements (graph on the left) and the number of superfamilies grouped in each fold (graph on the right).

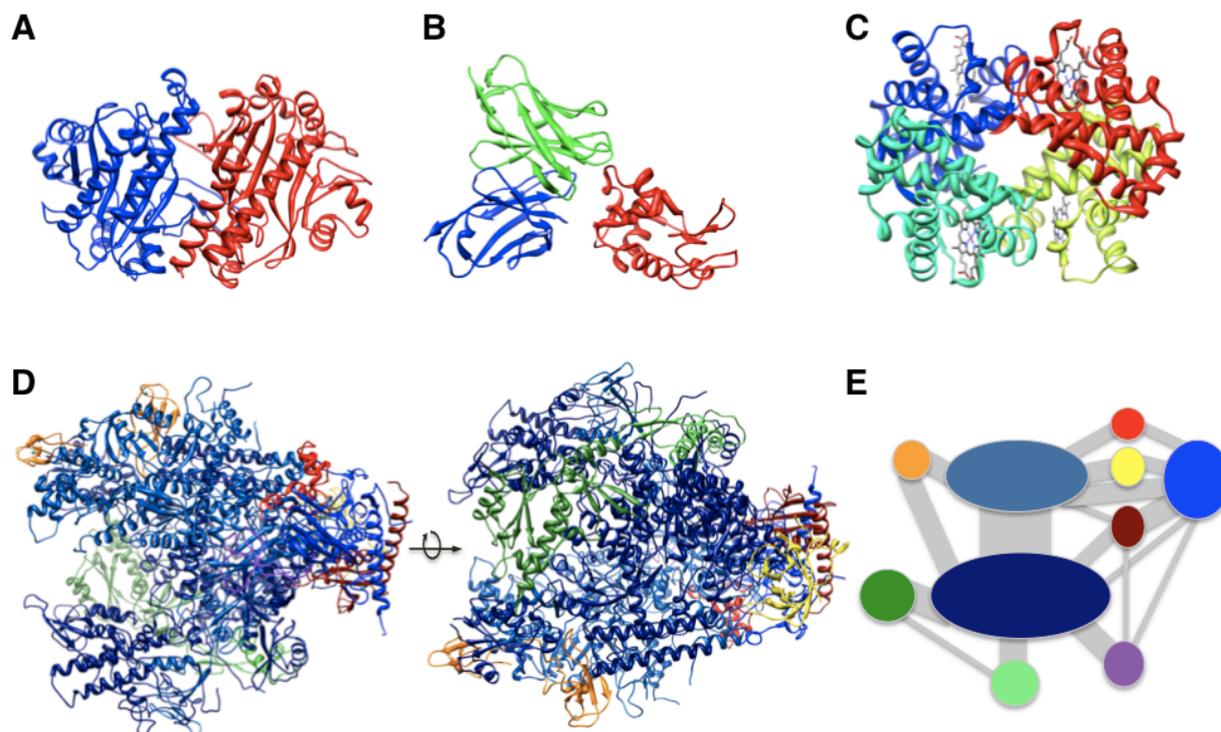
From this data, it becomes clear that structure variability is much lower than sequence diversity, suggesting that the three-dimensional architecture was more highly conserved through evolution. In particular, it seems that proteins prefer a limited number of peculiar folds, probably due to stability and folding efficiency reasons; indeed some folds were selected among the others simply because these structures represent better scaffolds for the establishment of active sites.

Finally, when talking about protein diversity, we have also to consider the last level of structural organization, which is represented by complexes of functionally related proteins. The tremendous relevance of protein assembly is given by the fact that every major process in the cell is carried out by multi-protein complexes of ten or more molecules, each of these interacting with several other large protein assemblies [10]. For instance, mutations that cause lacking of correct interactions

between related proteins lead to pathology in human; this is the case of neurological diseases such as Creutzfeldt-Jacob and Alzheimer's [11], familial Mediterranean fever [12], immunodeficiency-centromeric instability-facial anomalies syndrome (ICF) [13], adult respiratory stress syndrome (ARDS) [14] and emphysema [15], to name a few. In this view, it is consistent to assume that proteins must co-evolve: any divergent changes in one protein surface are complemented at the interface by their interaction partner(s) [16]. On this basis, it is reasonable to hypothesize that protein differentiation has occurred as a structure- and function-driven process by which polypeptides with different amino acid sequence converged to a little number of recurrent folds, constituting the protein structure vocabulary. This restricted subset of folds was maintained through selective pressure played by their interaction partner(s), following the criteria that a missing interaction could lead to loss of function. In this scenario, protein-protein interactions (PPIs) evolve to optimize functional efficacy: large complexes may reflect the need for stability, especially in the case of weak interactions which are strictly controlled by the establishment of a highly structured assembly.

## 1.2. Classification of protein-protein interactions

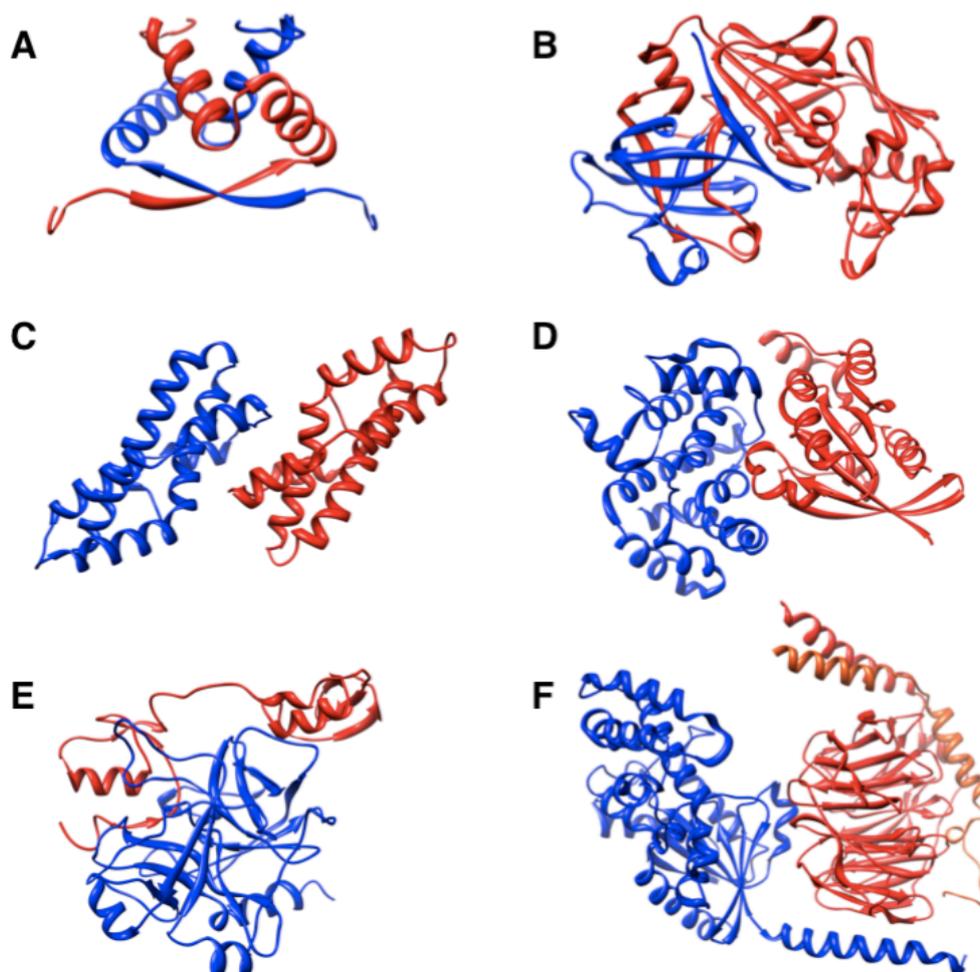
PPIs constitute the basis of the quaternary structure of multi-meric proteins, and represent one of the highest levels of structural organization in biological molecules [17]. Experimental techniques such as protein X-ray crystallography, nuclear magnetic resonance spectroscopy and cryoelectron microscopy have provided atomic details of many protein-protein assemblies. Among these structural architectures it is possible to distinguish binary and multi-subunit protein complexes (Figure 2). Depending on the amino acid sequence of the polypeptide chain(s), the protein complexes can be further subdivided into homo-oligomers (Figure 2A,C), featuring identical chains, and hetero-oligomers (Figure 2B,D), possessing non-identical chains. Moreover, oligomers of identical proteins can be organized in an iso-logous or hetero-logous way [18]; iso-logous associations involve the same surface on both monomers, related by a two-fold symmetry axis, while the hetero-logous assemblies use different interfaces, that can lead to infinite (non-cyclic) aggregation patterns. In the higher order of the homo-oligomeric assembly, such as tetramer, hexamer, octamer, and dodecamer, etc., more than two identical subunits come in contact. Oligomerization occurs due to strong selection pressure for the evolution of monomeric proteins into oligomeric complexes, driven by benefits such as reduction of surface area, increased stability and novel function through inter-subunit communication.



**Figure 2 - Assembly of protein-protein complexes**

(A) Binary homo-oligomeric assembly of asparagine synthetase (PDB entry: 12as). (B) Hetero-trimeric assembly of FV fragment of monoclonal antibody D1.3 (light chain in blue, heavy chain in green), complexed with hen egg lysozyme (in red) (PDB entry: 1vfb). (C) Hemoglobin homo-tetrameric assembly (PDB entry: 1fsx). (D) Hetero-multimeric assembly of RNA polymerase II complex (PDB entry: 1i50); the structure is rotated by 180° around the horizontal axis for complete illustration of the 10 subunits. (E) Schematic interaction diagram of RNA polymerase II complex (color of subunits are the same as in panel D). The thickness of the connecting lines is proportional to the buried surface area in the corresponding subunit interface. For each structure, chains are distinguished using different colors.

Another classification of protein complexes can be made by considering whether the association is obligate or non-obligate [18]. Typically, the protomers that form homo-oligomers are not found as stable structures inside the cell and the complex formation occurs simultaneously during the folding process; this is the case of obligate complexes, that are generally also functionally obligate. An example is represented by the bacteriophage P22 *Arc* repressor (Figure 3A) [19], which is constituted by two identical chains that associate and fold simultaneously, and exists only as dimers in solution. On the other hand, the subunits of hetero-oligomers are often, but not always, stable as independent structural units inside the cell and they specifically interact only to carry out a precise function. Many of these complexes involve non-obligate interactions, such as intracellular signaling assemblies (Figure 3D) and antibody-antigen, receptor-ligand or enzyme-inhibitor complexes (Figure 3E). However, some homo-oligomers can also arrange into non-obligate assemblies, as in the case of sperm lysin (Figure 3C).



**Figure 3 - Protein-protein interaction classes**

Examples of different protein complexes, as described in the text, are shown: P22 Arc repressor (PDB entry: 1arr) as obligate homo-oligomer (A); human cathepsin D (PDB entry: 1lya) as obligate hetero-oligomer (B); sperm lysin (PDB entry: 3lyn) as non-obligate homo-oligomer (C); RhoA-RhoGAP complex (PDB entry: 1ow3) as non-obligate hetero-oligomer (D); thrombin-rhodniin inhibitor complex (PDB entry: 1tbr) as permanent hetero-oligomer (E); bovine G $\alpha$ -G $\beta$  $\gamma$  trimer (PDB entry: 1GP2) as non-obligate transient hetero-oligomer (F). For each structure, chains are distinguished using different colors (blue, red and orange).

Protein complexes can also be classified considering the lifetime, or the strength, of the interaction; some associations are permanent, usually very stable and thus only existing in their complexed form, whereas other are transient, continuously forming and dissociating *in vivo* [18, 20]. As a general statement, the tightness of a PPI is strictly related to the functional role played by the involved protein partners. Consequently, structurally and functionally obligate interactions are usually permanent, while non-obligate associations may also be transient. Furthermore, it is possible to distinguish between weak transient interactions, that feature dynamic oligomeric equilibrium in solution, and strong transient interactions, in which the binding to a molecular compound triggers an equilibrium shift stabilizing the oligomeric form. It is also important to note that many PPIs cannot be ascribed to a distinct type of association, especially because the viability of all protein complexes highly depends on the physiological conditions and environment. For

example, an interaction may be transient *in vitro* but become permanent under certain cellular conditions. Basically, all interactions are driven by the concentration of the protein partners and the free energy of the complex formation. As a consequence, PPIs can be controlled by altering the local concentration of the components or changing their binding affinity, determined by chemical and geometrical interface properties. In this view, concentration, pH, ligand binding, and other parameters play a crucial role in modulating the association equilibrium. In particular, three general mechanisms of PPIs regulation could be identified: *i*) control of the encounter between proteins or protomers, because association relies on the match of the interacting surfaces and requires co-localization in time and space; *ii*) control of the local concentration modulating gene expression or secretion levels, protein degradation, temporary storage, diffusion or viscosity; *iii*) control of the affinity of the complex components by the presence of an effector molecule (e.g. chemical ligands, metal ions, other proteins), covalent modifications (e.g. phosphorylation, glycosylation) or changes in physiological conditions (e.g. pH, temperature, ionic strength).

Actually, protomers that organize in obligate homo-oligomers, are often expressed simultaneously and are thus co-localized upon synthesis. This may also occur in the case of hetero-oligomers as reported for cathepsin D (Figure 2B), in which the genes encoding for the two non-identical subunits are controlled by the same promoter [21]. PPIs regulated by co-localization, such as receptor-ligand, enzyme-inhibitor and antibody-antigen interactions, are usually characterized by high affinity, as in the case of thrombin-rhodniin complex (Figure 3E) that features a dissociation constant in the nanomolar range [22]. These strong associations, once made, are permanent and irreversible, only perturbed by proteolysis. On the other hand, regulated transient interactions are dynamic and allow the precise control of protein networks in biology, changing the binding affinity between protomers or subunits by orders of magnitude. An appropriate example is represented by the G protein hetero-trimer (Figure 3F), which dissociates into  $G\alpha$  and  $G\beta\gamma$  subunits upon GTP binding but forms a stable complex in the GDP-bound form, exhibiting a 1000-fold increase of binding affinity [23].

### **1.3. Structural determinants of protein complexes: looking at interfaces**

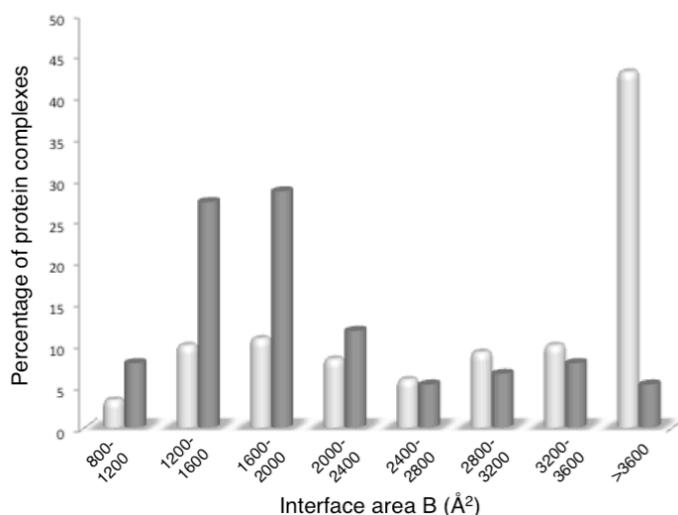
Given the functional diversity of PPIs, it is also possible to distinguish and classify protein assemblies from the knowledge of the structure. Proteins interact through their interfaces, which consist of interacting residues, belonging to different chains, together with some isolated residues in spatial proximity. In order to evaluate PPIs, it 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.

### 1.3.1. Interface size and shape

The size of protein-protein interface can be determined calculating the interface area in terms of solvent accessible surface area (SASA) of the interacting proteins, using the following equation:

$$B = \text{SASA}_{\text{protein1}} + \text{SASA}_{\text{protein2}} - \text{SASA}_{\text{complex}}$$

where B is the surface involved in the interaction, corresponding to the buried area after protein association, and the other terms are SASA value for the two proteins and for the protein complex respectively. This kind of calculations can be easily made for the crystal structures of available protein complexes. Many studies defined the size of the interface area in dimeric proteins, which ranges from 670 to 4760 Å<sup>2</sup>, representing the 6.6-23.3% of the accessible surface area of the individual monomers [24]. In particular, homo-dimers are on average 2-fold larger in size than hetero-dimeric complexes (Figure 4) [25]. Indeed, in multi-meric protein complexes the surface area involved in the interaction tends to be at the high end of, or above, the range for dimers. On average, trimers and tetramers contribute 17.4% and 20.9% of their accessible surface area to the contact interface respectively, while the mean contribution for dimers is 12% [26].



**Figure 4 - Size distribution of the protein-protein interfaces**

Interface distribution in homo-dimers (white columns) and hetero-dimers (grey columns). Data sets reported refer to the results of Bahadur and coworkers (see text and [25] for details).

In principle, specific PPIs would involve larger interfaces compared to non-specific associations, but the size of interacting surfaces on average range from 800 Å<sup>2</sup> to 10.000 Å<sup>2</sup> [17]; this implies that the extension of the resulting buried surface area alone is not yet sufficient to discriminate between specific and random contacts in protein complexes. However, the lack of protein complexes that feature interfaces with a size below 800 Å<sup>2</sup> suggests that the establishment of a stable interaction requires a defined number of contacts (on average, each partner contributes by 20 residues) and the removal of solvent molecules from a portion of the protein interacting surface.

In addition to the size of a protein-protein interacting surface, the shape of the interface could be used to better distinguish between specific and non-specific PPIs. Generally the majority of the interfaces involved in PPIs are more or less flat; with few exceptions the interfaces are

approximately circular areas on the protein surface in both permanent and transient complexes. Usually the interfaces in obligate associations tend to be less planar and closer packed with respect to non-obligate assemblies. At first, in order to assess the curvature of an interface, the planarity, a spatial derived parameter, could be analyzed [27]. The planarity of two interacting surface is obtained by calculating the root mean square deviation of all the interface atoms from the least square plane through the atoms. In other words, if all the atoms would exactly fit to a plane the planarity index would be zero. The average value of planarity index ( $3.5\pm 1.7$  Å for homo-dimers and  $2.8\pm 0.9$  Å for hetero-dimers, respectively) confirms that protein-protein interfaces are generally flat in shape. Moreover, another parameter that could be used to evaluate the shape of protein binding surface is the circularity [27]. A circularity value near 1.0 is obtained for an interface approximately circular, but in general this is not perfectly the case of the interacting surfaces, as reported for homo-dimeric and hetero-dimeric complexes (circularity index values of  $0.71\pm 0.17$  and  $0.73\pm 0.05$  respectively).

Finally, is it possible to evaluate the specificity of PPIs through the shape complementarity, as a measure of interfacial packing in protein complexes; this could be done by relating the volumes of the interface cavities to the surface area involved in the interaction. At this purpose, Laskowski defined the GV (Gap Volume) index, which corresponds to the volume of interface cavities normalized with respect to the buried surface area. The average value of GV index is  $2.1\pm 1.2$  Å for homo-dimers and  $2.5\pm 1.0$  Å for hetero-dimers. Small GV values indicate that proteins in the complex are well-packed, as in the case of dimeric alkaline phosphatase that features a value of  $1.09$  Å. A combination of interface area and GV index could be used to effectively distinguish specific interface from non-specific crystal-packing interface, the latter showing much higher GV index value (on average  $4.4\pm 1.9$  Å). Moreover, other two packing indexes, namely LD (Local Density) and GD (Global density), were proposed by Bahadur and coworkers to evaluate shape complementarity in protein complexes [25]. LD index measures the packing density at each point of the interface and it is defined as the number of interface atoms within  $12$  Å of another interface atom. GD index measures the atomic density at the interface atoms normalized to the size of the interface. The average values of LD index for specific interactions range from 42 in hetero-dimers to 45 in homo-dimers (neighboring interface atoms), and is almost 30% lower in non-specific contacts such as crystal-packing interfaces. Analogously, GD index in specific protein assemblies is significantly higher ( $1.30$  and  $1.32$  for homo-dimers and hetero-dimers respectively) compared to non-specific crystal packing ( $0.96$ ).

### **1.3.2. Interface composition: type of amino acids and interacting forces**

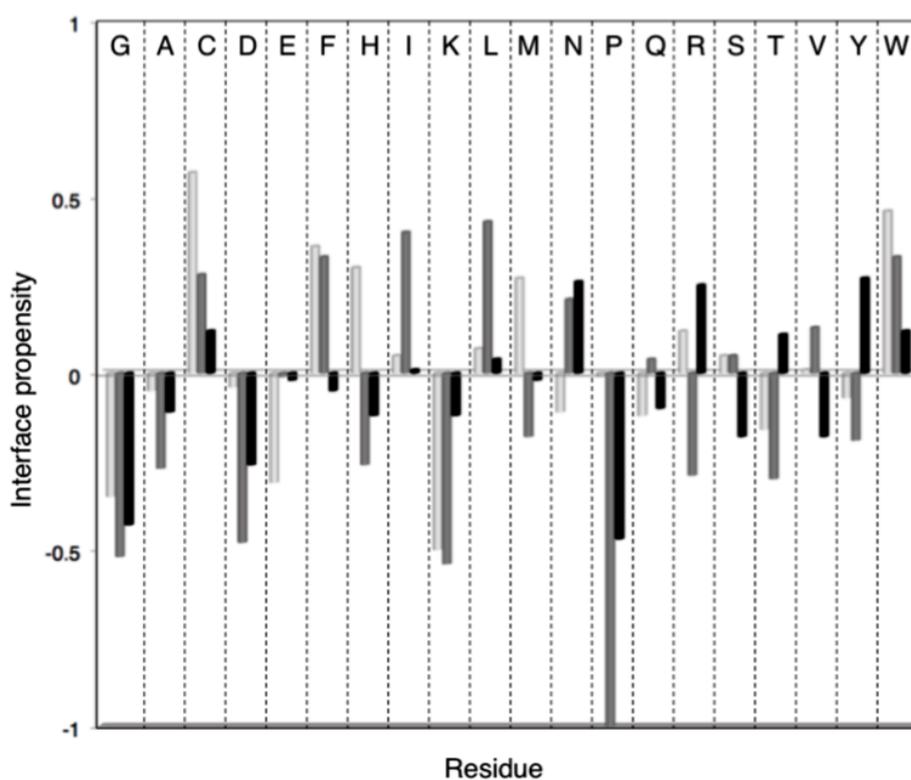
In general, PPIs are frequently mediated by hydrophobic effects [28], but also hydrogen bonds, electrostatic interactions and van der Waals attractions play a considerable role. In particular, it has

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 [26]. Hydrogen bonds between protein chains are more favorable with respect to those made with water molecules. It has been estimated that, upon association, about one hydrogen bond is formed per 170 Å<sup>2</sup> buried surface, meaning that a standard size interface mediates  $10 \pm 5$  H-bonds [29]. The occurrence of salt bridges in protein complexes is lower; for instance, only 56% of homo-dimers were found to possess electrostatic interactions, many of them having none, and at the most five. However, although high ionic concentration can screen electrostatic contacts, salt bridges can influence the rate of collision between protein partners by pre-orienting the molecules to selectively promote a correct assembly. On the other hand, at the interface level Van der Waals interactions are no more energetically favorable than those made with the solvent; however these weaker contacts are more numerous at the interface than H-bonds and can contribute significantly to the binding energy of the association [17].

The average amino acid composition of a protein complex interface (47% hydrophobic, 31% polar and 22% charged) reflects the character of the forces which guide protein-protein associations and greatly differs from the rest of the protein surface [26]. As a general statement, both homologous and heterologous assemblies show interfaces enriched in aliphatic (Leu, Val, Ile, Met) and aromatic (His, Phe, Tyr, Trp) residues, and depleted in charged amino acids (Asp, Glu, Lys) other than arginine [17]. In fact, despite of the high abundance of lysine on the protein surface, Lys is largely excluded from interfaces which mediate specific interactions. This preference is probably dictated by the greater capability of the guanidinium group in Arg to form H-bonds compared to the amino group of Lys. Moreover, there are some clear preferences for certain amino acids; tyrosine, after arginine, seems to be the favored type of residue. This fact can be explained considering that aromatics, in particular Tyr, provide an efficient way to cover large amount of solvent exposed surface without paying too much of a price in destabilization of the protein native state. However, this preference appears to be related to the type of interface. For example, this is less clear in homo-dimers than in antibody-ligand complexes, probably because homo-dimers are associated as a dimers for most of their time whereas both the antibody and the protein ligand it binds need to fold as a stable monomer. Similarly, methionine is much more present at the interface than elsewhere in the protein structure. This prevalence is clear for calmodulin, which binds to many different proteins and shows eight exposed Met residues in its binding site [30]. It has been proposed the flexibility of the Met side chain and the polarizability of the sulfur group allow the interaction of calmodulin with different protein partners [31]. This evidence that methionine-aromatic interactions are particularly favored due to the presence of sulfur group, which increase the favorable enthalpy when interacting with non-polar surfaces.

Finally, the relative propensity of the 20 amino acids to be at the interface of PPIs has been reported by Keskin and coworkers [20]. In particular the authors derive these values by clustering a

data set of structurally and sequentially non-redundant protein complexes into three different classes (Figure 5): *i*) Type 1 (358 complexes), in which the global folds of the parent chains are similar and the functions of the members of the cluster is also similar; *ii*) Type 2 (94 complexes), including members that often do not share similar functions and do not have globally similar structures; *iii*) Type 3 (367 complexes), grouping not-functionally related members with similar binding sites on one side of the interface but different protein partners.



**Figure 5 - Propensity of the contacting residues in different interface types.**

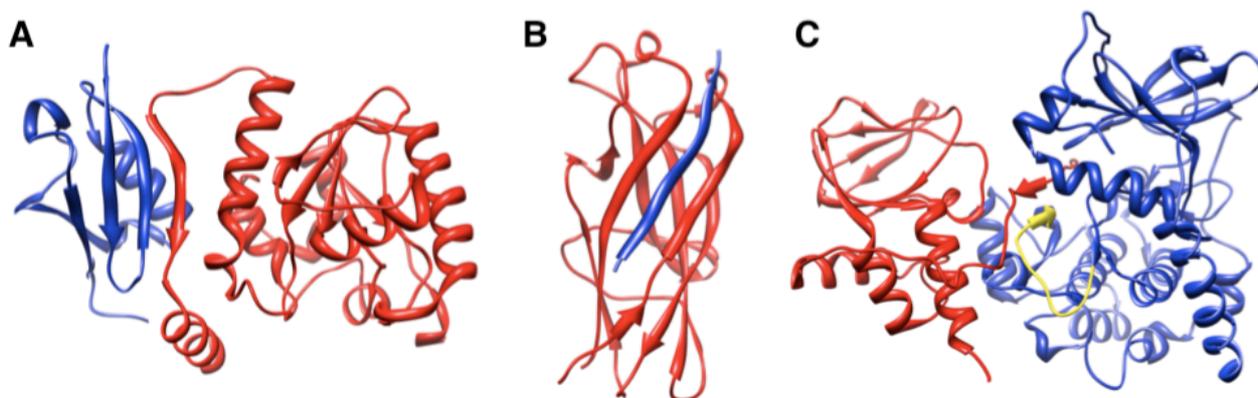
The logarithmic propensity of the 20 residues for the different interface types is shown as bars (Type 1 light grey, Type 2 grey, Type 3 black). A positive value indicates a favorable propensity in the interface as compared to the rest of the protein, whereas a negative score that is less likely to find the residue in the interface. The three types refers to different classes of protein complexes, as described in the text.

Overall, these data suggest that is not possible to find recurring chemical features that dictate PPIs. At the same time there is no residue composition at the interface differentiating between specific and functionally related protein associations versus loners. Nevertheless, some general trends are confirmed; aromatic and aliphatic residues, especially Phe, Trp, Ile and Leu, feature positive propensity in all classes of complexes, while polar and charged amino acids, except Arg, are not favored as interacting residues at the interface. Indeed a preference for certain residues is shown by particular classes of complexes, such as His for Type 1 and Arg for Type 1 and 3.

### 1.3.3. Structural motifs and conformational changes at interface

A large number of different PPIs motifs are possible but no strong tendency for certain types of secondary structure elements has been clearly found [32]. Independent studies reported contradictory data about it: in one case the average contribution of loop interactions was estimated as 40% of the interface contacts while other analysis reported that 53% of the interface residues were  $\alpha$ -helices, 22%  $\beta$ -sheets and 12%  $\alpha\beta$  elements, with the rest being coils [26]. However,

clustering analysis identified closely structural related patterns in a large number of proteins, with leucine zipper and helix-loop-helix being two of the better known motifs [33]. Another well characterized example is represented by PPIs that involve association of a  $\beta$ -strand from a protein ligand with a  $\beta$  element in the binding protein partner, also called “ $\beta$ -strand additions” [34]. These associations, that do not include  $\beta$ -sheet- $\beta$ -sheet juxtaposition, have been classified in three different classes (Figure 6): *i*)  $\beta$ -sheet augmentation, where the interaction is mediated by a strand from one of the proteins to the edge of a sheet in the other partner; *ii*)  $\beta$ -strand insertion and complementation, where the interaction is mediated by a strand from one protein inserting itself into the fold of another; *iii*)  $\beta$ -strand zipping, where unstructured loop regions from each binding partner come into contact to fold a two-stranded  $\beta$ -sheet or  $\beta$ -zipper.



**Figure 6 - PPIs through  $\beta$ -strand addition**

Examples of the three classes of  $\beta$ -strand addition are reported: (A)  $\beta$ -sheet augmentation, in SUMO1-thymine DNA glycosylase complex (blue and red respectively)(PDB entry: 1wyw); (B)  $\beta$ -strand insertion and complementation, in the complex between PapK (red) and the N-terminal extension peptide of PapE (blue) (PDB entry: 1n12); (C)  $\beta$ -strand zipping, in the complex of PKA (blue with A-loop highlighted in yellow) with its regulatory subunit RI $\alpha$  (red) (PDB entry: 1u7e).

More generally, the secondary structural distribution of interface in protein complexes resembles the structural content found in the exterior rather than interior residues of proteins [32]. Moreover, even if sheet-sheet, helix-helix and/or helix-sheet interactions appear in virtually all interfaces of protein complexes they do not make up the majority of residues or the majority of the surface area of the interface.

PPIs could also be affected by the conformation of the protein components: structural changes may mediate signaling events or trigger allosteric effects [17]. Large rearrangements in backbone and side chain conformation can occur upon complex formation but the extent of these changes can be evaluated only when three-dimensional structures of all individual proteins and that of the final assembly are available. A reasonable measure of moderate changes is the root-mean square distance (RMSD) of the main chain atoms after superimposition of free and bound structures. The majority of protein complexes feature RMSD values in the range 0.5 - 1.0 Å, with the only exception of antibody-antigen complexes, which undergo main chains movement of 1 - 2 Å; at the

interface the main chain can often move by 1 - 2 Å and a few surface side chains reorient [29, 35, 36]. However, there are associations where large movements take place (RMSD between bound and unbound protein partners over 2 Å). In that case, conformational changes modify the shape and the chemical features of limited areas on the proteins surface. This phenomena, often called “induced fit”, is localized close to loop regions at the interface or it may affect the whole protein architecture, like the movement of entire domains upon interaction. Generally, it is possible to assume that all complexes with interface area larger than 2000 Å<sup>2</sup> undergo large conformational changes upon association [17]. This can be explained considering that, for larger interfaces, the pre-formation of a stable protein interface that exactly fits the protein partner is more difficult than for a smaller interface. As a consequence, larger interfaces need a greater capability for conformational adaptation to generate a stable protein complex. This is the case of intertwined homo-dimers, non-obligate complexes that form an induced-fit permanent association such as the thrombin-rhodniin complex (contact area of 1740 Å<sup>2</sup>; Figure 3D) and the transient hetero-trimeric G protein assembly (contact area of 1160 Å<sup>2</sup>; Figure 3F).

#### **1.3.4. Disorder as major component of PPIs**

A large fraction of cellular proteins are estimated to be natively disordered [37, 38]; they play crucial roles in cell-cycle control, signal transduction, transcriptional and translational regulation, and large macromolecular complexes [39, 40]. Disordered proteins, also named IUPs (intrinsically unstructured proteins), lack a stable defined structure and exist in a series of conformations, from the less to the more structured states. Natively unstructured proteins undergo a disorder-to-order transition upon binding their physiological partner, but the global fold of disordered proteins does not change upon binding [37]. In fact, in many cases these unstructured regions constitute only delimited parts or domains of a whole protein.

Disorder is often implicitly associated with lack of information, considering that loss of structure means loss of the necessary information to perform a specific function. However, this paradigm is denied by the occurrence of IUPs performing key functions, and implies that the information content required to be functional is compatible with the co-existence of multiple conformers or structures. The question then arises as to how the information is stored in IUPs. Considering that many, although not all, disordered proteins adopt a folded conformation in the presence of a physiological partner, the real information is the one that the protein complex explicate after association.

Interestingly, a growing number of PPIs are found to be mediated by a large globular region in one protein binding to a comparatively short, peptide stretch, named “linear motif” in another [41]. Linear motifs are short patterns of around 10 residues. They are frequently found in disordered or unstructured regions, which are now known to be not simply loops or linkers, but serve a variety of

functions, and adopt a well-defined structure only upon binding [42]. Their short length and the fact that they often reside in disordered regions in proteins makes them difficult to detect through sequence comparison or experiment. Nevertheless, each new motif provides critical molecular details of how interaction networks are constructed, and can explain how one protein is able to bind to very different partners. Some examples are the SH3 (Src-homology-3) and WW domains which bind to proline-rich regions, the SH2 (Src-homology-2), PTB (phosphotyrosine binding) and 14-3-3 domains which bind to phosphorylated peptides, the PDZ domain and many others [43].

#### 1.4. Kinetics and thermodynamics of PPIs

Specific, rapid protein interactions are fundamental steps that guide many processes in life. The association between proteins begins, as for any other biological molecules, by a random search of the two interacting partners for each other within the space of solution, followed by a precise matching of their interfaces. Random interaction is dictated by Brownian motion, featuring a constant rate of  $7 \cdot 10^9 \text{ M}^{-1}\text{s}^{-1}$ , given by the Smoluchowski-Einstein equation [44]. However, a random collision between two proteins doesn't lead to complex formation, as their relative orientations add spatial restraints that slow down the reaction rate by three to five orders of magnitude [45]. In general, typical association rates are in the order of  $10^5$ - $10^6 \text{ M}^{-1}\text{s}^{-1}$  [46], but rate constants of  $>10^9 \text{ M}^{-1}\text{s}^{-1}$  have been measured for PPIs involving favorable electrostatic forces [47].

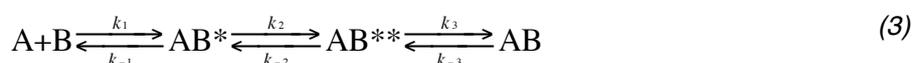
PPIs can be simply described as chemical reactions of the form (1):



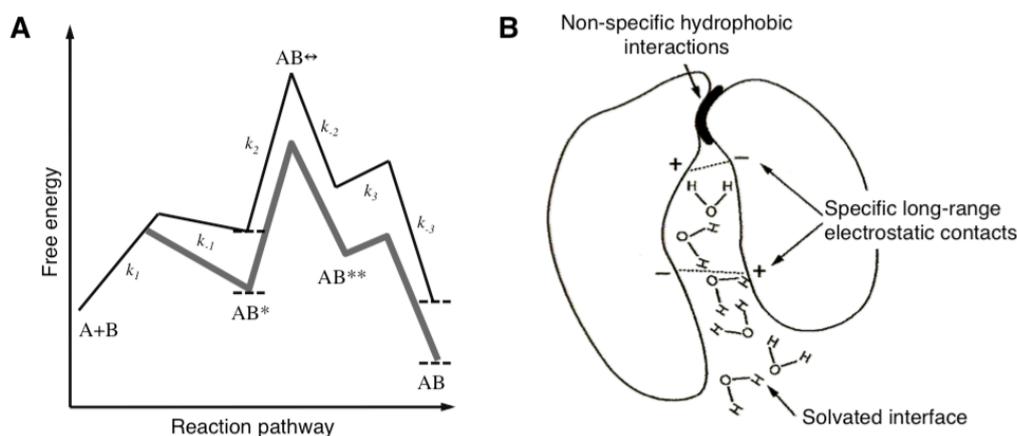
where A and B represent two proteins, or monomers, and AB the final complex. Multi-protein complexes are usually thought to be formed by adding subunits successively: therefore the establishment of multi-meric assembly could be viewed as a stepwise process including several binary interactions. Quantitatively, the association follows the mass action law (2):

$$\frac{[A][B]}{[AB]} = \frac{k_d}{k_a} = \frac{1}{K_a} = K_d \quad (2)$$

with  $k_a$  as second-order rate constant for the association reaction,  $k_d$  as first-order rate constant for the dissociation reaction,  $K_a$  and  $K_d$  as equilibrium constants for the association and dissociation process, respectively. The association of a protein complex can be best described using a four-state model (3) (Figure 7A):



In this scheme A and B represents two proteins that form in solution an initial unstable encounter complex by diffusion ( $AB^*$ ), which tends to re-dissociate ( $k_{-1} \gg k_2$ ). The protein complex  $AB^*$  evolves into an intermediate ( $AB^{**}$ ) that is already committed to form the final complex ( $k_3 \gg k_{-2}$ ) [46]. Describing the reaction in this way allows the existence of a mostly solvated encounter complex before the main transition and an intermediate state ( $AB \leftrightarrow$ ) past the transition.



**Figure 7 - Schematization of protein-protein complex formation**

(A) Free energy profile describing the of protein-protein complex ( $AB$ ) from the free protein partners  $A$  and  $B$ , via the encounter complex  $AB^*$ , the transition state  $AB \leftrightarrow$  and the intermediate assembly  $AB^{**}$  (see scheme 3). Thin black and thick grey lines represent PPIs in the absence and in the presence of favorable electrostatic forces, respectively. (B) Representation of the transition state for the interaction. Specific electrostatic interactions that guides the formation of the final protein complex are schematized.

Favorable electrostatic contacts has long been recognized as a driving force for fast association [48] (see the free-energy profile in the absence and in the presence of electrostatic forces in Figure 7A). The magnitude of these attractions can be altered by changing the ionic strength of the solution. The relationship between ionic strength and  $k_{on}$  was shown to follow the equation (4) [49]:

$$\ln k_{on} = \ln k_{on}^0 - \frac{U}{RT} \left( \frac{1}{1 + \kappa a} \right) \quad (4)$$

where  $k_{on}$  and  $k_{on}^0$  are the rates of association in the presence and absence of electrostatic forces, respectively,  $U$  is the electrostatic energy of interaction,  $\kappa$  is the inverse Debye length and  $a$  is the minimal distance of approach. As a consequence,  $k_{on}$  is the sum of the basal rate of interaction in the absence of electrostatic forces ( $k_{on}^0$ ) and the contribution of the electrostatic forces between proteins. This kind of linear relation was demonstrated to hold for the interaction between TEM1 and BLIP ( $\beta$ -lactamase inhibitor protein) [50], interferon-receptor association [51], hirudin-thrombin complex [49], barnase-barnstar assembly [52] and a hetero-dimeric leucine zipper [53], for all salt concentrations investigated.

For PPIs both encounter and intermediate complexes have been observed, even if these pre-complexes are often difficult to track experimentally. However some interactions follow a three-state or a two-state model. Sydor and coworkers have demonstrated the existence of a diffusion encounter complex for the interaction between Ras and the Ras-binding domain of c-Raf1 [54]. In that case, a two-step dissociation process was proposed, and the Ras-Raf1 intermediate ( $AB^*$ ) was suggested to dissociate faster than complexation occurs ( $k_{-1} \gg k_2$ ). Generally, for most PPIs a diffusion encounter complex was not detected but this doesn't exclude its existence at high protein concentrations (at least hundreds of  $\mu\text{M}$ ). On the other hand, the intermediate ( $AB^{**}$ ) is formed after the rate-limiting step for association and could be considered as a partially formed complex that has to reorganize in order to establish the final complex. In that state chemical bonds are in the process of being made and broken: the transition assembly is stabilized by electrostatic interactions and its structure resembles that of the final complex, but is mostly solvated (Figure 7B). In this case, the high-energy barrier for association is represented by the establishment of specific short-range interactions, which is accompanied by structural rearrangement and desolvation. Transition states have been characterized for the interaction between cystatin A and papain, where the reorganization step is fast ( $230 \text{ s}^{-1}$ ) [55], and for the interaction involving HEL (hen-egg lysozyme) and the antibody fragments HyHEL-10 or HyHEL-26, which show a slower rearrangement process ( $\sim 10^{-3} \text{ s}^{-1}$ ) [56].

#### 1.4.1. Determination of binding constant and thermodynamics parameters for PPIs

The determination of the association/dissociation constant is probably the first aim in the detailed study of any protein-protein assembly.  $K_d$  values, more often used compared to  $K_a$ , are very useful to determine at what protein concentration a complex might be formed and represents a common parameter to classify the nature and the strength of a considered interaction. The range of  $K_d$  values observed in biologically relevant processes that rely on PPIs is extremely wide and extends over at least 12 order of magnitude, from up to micro-molar ( $10^{-4} \text{ M}$ ) to less than picomolar ( $10^{-16} \text{ M}$ ) [26]. The latter is the case of many dimeric interactions in which monomers have to be denatured to dissociate the protein complex. More generally, dissociation constants in the milli- or micro-molar range are typical of weak interactions whereas  $K_d$  values in the nano-molar range or below define strong associations. However, the biological strength may depend on other effects such as protein cooperativity; for example several weak interactions between the subunits or the protein partner of a complex may still result in a highly stable assembly.

An association constant is dictated by the Gibbs free energy difference ( $\Delta G$ ) between the bound and unbound states of the proteins at the equilibrium, but the  $\Delta G$  of complex formation is only one part of the thermodynamics. The change in enthalpy ( $\Delta H$ ), entropy ( $\Delta S$ ) and heat capacity ( $\Delta C_p$ ) all provide useful information about the importance of several factors involved in PPIs [32]. The

free energy difference is defined by equations (5):

$$\Delta G = -RT \ln K_a \quad \Delta G = \Delta H - T\Delta S \quad (5)$$

In combination  $\Delta G$  and  $\Delta H$ , yielded experimentally, allow the calculation  $\Delta S$  at given temperature. The formation of the protein-protein complex is said to be entropy-driven if  $\Delta H$  is negative (favoring association) and  $\Delta S$  positive (disfavoring association) and entropy-driven otherwise [26]. Known values of free energy change for PPIs ranges from -6 to -19 kcal mol<sup>-1</sup> [20], i.e. 19 kcal are required to separate 1 mole of trypsin-pancreatic trypsin inhibitor [57]. From a thermodynamic point of view, a single pairwise interaction between amino acids may account for as much as 6 kcal mol<sup>-1</sup>. Particularly, residue pairs which form salt bridges and charged hydrogen bonds yield the largest contribution; pairs making neutral hydrogen bonds or non-polar interactions are in the 0 - 3 kcal mol<sup>-1</sup> range [26]. This value is much smaller than the energy of a hydrogen bond and implies that the interaction between the two residues in the complex is only marginally stronger than the interactions with water that occur in the free proteins.

Moreover, if binding enthalpies are determined at different temperatures, the change in heat capacity associated with the binding reaction can be calculated using the following equation (6):

$$\Delta C_p = \frac{\delta \Delta H}{\delta T} \quad (6)$$

As a general statement, the transfer of a hydrophobic molecule from water to a non-polar liquid at room temperature is associated an unfavorable enthalpy but a favorable entropy. This fact does not imply that the relative values of  $\Delta H$  and  $\Delta S$  of PPIs are not sufficient to determine the nature of the forces that guide the association. However, it has been estimated from empirical correlations that for non-polar surfaces (i.e. hydrophobic interactions), there is an energy gain of approximately 25 to 50 calories per Å<sup>2</sup>, or up to 72 calories per Å<sup>2</sup> based on other studies [26, 58]. Nevertheless, a more distinctive thermodynamic signature for the burial of hydrophobic surface is a large negative  $\Delta C_p$ , commonly found for protein-protein complex formation [59, 60]. On the contrary, the burial of polar surface in a non-polar environment features a positive  $\Delta C_p$  [59], but is lower in magnitude and it has been widely accepted that the hydrophobic effect dominates the heat capacity for protein folding and binding. Pertinent examples are represented by the endothermic and enthalpy driven interaction between bacterial neurotoxin from *C. botulinum* and synaptotagmin II, featuring a heat capacity of -326 cal mol<sup>-1</sup> K<sup>-1</sup> [61], and by the binding of xanthine oxidase to Cu,Zn-superoxide dismutase, showing a large positive  $\Delta C_p$  equal to 3.02 kJ mol<sup>-1</sup> K<sup>-1</sup> [62].

### 1.4.2. Calorimetrically determined thermodynamics of PPIs

ITC (isothermal titration calorimetry) is considered as the most quantitative technique available for measuring the thermodynamic properties of PPIs and is becoming a necessary tool for protein-protein complex structural studies [63]. ITC relies upon the accurate measurement of heat changes that follow the interaction of protein molecules in solution, without the need to label or immobilize the binding partners, since the absorption or production of heat is an intrinsic property of virtually all biochemical reactions. Measurement of heat allows the determination of binding constants ( $K_d$  or  $K_a$ ), reaction stoichiometry ( $n$ ) and all thermodynamic parameters ( $\Delta G$ ,  $\Delta H$ ,  $\Delta S$ ,  $\Delta C_p$ ) [64, 65]. Thermodynamic data of PPIs are available in dedicated databases, such as PINT [66].

Listed in Table 1 are some representative thermodynamics of association for 43 protein-protein (Section A) and 26 protein-peptide (Section B) complexes; the enthalpies were all calorimetrically determined at a constant temperature, most of them using ITC. Wesley Stites, in his work on *Chemical Reviews*, provided an extensive analysis of these results [32]. The values, most determined at ca. 25°C, reveal that neither entropy or enthalpy primary drive PPIs at this temperature: in 31 cases the enthalpy is favorable but the entropy is unfavorable; in 18 cases the association is entropy-driven and enthalpically opposed; in the remaining 20 cases both entropy and enthalpy favor the interaction. To sum up, in the 74% of the cases enthalpy favors association while entropy promotes PPIs in 55%. Despite no broad generalization is possible, there are no clear correlations between values of  $\Delta H$  or  $\Delta S$  with  $\Delta G$  for PPIs; neither do  $\Delta G$ ,  $\Delta H$  or  $\Delta S$  correlate with  $\Delta C_p$ .

More interesting considerations could be made comparing the thermodynamic parameters of PPIs (Table 1, Section A) with those reported for protein-peptide studies (Table 1, Section B), or peptide-peptide systems in two instances. Looking at the reported values, there is no great difference between the average thermodynamics of PPIs ( $\Delta G = -10.4 \pm 2.5 \text{ kcal mol}^{-1}$ ,  $\Delta H = -8.6 \pm 13.6 \text{ kcal mol}^{-1}$ ,  $\Delta S = 6.1 \pm 43.7 \text{ cal mol}^{-1} \text{ K}^{-1}$ ) and protein-peptide ( $\Delta G = -8.5 \pm 1.9 \text{ kcal mol}^{-1}$ ,  $\Delta H = -8.9 \pm 11.2 \text{ kcal mol}^{-1}$ ,  $\Delta S = -1.1 \pm 37.9 \text{ cal mol}^{-1} \text{ K}^{-1}$ ) interactions [32]. This could seem inconsistent with the fact that PPIs differ from protein-peptide interactions in the size of one of the partners in the association. However it is not clear how large the interacting domain is in a protein-protein complex, so PPIs might involve no more residues than a protein-peptide interactions. From a theoretical point of view, the change in entropy upon binding should be much less favorable for protein-peptide associations than for PPIs. Experimental data support this hypothesis. In fact, only in the 46% protein-peptide interactions are entropy favored (Table 1, Section B).

Protein complex	$\Delta G$ (kcal mol <sup>-1</sup> )	$\Delta H$ (kcal mol <sup>-1</sup> )	$\Delta S$ (cal mol <sup>-1</sup> K <sup>-1</sup> )	$\Delta C_p$ (cal mol <sup>-1</sup> K <sup>-1</sup> )	T (°C)	Ref.
<i>Section A: Protein-protein interactions</i>						
Trypsin-soybean inhibitor	-12.3	8.6	69.8	-442	25	[57]
Trypsin-cleaved soybean inhibitor	-10.8	12.6	78.6	-387	25	[57]
Trypsin-ovomuroid	-10.2	5.6	53.1	-270	25	[57]
Trypsin-lima bean inhibitor	-12.7	2.1	49.7	-430	25	[57]
Trypsin-pancreatic trypsin inhibitor	-10.7	2.5	44.7		22	[382]
Subtilisin inhibitor-chymotrypsin	-7.1	4.5	38.2	-260	25	[334]
Subtilisin inhibitor-subtilisin	-13.8	-4.7	31.1	-240	25	[381]
Calmodulin Ca <sup>2+</sup> -myosin light chain kinase	-11.5	-20.3	-29.2		25	[335]
Calmodulin Ca <sup>2+</sup> -seminal plasmin	-12.0	-12.0	0.0	0	25	[335]
Calmodulin-myosin light chain kinase	-7.2	0.0	24.1		25	[335]
Calmodulin-seminal plasmin	-8.1	0.0	27.2		25	[335]
ch4D5 Fab-p185HER2-ECD	-13.5	-17.2	-12.0	-400	25	[336]
HyHEL 5-Hen egg lysozyme	-14.5	-22.6	-27.2	-340	25	[380]
HyHEL 10-Hen egg lysozyme	-12.0	-21.9	-32.6	-335	30	[337]
D1.3-Hen egg lysozyme	-11.5	-21.7	-34.4	-380	24.2	[379]
F9.13.7-Hen egg lysozyme	-12.0	-11.1	3.3	-650	23.9	[338]
D44.1-Hen egg lysozyme	-9.7	-10.3	-2.3	-280	24.2	[338]
D11.15-Hen egg lysozyme	-12.1	-19.0	-23.0	-240	24.9	[338]
D1.3-E5.2	-10.4	-66.7	-188.4		25	[339]
D1.3-E225	-7.3	1.8	30.4	0	28.3	[339]
Ferredoxin-ferredoxin:NADP <sup>+</sup> reductase	-9.3	-0.3	30.1	-160	27	[378]
Colicin N-OmpF	-7.7	-12.3	-15.3		25	[340]
Colicin N-OmpC	-7.1	-3.7	11.2		25	[340]
Colicin N-PhoE	-7.4	-6.0	4.8		25	[340]
Barstar-barnase	-17.2	-13.9	12.2	-190	25	[377]
Human tissue factor-coagulation factor VII	-11.2	-32.0	-70.0	-730	25	[341]
Cytochrome <i>c</i> peroxidase-cytochrome <i>c</i>	-7.0	2.3	31.0	2	26	[376]
Cytochrome <i>b5</i> -cytochrome <i>c</i>	-9.1	1.0	33.9		25	[342]
Ab E3-cytochrome <i>c</i>	-9.7	-7.3	8.8	-350	25	[343]
Ab E8-cytochrome <i>c</i>	-9.5	-9.5	0.3	-165	25	[343]
Ab 2B5-cytochrome <i>c</i>	-12.6	-21.0	-28.2	-580	25	[344]
Ab 5F8-cytochrome <i>c</i>	-13.9	-21.7	-26.3	-172	25	[344]
CheY-CheA <sub>1-233</sub>	-8.1	-12.3	-14.4	-230	28	[345]
CheB-CheA <sub>1-233</sub>	-7.5	-10.1	-8.7	-450	28	[345]
Interleukin 5-IL5 receptor $\alpha$ subunit	-11.6	-11.4	0.7	-650	25	[375]
Erythropoietin-EPO receptor site 1	-11.7	-1.5	34.2		25	[346]
Erythropoietin-EPO receptor site 2	-8.4	-3.4	16.8		25	[346]
Human growth hormone G120R-hGHbp	-11.7	-9.4	7.7	-767	26.2	[347]
Phosphocarrier protein-enzyme I N-domain	-7.0	8.8	53.0		25	[374]
Stem cell factor-kit extracellular domain	-9.0	-13.0	-13.4		25	[348]
Hck SH3 domain-HIV1 Nef	-9.2	-12.8	-12		25	[373]
Fyn SH3 domain-P13 kinase p85 subunit	-7.6	10.6	60.2		30	[349]
Elastase-ovomuroid third domain	-14.5	-1.0	45.3		25	[372]

Section B: Protein-peptide interactions						
Ab 131-angiotensin II	-11.0	-8.9	6.9	-240	30	[350]
Endothiapepsin-pepstatin A	-9.1	-2.5	22.8		16.1	[371]
HK 565 peptide-PK 262 peptide	-5.9	0.4	21.1		30	[351]
human Grb2 (SH3)-human Sos peptide	-6.4	-6.2	0.7		25	[370]
Fyn SH3 domain-P2L peptide	-6.6	-12.3	-18.6		30	[349]
Lek SH2 domain-Lck phosphopeptide	-7.3	-8.4	-3.5		25	[352]
p85 SH2 domain-PDGFR phosphopeptide	-8.7	-9.4	-2.4		25	[352]
Src SH2 domain-pYHmT phosphopeptide	-8.6	-8.4	0.4		25	[352]
Fyn SH2 domain-pYHmT phosphopeptide	-6.3	-4.3	6.6	-100	25	[353]
Fyn SH2 domain-pY531 phosphopeptide	-8.4	-8.7	-1.3	-270	25	[353]
SHC Nterm. domain-EGFR1148 peptide	-10.3	-5.46	16.2	-185	25	[354]
SHC Nterm. domain-Trk490 peptide	-10.1	2.4	41.7	-207	25	[354]
Ab 13AD-peptide LZ	-10.5	-12.6	-7.1	-251	27	[355]
Ab 13AD-peptide LZ(7P14P)	-9.5	-17.4	-26.0	-366	27	[355]
Ab 29AB-peptide LZ	-11.0	-13.8	-9.5	-335	27	[355]
Ab 29AB-peptide LZ(7P14P)	-9.5	-17.1	-25.8	-392	27	[355]
Ab 42PF-peptide LZ(7P14P)	-10.4	-13.4	-10.0	-691	27	[355]
Calmodulin Ca <sup>2+</sup> -melittin	-11.7	7.2	63.3		25	[356]
Calmodulin-melittin	-8.1	4.8	43.3		25	[356]
Ribonuclease S-truncated S peptide	-9.4	-41.9	-109.0	-1200	25	[369]
Streptavidin-FSHPQNT peptide	-5.3	-19.3	-47.0		25	[357]
Streptavidin-pStrep tag	-6.1	-12.6	-21.8		25	[368]
CheR-receptor pentapeptide	-7.9	-13.6	-18.9		28	[358]
Profilin-Pro <sub>11</sub>	-5.4	-5.1	-1.1		28	[367]
A-B hetero-dimeric coiled coil	-10.6	-24.7	-48.3	-720	20	[359]
GroEL-unfolded subtilisin BPN' mutant	-7.6	-19.9	95.7	-850	14.3	[360]

**Table 1 - Thermodynamics of protein-protein (Section A) and protein-peptide (Section B) associations**

Values listed refer to the analysis conducted by Stites (see text and [32] for details).

Concerning  $\Delta C_p$ , its experimental determination requires carrying out titrations at different temperatures and not all workers have performed this, as shown in Table 1. In those systems where  $\Delta C_p$  has been determined, this parameter is usually large and negative; particularly the average value of  $\Delta C_p$  for PPIs and protein-peptide association were  $-333 \pm 202 \text{ cal mol}^{-1} \text{ K}^{-1}$  and  $-447 \pm 309 \text{ cal mol}^{-1} \text{ K}^{-1}$  respectively [32]. Although the magnitudes of heat capacity in protein folding reactions are generally larger, the great negative value of  $\Delta C_p$  indicate that hydrophobic effects are usually quite important in PPIs. However, Table 1 reports three exceptions represented by calmodulin Ca<sup>2+</sup>-seminal plasmin and ferredoxin-ferredoxin:NADP<sup>+</sup> reductase interactions, both showing a  $\Delta C_p$  approximately equal to zero, and the association between cytochrome *c* peroxidase and cytochrome *c*, that features a  $\Delta C_p$  value of  $2 \text{ cal mol}^{-1} \text{ K}^{-1}$ . These evidences, as many others like the previously described xanthine oxidase-Cu,Zn-superoxide dismutase interaction, support the idea that hydrophobic contacts do not completely dominate PPIs as they do protein folding

processes. However, this doesn't mean that the contributions of hydrophobic interactions to PPIs are negligible, even in cases of positive  $\Delta C_p$ . This is properly the case of the association involving cytochrome *c* for which the binding to four different antibodies has also been characterized. All four associations showed large negative values of  $\Delta C_p$  (Table 1, Section A), indicating that hydrophobic interactions are important for the protein-protein assembly of cytochrome *c*.

## **1.5. Mutational investigation and prediction of PPIs: definition of “hotspots”**

In order to fully understand PPIs, and to manipulate them, scientists need to identify the residues that account for binding of the proteins and stabilizing the final assembly. It has been proposed that only very few of the residues in protein-protein interfaces are absolutely required for the association: for an average interface (1200 - 2000 Å<sup>2</sup>), less than 5% of surface residue contribute effectively to binding; in small interfaces, this can mean as few as one amino acid on each protein partner [67]. These energetically important residues were defined “hotspots” by Clackson and Wells, in their pioneering alanine scanning work on the binding of hGH (human growth hormone) to its receptor (Figure 8A) [68]. Alanine scanning mutagenesis is a powerful experimental method for mapping functional epitopes through systematic substitution of the residues in the interface with alanine [69]. In fact, alanine substitutions remove side-chain atoms past the  $\beta$ -carbon without introducing additional conformational freedom [70] and allow to infer the role of chemical functional groups at specific positions to protein binding. Glycine would also nullify the side chain but could introduce conformational flexibility into the protein backbone, and therefore is not commonly applied [71]. In particular the difference in the binding free energy ( $\Delta\Delta G$ ) between the wild type and each mutant is measured; hotspots are defined as those substitutions that lead to a significant ( $\Delta\Delta G \geq 2 \text{ kcal mol}^{-1}$ ) drop in the binding free energy [68].

### **1.5.1. Disruption of protein-protein interfaces**

The first attempts to study PPIs through mutagenesis were aimed at the disruption of the interface, usually by removal or introduction of charged groups. Disruption of an interface itself is relatively easy; much more difficult can be to find the mutation, or the set of mutations, that result into a folded and stable protein. One application in this field has been the creation of stable proteins with reduced tendency to associate, through the abolishment of PPIs. For instance, since the 1980s the effect of introducing isolated charges [72] and ion pairs [73] into the interface of tyrosyl-tRNA synthetase were analyzed: isolated charges increased the propensity to dissociate while complementary charges did not. Other example are represented by ribulose 1,5-biphosphate carboxylase, where a lysine to aspartate mutation prevented hetero-dimeric assembly [74], and insulin, in which charges were introduced at the interface level to increase oligomer dissociation

[75]. In the latter case, it has been demonstrated that oligomer destabilization doesn't require introduction of large chain or charges, but a range of neutral and charged substitutions at defined positions (Pro28 and Lys29) are extremely disruptive to insulin oligomerization.

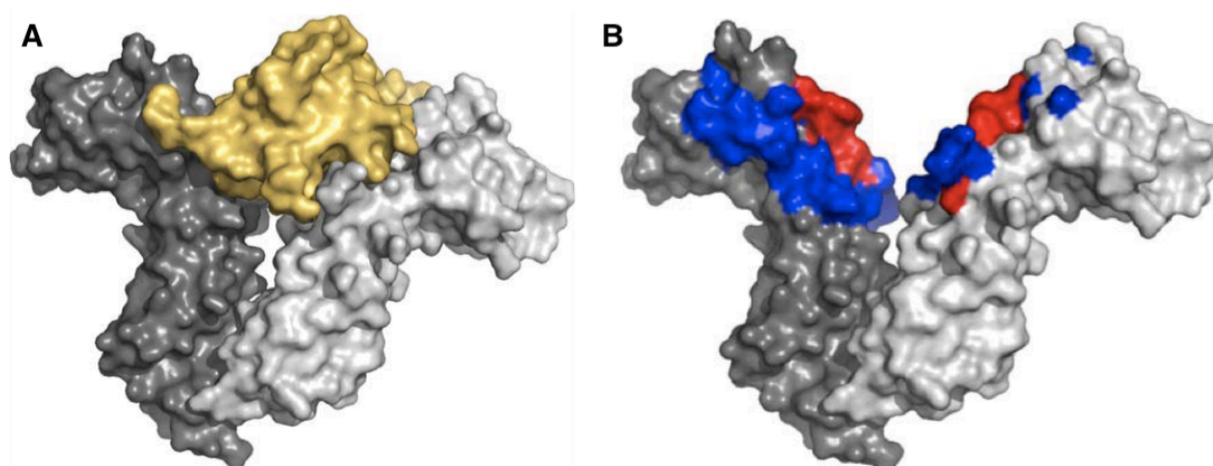
In other cases, when the interacting interface is separated from the rest of the protein it can be removed. This strategy has been applied to create a stable monomeric form of triosephosphate isomerase by replacement of a 15 amino acids loop, containing most of the interface residues, with an eight residue stretch [76]. Removal of binding domains is not always practicable, or advantageous. Alternatively it is possible to disrupt the interface by removal of side chains that stabilize association and/or by introduction of side chains that destabilize interaction. Substitution of an interface aspartate residue has been carried out to convert the tetrameric fructose-1,6-bisphosphate aldolase into a dimeric protein, disrupting quaternary structure of the enzyme [77]. However, altering PPIs through relatively simple mutagenesis approaches can yield unexpected results, as in the case of the HEL-Gloop2 antibody interaction [78]. Surprisingly, the abolishment of electrostatic interactions, thought to be essential for PPIs, through substitution of a glutamate and a lysine with hydrophilic residues strengthened the association with the antibody.

### 1.5.2. Alanine scanning as a tool to discriminate hotspots: specificity and cooperativity

Another common theme in the mutagenesis literature is the systematic analysis of the interfaces through alanine scanning, in order to demonstrate the contribution of specific residues to the overall stability of the complex. This experimental approach has been used to map key residues in a large number of protein-protein interfaces and its employ in this capacity was reviewed since the 1990s [69]. ASEdb, a complete depository of alanine scanning data of interfaces involved in PPIs, including their effects on the free energy of binding, has been developed by Thorn and Bogan [79].

Historically, the hGH-hGH receptor interaction has been the first subject of intense investigation using mutagenesis directed on residues of both protein partners to identify the relative importance of their contributions to binding affinity. The crystal structure of the complex (Figure 8A) shows that 30 residues on the surface of the receptor are in physical contact with 31 residues of the hormone. However, mutagenesis studies demonstrated that only six of these amino acids on the surface of the hGH receptor are energetically crucial for the interaction (Figure 8A), accounting for over 85% of the binding energy. On the hormone side of the interface the situation is very similar. In particular, two hotspot residues are tryptophan and feature a relevant change in the free energy of binding ( $\Delta\Delta G > 4.5 \text{ kcal mol}^{-1}$ ). This is not surprising; it has been widely demonstrated that Trp residues plays a unique function, probably owing to its large size and aromatic nature [80]. Moreover, tryptophan mutation to alanine generates a large cavity, due to the large difference in sizes [67], which can create a highly complex destabilization. This fits with the scenario of the hGH-hGH receptor association, where the hotspot core is largely hydrophobic and is surrounded

by less important polar residues.



**Figure 8 - Hotspots residues for PPIs**

(A) Human growth hormone (yellow) bound to the extracellular portion of its homo-dimeric receptor (gray) (PDB entry: 3hrh). (B) Localization of the protein-protein interface of the receptor (blue and red) involved in the interaction. The hotspots (red) were identified with alanine scanning mutagenesis.

Alanine scanning is a very arduous and time-consuming technique to perform, as different mutants must be treated separately [81]. Each alanine-substituted protein needs to be individually constructed, expressed, purified and sometimes refolded, and the loss of the side chain contribution is then assessed through *in vitro* activity/affinity assays [71]. Combinatorial libraries of alanine substitutions represent an alternative to the laborious process of scanning single positions in a protein. Through a single round of site-specific oligonucleotide-directed mutagenesis, “binomial substitutions” of either alanine or a wild-type amino acid residue are readily accessible by conventional oligo-nucleotide synthesis for seven amino acids (Asp, Glu, Gly, Pro, Ser, Thr, Val), for which altering a single encoding nucleotide can result in a codon for Ala [71]. Another method is “shotgun scanning”, which implements a simplified format for combinational alanine scanning and uses phage-displayed libraries of Ala-substituted proteins for high-throughput analysis. The rapidity and general applicability of the shotgun-scanning scheme should accelerate the investigation of PPIs [82].

Interestingly, an experimental approach called “alanine shaving”, which relies on making multiple simultaneous alanine mutations, could be used to investigate cooperativity between inert side chains [67]. Cooperativity can be detected by multiple mutation cycles, in which the free energy change caused by the simultaneous mutations at residue positions in a protein is compared with the sum of the free energy changes associated with single mutations at each of the residues positions [83]. Considering that additivity of mutational effects is very common for PPIs, deviations from additivity are indicative of cooperative interactions [79, 84]. Nonetheless, exceptions to this rule have been observed [85, 86]. As an example, the role of residues surrounding the hotspot of

BPTI (bovine pancreatic trypsin inhibitor) in interaction with two proteinases, trypsin and chymotrypsin, was analyzed by alanine shaving [87]. The replacement of up to six residues with alanine has been demonstrated to be fully additive at the level of protein stability. Shaving of two binding loops led to a progressive drop in the association energy, more pronounced for trypsin (decrease up to 9.6 kcal mol<sup>-1</sup>) than chymotrypsin (decrease up to 3.5 kcal mol<sup>-1</sup>).

### 1.5.3. Computational alanine scanning to probe PPIs

Given the fact that experimental hot spots determination is time-consuming and involves a high cost, accurate, predictive computational methodologies for alanine scanning has been developed, in order to reproduce the experimental mutagenesis values [88]. For that purpose it is important to accurately calculate the binding free energies of known three-dimensional structures and the effect of mutations on these affinities. A large amount of algorithms of increasing complexity has been employed to address the binding energy between biological molecules, and can be divided essentially in two types. First, empirical functions or simple physical methods that use knowledge-based simplified models are used to evaluate binding. Second, fully atomistic methods that estimate the free energy of association directly or changes in the binding free energies as a result of mutating the residues of the interacting molecules.

An example is provided by Kortemme and coworkers [88], who developed a computational alanine scanning protocol that, given the three-dimensional structural of a protein-protein complex, allows the automatic scanning of a complete protein-protein interface. The approach in question uses a simple free energy function to calculate the effects of Ala-substitutions on  $\Delta\Delta G$ . The function consists of a linear combination of a Lennard-Jones potential to describe atomic packing interactions, an implicit solvation model, an orientation-dependent hydrogen-bonding potential derived from high-resolution protein structures, statistical terms approximating the backbone-dependent amino acid-type and rotamer probabilities, and an estimate of unfolded reference state energies [89]. The method was tested on 19 protein-protein complexes with 233 mutations: 79% of the energetic hotspots, previously defined experimentally, were identified by the free energy function. Moreover, the computational methodology has been validated by the successful design of protein interfaces with novel specificity and activity for the NKG2D immunoreceptor-MHC complex [90] and the assembly of gp130 shared signalling receptor with different cytokines [91].

Other methodological approaches has been proposed to identify the interfacial hot spots, such as MM/PBSA (molecular mechanics/Poisson-Boltzmann surface area) [92],  $\lambda$ -dynamics [93], chemical Monte-Carlo/molecular mechanics [94] or ligand interaction scanning [95]. More time-consuming methods that involves fully atomistic simulations and include both the rigorous free energy perturbation and thermodynamic integration [96], are also available.

## **Section A2 - Metal-mediated PPIs: role of cations in protein assembly**

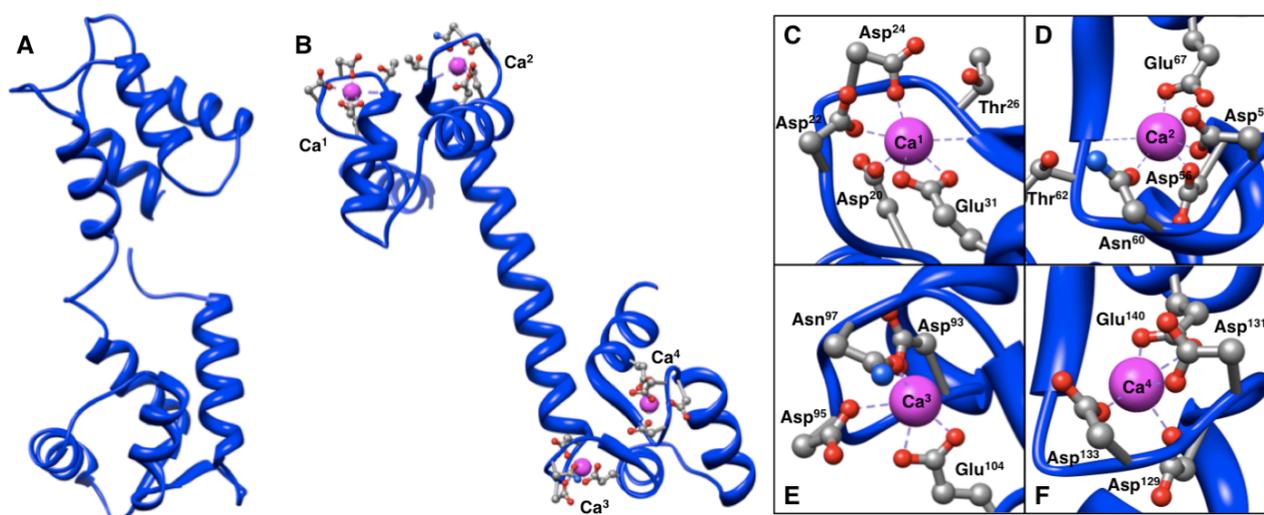
PPIs involve direct interactions between proteins or can be mediated through conformational changes induced by cofactor binding. In particular, exogenous cations can be involved in protein associations and offer an alternative route to control the geometry of the final assembly [97]. This is consistent with the intimate involvement of metals and metalloids in the regulation of genes and proteins involved in their uptake, utilization and detoxification. The regulation occurs during transcription, translation and directly at the protein level. The latter mechanism is the main subject of this chapter, which is aimed at investigating metal-binding proteins that mediate PPIs and the role of metal ions, such as  $\text{Ca}^{2+}$ ,  $\text{Cu}^+$  and  $\text{Zn}^{2+}$ , in protein-protein associations. However, the role of PPIs in metal homeostasis circuits is also discussed.

### **1.6. PPIs involving $\text{Ca}^{2+}$ -binding proteins: the cases of calmodulin and cadherin**

The divalent cation calcium ( $\text{Ca}^{2+}$ ) plays a crucial role in the metabolism and physiology of eukaryotes [98]. In prokaryotic cells, an equivalent important role for calcium has also been demonstrated [99]. The intracellular free calcium concentration in the model organism *Escherichia coli* falls in the range 0.1 - 1 mM [100, 101]. Such levels are similar to those in eukaryotic cells and are thousand times less than those typically found outside the cell. In fact, calcium exists as a gradient across the plasma membrane, with extra-cellular concentrations being about 10.000 times higher than intracellular ones. Interestingly, despite its critical nature in the cell,  $\text{Ca}^{2+}$  is highly reactive, and therefore toxic to cells upon prolonged exposure to high levels, so it is essential that cells preserve a low intracellular concentration of  $\text{Ca}^{2+}$  (50 - 100 nM) in their basal state [102]. On the other hand, it is this highly reactive nature of  $\text{Ca}^{2+}$  and the low intracellular concentration that make it such a potent metal ion for use in cellular signaling. External signals, such as hormones, light, stress or pathogenesis, can often lead to transient increases in calcium concentrations within the cell. In this way, calcium participates in an intracellular signaling system by acting as a diffusible second messenger to the initial stimuli. Increased calcium concentrations lead to calcium binding by regulatory proteins, which turn the calcium signal into a biological response.

There are many regulatory proteins that bind calcium, which together form an intricate network of feedback loops to control the location, amount and effect of  $\text{Ca}^{2+}$  influx. Calmodulin (CaM), a small acidic protein of 16.7 kDa, is considered a major transducer of calcium signals [103]. It is an ubiquitous protein that can bind to and regulate a multitude of different protein targets, thereby modulating many different cellular processes such as signal transduction, gene transcription, ion conductivities, vesicular fusion and cytoskeleton functions [104]. Interestingly, many of the proteins that CaM interacts with are unable to bind  $\text{Ca}^{2+}$  ion themselves, and as such use CaM as a

“calcium sensor” and signal transducer. CaM is basically constituted by two globular domains, each one containing two helix-loop-helix EF-hand  $\text{Ca}^{2+}$  binding motifs, connected by a flexible central helix (Figure 9B). Consequently, CaM can bind up to four metal ions. The structure shown on Figure 9A displays calmodulin in the absence of metal ions, while the structure on Figure 9B shows calmodulin after calcium binds. The four nearly identical high-affinity  $\text{Ca}^{2+}$  binding sites are shown on Figures 9C-F: the calcium ion is bound to the negatively-charged side chains of three aspartate residues, or two aspartates and one asparagine, and one glutamate, as well as one oxygen atom from the backbone of the protein chain. Upon binding to calcium ions,  $\text{Ca}^{2+}$ -CaM undergoes a large conformational change that involves the flexible central  $\alpha$ -linker. The transition from the closed state to an open conformation exposes a large hydrophobic binding pocket and enables CaM to interact with specific protein partners, for a specific response. In this way  $\text{Ca}^{2+}$  binding to CaM can trigger the interaction with CaM’s effectors by modulating the conformational transition of the protein. Indeed, spectroscopy studies clearly showed that the connector between the two calcium binding globular domains is flexible, even when it is not bound to its target proteins [105]. However, the full range of flexibility can be seen in CaM's interactions with its target proteins.

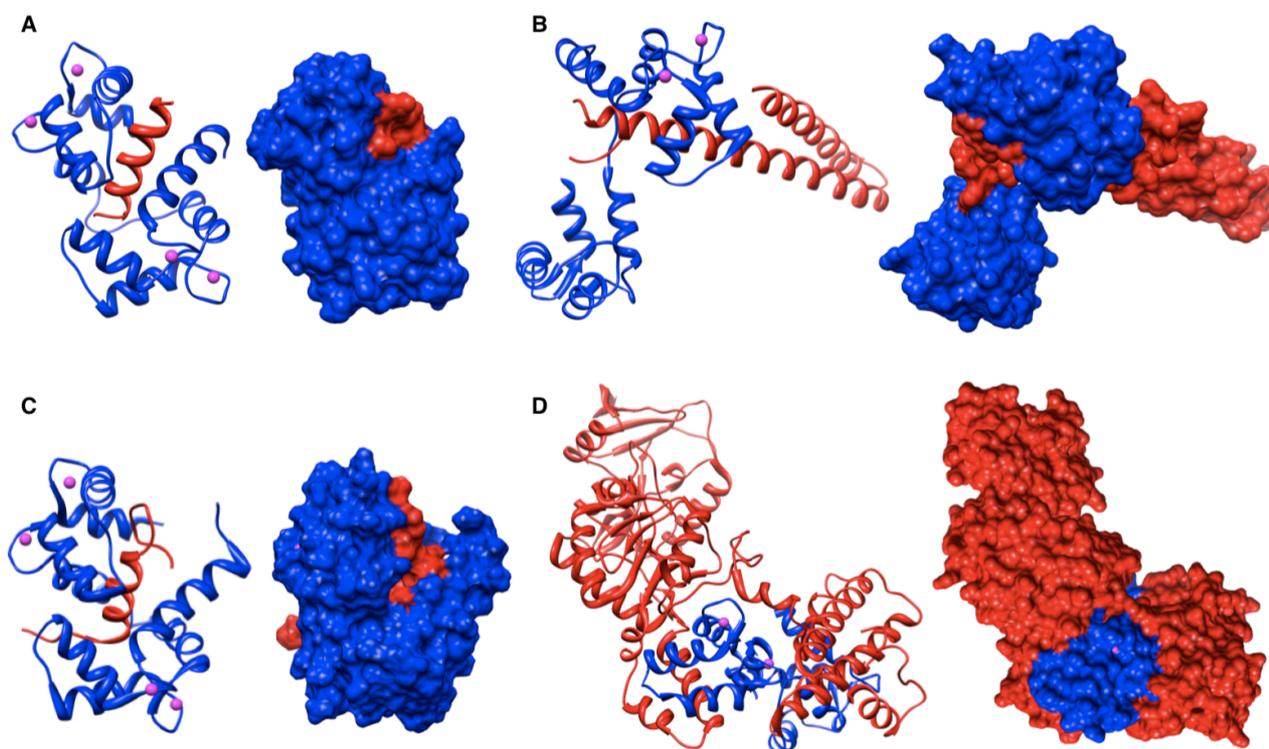


**Figure 9 - Calmodulin three dimensional structures:  $\text{Ca}^{2+}$  binding and conformational changes**

Ribbon structural diagram of calmodulin in the apo- (A) and  $\text{Ca}^{2+}$ -bound state (B), that represent the closed and the open state, respectively (PDB entries: 1cfd, 1cll). Details of the four calcium binding sites are shown (C,D,E,F). Residues that coordinates metal ions are represented in “ball and stick” and colored using the following atomic scheme: carbon, grey; oxygen, red; nitrogen, blue; calcium, magenta.

CaM's target proteins come in various shapes, sizes and sequences and are involved in a wide array of functions. For example,  $\text{Ca}^{2+}$ -CaM binds and activates a large number of kinases and phosphatases that play significant roles in cell signaling, ion transport and cell death. One common theme in the contact between CaM and its different target proteins is the use of non-polar contacts, in particular, through the interaction with the unusually abundant methionine residues of CaM.  $\text{Ca}^{2+}$  binding exposes these non-polar surfaces of CaM forming two neat grooves, which then bind to

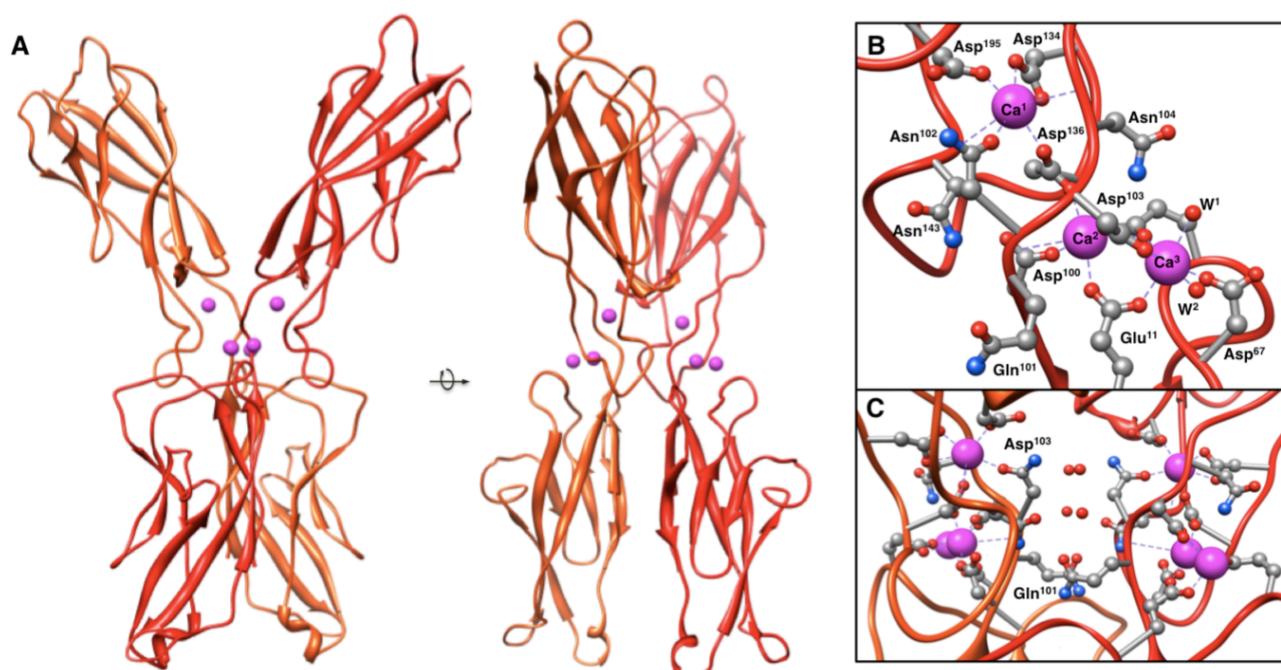
non-polar regions on the target proteins. Because these non-polar grooves are generic in shape, CaM acts as a versatile regulatory protein and its targets are not required to possess any specific amino acid sequence or structural binding motifs. CaM typically wraps around its target, with the two globular domains gripping either side of the interacting protein. This is the case of the Ca<sup>2+</sup>-sensitive potassium channel [106] and other target enzymes such as CaM KII- $\alpha$  (calmodulin-dependent kinase II-alpha) [107] and CaM kinase kinase [108]. The crystal structures of these complexes between CaM and peptides corresponding to a small portion of the target proteins are shown in Figure 10A-C. All the three structures define the canonical view of CaM interaction with the effector, in which the hydrophobic faces of the two CaM lobes act as a clamp surrounding a target helix or helices. A different binding architecture is seen in the complex of CaM with the edema factor toxin from *Bacillus anthracis* (Figure 10D). Since CaM is absent in bacteria, *B. anthracis* have cleverly evolved to exploit the abundance of CaM in their hosts in order to trigger the toxin and take control of their cellular machinery. In fact, once CaM binds to the toxin a conformational change in the toxin activates its adenylyl cyclase activity, which then depletes the host cell's energy stores [109]. In this assembly, the CaM lobes do not act as a clamp but also in this case the overall conformation of the protein is extended. Interestingly, the hydrophobic binding interface is buried within CaM rather than oriented towards the effector target.



**Figure 10 - Comparison of CaM conformations in complex with different protein partners**

Ribbon diagrams and solvent excluded surfaces of CaM (blue) in complex with CaMKII- $\alpha$  peptide (A) (PDB entry: 1cm1), Ca<sup>2+</sup>-activated K<sup>+</sup> channel (B) (PDB entry: 1g4y), CaM kinase kinase (C) (PDB entry: 1ckk) and edema factor toxin (D) (PDB entry: 1k93), respectively. All the CaM binding targets are colored in red. Calcium ions are shown as magenta spheres.

$\text{Ca}^{2+}$  ions can also directly induce PPIs as in the case of E-cadherin dimerization. Cadherins constitute a large family of cell surface proteins, many of which participate in cell adhesion that plays a fundamental role in the formation of solid tissues [110]. Calcium binding is required for cadherin-mediated cell adhesion and is also central to E-cadherin dimer formation. Dynamic light scattering and sedimentation analysis has proven that E-cadherin dimerizes only in the presence of calcium ion with high affinity ( $K_d = 0.6 \mu\text{M}$ ) [111]. The crystal structure of a fragment of murin E-cadherin (Figure 11A) provide structural evidences for the role of this metal in dimer stabilization. Each molecule of the dimer is composed of two seven stranded  $\beta$ -barrels which are connected by a flexible linker and bridged by an arrangement of three contiguous  $\text{Ca}^{2+}$  binding sites (see  $\text{Ca}^1$ ,  $\text{Ca}^2$ ,  $\text{Ca}^3$  in Figure 11B). In particular,  $\text{Ca}^1$  and  $\text{Ca}^2$  ions are each coordinated by seven oxygen atom coming from aspartate, asparagine or glutamate residues, whereas  $\text{Ca}^3$  is coordinated by six, two of which are from water molecules (see  $\text{W}^1$  and  $\text{W}^2$  in Figure 11B). Of particular relevance for dimer stabilization are the direct hydrogen bonds made between molecules of cadherin by the side chains of  $\text{Gln}^{101}$  and  $\text{Asn}^{143}$ , residues that bind calcium through their backbone carbonyl oxygen atoms (Figure 11C). Conversely, the side chains of  $\text{Asp}^{100}$  coordinates calcium, while its carbonyl oxygen atom mediates a direct hydrogen bond across the dimer interface. In addition  $\text{Gln}^{101}$ ,  $\text{Asn}^{143}$  and other residues at the interface stabilize dimer formation through water-mediated hydrogen bond interactions.



**Figure 11 -  $\text{Ca}^{2+}$ -mediated E-cadherin dimerization**

Ribbon representations of E-cadherin dimer (A): the structure is rotated by  $180^\circ$  around the vertical axis (PDB entry: 1edh). Details of calcium binding sites, in the monomer (B) and in full dimeric structure (C). Residues that coordinate metal ions are represented as *ball and stick* and colored using the following atomic scheme: carbon, grey; oxygen, red; nitrogen, blue; calcium, magenta. Calcium ions and water molecules are indicated with  $\text{Ca}^{1-3}$  and  $\text{W}^{1-2}$ , respectively.

## **1.7. Cu<sup>2+</sup>-binding proteins: PPIs involved in metal trafficking**

Copper is an essential trace element for all organisms mainly because serves as a cofactor in over 30 enzymes, due to its biologically suitable redox potential. In fact, the ability of this metal ion to cycle between oxidized Cu<sup>2+</sup> and reduced Cu<sup>+</sup> is utilized by enzymes that catalyze redox reactions. However, free copper ions are strongly toxic for the cell due to its ability to form radicals. Therefore, immediately after uptake the vast majority of Cu ions are sequestered by scavenging proteins like metallo-thioneins to prevent copper from accumulating in a toxic form. Indeed, the estimated values of concentration of free copper in the cell ( $>10^{-18}$  M) [112], demonstrate the efficiency of these scavenging systems.

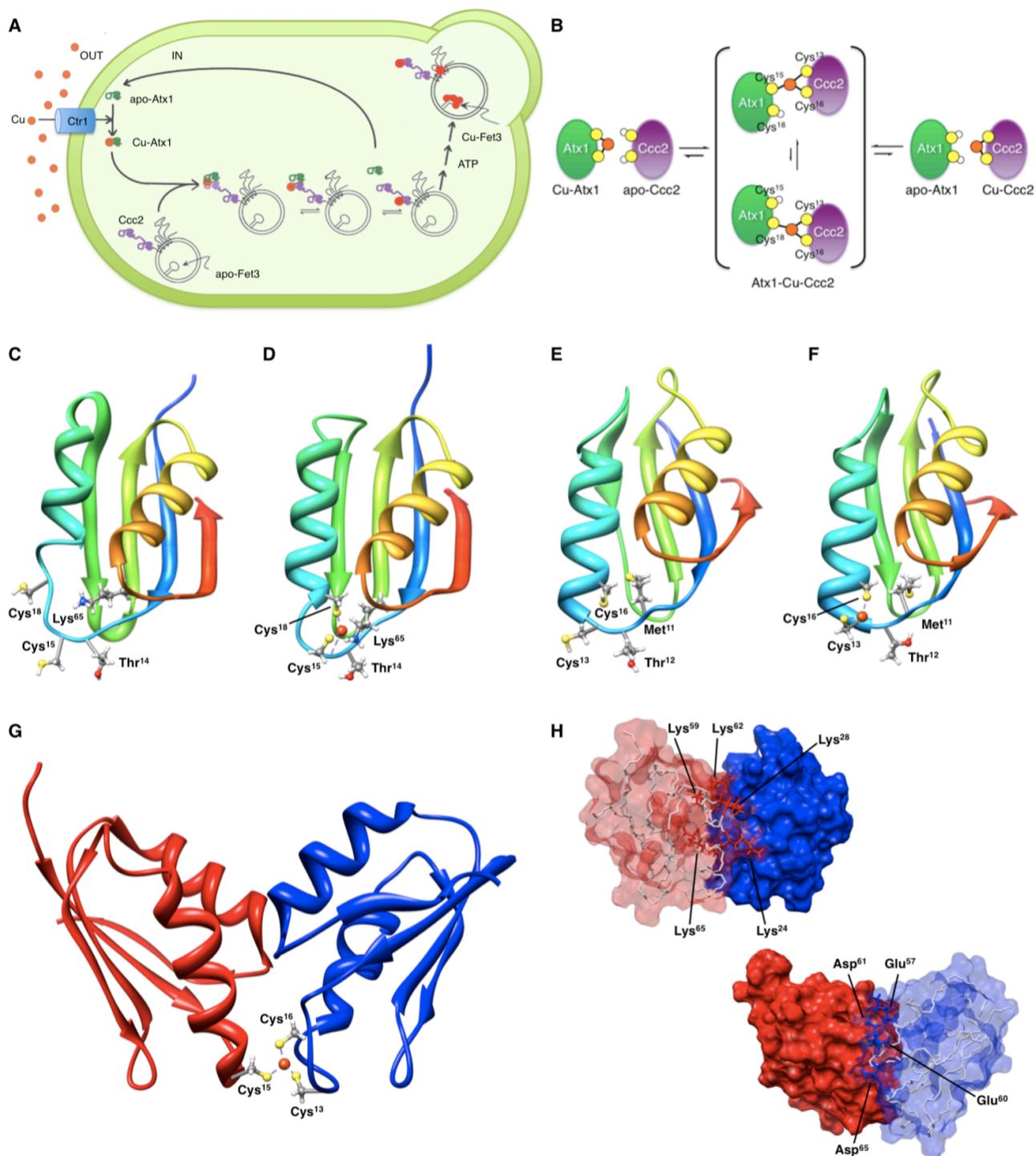
Copper biological centers generally occur as one of three main types and while other copper sites do exist, most are based on at least one of these [113]. Type 1 copper centers are found in “blue copper proteins”, such as azurin and plastocyanin, which play a functional role in electron transfer. The core of the centre is a copper ion trigonally ligated by two histidine nitrogen atoms and one cysteinyl sulfur. Type 2 copper centers contain a copper ion coordinated by at least three histidine residues and usually has square planar or pyramidal geometry. Examples are represented by the Cu<sub>B</sub> center of cytochrome *c* oxidase, which catalyses the four electron reduction of oxygen to water and the copper site of Cu/Zn SOD (superoxide dismutase), which protects the cell from oxidative damage by catalyzing the disproportionation of the superoxide anion radical into oxygen and hydrogen peroxide. Type 3 copper centers instead contain two Cu<sup>+</sup> ions, each of which are ligated by three histidine residues, such as the dioxygen binding site in the oxygen carrier hemocyanin. In addition, an important class of copper containing enzymes, the multi-copper oxidases, contain all three types of copper center. Ceruloplasmin, the largest source of copper in human plasma, and its yeast homologue Fet3 are examples of this class. Other two important centers, Cu<sub>A</sub> in cytochrome *c* oxidase and nitrous oxide reductase and the Cu<sub>Z</sub> in nitrous oxide reductase, cannot be ascribed to any of the main three types. Cu<sub>A</sub> centers are directly coordinated by four ligands: two cysteines bridge the two copper ions and each copper ion is coordinated by a terminal histidine. This arrangement displays a formal charge on each copper ion of +1.5. The Cu<sub>Z</sub> center is instead a unique tetranuclear copper cluster bound by seven histidine residues and with an inorganic sulphide ligand which bridges all four copper ions.

Regulation through the sensing of the intracellular copper levels by regulatory proteins has been extensively studied. These control pathways often feature copper responsive transcriptional factors in which the metal, in its cuprous form, plays structural and regulatory roles. Cu<sup>+</sup>-responsive transcriptional repressors/activators are used to sense changes in the intracellular copper concentration and consequently up- or down-regulate the transcription of metal acquisition, distribution and sequestration genes. In this scenario, Cu<sup>+</sup> ions can also directly act at protein level

regulating PPIs between chaperones and effectors involved in copper homeostasis and utilization. A structurally characterized example is represented by the well established metal-mediated interaction between Atx1 and Ccc2 from *Saccharomyces cerevisiae* [114]. Atx1 is a copper chaperone that transports  $\text{Cu}^+$  into the TGN (trans-Golgi network), the secretory compartment of the eukaryotic cell, which traffics proteins towards the cell membrane or beyond [115]. Atx1 directly delivers copper to the  $\text{Cu}^+$ -transporting P-type ATPase Ccc2, which transports the metal ion across the membrane into the TGN. Here, copper is incorporated into the multi-copper oxidase Fet3, which is located in the cell membrane and is required for high-affinity iron uptake into the yeast cell (Figure 12A) [116]. The thermodynamic gradient for metal transfer is shallow ( $K_{\text{eq}} = 1.5$ ), establishing that transfer of copper from  $\text{Cu}^+$ -Atx1 to Ccc2 is not based on a higher copper affinity of the target domain [117]. Instead, Atx1 protects  $\text{Cu}^+$  from non-specific reactions and allows rapid metal transfer to its partner ( $K_{\text{ex}} > 10^3 \text{ s}^{-1}$ ) [118].

Solution structures of the native  $\text{Cu}^+$ -bound and the reduced apo forms of both Atx1 (Figure 12C,D) [119] and the first soluble domain of Ccc2 (Figure 12E,F) [120] have been solved. These two different 72 residue polypeptides share a classical “ferredoxin-like”  $\beta\alpha\beta\beta\alpha\beta$  fold, in which the antiparallel strands form a  $\beta$ -sheet, on which the two  $\alpha$ -helices are superimposed. Both proteins display a conserved MXCXXC Cu-binding motif located on a flexible solvent-exposed loop at the beginning of the first  $\alpha$ -helix, which is more extended in Atx1 compared to Ccc2. In Atx1 the  $\text{Cu}^+$  ion is bound in a trigonal environment, by Cys<sup>15</sup> and Cys<sup>18</sup>, from the motif, and Thr<sup>14</sup> and Lys<sup>65</sup> are found in close proximity to the metal (Figure 12D). It has been suggested that this positively charged lysine side chain might help to stabilize the negative charge associated with  $\text{Cu}^+$  bound by two cysteine thiolate groups. In the Cu-Ccc2 structure, the copper ion coordinates two of the six cysteine residues, Cys<sup>13</sup> and Cys<sup>16</sup>, and Met<sup>11</sup> and Thr<sup>12</sup> are in proximity to the metal binding site (Figure 12F). Notably, Atx1 undergoes structural rearrangements at the N-terminus of the first helix upon copper binding, whereas Ccc2 structure is essentially unaffected by metal binding.

The  $\text{Cu}^+$ -dependent interaction between Atx1 and the first N-terminal domain of Ccc2 was directly elucidated by NMR chemical shift mapping experiments [114], which revealed the formation of a complex, although a relatively weak one, between the proteins. In fact, the driving force for the association is  $\text{Cu}^+$ -binding, as demonstrated by the observation that, despite their electrostatic complementarity, apo-Atx1 and apo-Ccc2 do not form a complex. In the complex structure, the metal-binding site, build up using Cys<sup>15</sup> of Atx1 and Cys<sup>13</sup> and Cys<sup>16</sup> from Ccc2, is relatively adjacent to the region involved in the intermolecular contact between the two proteins (Figure 12G). Upon the assembly some structural rearrangements for both Atx1 and Ccc2 are observed in the first  $\alpha$ -helix, compared to the two isolated protein structures. In particular, Cys<sup>18</sup> from Atx1 is in a helical conformation that is similar to that observed in Cu-Atx1 and the two cysteine residues from Ccc2 partially unwound and reoriented to accommodate the proper coordination geometry.



**Figure 12 - Atx1-Ccc2 complex: mechanism and structural details of the association**

Schematization of the Cu traffic in *Saccharomyces cerevisiae*, involving the Ctr1 ionic pump, the two copper chaperones Atx1 and Ccc2, and the Fec3 cuproenzyme (A) and mechanism for copper transfer between Atx1 and Ccc2 (B). Ribbon diagrams of the apo- and Cu-bound forms of Atx1 (C, PDB entry: 1fes; D, PDB entry: 1fd8) and Ccc2 (E, PDB entry: 1fvs; F, PDB entry: 1fvq). Structural elements are colored from blue, in the proximity of the N-terminal, to red, at the C-terminus. Ribbon representation (G) and solvent excluded surfaces (H) of the Cu-mediated complex between Atx1 (red) and Ccc2 (blue) (PDB entry: 2ggp). Residues involved in metal coordination are represented as a *ball and stick* and colored using the following atomic scheme: carbon, grey; nitrogen, blue; oxygen, red; sulfur, yellow; hydrogen, white; copper, orange. In panel H, residues involved in PPIs from both protein partners are shown using a *stick* model.

The coordination of copper ion at an intermolecular binding site provides significant evidence in favor of the generally accepted mechanism of  $\text{Cu}^+$  transfer between Atx1 and Ccc2, in which an intermediate complex confirms that PPIs serve to enhance the rate of copper exchange. In particular, mutagenesis of cysteine residues has demonstrated that Cys<sup>15</sup> of Atx1 and Cys<sup>13</sup> of Ccc2 are essential for complex formation, whereas Cys<sup>18</sup> from Atx1 and Cys<sup>16</sup> from Ccc2 are not [114]. This evidence proves that the coordination of copper involves a rapid equilibrium of many Cu-bridged intermediates, including the dominant species where  $\text{Cu}^+$  is tricoordinated by Cys<sup>15</sup> of Atx1 together with Cys<sup>13</sup> and Cys<sup>16</sup> of Ccc2 (Figure 12B). In fact, the creation of one additional coordination bond constituted the energetic factor that permitted complex formation between the  $\text{Cu}^+$ -loaded donor and the acceptor in its apo-form.

However, PPIs also have a crucial role. This is demonstrated by the homo-dimeric assembly reported for Hah1 (human Atx1 homologue), a 68 residue polypeptide containing a MXCXXC copper binding motif that can functionally substitute for Atx1 in yeast mutant *atx1* $\Delta$  [121]. The homologous protein chaperone plays delivering copper to the  $\text{Cu}^+$ -transporting P-type ATPases ATP7A, also known as Menkes protein, and ATP7B, named Wilson protein, for transfer into the TGN [122]. Surface plasmon resonance and *in vivo* FRET experiments indicated the copper-mediated dimerization of Hah1, that features a dissociation constant in the micro-molar range [123]. Indeed, the crystal structure of Cu-Hah1 reveals a fold very similar to that of Atx1 but in a dimeric form (Figure 13A) [124]. Specifically, the structure displays a metal binding site in which  $\text{Cu}^+$  is found in a tetrahedral environment originated by four cysteine residues, two from each monomer (Figure 13B). In addition, bacterial systems behave similarly [125]. Taken together, these evidences support the idea that the formation of copper bridged intermediates, whether detectable or not, is presumably a general feature in the transfer mechanism of  $\text{Cu}^+$  in all homologous systems.



**Figure 13 - Structure of  $\text{Cu}^+$ -bound Hah1 dimer**

Ribbon representation of the  $\text{Cu}^+$ -bound Hah1 (A), with detail of its metal binding site (B) (PDB entry: 1fee). Residues involved in metal coordination are represented as a *ball and stick* and colored using the following atomic scheme: carbon, grey; sulfur, yellow; copper, orange.

## **1.8. Zn<sup>2+</sup>-binding proteins and PPIs: interface zinc sites and zinc finger domains**

Zinc is known to be essential for growth, development and transmission of the genetic message. Zinc carries out these roles through a remarkable mosaic of zinc binding motifs that mediate a wide variety of metabolic processes, including carbohydrate, lipid, protein and nucleic acid synthesis and degradation, as well as transcription and translation of the genetic message [126]. Of the transition metals, zinc is particularly prevalent in protein structures. The large utilization of zinc is probably due to its chemical properties, related to its electronic configuration, in particular, its redox inactivity. A search of the PDB (Protein Data Bank) identifies some 3400 zinc-containing entries [127], due also to the presence of zinc salts as a common component of protein crystallization solutions. However, it has been estimated that nearly 200 unique structures for zinc proteins are available, representing all six classes of enzymes and covering a wide range of phyla and species. Three primary types of zinc sites are apparent from examination of these structures: structural, catalytic and co-catalytic [126]. The most common amino acids that supply ligands to these sites are histidine, glutamate, aspartate and cysteine. In catalytic sites zinc generally forms complexes with water and any three nitrogen, oxygen and sulfur donors with His being the predominant amino acid chosen. Structural zinc sites have four protein ligands, with Cys as the preferred ligand, and no bound water molecule. Co-catalytic sites contain two or three metals in close proximity with two of the metals bridged by a side chain moiety of a single amino acid residue, such as Asp, Glu or His and sometimes a water molecule. Asp and His are the preferred amino acids for these sites. Moreover, the influence of zinc on quaternary protein structure has led to the identification of a fourth type of binding site called “protein interface zinc site” [128]. In this case, the ligands are supplied from amino acid residues residing in the binding surface of two interacting proteins. This novel class of zinc sites establish permanent, and modulate transient, PPIs, determining crucial functions such as catalysis, inhibition of enzymatic or other activities, protein packing, assembly or disassembly of multi-subunit macromolecular complexes, setting up of molecular scaffolds and protein-receptor association. The dependence of these PPIs on zinc availability suggests new mechanisms of regulating protein supramolecular assembly and quaternary structure.

The occurrence of different topologies of protein interface zinc sites has been reported for SAg (superantigen) proteins, toxins from viral or bacterial sources which form complexes with both MCH-II (major histocompatibility complex class II) and T-cell receptors [129]. The SAg are not processed to small peptides in APCs (antigen presenting cell) but they form a brace that links the MCH-II and the T-cell receptor. The best-characterized are the SEs (staphylococcal enterotoxins) and SPEs (streptococcal pyrogenic exotoxins), a family of bacterial toxins produced by *Staphylococcus aureus* and *Staphylococcus pyogenes* [130], respectively. These toxins are

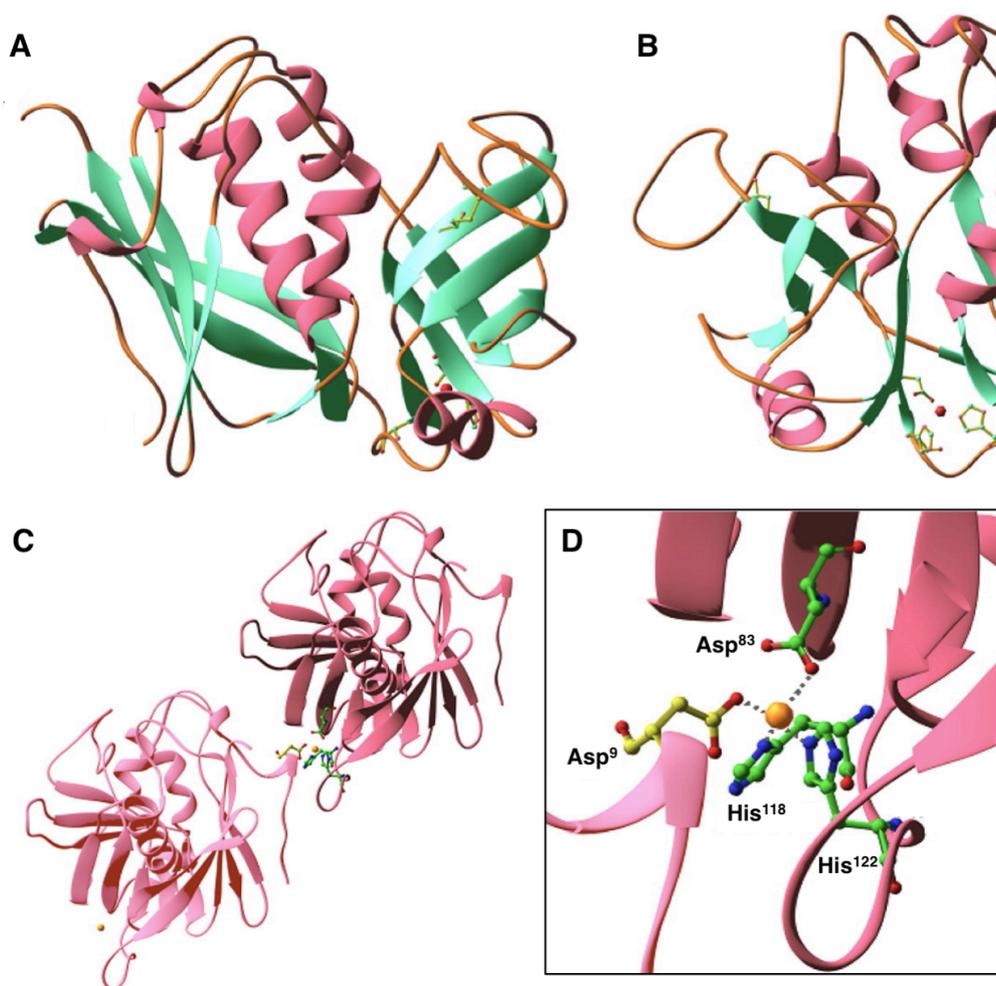
globular proteins of 22 - 29 kDa featuring a dimeric structural architecture and can be divided in three subfamilies based on the degree of sequence homology: SEA, SED, and SEH form one subgroup, while SEC and SPEC other two distinct subclasses. One common property of SEs and SPEs, but not of all Sags, is their ability to bind zinc and undergo homo-dimerization or hetero-dimerization with MHC-II. The Zn<sup>2+</sup>-dependent homo-dimerization of Sags can involve two distinct binding sites, while the hetero-dimerization process leads to the establishment of a zinc binding site where the fourth ligand comes from an MHC-II molecule. In particular, three different binding types are observed (see Table 2): *i*) site 1, involving residues located at the bottom of the cleft between two Sags domains; *ii*) site 2, constituted by amino acids on the surface of the  $\beta$ -grasp motif at the C-termini *iii*) site 3, formed by residues on the edge of the  $\beta$ -barrel small domain.

Superantigen	Zinc site	Zinc ligands (monomer A)	Zinc ligands (monomer B)
<i>Subfamily I</i>			
SEA	2	His <sup>187</sup> , His <sup>225</sup> , Asp <sup>227</sup>	His <sup>61</sup> (or His <sup>81</sup> from MHC-II)
SED	2	Asp <sup>182</sup> , His <sup>220</sup> , Asp <sup>222</sup>	His <sup>218</sup>
	1	His <sup>114</sup> , Lys <sup>118</sup>	His <sup>13</sup> Glu <sup>17</sup>
SEH	2	His <sup>206</sup> , Asp <sup>208</sup>	His <sup>81</sup> from MHC-II
<i>Subfamily II</i>			
SEC3	1	Asp <sup>83</sup> , His <sup>118</sup> , His <sup>122</sup>	Asp <sup>9</sup>
<i>Subfamily III</i>			
SPEC	2	His <sup>167</sup> , His <sup>201</sup> , Asp <sup>203</sup>	His <sup>81</sup> (or His <sup>81</sup> from MHC-II)
	3	His <sup>35</sup> , Glu <sup>54</sup>	His <sup>35</sup> , Glu <sup>54</sup>

**Table 2 - Sags' zinc binding sites involved in PPIs**

Site types (see text for definition) and ligands involved in the homo-dimerization of SAGs or the hetero-dimerization of SAGs with MHC-II are reported.

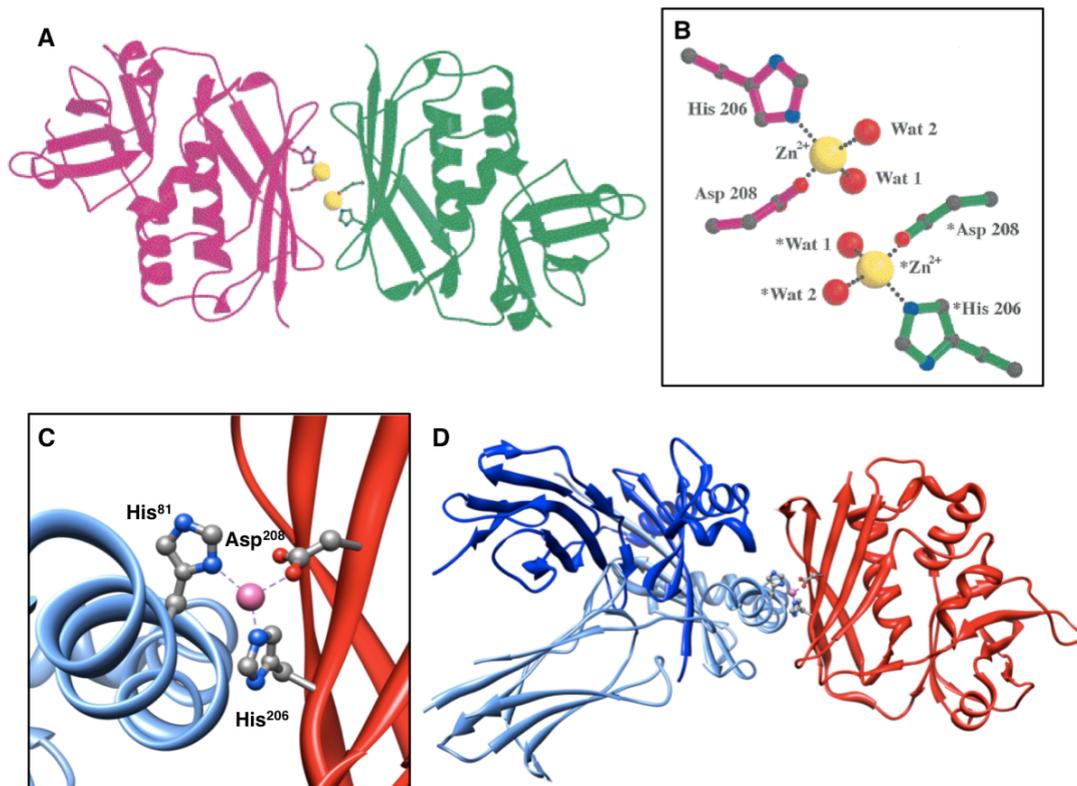
The monomeric crystal structure of SEC3 is shown in Figure 14A. The endotoxin displays structural features that are very similar to those found in other SAGs, characterized by two unequal sized domains. A zinc binding site is present in the crystal structure, located at the bottom of the cleft between the two domains [131]. The metal ion is coordinated by an aspartate (Asp<sup>83</sup>) from a  $\beta$ -strand of the small domain and two histidine residues (His<sup>118</sup>, His<sup>122</sup>) from the connecting loop (Figure 14B). The SEC3 dimerization occurs in the crystal environment in the presence of zinc; the dimeric structure is shown in Figure 14C. The process buries a large portion of solvent-accessible surface area (~590 Å<sup>2</sup> per monomer) and lead to the formation of two salt bridges at the interface of the dimer, involving Lys<sup>37</sup>, Lys<sup>56</sup> from one monomer and Asp<sup>10</sup>, Glu<sup>16</sup> from a neighboring molecule. The same Zn<sup>2+</sup>-mediated dimerization was observed in the SEC2 structure [132], which was grown in different conditions. The new structural arrangement of the dimeric zinc binding site is shown in Figure 14D; in this crystal, tail of Asp<sup>9</sup> from a neighboring SEC3 molecule complete the metal coordination shell, causing a conformational change of the N-terminal tail and protein dimerization.



**Figure 14 - Structure of staphylococcal enterotoxin C3 and Zn<sup>2+</sup>-mediated dimerization**

Ribbon diagrams of SEC3 monomer (A), with closeup of its monomeric zinc binding site (B), and dimer (C), with details of the Zn<sup>2+</sup> binding site in a dimeric configuration (D) (PDB entry: 1ck1). In panel D, Asp<sup>9</sup> is from the neighboring protein molecule. Residues involved in metal coordination are represented as a *ball and stick* model. Zinc ions (red, in panel A and B; yellow in panel D) and the disulfide bond are also shown.

A different binding site architecture is shown by SEH, which features two zinc ions at the interface of its dimeric structure (Figure 15A). In fact, in the crystal each SEH monomer binds one Zn<sup>2+</sup>, which is coordinated by only two protein ligands, His<sup>206</sup> and Asp<sup>208</sup>, and two water molecules (Figure 15B). His<sup>206</sup> forms a hydrogen bond with Ser<sup>205</sup> and an additional hydrogen bond stabilizes the zinc coordination environment. Finally, these metal ions are located at a protein interface but they are not properly interface zinc sites, since the ligands stem from only one polypeptide chain. This dimeric structure demonstrates that zinc can be bound by only two protein ligands. Moreover, in the complex with MHC-II, SEH displays a different coordination environment for zinc (Figure 15D). In fact, MHC-II binds to the SEH surface, that would otherwise serve as the dimerization interface with a second SEH molecule, and provides His<sup>81</sup> from its  $\beta$ -chain as a third protein ligand to one zinc ion (Figure 15C). In this case, the presence of water molecules as possible ligands has been excluded due to the low occupancy in the fourth coordination position in the crystal structure.

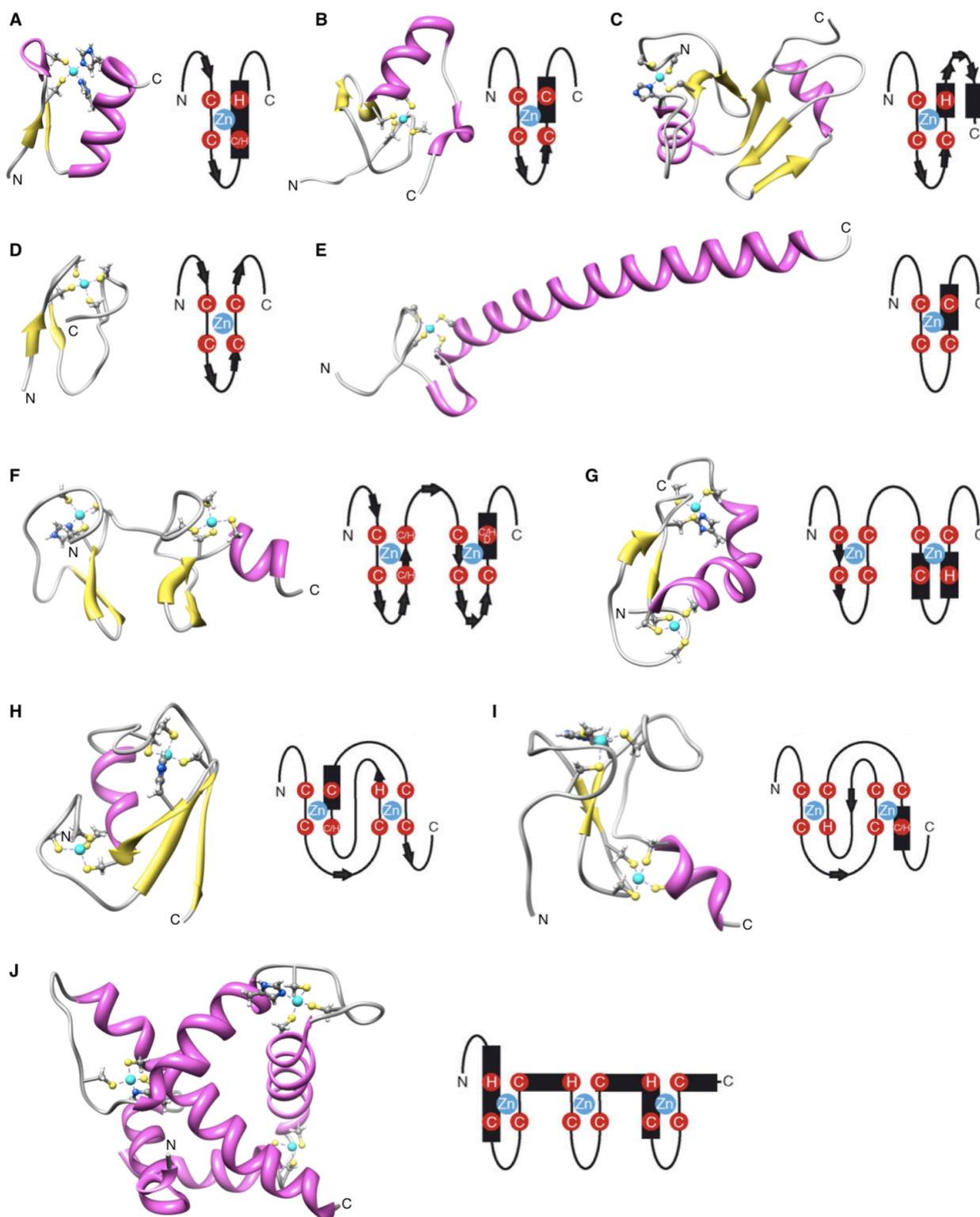


**Figure 15 - Zinc coordination of staphylococcal enterotoxin H**

Ribbon diagrams of SEH dimer (A) (the monomers are shown in pink and green, PDB entry: 1hxy) and the hetero-complex between SEH (red) and MHC-II (blue) (D) (PDB entry: 1ewc). Details of the two corresponding zinc binding sites (B,C). In panel B, Wat 1 and Wat 2 refer to the two water molecules that coordinate zinc in each toxin monomer. Residues involved in metal coordination are represented as a *ball and stick*. Zinc ions (yellow in panels A and B; pink, in panels C and D) are shown as *spheres*.

The examples of SAg proteins examined here demonstrate a biological role of protein interface zinc sites in immune functions and host-pathogen interaction and raise the possibility that such  $Zn^{2+}$ -mediated PPIs are important for yet additional immune functions. More generally, zinc binding sites of SAg exemplify the difficulty encountered in establishing the number of interface zinc sites and in determining their relevance and function. The physiological significance of such sites may be evaluated only for assembly of a protein with its correct biological partner. In fact, in the absence of the protein partner, zinc may coordinate differently or may be absent owing to a lack of stabilization.

However, apart from the role of interface zinc site in PPIs, other zinc structural arrangements has been demonstrated to actively mediate protein-protein association. At this purpose, an intriguing lesson could be learned by the examination of zinc fingers domains (ZnFs). ZnFs are common, relatively small (<100 residues) protein motifs that fold around one or more  $Zn^{2+}$  ions in a structural manner through cysteine, histidine, and occasionally aspartate, side chains. This motif was first identified as DNA-binding domain in the *Xenopus leavis* transcription factor TFIIIA [133], and subsequently the term has been extended to many classes of zinc ligating domains.

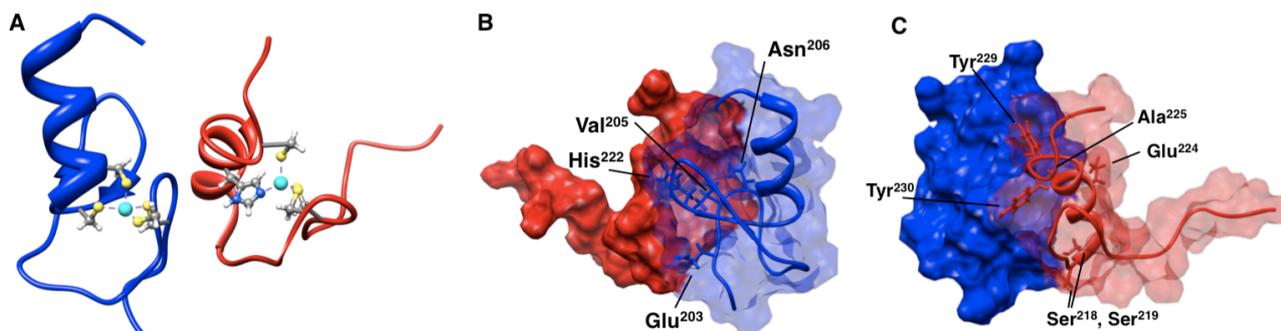


**Figure 16 - Topology and structures of ZnFs involved in PPIs**

Ribbon diagrams and topological representations of the most common ZnFs: classical C2H2 (A) (PDB entry: 1tf3); GATA-like (B) (PDB entry: 1gat); ZnF UBPF (C) (PDB entry: 2g43); RanBP zinc ribbon (D) (PDB entry: 1q5w); A20 treble cleft (E) (PDB entry: 2fid); LIM (F) (PDB entry: 1a7i); MYND (G) (PDB entry: 2jw6); RING (H) (PDB entry: 1chc); PHD (I) (PDB entry: 1xwh); TAZ domain-like (J) (PDB entry: 1r8u). In the reported structures, residues that coordinate metal ions are represented as *ball and stick* and colored using the following atomic scheme: carbon, grey; nitrogen, blue; sulfur, yellow; hydrogen, white; zinc, cyan.

ZnFs are typically assumed to mediate DNA binding but, in the last ten years, detailed structural and functional data have been reported for several PPIs involving ZnF proteins, proving unequivocally that these zinc binding structural elements can act as protein recognition motifs [134-137]. Here, the protein binding capabilities of different ZnF classes are described following a zinc coordination topology based classification (Figure 16).

The involvement of classical C2H2 ZnF in a large variety of PPIs (more than 100 protein-protein assemblies) it has been recently reviewed by Brayer and Segal [136]. This is not surprising considering that, through all kingdoms, C2H2 domain is not only ubiquitous but is also one of the most common protein motifs found in eukaryotic genomes. Classical ZnFs comprise a short  $\beta$ -hairpin and an  $\alpha$ -helix coordinated in either a Cys-Cys-His-His or a Cys-Cys-His-Cys manner (Figure 16A). Structural evidences for transcriptional regulator FOG-1 (friend of GATA-1), containing nine classical ZnFs, in mediating PPIs have been reported [138]. In particular four of these ZnFs can independently bind the N-terminal domain of GATA-1 (also termed “N-finger”), another ZnF from the GATA-type group (Figure 16B). The solution structure of the complex between the *Drosophila* FOG-like domain and GATA-1 N-finger is shown in Figure 17A. The two protein domains contact each other through a combination of hydrophobic contacts and polar interactions, that ensure specificity to the macromolecular recognition (Figure 17B,C). In particular, the binding surface of N-finger is centered on Val<sup>205</sup> which forms extensive hydrophobic contacts with proximal residues of FOG (Thr<sup>222</sup>, Ala<sup>225</sup>, Tyr<sup>229</sup> and Tyr<sup>230</sup>). The remainder of the interactions are largely polar in nature: Glu<sup>203</sup> and Arg<sup>202</sup> from N-finger interacts with Ser<sup>218</sup>, Ser<sup>219</sup> residues of FOG, whereas His<sup>222</sup> and Asn<sup>206</sup> from N-finger pair with Glu<sup>224</sup> and Tyr<sup>230</sup> of FOG, respectively.



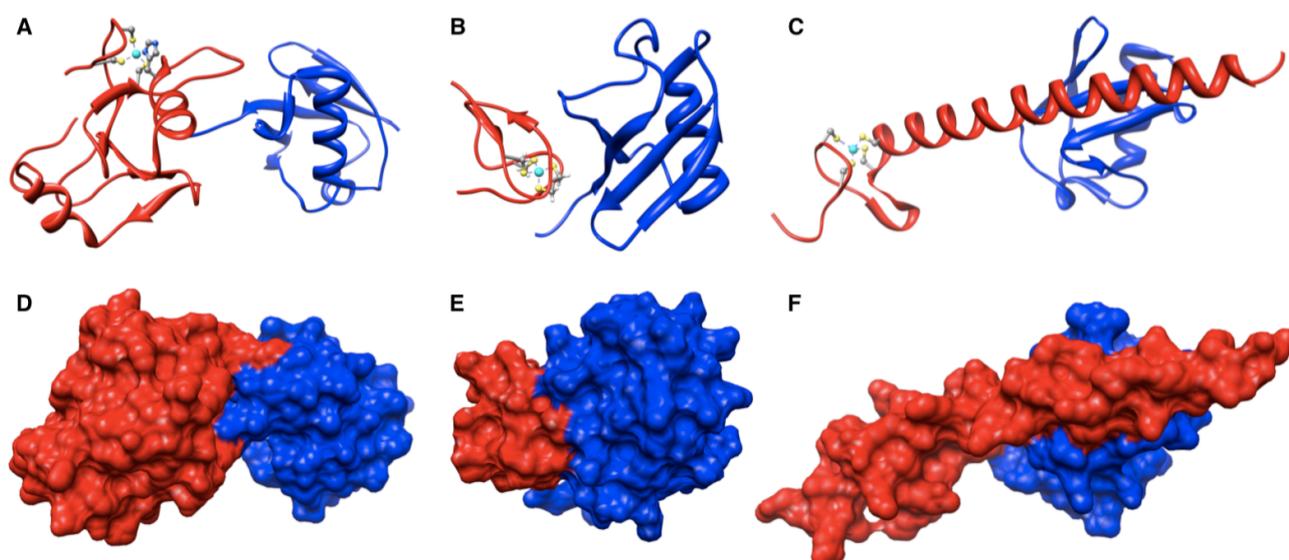
**Figure 17 - Protein binding capability of classical ZnFs**

Ribbon diagram of the complex between *Drosophila* FOG-like domain (red) and GATA-1 N-finger (blue) (A, PDB entry: 1y0j). Residues that coordinate zinc ions are represented as *ball and stick* and colored using the following atomic scheme: carbon, grey; nitrogen, blue; sulfur, yellow; hydrogen, white; zinc, cyan. Solvent excluded surface representations of the complex (B,C): transparency was alternatively applied on both ZnFs in order to better investigate interfaces involved in protein recognition. Residues involved in PPIs are highlighted using a *stick* model and colored according to the scheme used in panel A.

Interestingly, the complex structure shows that the DNA recognition motif of the GATA-1 N-finger is distinct from the FOG-interacting domain. By contrast, the protein binding domain of FOG

overlaps considerably with the highly conserved DNA recognition motif shown by classical ZnFs. This evidence, together with the fact that protein binding domains in the structure of FOG-1 are widely spaced, supports the idea that these ZnFs are not typical DNA recognition modules but instead they could play a crucial role in PPIs.

Another example of ZnFs involved in PPIs is constituted by RanBP domains, termed in this way due to their presence in the nuclear export protein RanBP2. These domains, found in proteins from one to eight repeats, consist of two hairpins that sandwich a single  $Zn^{2+}$  ion and display a zinc ribbon fold (Figure 16D). The structure of the RanBP domain of Npl4, a protein involved in endoplasmic reticulum-associated degradation and nuclear-envelope reassembly, in complex with ubiquitin has been recently reported [139]. The solution structure of this complex is shown in Figure 16B,E. In particular, the crucial determinant for the association was found in the Thr-Phe dipeptide located in the zinc ligand of Npl4, following a Cys-Thr-Phe-Cys arrangement. Interestingly, also the deubiquinating enzyme IsoT (isopeptidase T) contains a zinc binding motif termed ZnF UBP (Figure 17C) that is a structural elaboration of the GATA-like domain. This domain specifically binds the C-terminal Gly-Gly dipeptide, as displayed in the available structure of IsoT-ubiquitin complex (Figure 18A,D) [140], and correctly directs the recycling of polyubiquitin chains.



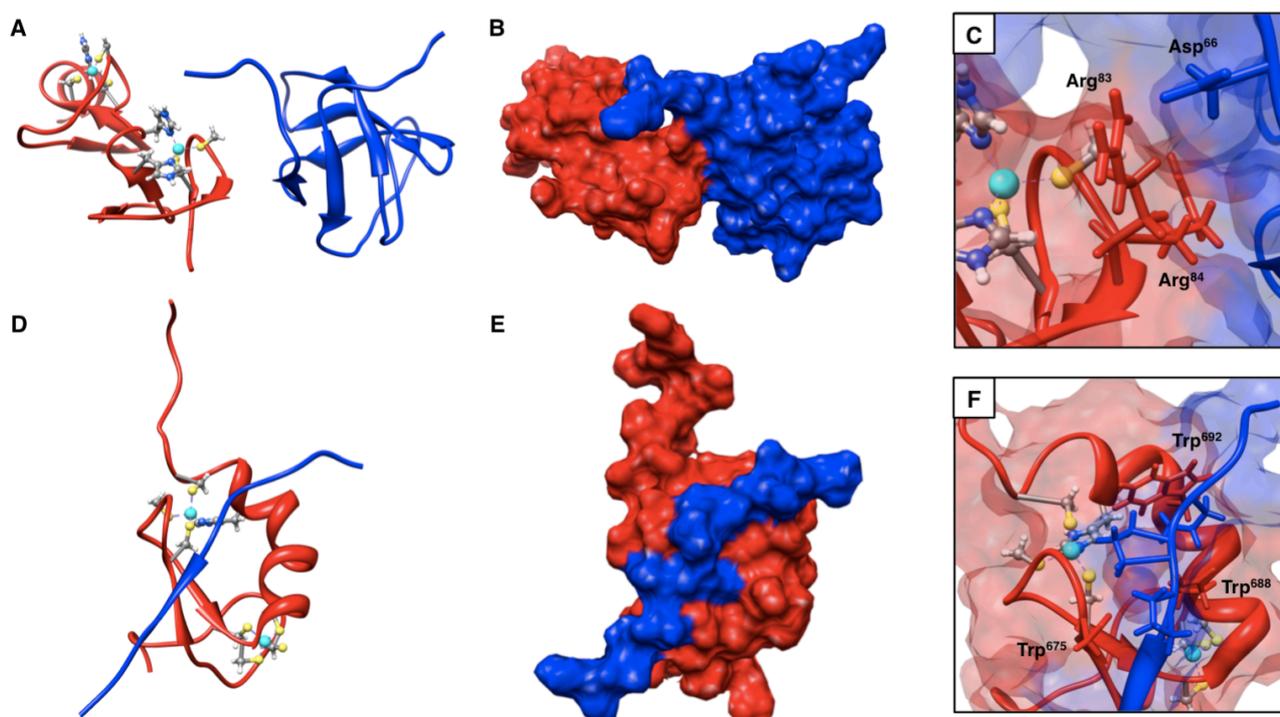
**Figure 18 - Binding of ZnFs to ubiquitin**

Ribbon diagrams and solvent excluded surfaces of ZnF proteins (red) in complex with ubiquitin (blue): ZnF UBP of IsoT (A,E) (PDB entry: 2g45), RanBP ZnF of Npl4 (B,D) (PDB entry: 1q5w), A20 ZnF of Rabex-5 (C,F) (PDB entry: 2fid). Residues that coordinate zinc ions are represented as *ball and stick* and colored using the following atomic scheme: carbon, grey; nitrogen, blue; sulfur, yellow; hydrogen, white; zinc, cyan.

A similar ubiquitin-binding activity has been also reported for another ZnF, the A20 zinc binding domain from the guanine nucleotide exchange factor Rabex-5. This domain belongs to a treble-cleft ZnF family that was originally identified in the protein A20 (Figure 16E) [141]. As shown in the structure of this domain in complex with ubiquitin, the A20 ZnF of Rabex-5 displays a completely

different binding mode to ubiquitin from those displayed by RanBP and UBP ZnFs. The existence of these two different binding sites imply the possibility that a single molecule of ubiquitin might be recognized simultaneously by two different ZnF proteins. This hypothesis is supported by the two opposing effects on the NF- $\kappa$ B signalling pathway. In fact, TAB2 and TAB3 proteins, which show a RanBP ZnF, bind to polyubiquitin activating NF- $\kappa$ B signalling whereas A20 protein polyubiquitinates the mediator protein RIP, targeting it for degradation and thereby down-regulating NF- $\kappa$ B signalling.

Other classes of ZnFs that coordinate more than one zinc ion have been also reported to mediate PPIs. This is the case of LIM (named after the three proteins, LIN-11, Isl1 and MEC-3, where it was first discovered [142]), and MYND (termed after the three proteins Myeloid translocation protein 8, Nery and DEAF-1 [143]) domains.



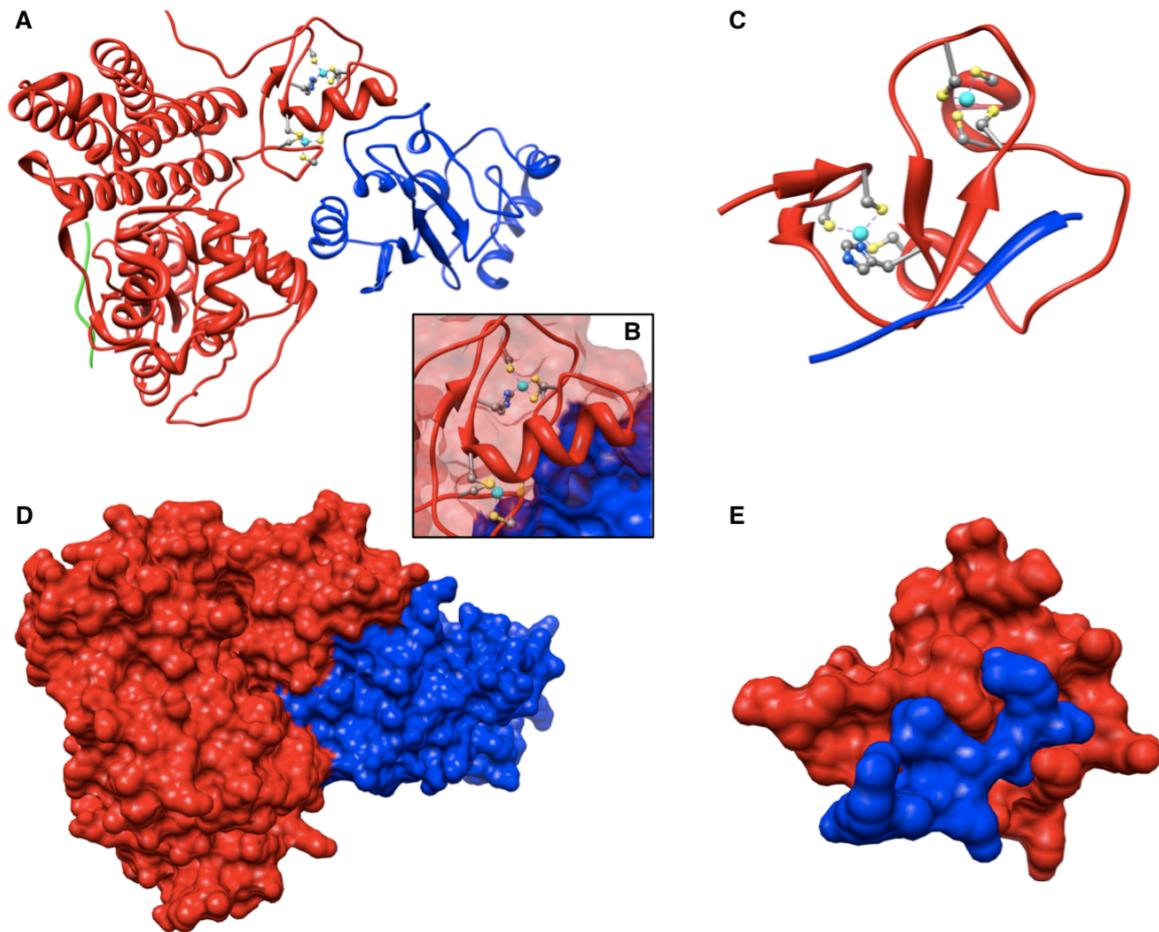
**Figure 19 - Protein binding capability of LIM and MYND ZnFs**

Ribbon diagram (A) and solvent excluded surface (B) of the LIM4 domain of PINCH (red) in complex with SH3 domain of Nck2 (blue)(PDB entry: 1u5s). Ribbon diagram (D) and solvent excluded surface (E) of the MYND domain from AML1/ETO (red) in complex with SMRT/N-CoR peptide (blue)(PDB entry: 2odd). Residues that coordinate zinc ions are represented as *ball and stick* and colored using the following atomic scheme: carbon, grey; nitrogen, blue; sulfur, yellow; hydrogen, white; zinc, cyan. Details of the two interactions are shown in panels C and E. Here, residues involved in PPIs are shown using a *stick* model.

LIM ZnFs comprise two sequential zinc binding domains that resemble the GATA-type fold and show cysteine, histidine, aspartate and glutamate as zinc coordinating residues in their consensus sequence (Figure 16F). The involvement of LIM domains in protein recognition is supported by currently available data, also suggesting a role as molecular bridges for proteins that contains

multiple LIM ZnFs [144]. A pertinent example is provided by the adaptor protein PINCH, which consists of solely five LIM domains and plays roles in cell adhesion, growth and differentiation. PINCH interacts with ILK (integrin-like kinase) and Nck2, and both associations are essential for integrin signalling. Specifically, the first domain of PINCH, LIM1, binds to the ankyrin repeat domain of ILK through its second zinc binding module, while the fourth ZnF of PINCH, LIM4, interacts with the third SH3 domain of Nck2. The structural details of the latter interactions are shown in Figure 19A,B. In particular two arginine residues from LIM4 are responsible for the formation of salt bridges and hydrophobic contacts with Nck2 (Figure 19C). MYND domains also coordinate two zinc atoms in a sequential way (Figure 16G), but they are smaller than LIM ZnFs and usually the first module contains only a  $\beta$ -hairpin whereas the second  $Zn^{2+}$ -binding site is formed by two short helices. Like LIM domains, also MYND ZnFs act as protein binding motifs and are mainly present in transcriptional regulators. Little structural information is available on MYND-mediated PPIs, but it is known that MYND domains recognize a Pro-X-Leu-X-Pro motif in their protein partner. However, the structure of the MYND domain from AML1/ETO, involved in the repression of cell proliferation in granulocyte differentiation, in complex with SMRT/N-CoR peptide has been recently released [145]. As in the case of other domains that bind to Pro-rich sequences, such as SH3 and WW, the MYND domain binding specificity is achieved through a combination of proline-tryptophan packing interactions and other van der Waals contacts via the complementary surfaces. In this complex, the first proline in the "PPPLIP" motif of SMRT/N-CoR peptide packs against Trp<sup>692</sup> from MYND domain (Figure 19F). In addition, the second and third proline residues form hydrogen bonds with Asn<sup>688</sup> and Ser<sup>675</sup>, respectively.

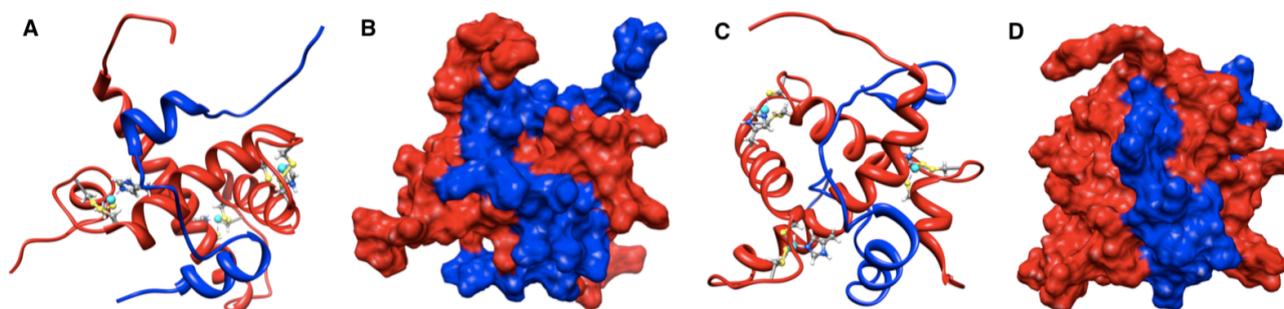
Also other domains that coordinate two zinc ions, such as RING (really interesting new gene) and PHD (plant homeodomain), function primarily as protein recognition motifs [137]. However, for these ZnFs the two set of zinc ligands are not sequential but instead are interdigitated or "cross-braced" (Figure 14H,I). RING ZnFs seem to participate in a wider range of functions, but a large majority of these domains is found in E3 ubiquitin ligases which catalyze the final step in the protein ubiquitination pathway. The crystal structure of the ternary complex of E3 ligase c-Cbl, E2 Ubch7 and phosphorylated ZAP-70 peptide, the latter being the substrate for ubiquitination, is shown in Figure 20A,D. The structure displays that the RING domain of c-Cbl makes direct contacts with Ubch7 through a conserved shallow groove on the surface of the former, that includes the RING ZnF (Figure 18B). On the other hand, PHD domains have long pointed a role in the regulation of chromatin structure. This hypothesis has been confirmed with the dissection of the interaction of PHD domain from ING2 (inhibitor of growth 2) with the N-terminal tail of H3K4me3 (histone H3 trimethylated on Lys4). The structure of this complex (Figure 20C,E) reveals that the side chain of the trimethylated lysine is recognized by a cage formed by two aromatic residues of the PHD domain. This evidence confirms the relevance of this Trp residue, located two residue N-terminal to the seventh metal ligand, in the protein binding mode of PHD domain.



**Figure 20 - Protein binding capability of RING and PHD ZnFs**

Ribbon diagram (A) and solvent excluded surface (D) of the ternary complex between E3 ligase c-Cbl (red), E2 UbcH7 (blue) and phosphorylated ZAP-70 peptide (green) (PDB entry: 1fbv). Detail of the two zinc binding sites is shown in panel B. Ribbon diagram (C) and solvent excluded surface (E) of the PHD domain from ING2 (red) with the N-terminal tail of H3K4me3 (blue) (PDB entry: 2g6q). Residues that coordinate zinc ions are represented as *ball and stick* and colored using the following atomic scheme: carbon, grey; nitrogen, blue; sulfur, yellow; hydrogen, white; zinc, cyan.

Finally, a last fold originated by the conjunction of two ZnFs is involved in protein-protein complex formation. This motif named TAZ (transcriptional adaptor zinc binding) seems to exist only in the transcriptional co-activator CBP; it displays three zinc ions coordinated in a triangular structure, where each  $Zn^{2+}$  is present in a loop region between two antiparallel  $\alpha$ -helices (Figure 16J). CBP is an acetyltransferase that is recruited to DNA through associations with a large number of transcription factors, using TAZ domains [146]. For example, the N-terminal TAZ1 domain binds to HIF-1 $\alpha$  (hypoxia inducible factor 1 $\alpha$ ) as shown by the available structure of TAZ1 bound to a peptide from HIF-1 $\alpha$  (Figure 21A,B) [147]. In this complex, the HIF-1 $\alpha$  peptide forms three helices that wrap almost completely around the TAZ1 domain. Also the structure of TAZ1 in complex with CITED2, which competes with HIF-1 $\alpha$  for TAZ1 in a feedback mechanism during the hypoxic response, has been solved (Figure 21C,D) [148]. The binding to CITED2 clearly occurs in a different way but involves an overlapping site on TAZ1.



**Figure 21 - Protein binding capability of TAZ ZnF**

Ribbon diagrams and solvent excluded surfaces of TAZ1 domain (red) in complex with HIF-1 $\alpha$  (A,B) (PDB entry: 1l3e) or CITED2 (C,D) (PDB entry: 1p4q). HIF-1 $\alpha$  and CITED2 are colored in blue. Residues that coordinate zinc ions are represented as *ball and stick* and colored using the following atomic scheme: carbon, grey; nitrogen, blue; sulfur, yellow; hydrogen, white; zinc, cyan.

To summarize, for many classes of ZnFs a functional role as protein recognition modules has been observed, even in the classes of ZnFs for which DNA-binding activity is well established such as GATA-type and classical ZnFs. This observation reinforces the idea that ZnF domains have evolved as stable molecular scaffolds, onto which evolution has grafted different binding functions as required. One conserved feature among these complexes is that ZnFs tend to undergo little, if any, structural rearrangement on recognizing their protein partners. This lack of rearrangement is presumably due in part to the stability afforded by zinc ligation. Another notable feature is that several of these interactions are modular in nature, similar to the mode of interaction observed for classical ZnFs binding to DNA. Moreover, as shown in Table 3, ZnF domains mediate interactions with a variety of different protein partners, featuring a wide range of affinity.

ZnF domain	Protein	Protein partner	Function	$K_d$ (M)	Ref.
GATA	GATA-1	FOG	Transcription regulation	$10^{-5}$	[138]
ZnF UBP	IsoT	Ubiquitin	Deubiquitination	$10^{-6}$	[140]
RanBP	Npl4	Ubiquitin	Ubiquitination and proteolysis	$10^{-4}$	[139]
A20	Rabex-5	Ubiquitin	Ubiquitination and proteolysis	$10^{-5}$	[361]
LIM	LMO4	Ldb1	Transcription regulation	$10^{-9}$	[366]
LIM	PINCH	Nck2 (SH3 domain)	Cell adhesion and migration	$10^{-2}$	[362]
RING	c-Cbl	UbcH7	Ubiquitination and proteolysis		[365]
PHD	ING2	H3K4me3	Chromatin regulation	$10^{-6}$	[363]
PHD	BPTF	H3K4me3	Chromatin regulation	$10^{-6}$	[363]
TAZ1	CBP	HIF-1 $\alpha$	Cellular hypoxic response	$10^{-7}$	[364]
TAZ1	CBP	CITED2	Cellular hypoxic response	$10^{-8}$	[148]

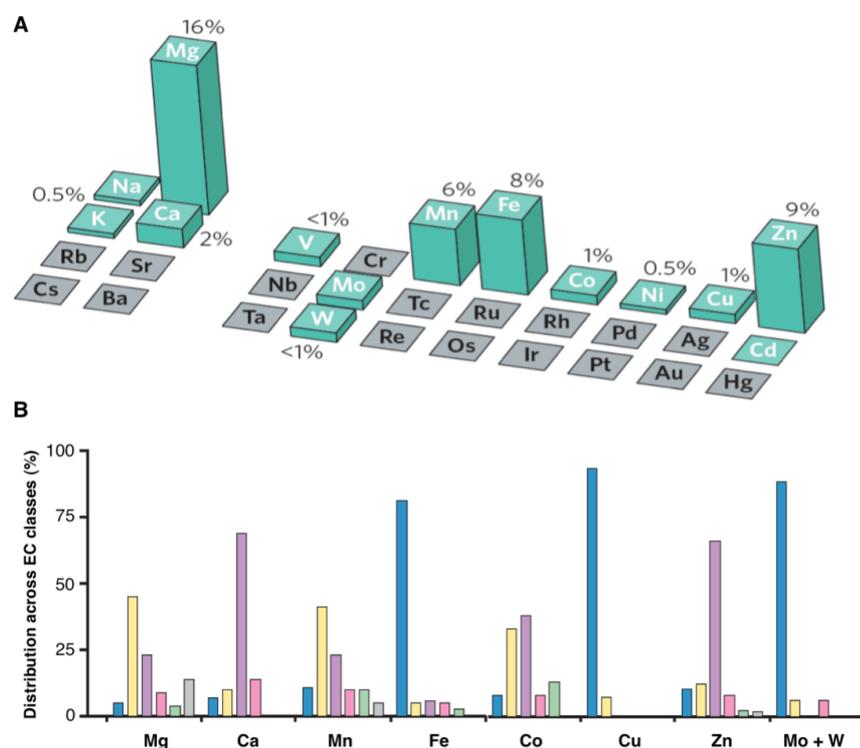
**Table 3 - ZnF-protein complexes: a wide range of functions and affinities**

These interactions tend to be of moderate to weak affinity (dissociation constant values range from micro-molar to nano-molar), which is perhaps not surprising considering the relatively small size of many ZnF domains. Nevertheless, the biological relevance of many of the weaker interactions, such as GATA-1 with FOG-1, Npl4 with ubiquitin and PINCH with Nck2, is indisputable, and despite their low affinity the associations are also highly specific. In fact, zinc binding could act as a control mechanism for these weak PPIs to trigger and strengthen the protein association reaction, or to promote the disassembly when zinc is subtracted.

## Section A3 - Metal homeostasis is regulated by PPIs: the accessory systems of nickel enzymes urease and hydrogenase

### 1.9 Metal trafficking systems: uptake, delivery and regulation

A commonly cited approximation is that one-third of proteins require metals. A systematic bioinformatics survey of 1.371 different structurally characterized enzymes estimated that 47% required metals, with 41% containing metals at their catalytic centers. Metallo-enzymes occur in all six EC (enzyme commission) classes, accounting for 44% of oxidoreductases, 40% of transferases, 39% of hydrolases, 36% of lyases, 36% of isomerases and 59% of ligases [149]. Magnesium is the most prevalent metal in metallo-enzymes, although it is often involved in loose partnerships with phosphate-containing substrates, such as ATP or ADP, and is sometimes interchangeable with manganese (Figure 22A). A catalogue of the principal type of enzyme that uses each metal reveals that iron (81%), copper (93%) and molybdenum plus tungsten (81%) are most commonly used as conduits for electrons in oxidoreductases (Figure 22B) [149]. Cobalt and molybdenum are found almost exclusively in association with cofactors in vitamin-B12-dependent and molybdopterin-dependent enzymes.

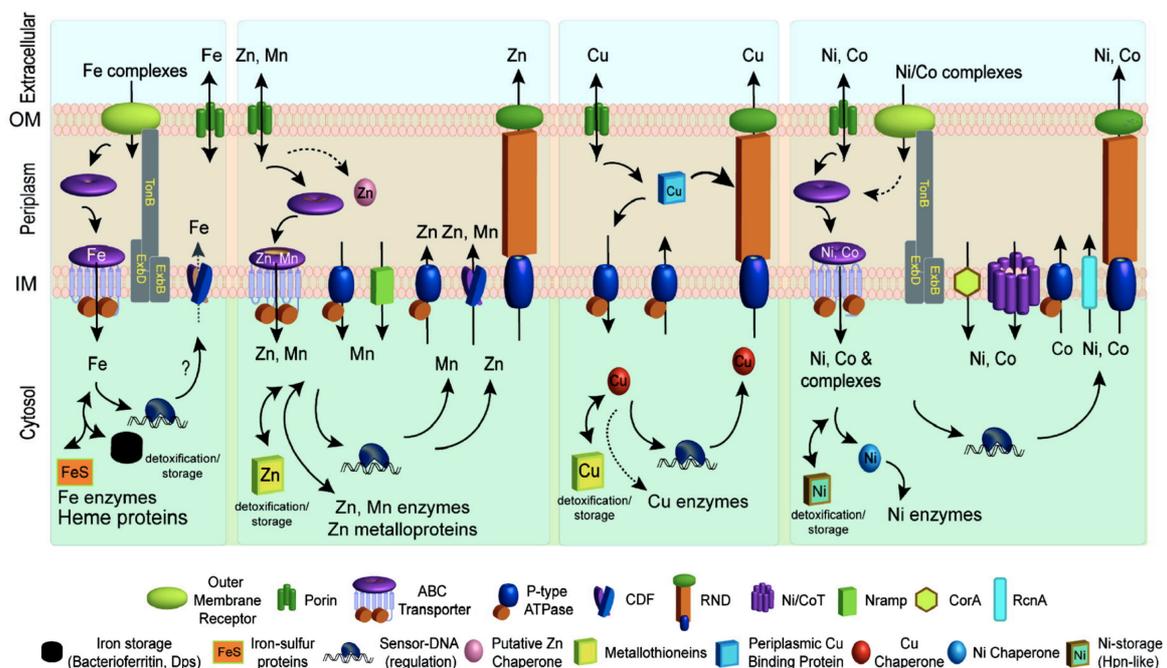


**Figure 22 - Metals in catalysis**

(A) The elements used as cofactors by enzymes are highlighted in cyan. The height of each column represents the proportion of all enzymes with known structures using the respective metal. A single enzyme uses cadmium. (B) The proportion of proteins using the indicated metals that occur in each of the six EC classes: oxidoreductases (EC1), blue; transferases (EC2), yellow; hydrolases (EC3), purple; lyases (EC4), pink; isomerases (EC5), green; ligases (EC6), grey.

On the other hand, the proportion of all proteins, not just enzymes, coordinating metals is expected to be less than 47% and the relative contributions of each metal may differ as the metals that perform structural roles, such as zinc in zinc fingers, are more fully accounted for. Of these, a

large part comprises proteins involved in the control of metal homeostasis, including specific protein-metal coordination complexes used to effect uptake, efflux, intracellular trafficking within compartments, and storage [150]. The acquisition, sequestration and homeostatic regulation in Gram-negative bacterial systems is schematized in Figure 23, which represents a general scenario for the principal biologically relevant metal ions. Metal transporters move metal ions across barriers, most of these being integral membrane proteins inserted in the inner or plasma membrane. Dedicated metallo-chaperones traffic metals within a particular cellular compartment and function to “hold” the metal and subsequently transfer it to the correct acceptor protein. This intermolecular transfer occurs through the establishment of transiently formed, specific PPIs that mediate intermolecular metal exchange. Finally, specialized transcriptional regulatory proteins, also named metallo-regulatory or metal sensor proteins, control the expression of genes encoding metal transporters, intracellular chelators, and/or other detoxification enzymes.



**Figure 23 - Metal homeostasis for Fe, Zn/Mn, Cu, and Ni/Co found in Gram-negative bacteria**

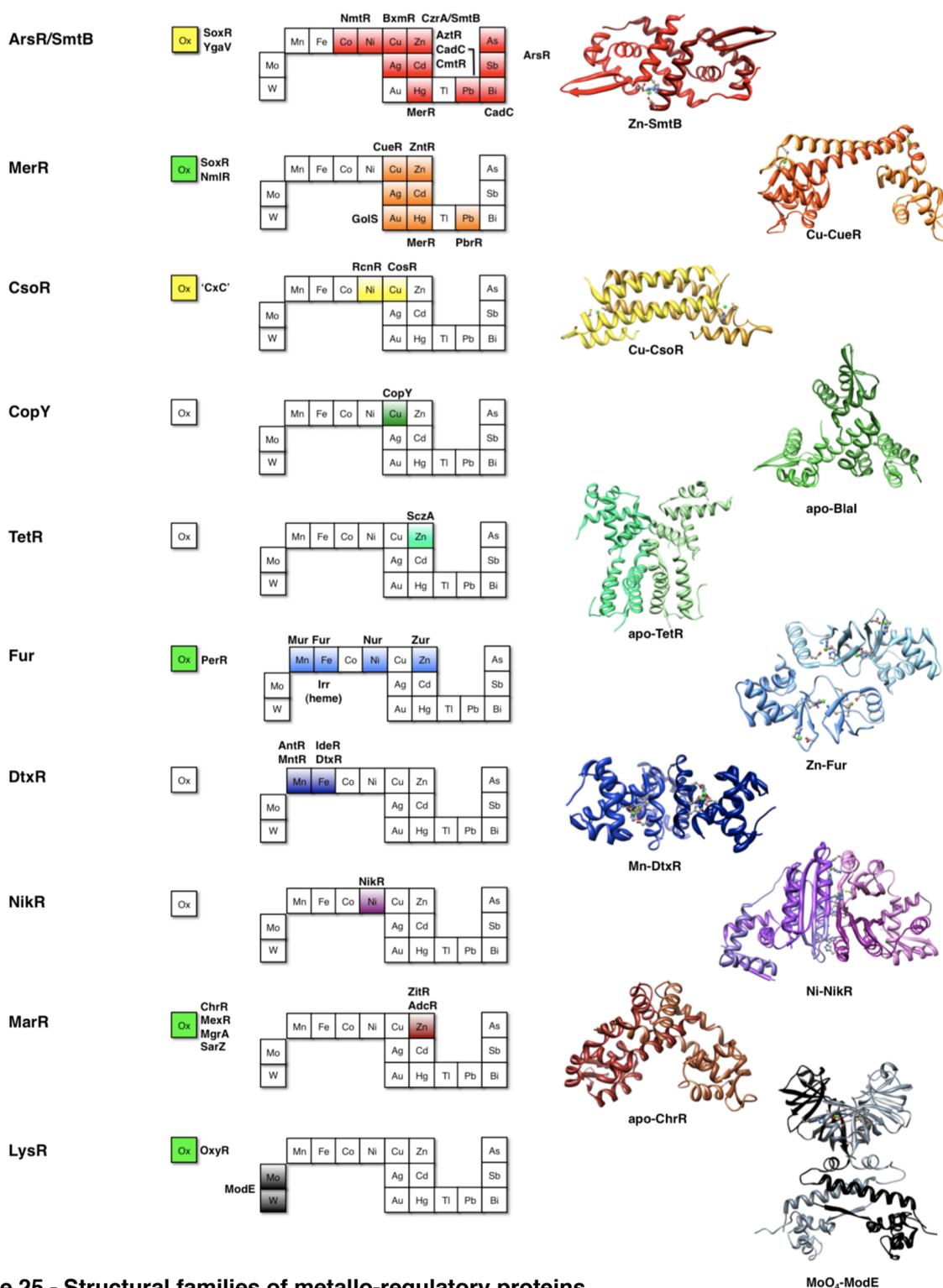
Homeostasis of molybdate and tungstate oxyanions are not shown, due primarily to a lack of knowledge of these systems, outside of uptake and cytosolic sensing. This schematic is not representative for any one bacterium nor is it meant to be exhaustive. Not all bacteria have all components of each homeostasis system indicated. The double-headed arrows are meant to illustrate that metals can move in and out of target protein targets in response to proteome remodeling. A putative chaperone shown for Zn is YodA/ZinT, while known metallo-chaperones for Cu (Atx1 and CopZ), and Ni-enzyme urease (UreE) or [NiFe]-hydrogenase (HypA) are also shown. Iron metallo-chaperones for Fe-S cluster assembly are not shown for clarity. The cytosolic transfer from a Cu chaperone is indicated by the dashed double-headed arrow. Outer membrane (OM) and inner membrane (IM) are indicated.

In bacterial systems, acquisition of essential metal ions from the extracellular milieu requires dedicated systems to pass membranes and finally entering the cytosol. In first instance, trimeric  $\beta$ -barrel porins allow for non-selective passive diffusion of metal ions across the outer membrane.



Precisely, ABC transporters and Nramp proteins mediate the accumulation of specific metal ions, while their export is largely carried out by cation diffusion CDFs, P-type ATPases, and tripartite RND (resistance-nodulation-cell division) transporters [155]. Individual members of each structural class of metal transporters are capable of transporting a variety of metals into and out of the cell, but some tend to be more selective for certain metals over others (Figure 24). For instance, Nramps have thus far only been identified as  $Mn^{2+}$  and  $Fe^{2+}$  transporters, whereas ABC transporters have been reported for nearly every biologically relevant metal ion. High-resolution structures of representative members of a number of multi-subunit ABC transporters [156], a single CDF protein, the  $Zn^{2+}$  transporter *E. coli* YiiP [157], and the  $Ca^{2+}$ -pumping P-type ATPase [158] have been solved, while a complete lower resolution model of an *Archaeal*  $Cu^{+}$ -translocating P-type ATPase, *Archaeoglobus fulgidus* CopA became available [159]. All these evidences have helped to clarify metal transport from a structural and mechanistic point of view.

Metallo-regulatory proteins have evolved metal coordination sites that “sense” specific metals ion(s) by forming specific coordination complexes; this, in turn, functions to activate/inhibit DNA binding or transcription activation, thereby controlling the expression of genes that mediate the adaptive response to environmental conditions. In general, metal sensor proteins that control metal uptake all bind metal ions as co-repressors, whereas metallo-regulatory proteins that regulate efflux and/or intracellular storage function via a transcriptional de-repression or an activation mechanism. Currently, ten major families of metal-sensing transcriptional regulators have been identified and three dimensional structures of at least one representative member has been reported for eight of these (Figure 25) [150, 160]. Sensor families are named referring to their founding members, that specifically are the  $As^{3+}/Sb^{3+}$  sensor *E. coli* ArsR [161], the  $Zn^{2+}$  sensor *Synechococcus* SmtB [162], *E. coli* MerR [163], the  $Cu^{+}$  sensor *M. tuberculosis* CsoR [164], the  $Cu^{+}$  sensor *E. hirae* CopY [165], the transcriptional factor *E. coli* TetR [166], the  $Fe^{2+}$  sensor *E. coli* Fur [167], the  $Fe^{2+}$  sensor *C. diphtheriae* DtxR [168], the  $Ni^{2+}$  sensor *E. coli* NikR [169], the transcriptional regulator *M. thermoautotrophicum* MarR [170] and the molybdate sensor *E. coli* ModE [171]. These metal sensors span the detection of the six primary biologically essential transition elements Mn, Fe, Co, Ni, Cu and Zn, as well as the heavy metals Ag, Au, Cd and Hg. In addition, ArsR and MerR family sensors are able to detect the heavy metal Pb and the three trivalent ions As, Sb and Bi, and the LysR family regulators are responsible for sensing Mo and W. Moreover, many metal sensors have been demonstrated or predicted to mediate resistance to oxidative and/or nitrosative stress (see green and yellow boxes in Figure 25). These proteins either exploit the reversible oxidation-reduction chemistry or intrinsic reactivity of cysteine thiols, or use a direct metal-mediated sensing of reactive oxygen species, as in the case of the MerR family regulator SoxR [172] and the Fur family regulator PerR [173].



**Figure 25 - Structural families of metallo-regulatory proteins**

For each family, boxes for metals sensed are shaded on the abbreviated periodic tables, and individual regulators that sense the particular metal(s) are indicated. Green and yellow boxes refer to metal- and nonmetal-sensing oxidative stress regulators, respectively. Ribbon representations of selected representative members are shown for each family: *S. elongatus* SmtB (PDB entry: 1r23); *E. coli* CueR (PDB entry: 1q05); *M. tuberculosis* CsoR (PDB entry: 2hh7); *S. aureus* Blal (PDB entry: 1sd4); *P. syringae* TerR (PDB entry: 3cdl); *P. aeruginosa* Fur (PDB entry: 1mzb); *B. subtilis* MntR (PDB entry: 1on1); *E. coli* NikR (PDB entry: 2hzv); *B. subtilis* OhrR (PDB entry: 1z9c); *E. coli* ModE (PDB entry: 1o7l). Where possible, metal ions are displayed as green spheres.

A central role in the intracellular trafficking of metal ions, briefly described above (see Figure 23), is played by metallo-chaperones, which specifically transport metal cofactors to diverse locations and subsequently sort them into correct metallo-enzymes via PPIs. The term metallo-chaperone is relatively new and refers to a family of soluble metal receptor proteins acting not as detoxification proteins but in a “chaperone-like” manner, guiding and protecting the metal ion while facilitating appropriate partnerships [174, 175]. Historically, the structural and functional knowledge of metallo-chaperone-mediated cofactor assembly derives primarily from studies of copper chaperones [176]. These chaperones belong to three functional groups: the Atx1-like chaperones, the copper chaperones for superoxide dismutase, and the copper chaperones for cytochrome *c* oxidase. Molecular information has also been collected on zinc [177], iron [178] and manganese [179] chaperones. Moreover, prokaryotic nickel-binding metallo-chaperones facilitating the insertion of Ni<sup>2+</sup> into nickel-dependent enzymes have been described. These proteins, acting in the urease and hydrogenase systems will be described in the next sections.

### 1.10 Urease and hydrogenase systems from *H. pylori*

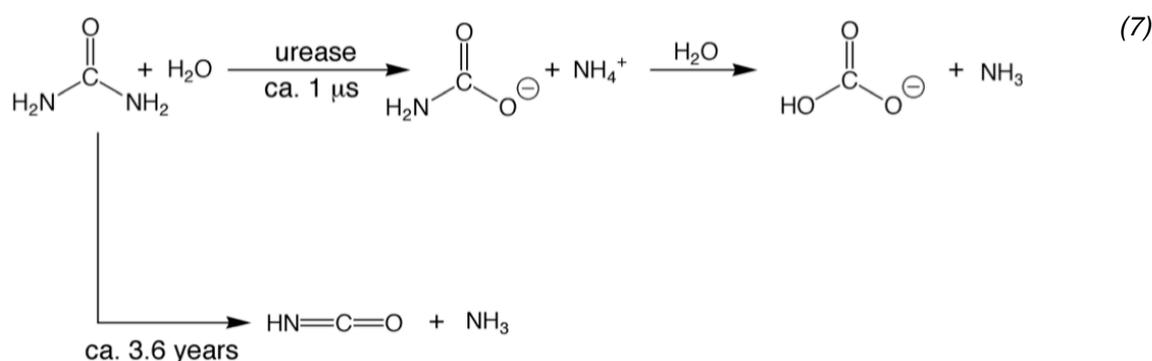
*Helicobacter pylori* is a spiral, micro-aerophilic, Gram-negative bacterium that colonizes the human gastric mucosa of approximately 50% of the world human population and is responsible for severe diseases, such as chronic gastritis, peptic and duodenal ulcers [180]. Moreover, this pathogen is considered as one of the risk factors for gastric neoplasms, including lymphomas and gastric cancers, and it has been classified as a “class 1” carcinogen by the WHO (world health organization) [181]. Its relevance has been recognized by the 2005 Nobel Prize for Medicine, awarded to Warren and Marshall, who discovered and isolated the pathogen in 1983 [182].

Actually, to date, *H. pylori* is the only known microorganism able to survive in the hostile niche of the human stomach [183]. It shows an optimal growth in a pH range around 6.0 and 8.0, but cannot survive at pH values below 4.0 or over 8.2 *in vitro* [184]. Indeed, its ability to colonize this acid environment (pH < 3.0) is related to the activity of two nickel-enzymes, urease and [NiFe]-hydrogenase [185-187]. In fact, the urea metabolism creates a micro-environment where pH conditions are suitable for bacterial survival and multiplication. The very high urease activity performed by *H. pylori* (this enzyme accounts for up to 10% of the total cellular protein) is found both in the cytoplasm and on the cell surface, constituting the essential protection from acid killing [188]. In addition, the bacterium is able to regulate its periplasmic pH at *ca.* 6.1 under acidic stress, and holds its cytoplasmic pH near 7.0 via the additional use of a membrane-linked carbonic anhydrase [189]. A key role for *H. pylori* survival is also played by the [NiFe]-hydrogenase, which provides the bacterium with a high-energy substrate yielding low potential electrons for energy generation [190]. In fact, use of H<sub>2</sub> by a bacterium that is otherwise metabolically deficient in carbon source catabolism (*Helicobacter* species use sugars poorly or not at all) may enable

*H. pylori* to take advantage of a readily available simple energy source, while obtaining carbon sources via peptides and amino acids [191]. *H. pylori* also suffers considerable oxidative stress *in vivo*, and some of the oxygen radical related detoxification enzymes require reducing power [192]. Perhaps a further role for H<sub>2</sub> oxidation is to facilitate production of such reductant. Here, the function, architecture and catalytic mechanism of both urease and [NiFe]-hydrogenase are discussed, whereas the metallo-chaperones required for nickel active sites assembly of these two enzymes are examined apart.

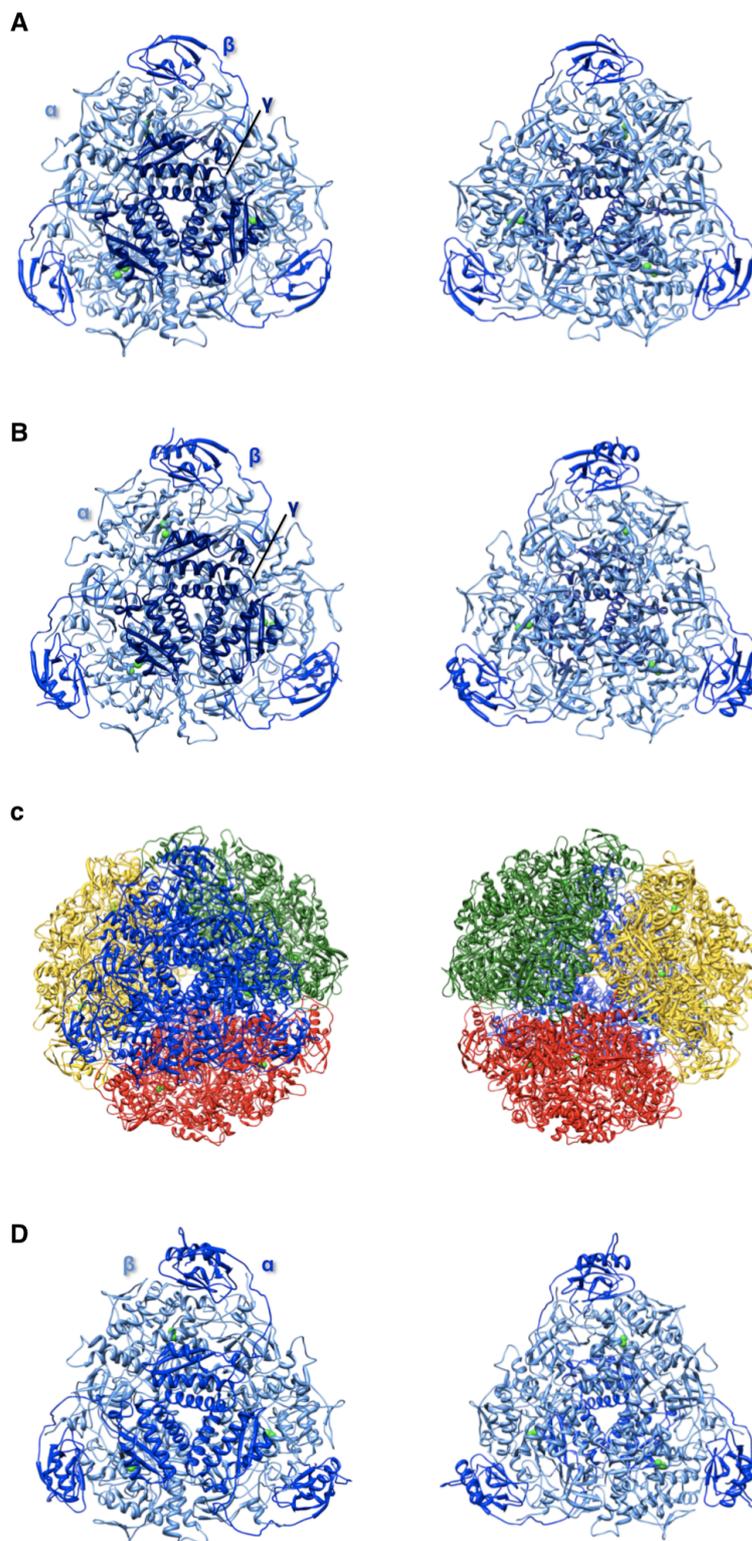
### 1.10.1. Urease: structure and activity

Urease plays a crucial role in the nitrogen cycle catalyzing the hydrolysis of urea, which is excreted by vertebrates, into ammonia and bicarbonate [193, 194]. Thus, urease is absent in vertebrates but facilitates nitrogen assimilation by plants, algae, and bacteria, a role that is underscored because urea is a major globally used soil fertilizer. This enzyme plays also an important role in many fundamental processes and it is a virulence factor for many ureolytic pathogens in the gut and urinary tract, promoting host colonization by neutralizing the low pH in the stomach. Urea is very stable, with a half-time of 3.6 years at 38 °C in solution [194]. Urease constitutes an efficient degrading system for this catabolite, providing a 3·10<sup>15</sup> fold enhancement of urea hydrolysis [195], with *k<sub>cat</sub>* values as high as 3·10<sup>3</sup> s<sup>-1</sup>. The reaction proceeds through a different mechanism with respect to the non-catalyzed degradation: urea is decomposed to ammonia and carbamate, which then spontaneously reacts at physiological pH, to give a second molecule of ammonia and bicarbonate [194, 196] (7):



Urease holds an important place in the history of protein discoveries. Jack bean urease was the first enzyme to be crystallized [197], and the first shown to contain nickel in its catalytic site [198]. Although the first urease enzyme has been crystallized more than 80 years ago, its structure has not been solved yet, due to the insufficient resolution of its protein crystals. This protein is a homo-hexamer (α<sub>6</sub>) and its molecular weight is about 590 kDa. Each monomer apparently contains two nickel ions [196]. In contrast to plant enzymes, microbial ureases are generally trimeric assemblies of hetero-trimers, showing molecular weights ranging from 200 to

250 kDa. Exceptions are those from *H. pylori*, which is considerably larger than the other bacterial ureases (MW = 1.1 MDa) [186], and the enzyme from *Prochlorococcus marinus*, showing the lower molecular weight (MW = 168 kDa) [199]. Many structures of urease from different bacterial sources, in the presence and absence of substrates and inhibitors, are available [194]. In particular, the crystal structures of urease from *Klebsiella aerogenes* [200], *Bacillus pasteurii* [201] and *H. pylori* [186], have been elucidated (Figure 26).

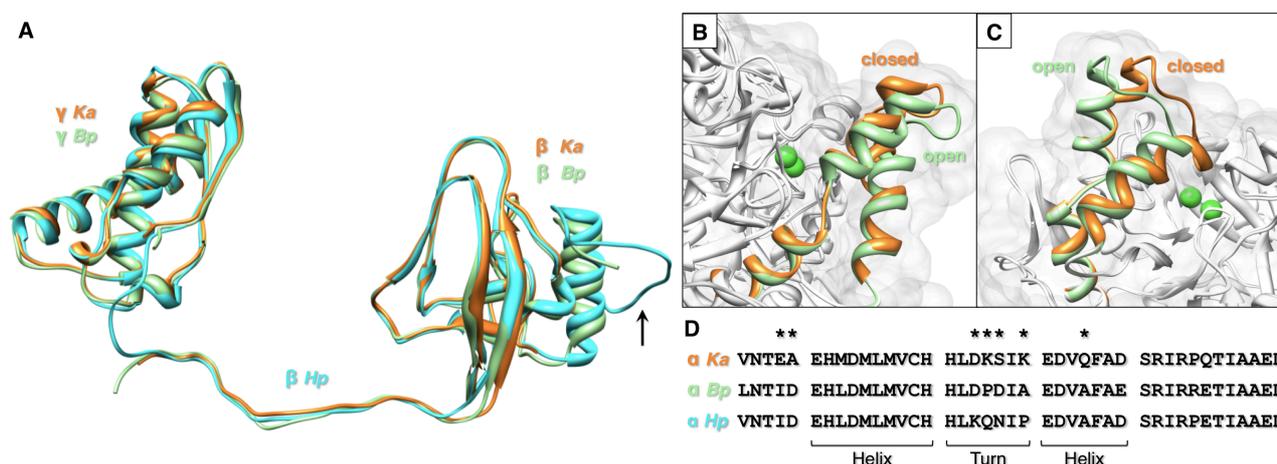


**Figure 26 - Crystal structures of bacterial ureases**

Ribbon representations of native ureases from *Klebsiella aerogenes* (A) (PDB entry: 1fwj), *Bacillus pasteurii* (B) (2ubp) and *Helicobacter pylori* (C,D) (PDB entry: 1e9z). The first two structures are composed by a trimer of hetero-trimers  $(\alpha\beta\gamma)_3$ .  $\alpha$  subunits are shown in sky blue,  $\beta$  subunits in blue and  $\gamma$  subunits in dark blue. The tetrameric superassembly of trimers  $[(\alpha\beta\gamma)_3]_4$  from *H. pylori* is shown in panel C (each trimer is differently colored). However, the trimer of hetero-dimers  $(\alpha\beta)_3$ , is similar to the functional trimer of hetero-trimers of the other two bacterial ureases. In panel D, a single trimer is shown, in the same representation used in panels A and B ( $\alpha$  subunit in blue,  $\beta$  subunit in sky blue). For each panel, a second picture, where the structure is rotated by  $180^\circ$  around the vertical axis, is shown for best illustration. Nickel ions are displayed as green spheres.

The first two enzymes show a very similar architecture and both are constituted by trimers of hetero-trimers  $(\alpha\beta\gamma)_3$ , with the three subunits codified respectively by *ureC*, *ureB* and *ureA* genes. In both structures, each  $\alpha$  subunit packs between the other  $\alpha$  subunits, forming the sides of a triangle, whose vertices are occupied by  $\beta$  subunits, whereas each  $\gamma$  subunit contacts two distinct  $\alpha$  monomers, forming the central core of the trimer (Figure 26A,B). Each trimer contains an independent bimetallic catalytic site, located in the  $\alpha$  subunit.

The *H. pylori* urease structure has revealed a much larger dodecameric assembly with a prominent central core (Figure 26C). The tetramer  $[(\alpha\beta)_3]_4$  is composed by four trimers of  $\alpha\beta$  hetero-dimers, similarly to the trimeric assembly shown by both *B. pasteurii* and *K. aerogenes* ureases. Here, the  $\alpha$  subunits contain the active sites and the  $\beta$  subunits are responsible for the trimer formation (Figure 26D). In fact, *H. pylori* urease possesses a major  $\alpha$  subunit and a minor  $\beta$  subunit, codified by the genes *ureA* and *ureB* respectively: the first corresponds to the  $\alpha$  subunit, while the second would result from the fusion of  $\beta$  and  $\gamma$  in other bacteria. In summary, *H. pylori* urease is a dodecamer, with twelve independent, nickel-containing active sites.



**Figure 27 - Structural comparisons of urease structures from different bacterial sources**

(A) Structural superimposition of *H. pylori* urease  $\beta$  subunit (cyan) with *K. aerogenes* (orange) and *B. pasteurii* (green)  $\beta$  and  $\gamma$  subunits. The additional loop is indicated by a black arrow. (B,C) Superpositions of *K. aerogenes* and *B. pasteurii*  $\alpha$  subunits: the open (green) and closed (orange) conformations of the flap are highlighted. Nickel ions, constituting the active site, are shown as green spheres. (D) Sequence alignment of the flap (helix-turn-helix) and flanking regions. The asterisks indicate the non-conserved residues in *H. pylori* urease compared to the other bacterial ureases. In all panels the following abbreviations are used: *Hp* for *H. pylori*, *Ka* for *K. aerogenes*, *Bp* for *B. pasteurii*.

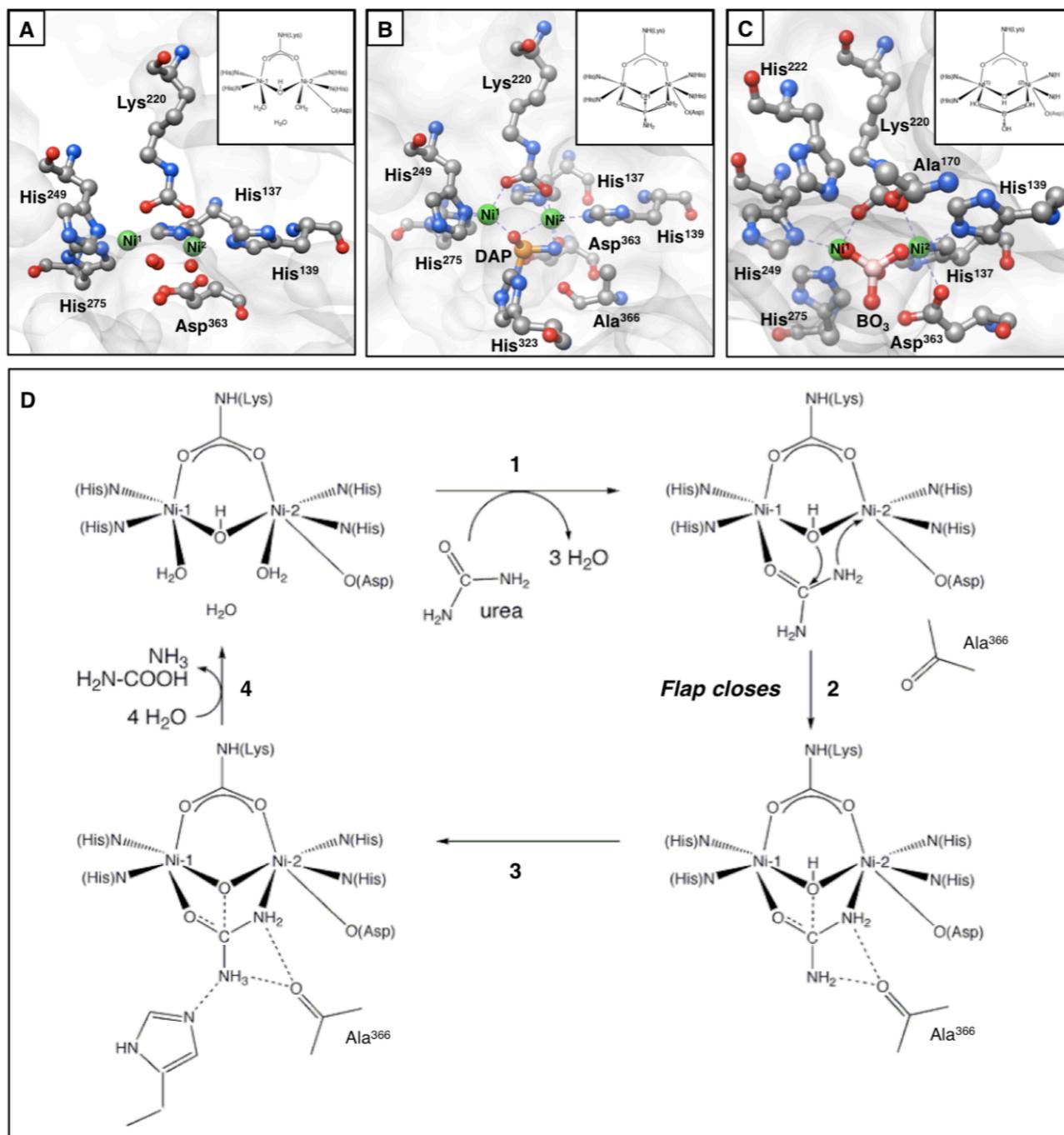
The atomic structure has also identified additional residues located in the C-terminal loop of the  $\beta$  subunit, which are responsible for tetramer formation (see structural comparisons on Figure 27A). This loop is present only in the ureases from gastric *Helicobacter* species, such as *H. acinocyclus*, whereas it is absent in the enzymes from other *Helicobacter* strains (e.g. *H. hepaticus*), suggesting that dodecamer formation would represent an advantageous characteristic for niche colonization. This advantage may result from the fact that the active sites face into the central core of the dodecameric assembly.

The crystal structure of the *B. pasteurii* urease has given insights into the catalytic mechanism [201]. The active site contains two essential Ni<sup>2+</sup> ions (Ni<sup>1</sup> and Ni<sup>2</sup>) bridged by a fully conserved, post-translationally carbamylated, lysine residue, and by a hydroxide ion (Figure 28A) [194]. In particular, two histidine residues and a water molecule coordinate Ni<sup>1</sup>, in a penta-coordinate, square-pyramidal geometry, whereas two histidine residues, a water molecule and one aspartate coordinate Ni<sup>2</sup>, in an esa-coordinate, octahedral geometry. A fourth water molecule is present in the active site, interacting with the others through hydrogen bonds, and completes a tetrahedral cluster of water/hydroxide molecules.

Despite having differences in their quaternary structure, ureases possess essentially identical folds, sharing a basic trimeric structure containing three catalytic centers (Figure 26). Moreover, the residues that coordinates nickel ions and the substrate are strictly conserved. Nevertheless, large differences in the measured  $k_{cat}$  and  $K_m$  for the three structurally characterized bacterial ureases has been reported. This was correlated with the differences observed in the flap, an HTH (helix-turn-helix) motif located in close proximity to the active site and connected to the rest of the protein by a flexible hinge. Interestingly, in *K. aerogenes* urease this segment displays a high degree of mobility, suggesting that it could modulate substrate binding and products release [202]. In fact, the structures of native urease from *K. aerogenes* and *B. pasteurii* show two different flap conformations, defined as “close” and “open” respectively (Figure 27B,C). In the structure from *H. pylori*, the flap seals the empty active site cavity; but weakly, so that it transiently opens and closes to wedge in the substrate. In fact, the flap sequence is considerably different in this bacterium (Figure 27D), probably influencing its flexibility and motions [186]. These observations correlate with the relative low  $K_m$  calculated for *H. pylori* enzyme (0.18 mM), compared to the values reported for *K. aerogenes* (2.3 mM) and *B. pasteurii* (55.2 mM) ureases [186]. However, the unusually low value of *H. pylori* urease is functionally explained considering the low available urea concentrations in the human gastric environment.

The fine structure of urease active site has also been determined in the presence of several inhibitors, in order to elucidate both the substrate binding mode and the catalytic mechanism. In fact, the high efficiency of the enzyme makes impossible to collect structural data in the presence of urea, because the hydrolysis reaction is too rapid. The most significant structures for the comprehension of the catalytic mechanism are those obtained in the presence of the transition state analogue DAP (diamidophosphoric acid) [201] and the substrate analogue boric acid [203]. In the first structure, a molecule of DAP is positioned on the active site, binding Ni<sup>1</sup> and Ni<sup>2</sup> respectively with an oxygen and a nitrogen atom. The other oxygen atom of DAP substitutes the bridging hydroxide, while the other amine group points away toward the cavity opening (Figure 28B). In this structure the flap is in “close” conformation, stabilizing the binding of DAP to the binuclear nickel center through an H-bond network. The second structure displays the borate molecule symmetrically placed between the two Ni<sup>2+</sup> ions, leaving in place the bridging hydroxide

and not perturbing the Ni<sup>1</sup>-Ni<sup>2</sup> distance. Two oxygen atoms are bound to the nickel ions, while the third oxygen points toward the cavity opening, away from the nickel ions (Figure 28C). In summary, the geometry and coordination number of the nickel ions are not significantly affected by boric acid binding.



**Figure 28 - Urease active site and catalytic mechanism**

Structure of the active site of urease from *Bacillus pasteurii* in the native (A) (PDB entry: 2ubp), DAP-inhibited (B) (PDB entry: 3ubp) and borate-bound (C) (PDB entry: 1s3t) states. For each structure, a schematization of the active site is shown. Catalytic residues and ligand molecules are represented as *ball and stick* and colored using the following atomic scheme: carbon, grey; nitrogen, blue; oxygen, red; phosphor, orange; boron, pink; nickel, green. (D) Proposed catalytic mechanism for urea hydrolysis.

These two structures suggest a probable urease catalytic mechanism (Figure 28D) that contrasts with the first described [204]. The novel catalytic mechanism proposed assumes two different roles for the nickel ions in the active site and deduces a direct function for the Ni-bridging hydroxide, as the nucleophile of the urea hydrolysis reaction. Borate and urea possess an isoelectronic structure and therefore they likely bind the active site in a similar way, replacing the three labile water molecules with a neutral trigonal molecule [203]. Therefore, the oxygen atom of urea would bind Ni<sup>1</sup> and one amine group binds Ni<sup>2</sup> (Figure 28D, step 1). 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 an H-bond network, activating the substrate molecule toward the nucleophilic attack, and positioning the carbon atom of urea near the nucleophile [205]. The bridging hydroxide attacks urea molecule, forming a tetrahedral intermediate, illustrated by DAP-inhibited urease structure. Concomitantly with the nucleophilic attack, the flap switches in the “close” conformation, creating an H-bond network that stabilizes the catalytic intermediate and lowers the activation energy for the reaction (Figure 28D, step 2,3) [205]. Transfer of a proton to the “free” distal amino group promotes attack of water on the urea carbonyl group, leading to formation of ammonia and carbamate, which spontaneously hydrolyzes into bicarbonate and another molecule of ammonia (Figure 28D, step 4). As water molecules bind, ammonia and carbamate dissociate to reconstitute the catalytic center.

### 1.10.2. [NiFe]-hydrogenase: structure and activity

Hydrogen gas is often referred to as an energy vector by chemists and technologists. In nature, H<sub>2</sub> is for many bacteria an energy source, the highest yield of chemical energy being provided by the oxidation of hydrogen by oxygen. This reaction in anaerobic and aerobic environments implicates hydrogenase enzymes as catalysts. In fact, hydrogenase catalyzes the reversible two-electron reduction of protons to H<sub>2</sub> (8) [206, 207].



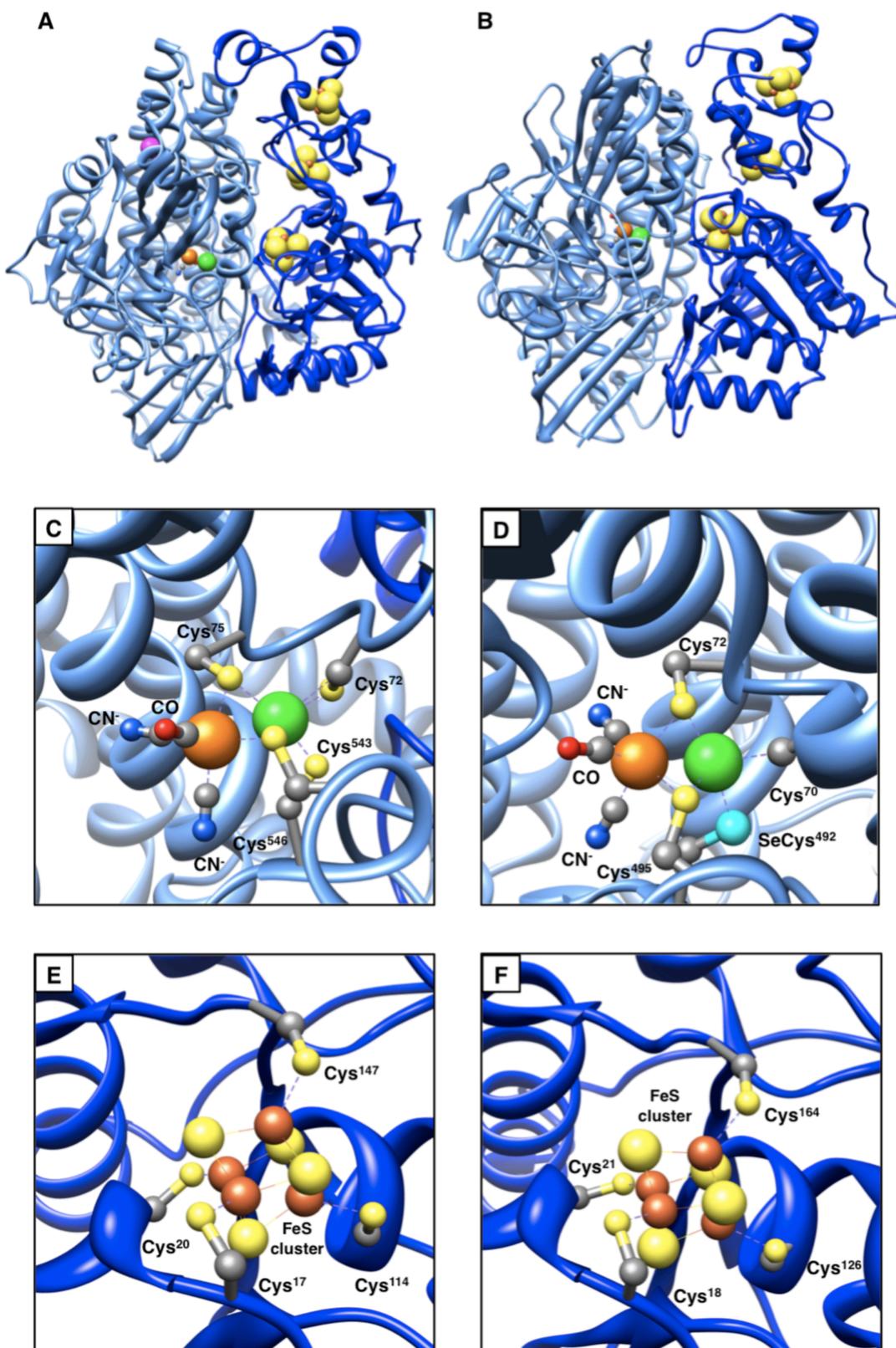
Anaerobic microbes, such as methanogenic archaea, acetogenic bacteria, sulfate- and nitrate-reducing bacteria, remove H<sub>2</sub> from the environment and couple its oxidation to the reduction of various terminal electron acceptors (e.g. O<sub>2</sub>, NO<sub>3</sub><sup>-</sup>, SO<sub>4</sub><sup>2-</sup>, CO<sub>2</sub>, and fumarate) [207]. In aerobic conditions, H<sub>2</sub> is produced by some photosynthetic bacteria, algae, azototrophus, during the dinitrogen reduction catalyzed by nitrogenase, and Knallgas cells that are able to live on H<sub>2</sub>, O<sub>2</sub> and CO<sub>2</sub> [196].

Actually to date, three phylogenetically unrelated classes of hydrogenase enzymes have been identified, comprising [FeFe]-hydrogenases, containing only iron cofactor in FeS clusters, [NiFe]-hydrogenases, which show both iron and nickel in their active site, and metal-free or Hmd

hydrogenases, that lack any FeS cluster [207, 208]. In the latter, hydrogen uptake is coupled to methenyltetrahydromethanopterin reduction and the active site contains a labile light-sensitive cofactor with a mononuclear low-spin iron that binds two CO ligands [209, 210]. Moreover, [NiFeSe]-hydrogenases, a variant of the [NiFe]-hydrogenase class found in sulfate-reducing bacteria which contain a selenium atom from a Se-Cys residue, have been also reported [211].

[NiFe]-hydrogenases constitute the most numerous and best studied class, that is found in aerobes, facultative anaerobes and some anaerobic bacteria. The crystal structures of five periplasmic [NiFe]-hydrogenases from sulfate-reducing bacteria *Desulfovibrio gigas* [212], *Desulfovibrio vulgaris* [213], *Desulfovibrio fructosovorans* [214], *Desulfomicrobium baculatum* [211], *Desulfovibrio desulfuricans* [215], have been reported. All these enzymes possess essentially the same fold and domain structure. Here, the representative crystal structures of [NiFe]-hydrogenase from *D. fructosovorans* (Figure 29A) and the [NiFeSe]-hydrogenase from *D. baculatum* (Figure 29B) are illustrated. Both enzymes are hetero-dimers, constituted by two subunits usually termed as L (large), which ranges 43 - 72 kDa, and S (small), showing a molecular weight between 23 - 38 kDa. The L subunit consists of five different domains, three showing an  $\alpha\beta$  structure, one largely unstructured and another one with four long  $\alpha$ -helices. This subunit contains the active site, deeply buried inside the protein, without any access for water molecules. The S subunit, containing one to three Fe-S clusters, is formed by an N-terminal domain, featuring an  $\alpha\beta$  twisted open-sheet, and a C-terminal domain, less structured with one to three  $\alpha$ -helices. In the *D. fructosovorans* structure, the bimetallic NiFe center is coordinated to the protein by two cysteine residues. In the active, reduced form, nickel is also coordinated to other two cysteines resulting four coordinate, while iron is bound to three exogenous non-protein ligands, one CO and two  $\text{CN}^-$ , and results five-coordinated (Figure 29C). The small subunit contains up to three linearly arranged FeS clusters of the [4Fe-4S] type (Figure 29E), which conduct electrons between the  $\text{H}_2$ -activating center and the physiological electron acceptor (or donor) of hydrogenase. Besides the substitution of the terminal  $\text{Cys}^{546}$   $\text{Ni}^{2+}$ -binding residue by a selenocysteine (Figure 29D), the [NiFeSe]-enzyme from *D. baculatum* shows some additional internal differences at the enzyme surface [211]. For example, the large subunit stretch between the fourth and the seventh  $\alpha$ -helices is much shorter in the *D. baculatum* enzyme and contains only one  $\alpha$ -helix, instead of the three observed in the other structures (Figure 29B). On the contrary, the small subunit of the [NiFeSe]-hydrogenases from *D. baculatum* possess three [4Fe-4S] clusters (Figure 29F), similarly to those observed in the *D. fructosovorans* structure.

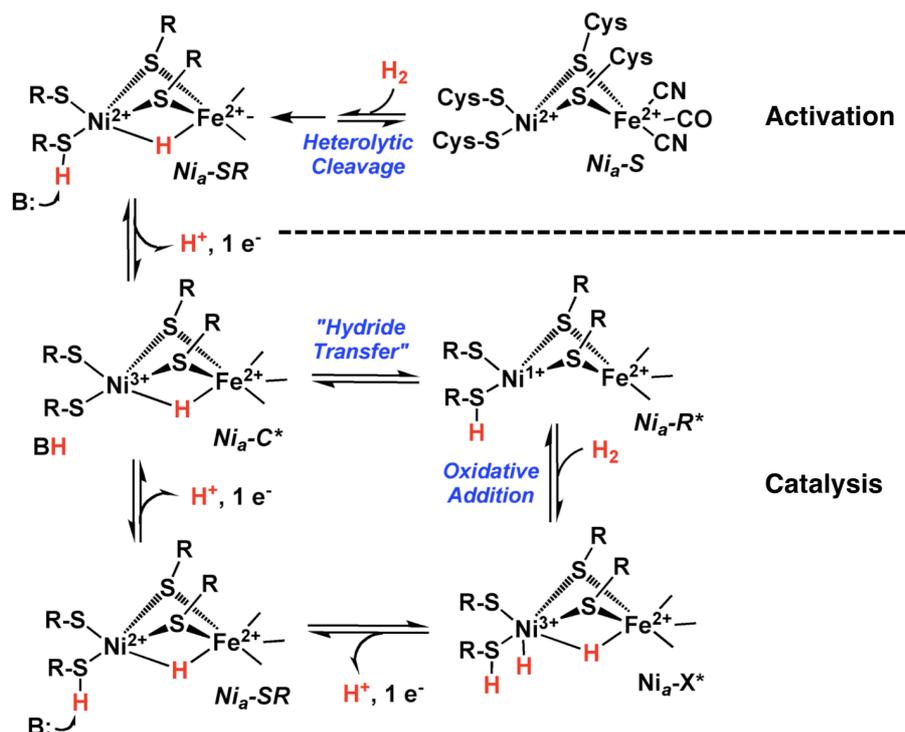
Several mechanisms have been proposed to explain the hydrogenase-catalyzed reaction [207, 216].  $\text{H}_2$  oxidation is diffusion-controlled ( $k_{cat}/K_m \sim 10^8 - 10^9 \text{ M}^{-1} \text{ s}^{-1}$ ), with a turnover number reaching  $9 \cdot 10^3 \text{ s}^{-1}$  at 30 °C. Because the NiFe catalytic center is buried 30 Å beneath the surface of the protein,  $\text{H}_2$  must travel through a tunnel in the protein to reach and react with the active site [214].



**Figure 29 - Structures of [NiFe]- and [NiFeSe]-hydrogenases**

Ribbon diagrams of [NiFe]-hydrogenase (A, PDB entry: 1yqw) and [NiFeSe]-hydrogenase (B, PDB entry: 1cc1). Large (sky blue) and small (blue) subunits are highlighted. For each structure, details of NiFe- or NiFeSe-active sites (C,D) and the first FeS cluster (E,F) are provided. Catalytic residues and ligand molecules are represented as *ball and stick* and colored using the following atomic scheme: carbon, grey; nitrogen, blue; oxygen, red; sulfur, yellow; selenium, cyan; iron, orange; nickel, green; magnesium, magenta.

The [NiFe]-hydrogenase requires activation, involving prolonged treatment with H<sub>2</sub> to generate the Ni<sub>a</sub>-C\* state, perhaps involving replacement of an OH ligand with a hydride bridge (in red) between the nickel and iron sites (Figure 30) [206]. Activation appears to involve heterolytic H–H bond cleavage. Catalysis ensues upon conversion of Ni<sub>a</sub>-C\* to a Ni<sup>1+</sup> oxidation state (Ni<sub>a</sub>-R\*) by a hydride transfer or proton-coupled electron transfer reaction, allowing productive binding of H<sub>2</sub>. H–H bond cleavage during the catalytic cycle is proposed to occur by an oxidative addition mechanism that would generate the Ni<sub>a</sub>-X\* intermediate, which undergoes two successive proton-coupled electron transfer steps to regenerate Ni<sub>a</sub>-C\*.



**Figure 30 - Proposed catalytic mechanism of [NiFe]-hydrogenase activation and catalysis**

The mechanism refers to the work of Lill and Siegbahn [216]. The asterisks indicate an EPR-active state.

## 1.11 Urease and [NiFe]-hydrogenase accessory systems

In analogy to other metal-dependent systems, the mechanisms that build nickel-dependent enzymes are complex processes and involve different chaperones. Actually, these enzymes are normally synthesized as precursors, requiring different steps and accessory proteins for maturation, metal incorporation and activation. However, intracellular nickel trafficking remains relatively obscure, because the role of nickel in fundamental biological processes has been demonstrated quite recently [194]. In this scenario, urease and [NiFe]-hydrogenase represent the more studied nickel-dependent systems, presumably because of their role in pathogenesis. Nevertheless, in both cases a complete picture of the activation mechanism has not been elucidated yet.

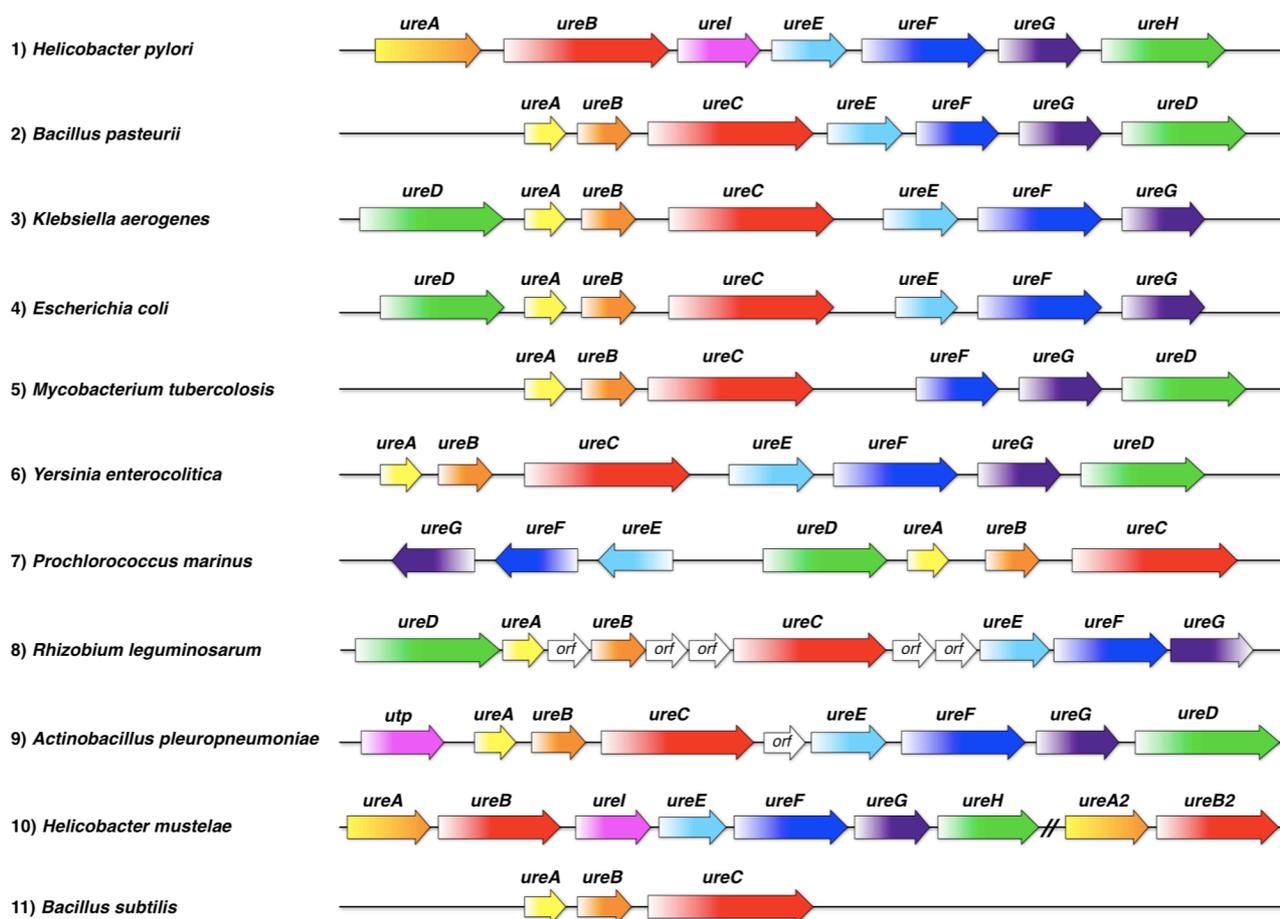
### 1.11.1. Urease genetic clusters in bacteria

Although pioneering reports on urease dealt with plant enzymes, the most detailed studies were carried out on bacterial ureases, probably because plants genes lack an organization in genetic clusters. In 1990, urease from *K. aerogenes*, grown in the absence of nickel, was expressed and purified, and the structural organization of urease was demonstrated not to need the presence of this metal. Indeed, the protein is assembled *in vivo* as an apo-enzyme, even in absence of nickel [217]. However, the building of nickel-containing active site is a multi-step process, requiring *in vivo* a precise sequence of events [218], including not only the presence of metal ions, but the simultaneous occurrence of different, non-physiological conditions. In particular, it has been reported that the activation of 12% of the enzyme is achieved in the presence of 100 mM NiCl<sub>2</sub> and 100 mM NaHCO<sub>3</sub>, the latter needed for lysine carbamylation [218]. This and other evidences strongly suggest the involvement of specific protein chaperones for urease.

The discovery of proteins associated to nickel containing site synthesis was first reported for the *K. aerogenes* system, and their genes were sequenced [219]. Homologous genes in regions flanking urease clusters were subsequently identified in many bacteria, demonstrating the presence of a unique genetic organization that possesses the characteristics of an operon. However, the number and order of bacterial urease structural and accessory genes are not universal across species. A large part of bacterial urease operons contains three structural genes, *ureA*, *ureB*, *ureC*, codifying respectively for the apo-urease subunits  $\gamma$ ,  $\beta$ ,  $\alpha$ , and four accessory genes, *ureE*, *ureF*, *ureG*, *ureD*, as in the case of *B. pasteurii* and *Y. enterocolitica* [220] (Figure 31, n.2, 6). Whereas the structural genes serve for the right building of the apo-enzyme, also in absence of metal, the accessory genes are required to obtain a functional enzyme, with the correctly assembled catalytic site. A similar organization is shown by *K. aerogenes* and *E. coli*, showing the structural genes *ureABC* flanked by the accessory genes, with *ureD* located upstream and *ureE*, *ureF*, and *ureG* found downstream (Figure 31, n.3, 4). This gene structure is retained in *R. leguminosarum*, but with several uncharacterized ORFs (open reading frames) inserted into this cluster (Figure 31, n.8) [221]; similarly in *A. pleuropneumoniae*, the *ureABCEFGD* cluster is interrupted by an ORF and is preceded by a possible urea permease (Figure 31, n.9) [222]. Interestingly, also the orientation of the genes in the urease cluster can differ, as reported for *P. marinus* (Figure 31, n.7) [223].

Moreover, peculiar exceptions are found in *Helicobacter* species. In particular, the cluster from *H. pylori* consists of *ureA*, a fusion of the small subunit genes from other bacteria, and *ureB*, encoding the large subunit, along with five downstream genes: *ureI*, encoding a proton-gated urea channel, *ureE*, *ureF*, *ureG* and *ureH*, homologous to *ureD* of other bacteria (Figure 31, n.1) [224-226]. *H. mustelae* contains an analogous complete urease gene cluster, as well as a second

set of structural genes not associated with any accessory protein genes (Figure 31, n.10) [227]. Furthermore, some bacteria lack one or more urease accessory genes; this is exemplified by *M. tuberculosis*, lacking *ureE* (Figure 31, n.5) [228], and *B. subtilis* which does not show any identifiable accessory genes (Figure 31, n.11) [229]. This observation suggests that accessory proteins are not always required for *in vivo* urease activation or that, more plausibly, genetically unlinked cellular maturation factors could be alternatively utilized in some cases.



**Figure 31 - Genetic organization of representative bacterial urease operons**

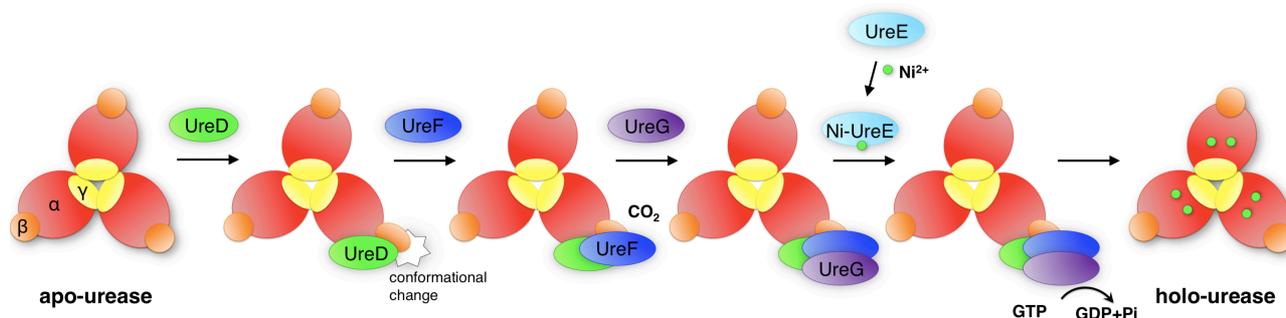
The urease gene cluster of *H. pylori* is compared to the gene organization found in other selected bacteria. This scheme is not representative for any urease operons nor is it meant to be exhaustive. Genes encoding urease subunits are shown in yellow, orange and red, *ureD* or *ureH* genes are green, *ureE* genes are cyan, *ureF* genes are blue, *ureG* genes are purple, genes encoding proteins involved in nickel uptake are orange, those encoding urea transporter (*ureI*, *utp*) are pink and unknown genes are white. The position and the orientation of each gene are assigned on the basis of its genomic sequence.

### 1.11.2. Maturation of urease enzyme

Although the activation of purified *K. aerogenes* urease apoprotein was accomplished *in vitro* by providing carbon dioxide in addition to nickel ion in a pH-dependent reaction [218], the urease activation *in vivo* was demonstrated to require specific accessory proteins. In particular, using

deletion analysis it has been demonstrated that *ureD*, *ureF*, and *ureG* are all essential for the production of functional urease in *K. aerogenes*, since mutations in these genes nearly abolish activity in cell extracts [230]. These early studies also reported that partial deletions in *ureE* resulted in only ~50% lower specific activities; however, subsequent *ureE* deletion data showed that urease activity is essentially eliminated [231]. Furthermore, *E. coli* cells expressing only the *K. aerogenes* structural genes and grown in the presence of 5 mM NiCl<sub>2</sub> possess very low level of urease activity [229]. Genetic experiments using deletion, insertional inactivation, and complementation approaches have been used to identify multiple non-urease subunit genes required for urease activity in *Proteus mirabilis* [232], *Klebsiella pneumoniae* [233], *Providencia stuartii* [234], and many other bacteria, including *H. pylori*. In the latter case, disruptions or deletions in *ureA*, *ureB*, *ureF*, *ureG*, or *ureH* genes expressed in *E. coli* result in a non-ureolytic phenotype [226]. Taken together these evidences strongly support the idea that each of the four accessory proteins, UreD, UreF, UreG and UreE, plays an essential role in the urease maturation process.

A mechanism for the maturation of a functionally active holo-enzyme was proposed on the basis of functional studies performed on *K. aerogenes* urease system. The current model for this process [235] entails the formation of a multi-meric complex between the apo-enzyme and the four protein chaperones, followed by the delivery of Ni<sup>2+</sup> concomitantly with the GTP-dependent transfer of a CO<sub>2</sub> molecule necessary for lysine carbamylation (Figure 32).



**Figure 32 - Schematization of the urease activation mechanism**

In the proposed model for urease activation the chaperones UreD, UreF and UreG bind sequentially to the apo-enzyme. In the last step, UreE delivers nickel ions to the UreDFG-apo-urease assembly. The color scheme used for the urease subunits and accessory proteins reflects that used in Figure 31 for the respective genes. The urease is represented as trimeric assembly ( $\alpha\beta\gamma$ )<sub>3</sub>.

In particular, the first protein supposed to bind apo-urease was UreD, on the basis of the recovery of UreD-urease complexes from *K. aerogenes*, containing one, two or three monomers of *KaUreD* (*K. aerogenes* UreD) [236]. Furthermore, cross-linking experiments revealed the occurrence of PPIs involving *KaUreD* and both the  $\alpha$  and  $\beta$  subunits of apo-urease [237]. Its structural properties and functional role remain largely obscure. However, UreD has been proposed to bind to apo-urease, thus inducing a conformational change required for the

subsequent steps of the activation process [238], as recently supported by small angle X-ray scattering results [239]. UreF was supposed to be the second accessory protein binding urease, due to the observation of an *in vivo* KaUreDF-apourease complex [240]. This ternary complex was resistant to inactivation by NiCl<sub>2</sub> in the absence of bicarbonate, and the bicarbonate concentration dependence for urease activation was significantly decreased, compared to that of urease alone and UreD-urease apoproteins. Yeast two-hybrids analysis and immunoprecipitation experiments suggested the presence of this interaction also in from *Proteus mirabilis* [241] and *H. pylori* [242]. Analogous studies on *P. mirabilis* proteins demonstrated the interactions between UreD and UreF and the association of these proteins with the α-subunit of apo-urease [241], whereas cross-linking results detected that KaUreF also contacts the β subunit of urease [237]. All these evidences led to suppose that UreF modulates the UreD-apourease activation properties by avoiding the binding of Ni<sup>2+</sup> ions to the active site when the coordinating lysine is not carbamylated. Recently, a functional role for UreF as GAP (GTPase activator protein) has been proposed on the basis of structural bio-modeling studies [243]. Furthermore, the presence of an *in vivo* KaUreDFG-urease complexes suggested that such large assembly constitutes the minimum core required for the urease activation [244]. UreG contains a fully conserved P-loop motif, characteristic of nucleotide-binding proteins, and is responsible for the GTP hydrolysis associated to the transfer of CO<sub>2</sub> to the active site lysine [245]. Finally, it has been reported that KaUreE binds the KaUreDFG-apourease complex, acting as a nickel-transporter that delivers Ni<sup>2+</sup> to the active site of the enzyme [246]. Detailed investigations of the four standard accessory proteins are discussed in Section 1.11.3.

### **1.11.3. Urease accessory proteins: structures and functions**

#### **UreG**

Among the four urease accessory proteins, UreG is the most conserved [247-249]. The generally accepted hypothesis for this chaperone *in vivo* is the hydrolysis of GTP concomitant with the activation of urease, as suggested by the presence of a fully conserved P-loop motif, which is also found in many nucleotide-binding proteins [245]. Furthermore, GTP is needed for activation of the KaUreDFG-apourease complex, although it has an inhibitory effect on the nickel reconstitution of the apo-urease, KaUreD-apo-urease and KaUreDF-apo-urease complexes [245]. Moreover, the importance of the P-loop motif for UreG activity has been reported for *K. aerogenes* [247], where it affects the formation of the UreDFG-apourease complex, and *H. pylori* system [250]. UreG seems to be also involved in delivering CO<sub>2</sub> necessary for the carbamylation of the nickel-bridging lysine: 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

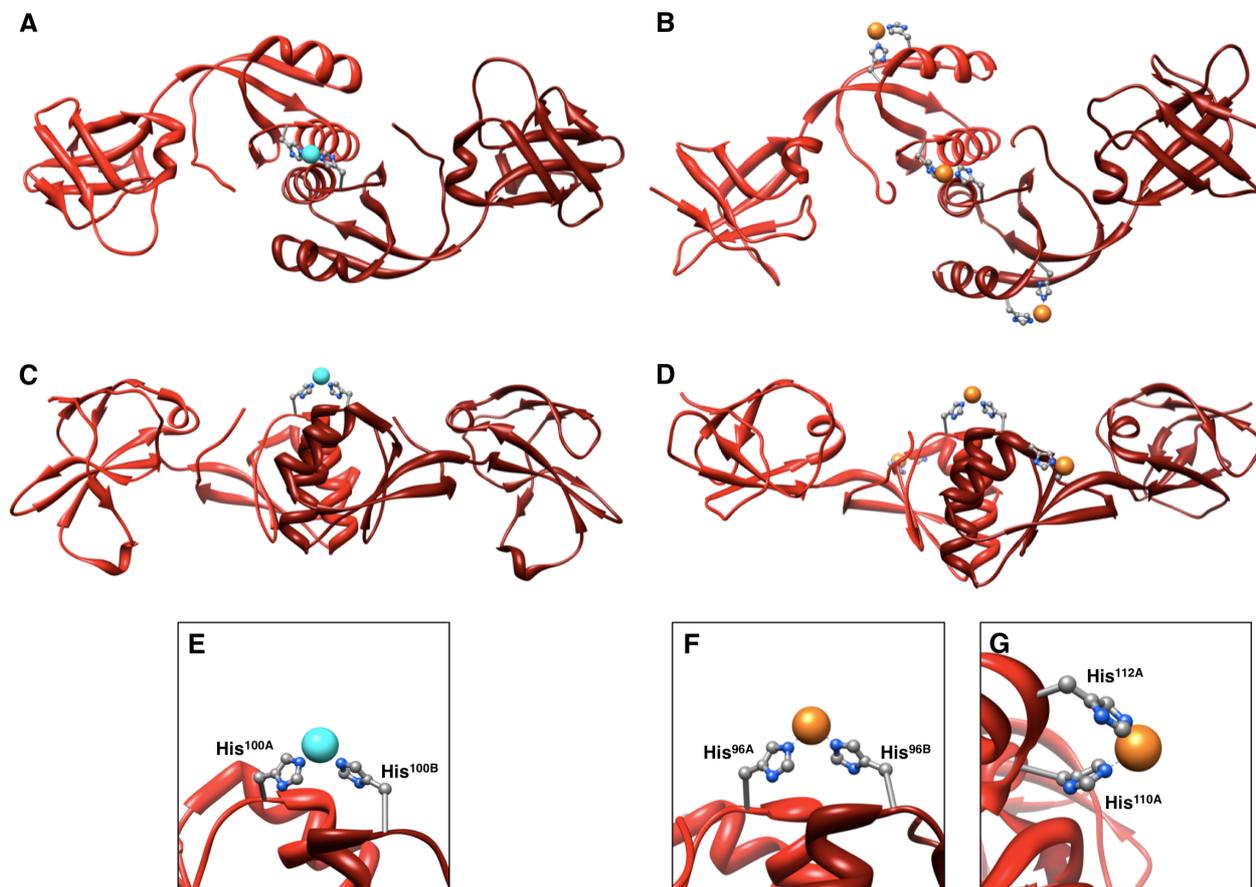
in the presence of UreG [245]. In particular, it has been proposed that UreG may induce GTP-dependent changes on the apo-urease structure, increasing accessibility of both nickel and CO<sub>2</sub> to the developing active site. Alternatively, UreG may use GTP and CO<sub>2</sub> to synthesize carboxyphosphate, which could serve as a CO<sub>2</sub> donor to the lysine residue. This hypothesis is supported by the observation of a 50% activity enhancement for *KaUreDFG*-apourease complex compared to the apo-enzyme only in the presence of nickel ions, bicarbonate and GTP [245].

UreG proteins from different organisms, including *B. pasteurii* [248], *K. aerogenes* [247], *M. tuberculosis* [249] and *H. pylori* [251], have been isolated and purified. However, structural information regarding this chaperone has not been available yet due to its unstructured behavior. This distinct feature includes UreG among the very few enzymes that so far can be ascribed to the class of intrinsically unfolded proteins. UreG from *K. aerogenes* (*KaUreG*) (21.8 kDa) has been the first protein of this class to be isolated as a monomer, as evinced by size-exclusion chromatography analysis [247]. It did not, by itself, hydrolyze GTP or ATP. Indeed, no nucleotide was found to be associated with isolated *KaUreG*, nor this protein could bind GTP or ATP in the absence of the other accessory proteins [247]. Similar results were obtained for UreG from *H. pylori* (*HpUreG*) (21.9 kDa), which showed negligible GTPase activity *in vitro* [252]. UreG from *B. pasteurii* (*BpUreG*) (23.1 kDa) is instead present in solution as a dimer [248]. It has been reported that *BpUreG* possess a GTPase activity ( $k_{cat} = 0.04 \text{ min}^{-1}$ ), and binds two Zn<sup>2+</sup> ions per dimer ( $K_d = 42 \mu\text{M}$ ) [248]. This chaperone is also able to bind four Ni<sup>2+</sup> ions per protein dimer, but with a ten-fold lower affinity compared to zinc [248].

The key importance of two conserved residues, Cys<sup>66</sup> and His<sup>68</sup> (Figure 36C), in *HpUreG* in zinc binding has been recently established by isothermal titration calorimetry and site-directed mutagenesis [251]. *HpUreG* specifically binds 0.5 equivalents of Zn<sup>2+</sup> per monomer ( $K_d = 0.33 \mu\text{M}$ ), whereas it displays a 20-fold lower affinity for Ni<sup>2+</sup>. Moreover, zinc binding causes protein dimerization, as confirmed by light scattering measurements [251]. A homology-based molecular model of dimeric *HpUreG* is shown in Figure 33A. The structural model shows a globular protein made of a seven-stranded parallel  $\beta$ -sheet flanked by nine  $\alpha$ -helices and largely reproduces the template used for homology modeling HypB from *Methanocaldococcus jannaschii* (*MjHypB*) [253] (see structural alignment in Figure 33B), not only in the overall architecture but also in the arrangement of key functional elements, such as the P-loop, switch-I and switch-II (Figure 33A). The observed zinc-induced dimerization of *HpUreG* does not involve an increase in the GTPase activity. This suggests that the presence of Zn<sup>2+</sup>, although inducing a conformational change, may not be sufficient, in the absence of additional factors possibly represented by the interaction with the enzyme or other urease chaperone(s), to bring the protein in a functionally active structure.



role of intracellular metal ion transport associated with UreE proteins. Finally, the elucidation of the crystal structures of UreE from *B. pasteurii* (*BpUreE*) [262] and *K. aerogenes* (*KaUreE*) [263], indirectly confirmed this function. The latter crystal structure refers to a truncated form of *KaUreE* lacking the last 15 His-rich residues, named H144\**KaUreE*. The two structures showed an analogous architecture, made up of two distinct domains (Figure 34 A-D).



**Figure 34 - Crystal structures of *BpUreE* and *KaUreE***

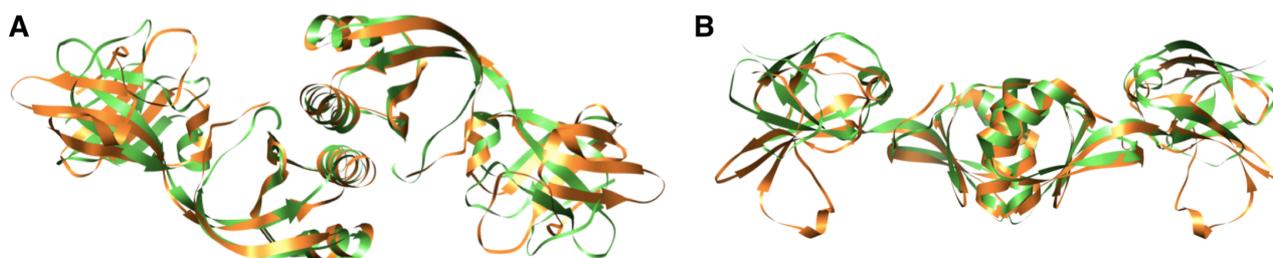
Ribbon representations of dimeric *BpUreE* bound to zinc ion (A,C) (PDB entry: 1ear) and *KaUreE* bound to copper ions (B,D) (PDB entry: 1gmw). The proteins are shown with the metal binding site toward the viewer (top panels) and rotated by 90° around the horizontal axis (bottom panels). Details of the conserved (E,F) and non-conserved (G) metal binding sites are also shown. Residues are represented as *ball and stick* and colored according to the following atomic scheme: carbon, grey; nitrogen, blue; zinc, cyan; copper, orange.

Precisely, the N-terminal domain is composed of two three-stranded  $\beta$ -sheets stacked upon each other in a nearly perpendicular fashion, with a short helical region between the two sheets, whereas the C-terminal domain displays a  $\beta\alpha\beta\beta\alpha\beta$  fold, similarly to the structural organization shown by the copper chaperone Atx1 [119]. The functional dimers of both *BpUreE* and H144\**KaUreE* are built by a head-head interaction, involving the hydrophobic face of an amphiphilic helix in the C-terminal domain with the metal-ion located at the interface between the two monomers [264]. Furthermore, *BpUreE* crystallized as a tetramer, built from the dimerization of two functional dimers, in two oligomerization forms with different orientations of one dimer respect to the other. The presence of the protein in this tetrameric form does not appear to be relevant for

the protein physiological role. In both structures, no electron density for the last few residues at the C-termini is observed, because of disorder, caused, in the case of *BpUreE*, by the formation of the dimer of dimers, while in the case of H144\**KaUreE* presumably due to the non-natural form of the truncated protein [264]. The flexibility of this C-terminal region is possibly related to metal binding and release, as described below.

A metal-ion-binding site is found at the protein dimerization interface, involving two histidines, one from each monomer (His<sup>100</sup> in *BpUreE*, Figure 34E; His<sup>96</sup> in *KaUreE*, Figure 34F). This residue is strictly conserved among all the UreE sequences and appears to carry out a relevant role in metal delivery. The metal binding site is occupied by Zn<sup>2+</sup> in the structure of *BpUreE* and Cu<sup>2+</sup> in that of H144\**KaUreE*, but is generally assumed to coordinate Ni<sup>2+</sup> *in vivo*, a hypothesis supported by anomalous difference X-ray diffraction maps of *BpUreE* crystals soaked in a Ni<sup>2+</sup> solution [262]. The structure of H144\**KaUreE* also contains two additional Cu<sup>2+</sup> ions bound to a pair of histidines on the surface of each monomer, His<sup>110</sup> and His<sup>112</sup> (Figure 34G), but these residues are not conserved in other sources. This second metal binding site is characteristic of *KaUreE* and few other UreE proteins, and cannot be considered as a general feature [264].

However, the most evident difference between the overall fold of the two homologous proteins resides in the different relative orientations between the N-terminal and the C-terminal domains (Figure 35). 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 [264]. This protein flexibility is likely related to induced-fit processes during formation of protein complexes involving the other urease chaperones.



**Figure 35 - Superimposition of *BpUreE* and *KaUreE* crystal structures**

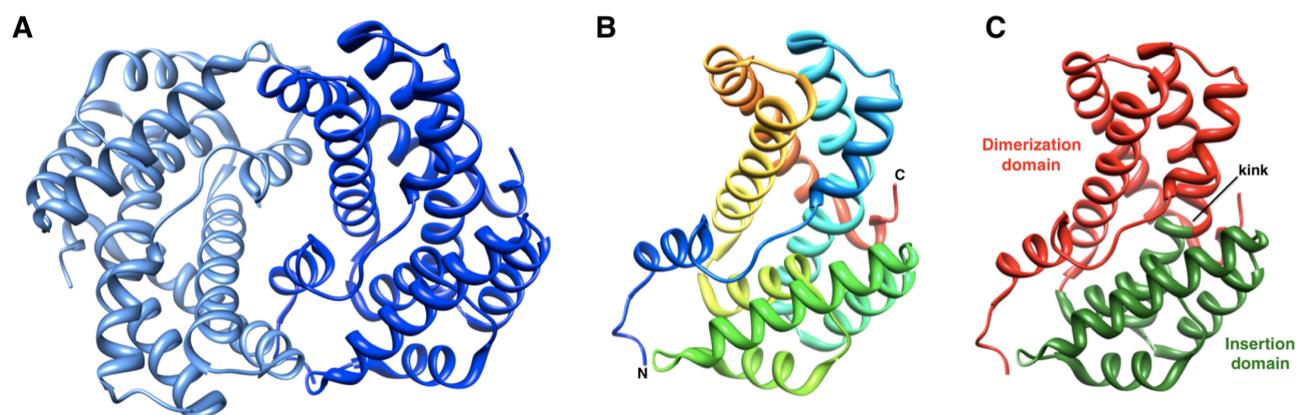
*BpUreE* is shown in orange and *KaUreE* in green. In panel B the structures are rotated by 90° around the horizontal axis.

The presence of an extensive hydrophobic surface in *BpUreE* zinc-binding side suggests that this part is possibly involved in the interaction with the UreDFG-apourease complex. These observations were confirmed by computational analysis that generated structural models for all the available UreE sequences on the basis of the two known structures [264].

## UreF and UreD

Whereas both UreE and UreG have been extensively investigated, progress in the elucidation of UreD and UreF has been hampered by their insolubility. Encouragingly, the translational fusions of *K. aerogenes* UreF (*KaUreF*) with MBP (maltose binding protein) [265] and UreE [266] were soluble and subsequently isolated, and the latter was shown to be able to facilitate urease activation *in vivo*.

Recently, a soluble and crystallizable form of UreF from *H. pylori* (*HpUreF*) has been isolated, due to fortuitous limited proteolysis during the purification, and a partial three dimensional structure have been released, representing the first structural characterization of this protein from any species [267]. The crystal structure shows a dimeric architecture (Figure 36A) with an all-helical topology for the crystallographically observed protein portion (residues 25 - 233), where each monomer consists of nine  $\alpha$ -helices arranged in an antiparallel fashion and a  $3_{10}$  helix, at the C-termini (Figure 36B). A pronounced kink in the middle of a central helix defines two domains (Figure 36C). The first domain is largely involved in dimerization of *HpUreF* (shown in red in Figure 36C) and comprises a four-helix bundle together with an N-terminal helix and a short C-terminal  $3_{10}$  helix. The second domain is an insertion of a 5-helix bundle (shown in green in Figure 36C) and participates minimally in the dimerization of *HpUreF*.

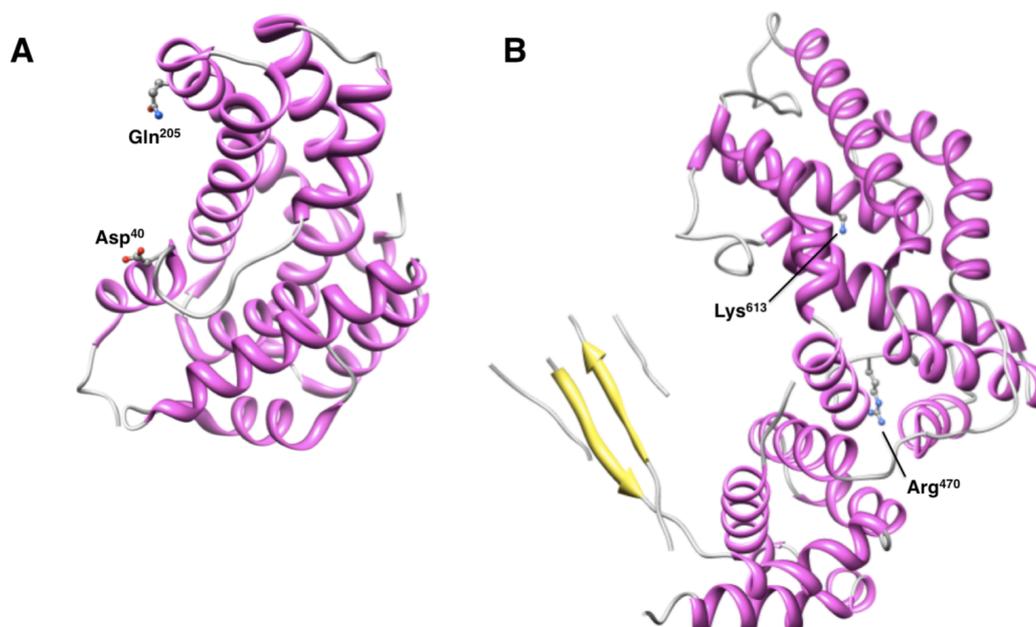


**Figure 36 - Crystal structure of *HpUreF***

Ribbon schemes of dimeric (A) and monomeric (B,C) *HpUreF* structure (PDB entry: 3xcn). In panel A, the two monomers are differently colored. In panel B, the ribbons are colored from blue, in the proximity of the N-terminal, to red at the C-terminus. In panel C, the two individual domains of *HpUreF* are indicated.

This architecture shows a weak, but intriguing, structural similarity between the dimerization domain and the GAP domains of SynGAP (Figure 37), confirming the previously hypothesized role for *BpUreF* as GTPase activator [243]. Interestingly, the catalytic interface of SynGAP coincides with the dimer interface in *HpUreF*. However, the catalytic Arg finger (Arg<sup>470</sup> in SynGAP) and an invariant Lys (Lys<sup>613</sup> in SynGAP), required for substrate binding, are lacking in *HpUreF* (Figure 40). Moreover, the fingerprint of functional Ras-GAPs, a conserved Phe-Leu-Arg sequence motif, is absent in *HpUreF*, providing further evidence for the divergence in function of these proteins

despite any structural similarity. Although distantly related structurally to GAPs, the proposed function of UreF as a GAP acting upon UreG [243], still remains an attractive hypothesis. In particular, it is possible to speculate that an extremely conserved Lys<sup>195</sup> in UreF proteins could replace the catalytic Arg residue. Alternatively, it could also be hypothesized that Lys<sup>195</sup> prevents nickel binding to the non-carboxylated urease apo-enzyme by regulating the sequential incorporation of bicarbonate into the active site before nickel can gain access.

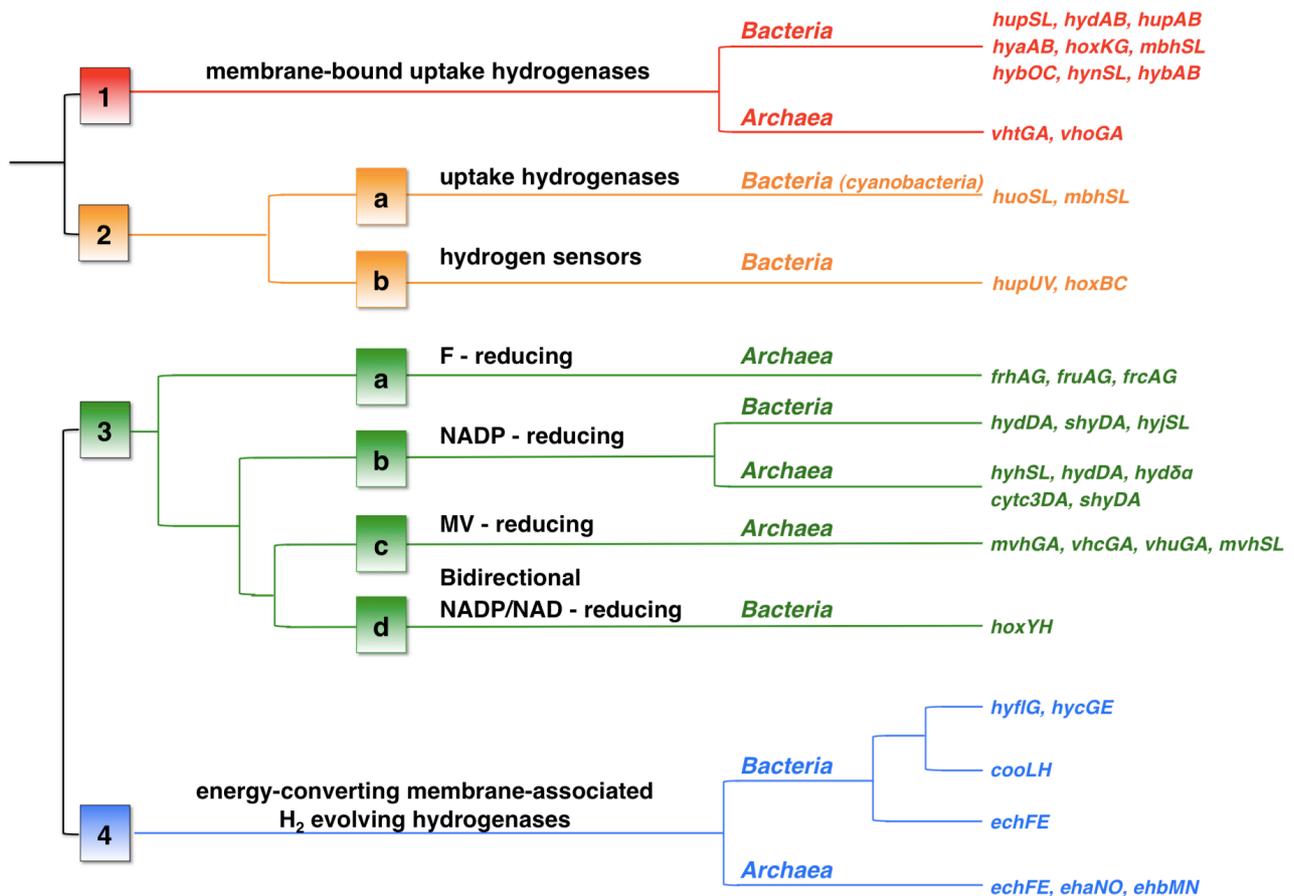


**Figure 37 - Structural comparison between *HpUreF* and SynGAP**

Ribbon diagrams of *HpUreF* monomer (A) and C2-GAP monomeric structure of SynGAP (PDB entry: 3bx). Structures are colored following secondary structure indication ( $\alpha$ -helix, magenta;  $\beta$ -strand, yellow; turn, grey). Catalytic residues of SynGAP and corresponding residues of *HpUreF* are represented as *ball and stick* and colored according to the following atomic scheme: carbon, grey; nitrogen, blue; oxygen, red.

#### 1.11.4. [NiFe]-hydrogenase genetic organization in bacteria

In proteobacteria, also the genes that encode [NiFe]-hydrogenase, and the protein required for its activation, are clustered. Nevertheless, [NiFe]-hydrogenase are widely distributed among microorganisms and are involved in many biological processes; indeed they are located in either the cytoplasm or the periplasm of bacteria (Figure 38). These enzymes are present in almost all the living *Bacteria* and *Archaea*, even in the most ancient branch of *Bacteria* such as *Aquifex aeolicus* [268], showing a high degree of similarity. This suggests that the microbial ability to metabolize hydrogen is of great importance and ancient origin.



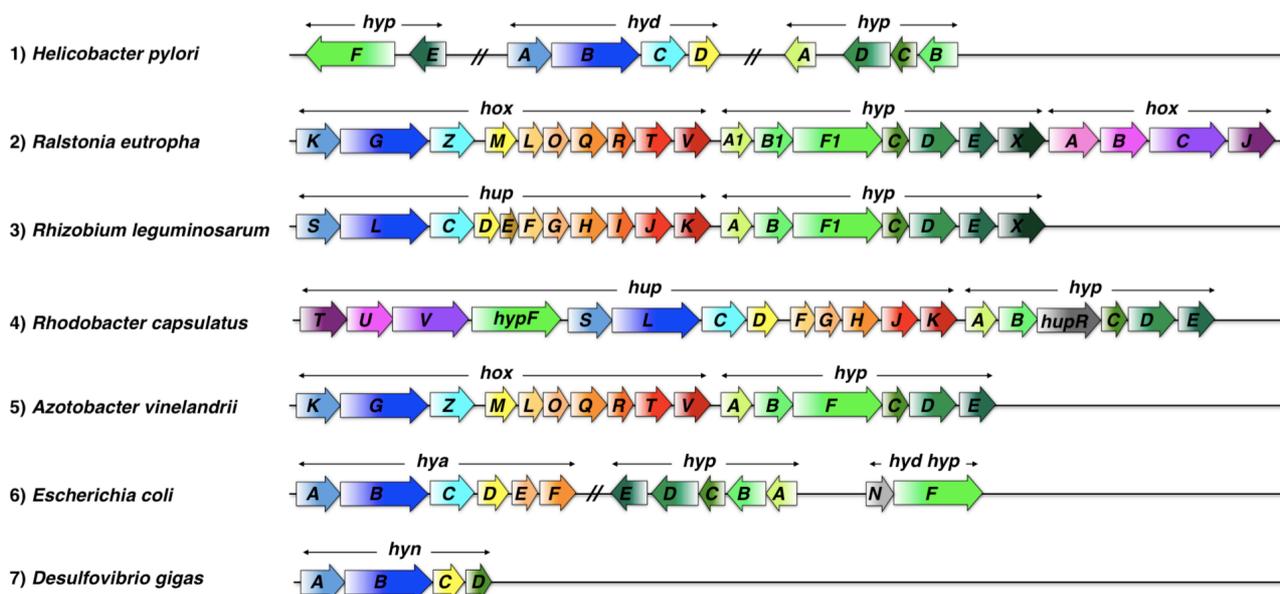
**Figure 38 - Schematic representation of the phylogenetic tree of [NiFe]-hydrogenases**

The reported phylogenetic tree is based on the complete sequences of the small and the large subunits (the same tree was obtained with each type of subunit), originally established by Vignais and coworkers [383]. Four classes of enzymes are colored differently, based on their function and/or cellular localization. The names of structural genes encoding for the two subunit of the enzymes are reported.

The hypothesis that the [NiFe]-hydrogenase operons were derived from one common ancestor is also supported by the strong conservation of the gene composition and organization in the various transcriptional units throughout the microbial kingdoms [269]. These operons contain many genes organized in several transcription units. The structural genes, encoding the small and large subunits, are usually present at the beginning of the upstream operonic structure followed by a set of accessory genes for maturation and the insertion of Ni, Fe, CO and CN<sup>-</sup> at the active site (Figure 39). Some organisms also comprises regulatory genes, that control expression of the structural genes in the hydrogenase gene cluster.

Several mutational analyses have been performed on hydrogenase operons from *E. coli*, *R. eutropha*, *R. capsulatus*, *R. leguminosarum*, *B. japonicum*, *A. vinelandii* and *A. chroococcum* [270, 271], which have led to the identification of two main groups of maturation genes, on the basis of the phenotypes resulting from their mutation. One group of genes that is mainly located on the same transcription unit as the structural genes specifically impairs the processing or activity of the hydrogenase encoded *in cis* in the operon when disrupted (Figure 39, n.3-5). One exception,

however, has been found to exist in the case of *hybG*, which belongs to the *E. coli* hydrogenase 2 operon and is involved in both hydrogenase 1 and 2 maturations [272]. Mutations occurring in this family of accessory genes cannot be complemented *in trans* by homologous genes belonging to other isoenzyme operons [272-274]. Even though hydrogenase operons are very conserved and exhibit a high degree of similarity, each *cis*-acting maturation system is specific to the corresponding enzyme, probably because of intimate PPIs occurring during processing. Mutations introduced in the second group of maturation genes, *hyp* (hydrogenase pleiotropic genes) from *E. coli* [275] and *R. eutropha* [276] resulted in a different phenotype, as synthesis and activity of all the hydrogenase isoenzymes were affected pleiotropically. In the case of *E. coli*, a single copy of the *hyp* operon is present in the genome and is responsible for the maturation of all three active isoenzymes (Figure 39, n.6), with the exceptions of *hypC*, which is not involved in hydrogenase 2 maturation and *hypA*, which is involved in hydrogenase 3 maturation only. A similar genetic organization is shown by *H. pylori*, where some of the *hyp* genes are found upstream the structural genes (Figure 39, n.1). In the case of *R. eutropha*, one part of the *hyp* operon is duplicated and mutations in both alleles are required to observe a hydrogenase-negative phenotype (Figure 39, n. 2) [276]. The case of complementation of mutations in the *hyp* locus might be explained by assigning the Hyp proteins a broader specificity towards the structural subunits.

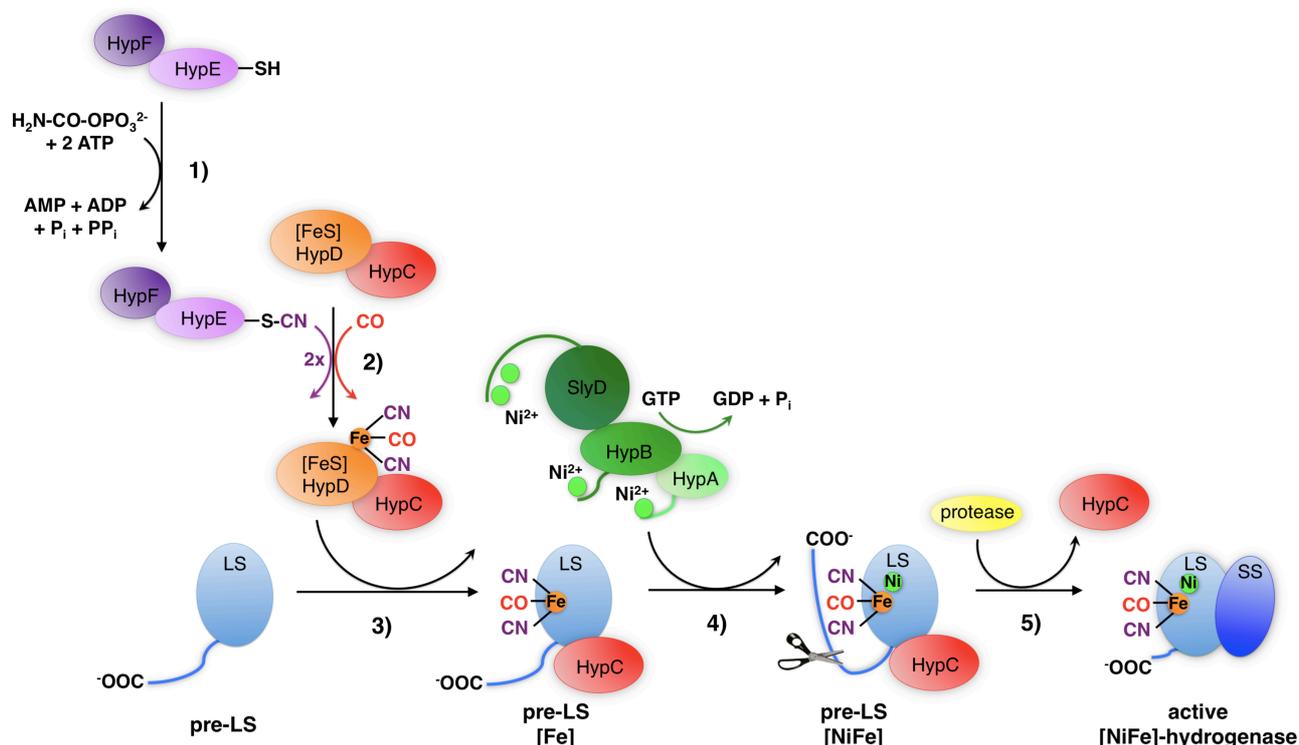


**Figure 39 - Genetic organization of representative bacterial [NiFe]-hydrogenase operons**

Examples of selected [NiFe]-hydrogenase gene clusters are reported. This scheme is not representative for any hydrogenase operons nor is it meant to be exhaustive. For *E. coli* only the operon of hydrogenase 1 is shown. Functionally homologous genes are shaded the same color. The arrows indicate the limits of the transcription units. The position and the orientation of each gene are assigned on the basis of its genomic sequence.

### 1.11.5. Biosynthesis and activation of [NiFe]-hydrogenase

The maturation of [NiFe]-hydrogenase involves several steps: *i*) synthesis of the apo-enzyme; *ii*) transport and storage of nickel and iron; *iii*) ligand synthesis and partial active site assembly; *iv*) insertion of nickel, proteolysis of a C-terminal amino acid stretch, and folding of the nascent C-terminal portion into the rest of the large subunit. This process requires at least six auxiliary proteins encoded by the *hyp* genes, HypA, HypB, HypC, HypD, HypE, HypF, and a nickel-dependent endopeptidase [275], although several are replaced by alternative homologous proteins specific for the biosynthesis of individual isoenzymes [277]. This set of proteins directs the synthesis and incorporation of the metal center into the large subunit, checks the fidelity of insertion of the correct metal, maintains a folding state of the protein competent for metal addition, and allows protein conformational changes for internalization of the assembled metal center. The accepted model implies that the two metal centers are sequentially delivered to the hydrogenase large subunit, with the iron inserted before the nickel ion (Figure 40), as demonstrated by the purification of a tagged hydrogenase large subunit from *R. eutropha* that contained a completed iron center but that was only partially loaded with nickel [278].



**Figure 40 - Schematic model for the activation of the [NiFe]-hydrogenase**

Proposed mechanism for the assembly of the bimetallic center on the large subunit (LS) of the [NiFe]-hydrogenase. Final assembly with the small subunit (SS) is also shown.

In particular, the synthesis of hydrogenase starts from a precursor of the large subunit, generally identified as pre-LS, showing an extension at the C-terminal. This precursor is associated with

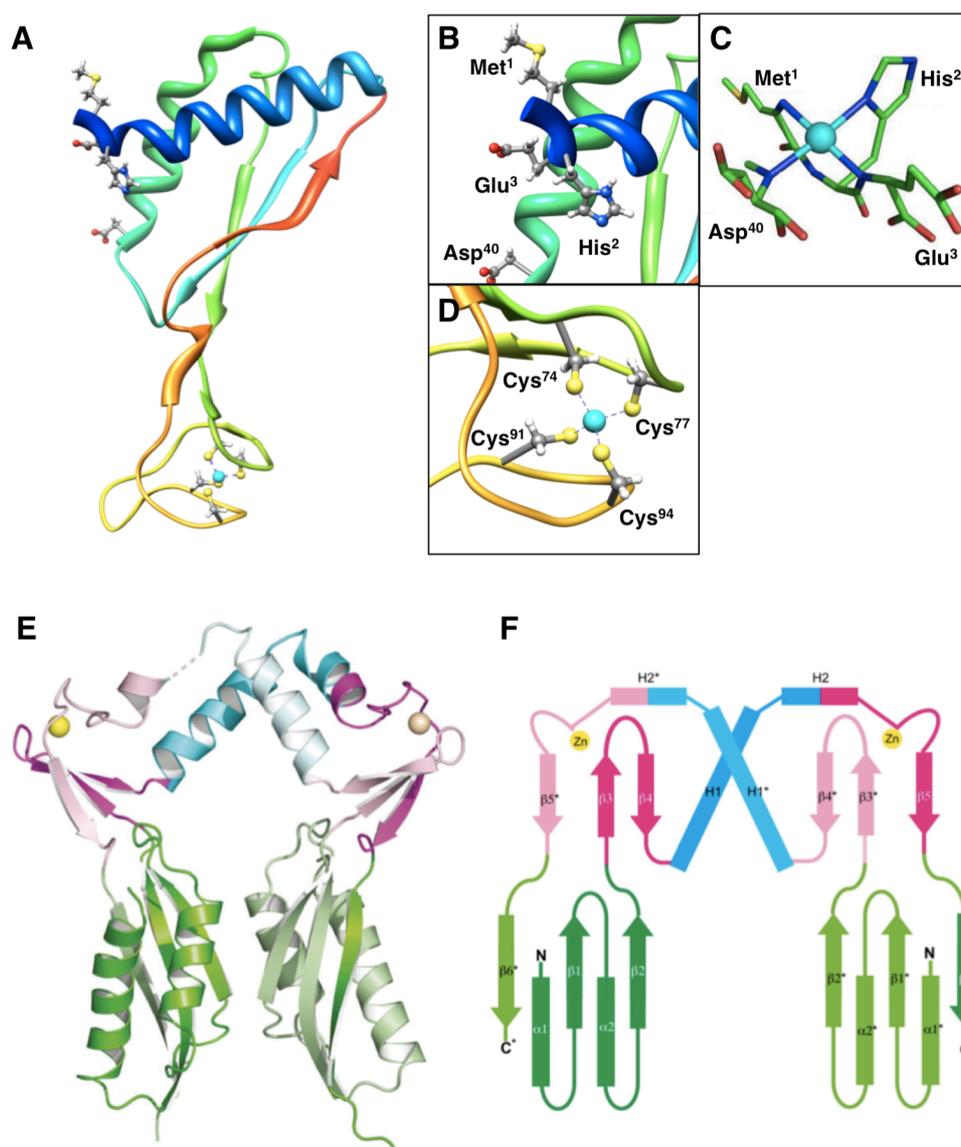
specific chaperones that confer to the protein a functional conformation for the subsequent activation. The iron atoms at the active site of hydrogenases are linked to the non-biological ligands, carbon monoxide and cyanide. Recently, it has been shown that  $\text{CN}^-$  synthesis in *E. coli* depends on two hydrogenase maturation proteins, HypF and HypE [279]. HypF possess an acylphosphatase domain at the N-terminal and a carbamoyltransferase domain located at the C-termini. HypE shows instead high sequence similarity with PurM, an enzyme that catalyzes the dehydration of aminoimidazole ribonucleotide in an ATP-dependent manner [280]. Also HypE displays a conserved ATP-binding motif and additionally features a C-terminal cysteine [281]. Briefly, HypF hydrolyzes carbamoyl phosphate and transfers the carbamoyl group to the HypE C-terminal cysteine, where it is dehydrates to form a HypE-SCN adduct (Figure 40, step 1). The mechanism has been experimentally confirmed, demonstrating the interaction occurring between HypF and HypE [282, 283] and identifying by mass spectrometry the thiocarbamate form of HypE [281]. Furthermore, it has been demonstrated that HypE and HypF form a dynamic complex with HypC and HypD. HypC possess a conserved CxxxP motif in its N-terminal region [284], whereas HypD contains a  $[\text{Fe}_4\text{S}_4]$  cluster [285]. It is thought that the HypCD complex gets the  $\text{CN}^-$  ligands from HypE and then transfers them to the hydrogenase (Figure 40, steps 2, 3), as demonstrated by the isolation of a HypCDE complex and the strong interaction reported for HypC and the hydrogenase large subunit [284], that probably occur through its C-terminal cysteine. It is known that the biosynthetic route for carbon monoxide to the NiFe active site is different from that for cyanide, but the mechanism by which  $\text{CN}^-$  ligands are transferred to the prospective active site has not been elucidated. Also, it is not clear whether the  $\text{Fe}^{2+}$  ion comes from the  $[\text{Fe}_4\text{S}_4]$  cluster in HypD or from some other source. HypC remains in a complex with the hydrogenase precursor until after the nickel is inserted, a step that requires the GTPase activity of HypB and either HypA (Figure 40, step 4) [277]. This is supported by the hetero-dimerization shown *in vitro* by HypA and HypB [252, 286]. Both HypA [287, 288] and HypB [289] feature nickel-binding properties, the latter containing a His-rich motif in its N-terminal part. In addition, these two proteins can be partially replaced *in vivo* by adding high concentrations of nickel to the media [290], suggesting that they cooperate to ensure that the nickel is delivered in the competing cytosolic environment. Also SlyD, a proline cis/trans isomerase interacting with HypB [291], seems to play a role in nickel delivery to the nascent active site, as confirmed by the reduction of hydrogenase activity consequently to its mutation [292]. Finally, the last step in hydrogenase biosynthesis includes chaperones dissociation, cleavage of the C-terminal extension and eventually LS and SS association (Figure 40, step 5) [293]. In particular, depending on the hydrogenase, the size of the extension can vary from only 5 to as many as 32 amino acid residues, but the HybD and Hycl endopeptidases involved in proteolysis are extremely specific [294]. However, some  $[\text{NiFe}]$ -hydrogenases, such as cytoplasmic  $\text{H}_2$  sensors and energy-converting hydrogenases [295], do not have a C-terminal extension in the large subunit, so they do not require proteolytic processing. The hydrogenase

accessory proteins HypA and HypB, required for nickel delivery and active site assembly are analyzed more exhaustively in Section 1.11.6.

### 1.11.6. Accessory proteins for Ni<sup>2+</sup>-delivery and active site assembly of [NiFe]-hydrogenase

Although the mechanism of nickel insertion is not clearly established, both HypA and HypB are reported to bind nickel [286, 287, 291, 296]. Specific nickel binding properties have been observed for HypA from *H. pylori* (*HpHypA*) and *E. coli* (*EcHypA*), which coordinate one nickel per monomer [286], suggesting a possible role in hydrogenase biosynthesis as nickel transporter. Similarly, *EcHybF*, the *EcHypA* homologue for hydrogenase-1 and -2, binds one nickel per monomer. Both proteins show a micro-molar binding affinity for nickel, that is bound to the conserved His<sup>2</sup>, essential for *in vivo* activity [286, 287, 296]. In this context, it has been suggested that HypA acts as a bridging protein between HypB and the hydrogenase enzyme [287], a hypothesis that would explain why the two homologues, *EcHypA* and *EcHybF*, are required for different hydrogenases [296], but that has not yet been proven experimentally [278]. Furthermore, both *EcHypA* and *EcHybF* also contain a zinc ion that, because it is bound in a tetrathiolate coordination sphere [287], is likely a structural cofactor such as those often found in protein domains that mediate PPIs. Interestingly, mutation of the zinc finger in *EcHypF* does not abolish the protein activity, confirming a structural role for this metal binding site [296].

Recently two structures of HypA from different bacterial sources have been solved. The first is a NMR solution structure of monomeric *HpHypA* (Figure 41A) [297], whereas the second is a crystal structure of both monomeric and dimeric HypA from *Thermococcus kodakaraensis* (*TkHypA*) (Figure 41E) [298]. The protein displays a mixed  $\alpha/\beta$  fold, consisting of three  $\alpha$ -helices and five  $\beta$ -stands and an exposed loop where is located the Zn<sup>2+</sup>-binding site. Both studies confirmed that the zinc ion is coordinated by the four cysteine residues in the two conserved CxxC motifs (Cys<sup>74</sup>, Cys<sup>77</sup>, Cys<sup>91</sup>, Cys<sup>94</sup> in *HpHypA*), exhibiting a unique topology (Figure 41D). The nickel binding site of *HpHypA* is situated near the N-terminus and involves the conserved His<sup>2</sup> residue. In particular, the *HpHypA* structure reported a planar diamagnetic Ni<sup>2+</sup> site that is separated by ~30 Å from the zinc site (Figure 41C). On the other hand, the crystal structure revealed two different forms of *TkHypA* that were purified separately. The first form is a monomer structure, analogous to that reported for *HpHypA*, while the second form is a domain-swapped dimer involving a zinc site that is coordinated by four cysteine residues (Figure 41E,F), but where two cysteine residues arise from one monomer and the other two are from the adjacent monomer. The domain swapping dimerization is mediated by two additional linker helices (see topology scheme in Figure 41F) that are not present in the large majority of HypA proteins, but this seems to be a specific structural feature of *Archaea* species. Indeed, the observation that *HpHypA* does not possess the additional helices explains the occurrence of this protein in the monomeric form.



**Figure 41 - Structures of *HpHypA* and *TkHypA***

Ribbon diagrams of *HpHypA* monomer (A) (PDB entry: 2kdx) and *TkHypA* domain swapped dimer (E) (PDB entry: 3a44). In panel A, the structure is colored from blue, near the N-terminal, to red at the C-terminus. Details of Ni<sup>2+</sup>- (B,C) and Zn<sup>2+</sup>-binding (D) sites and a topology diagram of *TkHypA* dimer (F) are also displayed. In panel E, Ni-binding domains (pink), dimerization helices (blue) and Ni-binding domains (green) are shown. Residues that coordinate metals are shown as *ball and stick* and colored according to the following scheme: carbon, grey; nitrogen, blue; oxygen, red; sulfur, yellow; hydrogen, white; zinc, cyan.

These findings suggest a functional diversity of HypA proteins and provided additional evidence for a structurally flexible zinc site, as well as a possible mechanism for communication between the metal binding domains.

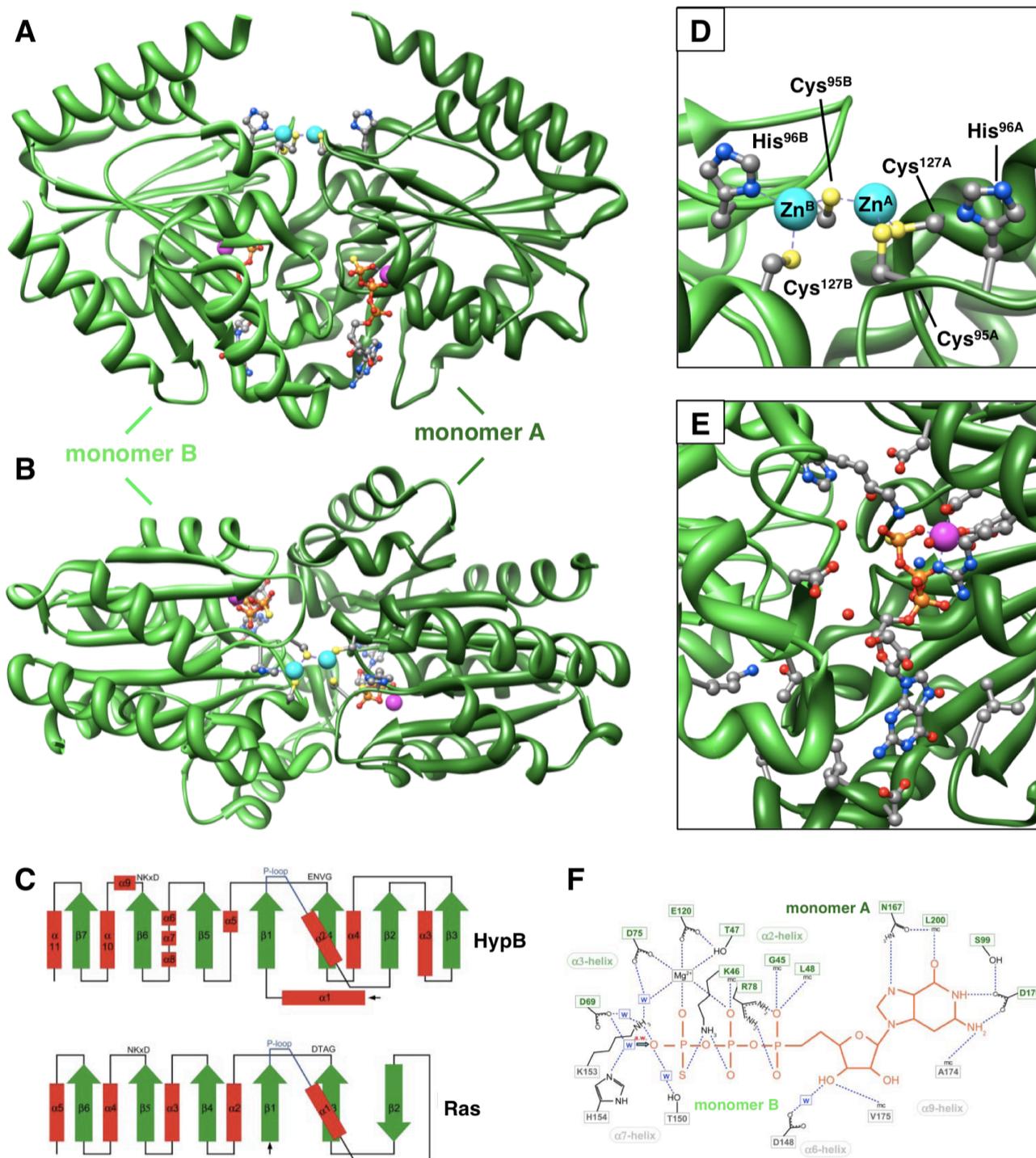
Nickel-insertion process at the [NiFe]-hydrogenase active site level requires GTP-hydrolysis, likely associated to the nucleotide-binding protein HypB, which has demonstrated a low GTPase activity ( $k_{cat} = 0.18 \text{ min}^{-1}$  for HypB from *B. japonicum* [299],  $k_{cat} = 0.17 \text{ min}^{-1}$  for HypB from *E. coli* [300]). Moreover, also HypB possesses nickel binding properties, and in some organisms it has an additional His-rich sequence that is thought to be involved in nickel storage [301]. In *B. japonicum*

this stretch is capable of binding 16 nickel ions per protein dimer with high affinity ( $K_d = 2.3 \mu\text{M}$ ) [299]. However, *E. coli* HypB (*EcHypB*) does not have such a motif but it still binds two nickel ions [291]. In particular, one metal ion is bound with pico-molar affinity in an N-terminal CxxCGC motif, that is not conserved in all species. The other is bound with micro-molar affinity by several completely conserved residues in the GTPase domain and can be competed out with zinc.

*Methanocaldococcus jannaschii* HypB (*MjHypB*), lacking the high affinity nickel binding domain, has been structurally elucidated. The crystal structure revealed an elongated dimeric architecture made of two globular monomers. Each monomer consists of a seven-stranded parallel  $\beta$ -sheet flanked on both sides by eleven  $\alpha$ -helices (Figure 42A,B). The structure shows that two GTP- $\gamma$ S molecules are bound at the dimer interface, one to each monomer, in a parallel orientation (Figure 42E,F). In fact, both monomers contribute to the active site around the  $\gamma$ -phosphate and shield it from solvent. Apart from the canonical interaction with the P-loop and the magnesium ion, the  $\gamma$ -phosphate is additionally fixed by Asp<sup>75</sup>, and a water-mediated contact to Thr<sup>150</sup>. The other monomer also supplies an invariant lysine Lys<sup>153</sup>, which directly contacts the  $\gamma$ -phosphate and is itself bound by two water molecules, one of which is a ligand to Mg<sup>2+</sup> (Figure 42G). Moreover, the protein shows an asymmetric di-zinc cluster, formed by residues Cys<sup>95</sup>, His<sup>96</sup>, and Cys<sup>127</sup> coming from both monomers and located directly in the dimerization interface (Figure 42C). These residues coincide with the predicted metal binding site of *EcHypB*. A conserved motif connects the metal site and the  $\gamma$ -phosphate of the nucleotide, suggesting that metal binding is modulated by a GTP-mediated switch.

The topology shown by HypB is different from that of Ras and other Ras-like proteins (Figure 42C), but it is found in the G-proteins belonging to the SIMIBI class [302]. This P-loop superfamily of GTPases and related ATPases comprises also the urease accessory protein UreG and the nitrogenase protein NifH [302]. Interestingly, all these NTPases possess conserved metal-binding residues in the same region as HypB relative to the NTPase motifs. The exact role of the NTPase and metal-binding activities of these proteins in their respective maturation process is not known but could include regulating protein conformational changes required for cofactor maturation, metal insertion, and sensing correct metallo-center assembly before triggering the next step in the pathway.

To sum up, HypA and HypB mediate nickel insertion into [NiFe]-hydrogenase enzyme through a cooperation process, as demonstrated by the observation of homo-dimers in solution, as well as hetero-dimers in *H. pylori* and *E. coli* [252, 287]. However, this process involves also a series of PPIs with other chaperones and nickel storage proteins. It has been reported that *EcHypB* interacts with SlyD [292], a peptidyl-prolyl isomerase with an unusual C-terminal domain rich in metal-binding histidines, cysteines and carboxylate amino acids [303]. Moreover, SlyD could be replaced by other factors such as the metal binding heat shock protein HspA [304] or histidine-rich peptide Hpn [305] from *H. pylori*.



**Figure 42 - Crystal structure of *MjHypB***

Ribbon representations of the *MjHypB* structure with monomer A in dark green and monomer B in light green (A,B) (PDB entry: 2hf8). In panel B the structure is rotated by 90° around the horizontal axis. Topology diagrams of *MjHypB* and Ras are illustrated. Strands are shown as green arrows, and helices as red bars. The P-loop is shown as a blue line; the black arrow marks the N-terminus of the structural models. A close up of the asymmetric zinc binding site (D), detail (E) and schematic overview (F) of the nucleotide-binding site are shown. Crucial residues are shown as *ball and stick* and colored according to the following atomic scheme: carbon, grey; nitrogen, blue; oxygen, red; sulfur, yellow; phosphor, orange; magnesium, magenta; zinc, cyan.

Finally, in an interesting connection, HypA and HypB also function in the synthesis of the urease enzyme in *H. pylori* [252, 306]. Preliminary cross-linking studies involving urease accessory proteins and HypA-HypB from *H. pylori* have indicated that *HpHypA* can recognize accessory proteins from the heterologous system [306]. Most likely, the *HpHypA-HpHypB* complex is able to donate or mobilize  $\text{Ni}^{2+}$  to a *HpUreE-HpUreG* complex to facilitate nickel donation ultimately to urease [307]. However, a direct crosstalk between urease and hydrogenase accessory systems has not yet been confirmed experimentally.



## Aim of the study

---

Although nickel is a toxic metal for living organisms in its soluble form, its importance in many biological processes recently emerged. In this view, the investigation of the nickel-dependent enzymes urease and [NiFe]-hydrogenase, especially the mechanism of nickel insertion into their active sites, represent two intriguing case studies to understand other analogous systems and therefore to lead to a comprehension of the nickel trafficking inside the cell. Moreover, these two enzymes have been demonstrated to ensure survival and colonization of the human pathogen *H. pylori*, the only known microorganism able to proliferate in the gastric niche.

The right nickel delivering into the urease active site requires the presence of at least four accessory proteins, UreD, UreE, UreF and UreG. Similarly, analogous process is principally mediated by HypA and HypB proteins in the [NiFe]-hydrogenase system. Indeed, *HpHypA* and *HpHypB* also have been proposed to act in the activation of the urease enzyme from *H. pylori*, probably mobilizing nickel ions from *HpHypA* to the *HpUreE-HpUreG* complex. A complete comprehension of the interaction mechanism between the accessory proteins and the crosstalk between urease and hydrogenase accessory systems requires the determination of the role of each protein chaperone that strictly depends on their structural and biochemical properties.

The availability of *HpUreE*, *HpUreG* and *HpHypA* proteins in a pure form is a pre-requisite to perform all the subsequent protein characterizations, thus their purification was the first aim of this work. Subsequently, the structural and biochemical properties of *HpUreE* were investigated using multi-angle and quasi-elastic light scattering, as well as NMR and circular dichroism spectroscopy. The thermodynamic parameters of Ni<sup>2+</sup> and Zn<sup>2+</sup> binding to *HpUreE* were principally established using isothermal titration calorimetry and the importance of key histidine residues in the process of binding metal ions was studied using site-directed mutagenesis. The molecular details of the *HpUreE-HpUreG* and *HpUreE-HpHypA* protein-protein assemblies were also elucidated. The interaction between *HpUreE* and *HpUreG* was investigated using ITC and NMR spectroscopy, and the influence of Ni<sup>2+</sup> and Zn<sup>2+</sup> metal ions on the stabilization of this association was established using native gel electrophoresis, light scattering and thermal denaturation scanning followed by CD spectroscopy. Preliminary *HpUreE-HpHypA* interaction studies were conducted using ITC. Finally, the possible structural architectures of the two protein-protein assemblies were rationalized using homology modeling and docking computational approaches.

All the obtained data were interpreted in order to achieve a more exhaustive picture of the urease activation process, and the correlation with the accessory system of the hydrogenase enzyme, considering the specific role and activity of the involved protein players. A possible function for Zn<sup>2+</sup> in the chaperone network involved in Ni<sup>2+</sup> trafficking and urease activation is also envisaged.



---

## **B - Materials and Methods**

---



## Section B1 - Experimental procedures

### 2.1. General molecular biology techniques

The principal molecular biology procedures applied to carry out cloning of *ureE* and *hypA* genes from *H. pylori* (*HpureE* and *HphypA*, respectively) are described in details.

#### 2.1.1. Standard DNA amplification of the genes of interest

According to the sequence of the urease operon from *H. pylori* 26695 strain (NCBI code NC\_000915) and from *H. pylori* G27 strain (NCBI code NC\_011333) oligonucleotides were designed and synthesized to amplify the *HpureE* and the *HphypA* genes, respectively. The used forward and reverse primers are shown in Table 4: they introduced restriction enzyme recognition sites, here underlined. The start and the stop codons are bold faced, while the introduced mutations are italics in the text. For the *HphypA* gene two different forward primers were used.

Primer name	Sequence	Lenght	Restriction enzyme
5'-UreE_BspHI	5'-CACCC <u>TCATG</u> ATCATAGAGCGTTTAGTTGGC-3'	31 bp	<i>Bsp</i> HI
3'-UreE_XhoI	5'-AC <u>TCGAGCTA</u> TTTTACGACCACTTTAAAATC-3'	31 bp	<i>Xho</i> I
5'-HypA_NcoI	5'-GGTTT <u>ACCATG</u> GATGAATACTCGGTCG-3'	27 bp	<i>Nco</i> I
5'-HypA_NdeI	5'-GGTTT <u>CATATG</u> CATGAATACTCGGTCG-3'	27 bp	<i>Nde</i> I
3'-HypA_BamHI	5'-CGTT <u>CCTAGG</u> TTTTTTATTCCGCTAAC-3'	26 bp	<i>Bam</i> HI

**Table 4 - Description of the primers used for the cloning procedure of *HpUreE* and *HpHypA***

Oligonucleotides for amplifying *HpureE* gene were synthesized by *Novartis* (Siena, Italy) whereas those required for amplifying *HphypA* gene were purchased from *Invitrogen*.

*Easy-A Taq* or *Pfu Turbo* polymerase enzymes employed in the PCR reactions were purchased from *Stratagene*, while dNTPs were from *Promega*. The genes of interest were amplified using the *Helicobacter pylori* G27 strain genomic DNA as template for the PCR reaction, in a final solution of 50  $\mu$ L, whose composition is reported in Table 5. The PCR reaction was performed in a thermocycler (*Hybaid*), according the amplification protocol reported in Table 6. When the reaction was complete, the PCR solution was stored at 4 °C before proceeding to the subsequent steps. The correct size of the amplicons was checked by gel electrophoresis in 1-2% (w/v) agarose (*Sigma*). The fragments with the right base-pair number were purified eluting them from agarose gel using *QIAEX II gel extraction kit* (*Qiagen*); alternatively *QIAquick Purification kit* (*Qiagen*) was used when the presence of a unique band for PCR products was detected.

Reagents	Concentration or volume
DNA template	10-100 ng
5' primer	50 pmol
3' primer	50 pmol
dNTPs mixture	200 $\mu$ M
MgCl <sub>2</sub>	1.5 mM
DMSO	3 $\mu$ l
Reaction buffer (10x)	5 $\mu$ l
Polymerase enzyme	2.5 U
Nuclease free water	until 50 $\mu$ l

Table 5 - Composition of standard PCR reaction mix

Step	Temperature	Time	Cycles applied
1. Initial denaturation	95 °C	5'	1
2. Denaturation	94 °C	30''	35
3. Annealing	55 °C	30''	35
4. Elongation	72°C	45''	35
5. Final elongation	72°C	6'	1

Table 6 - Standard amplification protocol

### 2.1.2. Standard DNA ligase reaction

PCR products and DNA restriction fragments were ligated into their destination vectors. For subcloning procedures into the *pGEM-T* vector (*Promega*), the purified PCR products were used without modifications. This vector is provided as an open DNA molecule, with an overhang T base placed at the cloning site position. The ligase reaction depends on the template-independent activity of *Taq* polymerase, which adds an overhang A base at the ends of the PCR products. Using this strategy, it is possible to obtain a specific ligase reaction without the needing of DNA restriction.

For gene cloning into *pET* expression vectors, the inserts were digested with a combination of two *FastDigest*<sup>®</sup> restriction enzymes (*Fermentas*), whose specific sequences were appositely included in the 5' and 3' PCR primers. Particularly, for the *HpureE* gene the PCR product was double digested with *Bsp*HI and *Xho*I enzymes and while the *pET15b* expression vector (*Novagen*) was digested with *Nco*I (generating compatible ends with *Bsp*HI) and *Xho*I endonucleases. All the reactions were performed at 37°C for 15 minutes in a final volume of 20  $\mu$ L or 50  $\mu$ L, according to the protocol in Table 7. The restriction fragments were then purified by electrophoresis on 1-2% (w/v) agarose gel and the bands of interesting were eluted using *QIAEX II Gel Extraction kit* (*Qiagen*). DNA concentration was measured comparing the intensity of the visible bands in the agarose gel, stained using *SYBR*<sup>®</sup> *Safe DNA gel stain* (*Invitrogen*), with those of a DNA marker *MassRuler*<sup>™</sup> *DNA Ladder Mix* (*Fermentas*). *Safe Imager*<sup>™</sup> blue light transilluminator (*Invitrogen*) was used for detecting DNA bands, at a wavelength of 470 nm.

Reagents	Volume
PCR fragment/ subcloning plasmid	5 - 25 $\mu$ L
Enzyme 1	1 $\mu$ L
Enzyme 2	1 $\mu$ L
Reaction buffer (10x)	5 $\mu$ L
Nuclease free water	18 - 38 $\mu$ L

**Table 7 - Composition of the DNA ligation reaction mix**

DNA fragments were mixed at 1:1 and 1:3 ratio of vector and inserts, according to the formula:

$$ng_i = \frac{ng_v \times bp_i}{bp_v} \times (i : v)$$

where  $i$  and  $v$  represent respectively the DNA fragment (insert) and the destination vector, ( $i:v$ ) the applied ratio,  $ng_i$  and  $ng_v$  are the amounts of insert and vector respectively,  $bp_i$  and  $bp_v$  are their base pairs numbers.

In both cases, ligase reaction was performed in a final volume of 20  $\mu$ L, using *T4* DNA ligase (*Promega*). When the cloning plasmids (*pET15b*, *pET3a*, *pETM-GB1*) were used, they were previously digested with the required restriction enzymes. The reaction was incubated at room temperature for one hour and at 4°C for 4 - 16 hours. The purification of the ligase products was achieved using *QIAquick PCR purification kit* (*Qiagen*), eluting DNA in 20  $\mu$ L of nuclease free water. *XL-10 Gold® Ultracompetent cells* (*Stratagene*), were transformed using 2 or 5  $\mu$ L of the purified ligase product.

### 2.1.3. Transformation of the *E. coli* cells for cloning and expression procedures

Aliquots containing the chemically-competent cells, stored at -80°C, were defrozen on ice and supplemented with 2 or 5  $\mu$ L of plasmid DNA. The cells were subsequently incubated on ice for 30 min. After this time, they were treated with a thermal shock at 42°C for 45 seconds and then incubated on ice for 2 min. NZY+ broth (0.5 ml) was added to the cells and they were grown at 37°C for 60 minutes. An aliquot of 50 or 250  $\mu$ L of the transformed cells were plated on LB-agar medium, containing the appropriate antibiotic for transformants selection, and incubated at 37°C for 16 hours.

### 2.1.4. Selection of the positively transformed cells

The transformed cells were selected in two different phases for the correct incorporation of the required DNA plasmids. The first procedure was based on the antibiotic resistance codified by the sequence of the vector: after transformation, cells were plated on agar plates containing the specific antibiotic, ampicillin or kanamycin, at lethal concentrations for wild-type strains. In the second phase, the plasmid DNA of each positive clone was purified using the *StrataPrep™*

*Plasmid MiniPrep Kit (Stratagene)* and digested with specific *FastDigest*<sup>®</sup> restriction enzymes (*Fermentas*) according to the protocol in Table 8.

Reagents	Volume
Plasmid DNA	5 $\mu$ L
Enzyme 1	0.5 $\mu$ L
Enzyme 2	0.5 $\mu$ L
Reaction buffer (10x)	5 $\mu$ L
Nuclease free water	12 $\mu$ L

**Table 8 - Composition of the DNA restriction reaction mix**

The presence of the correct pattern of DNA bands was checked by electrophoresis on 1-2 % (w/v) agarose gel. The positive constructs were confirmed also by PCR reaction using specific primer pairs, according to the protocol described above, and by double-strand DNA sequencing, performed at *BMR Genomics*, the DNA sequencing service at the University of Padova. The clones that had incorporated the correct DNA constructs were grown in 50 - 100 ml cultures, in order to purify larger amounts of plasmidic DNA using the *StrataPrep*<sup>™</sup> *EF Plasmid MidiPrep Kit (Stratagene)*.

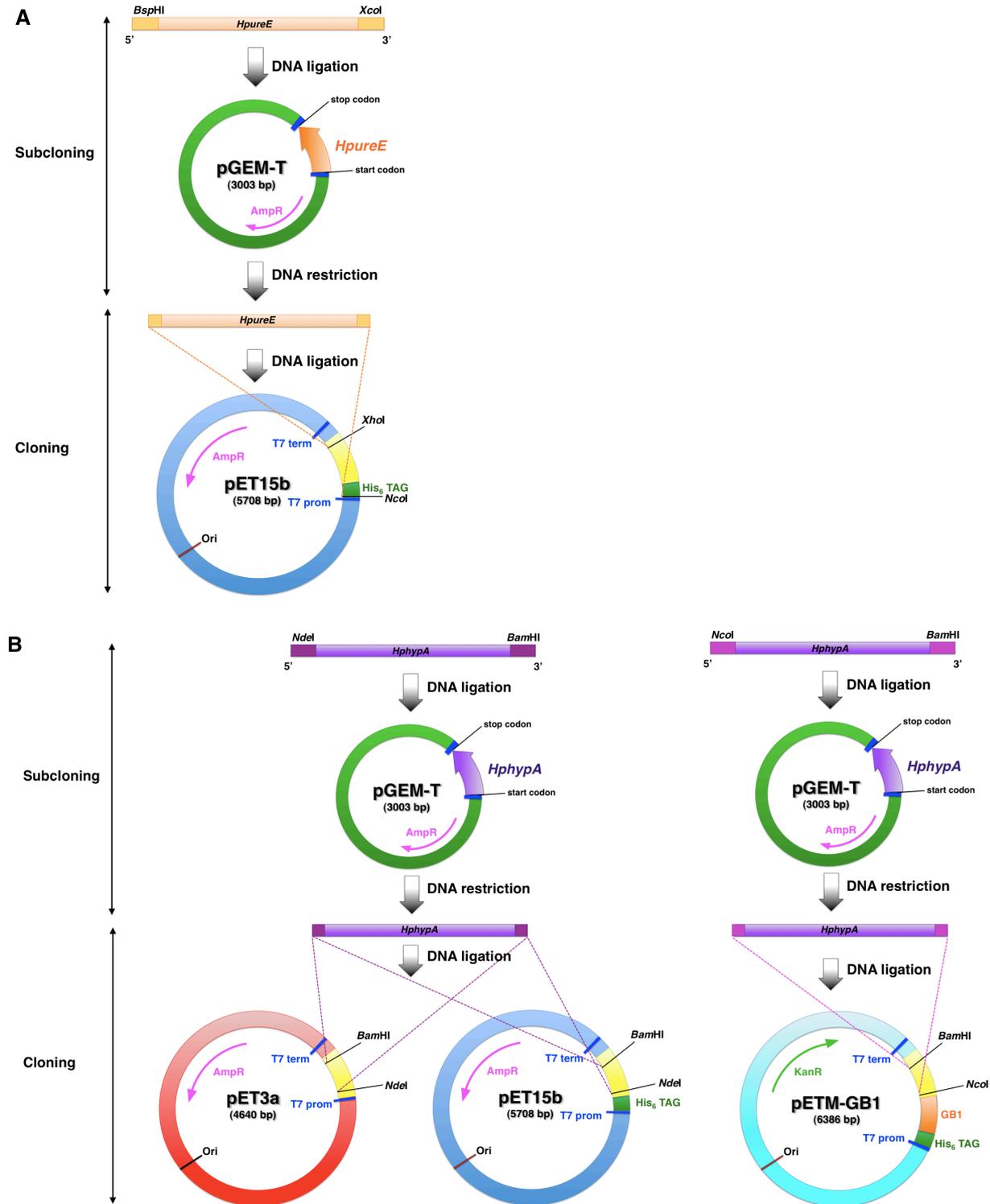
### 2.1.5. Cloning overview

Details of the cloning procedure for *HpureE* and *HphypA* genes are depicted in Figure 43, respectively. In both cases, ligase reaction was employed to build a subclone, where the gene of interest is inserted and can be excised using specific restriction enzymes, in order to clone it in the appropriate destination vector(s).

The *HpureE* gene was subcloned into the *pGEM-T* vector (*Promega*) and subsequently cloned into the *pET15b* expression vector (*Novagen*), which expresses the protein in its native form, because the chosen restriction endonucleases excise the nucleotides encoding the His<sub>6</sub> tag present immediately after the starting codon (see Figure 43A).

Also For the *HphypA* gene the *pGEM-T* vector (*Promega*) was used as subcloning vector. In this case, different expression vectors were tested, in order to evaluate the best system for the *HpHypA* protein. Two different PCR fragments, containing the recognition sites fro the appropriate restriciton enzymes, were directly inserted into the *pET3a* (*Novagen*), *pET15b* (*Novagen*) and *pETM-GB1*, expressing respectively the native, the His<sub>6</sub>-tagged protein and the GB1 fusion protein (see Figure 43B). The latter plasmid is a *pET* derived expression vector [308] which expresses the protein fused to an N-terminal His<sub>6</sub>-tag followed by the 56-aa GB1 (B1 domain of streptococcal protein G), a glycine-serine linker peptide (GSGSGS) and a TEV (tobacco etch virus) protease cleavage site (ENLYFQG). After protease digestion, two residue, Gly and Ala, will be left before the N terminus of target protein [309]. This expression vector is very similar to *pGBO* (plasmid of

His<sub>6</sub>GB1 domain fusion expression with thrombin digestion site), which has been reported to yield high expression levels in *E. coli* enhancing the stability (up to sixfold) and the refolding of fused target proteins, without impacting on their structure [310]. The *pETM-GB1*, kindly provided by Gunter Stier (EMBL, Heidelberg, Germany), requires kanamycin, and not ampicillin, for the selection of positive transformants.

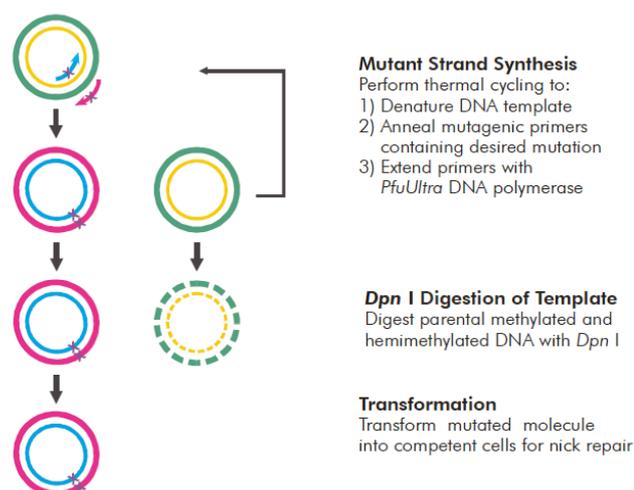


**Figure 43 - Overview of the cloning procedures**

Schematic representations of *HpUreE* (A) and *HpHypA* cloning (B).

### 2.1.6. Site-directed mutagenesis protocol

Generation of the H102A, H102K, H120Y and H152A *HpUreE* mutants was carried out by *in vitro* site-directed mutagenesis, using the *QuikChange® II Site-Directed Mutagenesis Kit* (Stratagene) and the *pET15b-HpureE* as template. This *in vitro* procedure allows site-specific mutation in virtually any double-stranded plasmid, by a three-step experimental protocol. The basic procedure utilizes a supercoiled dsDNA vector with an insert of interest and two synthetic oligonucleotide primers, both containing the desired mutation. The oligonucleotide primers, each complementary to opposite strands of the vector, are extended during temperature cycling by high fidelity *PfuUltra* DNA polymerase, without primer displacement. Extension of the oligonucleotide primers generates a mutated plasmid containing staggered nicks. Following temperature cycling, the product is treated with *DpnI* endonuclease, that is specific for methylated and hemimethylated DNA, allowing digestion of the parental DNA template. In this way it is possible to select for mutation-containing synthesized DNA (Figure 44).



**Figure 44 - Overview of the *QuikChange® II Site-Directed Mutagenesis Kit* procedure**

The mutagenesis primers were designed using the online available *QuikChange® Primer Design* software ([www.stratagene.com/sdmdesigner](http://www.stratagene.com/sdmdesigner)) and subsequently they were synthesized by *Invitrogen*. The oligonucleotide pairs are reported in Table 9.

Primer name	Sequence	Length
5'-H102A_UreE	5'-CTATGAAATAGGAAACCGCAAGGCGGCTTTATACTATGGCG-3'	41 bp
3'-H102A_UreE	5'-CGCCATAGTATAAAGCCGCCTTGCGGTTTCCTATTTTCATAG-3'	41 bp
5'-H102K_UreE	5'-CTATGAAATAGGAAACCGCGCTGCGGCTTTATACTATGGC-3'	40 bp
5'-H102K_UreE	5'-GCCATAGTATAAAGCCGCAGCGGTTTCCTATTTTCATAG-3'	40 bp
5'-H102Y_UreE	5'-CTATGAAATAGGAAACCGCTATGCGGCTTTATACTATGG-3'	39 bp
3'-H102Y_UreE	5'-CCATAGTATAAAGCCGCATAGCGGTTTCCTATTTTCATAG-3'	39 bp
5'-H152A_UreE	5'-CTTAACCGTGAGCATGCCCGCTAGTGAGCCTAATTTTAAGG-3'	41 bp
5'-H152A_UreE	5'-CCTTAAAATTAGGCTCACTAGCGGGCATGCTCACGGTTAAG-3'	41 bp

**Table 9 - Description of the primers used for the site-direct mutagenesis of *HpUreE***

Mutagenesis reactions were set up for each *HpUreE* mutant, as indicated in Table 10.

Reagents	Concentration or volume
Plasmid DNA	10 ng
5' mutagenesis primer	125 ng
3' mutagenesis primer	125 ng
dNTPs mix	200 $\mu$ M
<i>PfuUltra</i> DNA polymerase	2.5 U
Nuclease free water	until 50 $\mu$ L

**Table 10 - Composition of the PCR reaction mix for the site-direct mutagenesis of *HpUreE***

PCR amplification reactions were performed following the protocol reported in Table 11.

Step	Temperature	Time	Cycles applied
1. Initial denaturation	95°C	30''	1
2. Denaturation	95°C	30''	16
3. Annealing	55°C	30''	16
4. Elongation	68°C	6' 15''	16

**Table 11 - Amplification protocol applied for the site-direct mutagenesis of *HpUreE***

When the reaction was complete, 10 U of the *Dpn* I restriction enzyme were added directly to each amplification; each reaction was incubated at 37°C for one hour to digest the parental non-mutated supercoiled dsDNA. *XL-1Blue<sup>®</sup> Supercompetent cells (Stratagene)* were transformed using 2 or 5  $\mu$ L of each reaction solution. The plasmids containing the desired mutation were selected as described in section 2.1.4.

The sequences of the mutants were verified by sequencing the genes on both strands and finally the correct plasmids were extracted, purified and used to transform *BL21-(DE3) competent cells (Novagen)* to carry out protein expression.

## 2.2. Bacterial growth and heterologous protein expression

### 2.2.1. Culture media

The composition of the media used for bacterial growth and protein expression are reported. For all media the appropriate antibiotic concentration was added to maintain a selective pressure: ampicillin or carbenicillin at 100  $\mu$ g ml<sup>-1</sup>, kanamycin at 25  $\mu$ g ml<sup>-1</sup>.

#### Luria Bertani (LB)

In order to obtain LB medium, LB powder (*Amerham Pharmacia Biotech*) was dissolved in bi-distilled water to a final concentration of 20 mg l<sup>-1</sup>. The pH was adjusted to 7.6 and the solution was autoclaved. To produce agar plates, 1.5 % (w/v) agar was added to the medium before autoclaving it.

**NZY+**

NZY+ medium was prepared dissolving in bi-distilled water the components reported in Table 12. The pH was adjusted to 7.5 and the solution is autoclaved. The magnesium solutions were prepared and autoclaved separately. The glucose solution was prepared and filtered apart.

Reagents	Concentration
Casein hydrolysate	1 % (w/v)
Yeast extract	0.5 % (w/v)
NaCl	0.5 % (w/v)
MgSO <sub>4</sub>	10 mM
MgCl <sub>2</sub>	10 mM
Glucose	20 mM

**Table 12 - Composition of the NZY+ medium**

**M9**

M9 medium was prepared dissolving in MilliQ water the components reported in Table 13. The pH was adjusted to 7.5 and the solution was autoclaved. The MgSO<sub>4</sub> solution was prepared and autoclaved separately. The glucose solution was prepared and filtered apart. In case of <sup>15</sup>N-labeled protein production, the correspondent amount of <sup>15</sup>NH<sub>4</sub>Cl was added.

Reagents	Concentration
Na <sub>2</sub> HPO <sub>4</sub>	6 g l <sup>-1</sup>
KH <sub>2</sub> PO <sub>4</sub>	3 g l <sup>-1</sup>
NaCl	0.5 g l <sup>-1</sup>
(NH <sub>4</sub> ) <sub>2</sub> SO <sub>4</sub>	1.25 g l <sup>-1</sup>
MgSO <sub>4</sub>	0.236 g l <sup>-1</sup>
Glucose	4 g l <sup>-1</sup>

**Table 13 - Composition of the M9 medium**

**M9 auto-induction**

M9 auto-induction medium was prepared adjusting the protocol described above. The components used are listed in Table 14.

Reagents	Concentration
Na <sub>2</sub> HPO <sub>4</sub>	6 g l <sup>-1</sup>
KH <sub>2</sub> PO <sub>4</sub>	3 g l <sup>-1</sup>
NaCl	0.5 g l <sup>-1</sup>
(NH <sub>4</sub> ) <sub>2</sub> SO <sub>4</sub>	1.25 g l <sup>-1</sup>
MgSO <sub>4</sub>	0.236 g l <sup>-1</sup>
Glycerol	0.5 g l <sup>-1</sup>
Glucose	4 g l <sup>-1</sup>
Lactose	2 g l <sup>-1</sup>

**Table 14 - Composition of the M9 auto-induction medium**

The pH was adjusted to 7.5 and the solution was autoclaved. The MgSO<sub>4</sub> solution was prepared and autoclaved separately. The glucose and lactose solutions were prepared and filtered apart. In case of <sup>15</sup>N-labeled protein production, the correspondent amount of <sup>15</sup>NH<sub>4</sub>Cl was added.

### **2.2.2. HpUreE expression protocol**

In order to increase protein yield and solubility, large-scale expression of HpUreE was achieved in 2 l batches of M9 auto-induction medium, starting from a pre-inocule of 50 ml of a 16 hours cell culture. Cells were grown at 28 °C for 48 hours and then harvested by centrifugation at 8.000 g for 20 minutes, at 4 °C. The cellular pellet was re-suspended in 30 ml of 20 mM MES (2-(N-morpholino)ethanesulfonic acid), pH 6.5, containing 500 mM NaCl, 2 mM dithiothreitol, 10 mM MgCl<sub>2</sub>, and 20 µg ml<sup>-1</sup> DNase I. Cells were disrupted by two passages through a *French Pressure cell* (SLM-Aminco) at 20.000 psi (1 psi = 6.9 kPa). Cell debris was separated from the supernatant by centrifugation at 15.000 g for 30 minutes at 4 °C.

Heterologous expression of the H102KA, H102K, H102Y and H152A HpUreE mutants were carried out using the same protocol described for the wild-type protein.

### **2.2.3. HpUreG expression protocol**

Large scale expression of HpUreG in *E. coli* BL21(DE3) expression host (Novagen) was conducted as previously described [251]. Cells were grown in 2 l batches of M9 auto-induction medium at 28 °C for 48 hours, starting from a pre-inocule of 50 ml of a 16 hours cell culture. Cells were harvested by centrifugation at 8.000 g for 20 minutes at 4 °C. The cellular pellet was resuspended in 30 ml of 50 mM Tris HCl, pH 8, containing 2 mM EDTA, 2 mM DTT, 10 mM MgCl<sub>2</sub>, and 20 µg ml<sup>-1</sup> DNase I. Cells were disrupted 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 15.000 g for 30 minutes at 4 °C.

Heterologous expression of the C66A/H68A HpUreG double mutant were carried out using the same protocol described for the wild-type protein.

### **2.2.4. HpHypA expression protocol**

Small scale test for native, His<sub>6</sub>-tagged and His<sub>6</sub>GB1-tagged HpHypA expression were conducted using *E. coli* BL21(DE3) expression host (Novagen) transformed with *pET3a-HphypA*, *pET15b-HphypA* and *pETMGB1-HphypA* constructs, respectively. Transformed cells were grown in 20 - 50 ml of LB or M9 medium at 37 °C, until the OD<sub>600</sub> reached 0.5 - 0.6. Expression was induced by addition of IPTG at final concentrations ranging from 0.1 to 1 mM. Cells were harvested 4 hours after induction by centrifugation at 8,000 g for 10 min, at 4 °C and were resuspended in 1/100 of the

culture volume of 50 mM Tris HCl buffer, pH 8.0 and proper amount of *CellLytic™ Express* (Sigma) protein extraction formulation. After incubation at 37 °C for 30 min and sonication, the soluble and insoluble fractions were separated by centrifugation at 15.000 *g* for 15 minutes at 4°C. Pre-induction sample, and total cellular extract after induction were checked by SDS-PAGE.

Large-scale expression of His<sub>6</sub>GB1HpHypA was achieved in 2 l batches of LB medium, starting from a pre-inocule of 50 ml of a 16 hours cell culture. Cells were grown at 37 °C, until the OD<sub>600</sub> reached 0.7 - 0.8. Expression was induced by addition of IPTG at final concentration of 1 mM and the temperature was decreased to 28°C after induction. Cells were harvested 3 hours after induction by centrifugation at 8.000 *g* for 20 minutes at 4 °C. The cellular pellet was re-suspended in 30 ml of 20mM NaH<sub>2</sub>PO<sub>4</sub> pH 7.4, containing 500 mM NaCl, 20 mM imidazole, 10 mM MgCl<sub>2</sub>, and 20 µg ml<sup>-1</sup> DNase I. Cells were disrupted by two passages through a *French Pressure cell* (SLM-Aminco) at 20.000 psi. Cell debris was separated from the supernatant by centrifugation at 15.000 *g* for 30 minutes at 4°C.

### 2.2.5. TEV protease expression protocol

Expression of His<sub>6</sub>-tagged TEV protease was carried out adapting the protocol reported in past [311]. Cells were grown in 2 L batches of LB medium at 37°C until the OD<sub>600</sub> reached 0.8, starting from a pre-inocule of 50 ml of a 16 hours cell culture. Expression was achieved growing cells overnight at 28°C after addition of IPTG at final concentration of 1 mM. Cells were harvest by centrifugation at 4 °C (8.000 *g* for 20 minutes), resuspended in 30 ml of 50 mM NaH<sub>2</sub>PO<sub>4</sub>, 100 mM NaCl, pH 8.0 and lysed by passing twice through a *French Pressure cell* (SLM-Aminco) operating at 20.000 psi. The soluble fraction, was obtained after removal of the precipitated material by centrifugation at 4°C (15.000 *g* for 30 minutes).

## 2.3. Protein purification

### 2.3.1. Wild-type and mutated HpUreE purification protocol

Recombinant wild-type HpUreE was purified using a protocol adapted from a previous study [257]. According to the high isoelectric point (8.56) calculated for the HpUreE sequence using the, *Protparam* tool [312], the protein was isolated using cation-exchange chromatography followed by two size-exclusion separations. The soluble fraction obtained by cell lysis was dialyzed twice for 3 h at 4°C against 3 l of 20 mM MES, pH 6.5, containing 2 mM EDTA, using 3.5 kDa molecular mass cut-off *SnakeSkin™ Pleated Dialysis Tubing* (Thermo Scientific). Subsequently, it was applied on to a *SP-Sepharose XK 16/10* column (GE Healthcare) pre-equilibrated with the same buffer. The column was washed using a flow rate of 2 ml<sup>-1</sup> min<sup>-1</sup> with the starting buffer until the baseline was

stable. The protein was eluted with a linear gradient of 200 ml of 0 to 1 M NaCl. Fractions containing *HpUreE* were combined, concentrated using 10 kDa molecular mass cut-off *Centricon ultra-filtration units* (Millipore) and loaded onto a *Superdex 75 XK 16/60* column (*GE Healthcare*) equilibrated with two column volumes of 20 mM Tris HCl, pH 7.0, containing 150 mM NaCl. The eluted protein was further purified using high-resolution size-exclusion chromatography on *Superdex 75 HR 10/30* (*GE Healthcare*), in the same buffer. In each purification step, the purity of *HpUreE*, as well as its molecular mass under denaturing conditions, was estimated by SDS-PAGE.

Protein purification of the H102KA, H102K, H102Y and H152A mutants were carried out as described for the wild-type protein.

### **2.3.3. Wild-type and mutated *HpUreG* purification protocol**

Recombinant wild-type *HpUreG* was purified as previously described [251]. According to the isoelectric point (5.52) calculated for the *HpUreG* sequence using the *ProtParam* tool [312], the protein was isolated using a combination of anion-exchange chromatography and size-exclusion separations. The soluble fraction obtained by cell lysis was loaded onto a *Q-Sepharose XK 26/10* column (*GE Healthcare*) that had been pre-equilibrated with 20 mM Tris HCl buffer, pH 8, containing 2 mM EDTA. The column was washed with the starting buffer until the baseline was stable. The protein was eluted from the column with a 500 ml linear gradient of NaCl (0 - 1M). The fractions containing *HpUreG* were collected, and the protein was concentrated using 5 kDa MWCO *Amicon* and *Centricon ultra-filtration units* (Millipore) after addition of 2 mM DTT. The obtained sample was loaded onto a *Superdex 75 XK 16/60* column (*GE Healthcare*), previously conditioned with 20 mM Tris HCl buffer, at pH 8, containing 150 mM NaCl and 10 mM EDTA. The protein fractions were collected, stored in 2 mM DTT, and further purified by another step of size exclusion chromatography with *Superdex 75 HR 10/30* (*GE Healthcare*), using 20 mM Tris HCl pH 7.0, 150 mM NaCl. In each purification step, the purity of *HpUreG*, as well as its molecular mass under denaturing conditions, was estimated by SDS-PAGE electrophoresis.

The same purification protocol was followed to achieve purification of C66A/H68A *HpUreG* double mutant.

### **2.3.3. *HpHypA* purification protocol**

His<sub>6</sub>GB1-tagged *HpHypA* was purified using standard Ni-affinity chromatography protocol. The soluble fraction obtained by cell lysis was loaded onto a column containing 5 ml of the *Ni-NTA Superflow* affinity resin (*Qiagen*) pre-equilibrated with 25 ml of 20 mM NaH<sub>2</sub>PO<sub>4</sub> pH 7.4, containing 500 mM NaCl and 20 mM imidazole and washed with 10 ml of the same buffer containing 20 mM imidazole. His<sub>6</sub>GB1-*HpHypA* was eluted with 20 ml of 20 mM NaH<sub>2</sub>PO<sub>4</sub> pH 7.4, containing 500 mM NaCl and 250 mM imidazole. Immediately, after elution the protein solution was dialyzed twice for 2

h at 4°C against 2 l of 20 mM Tris HCl pH 7.0 and then a proper amount of TEV protease (200 µL of 1 mg ml<sup>-1</sup> enzyme stock), plus 0.5 mM EDTA and 1mM DTT, was added to the protein solution. Digestion was conducted at 20°C for 24 h. Protein was concentrated using *3.5 kDa MWCO Centricon ultra-filtration units (Millipore)*, until reaching a final volume of 2-3 ml. The sample was briefly centrifuged at 4°C (15.000 g for 2 minutes) to eliminate eventually precipitated material and then loaded onto a column filled with 5 ml of *Ni-NTA Superflow* affinity resin (*Qiagen*) pre-conditioned with 25 ml of f 20 mM Tris HCl pH 7.0 and the flow-through, containing the native form of *HpHypA* as a consequence of proteolytic cleavage, was collected. Additional elution was performed with 2-3 ml of the same buffer used for column equilibration. Protein size and purity, as well as the completion of proteolytic digestion, were checked by SDS-PAGE in each purification step. The fraction containing *HpHypA* were combined and concentrated using *3.5 kDa MWCO Centricon ultra-filtration units (Millipore)*. Protein buffer was exchanged to 20 mM Tris HCl pH 7.0, containing 150 mM NaCl, passing the sample onto a *PD-10 desalting* column (*GE Healthcare*).

#### 2.3.4. TEV protease purification protocol

Recombinant His<sub>6</sub>-tagged TEV protease was purified as previously described [311]. The supernatant after pellet separation was loaded onto a column containing 5 ml of the *Ni-NTA Superflow* affinity resin (*Qiagen*) pre-equilibrated with 25 ml of 20 mM NaH<sub>2</sub>PO<sub>4</sub> pH 7.4, containing 500 mM NaCl and 20 mM imidazole and washed with 10 ml of the same buffer. Elution was achieved using 20 ml of 20 mM NaH<sub>2</sub>PO<sub>4</sub> pH 7.4, containing 500 mM NaCl and 500 mM imidazole. Immediately, after elution the protein solution was extensively dialyzed at 4°C against 2 liters of 50 mM Tris HCl pH 7.5, containing 1mM EDTA and 5 mM DTT. Protein purity, as well as the correct molecular mass under denaturing conditions, was estimated by SDS-PAGE. Protein was concentrated using *5 kDa MWCO Centricon ultra-filtration units (Millipore)*, until reaching a final concentration of 1 mg ml<sup>-1</sup>. In order to preserve protease activity, glycerol (50% w/v) and Triton X-100 (0.1% v/v) were added to the sample for storage at -80°C.

## 2.4. Preliminary protein characterization

### 2.4.1. Evaluation of protein purity and molecular mass using SDS-PAGE electrophoresis

In order to check protein purity and size, SDS-PAGE electrophoresis was conducted on *NuPAGE Novex Pre-Cast Gel System (Invitrogen)* using *NuPAGE 4-12%* or *4-20% Bis-Tris* gels, for *HpUreE*, *HpUreG* and TEV protease sample or *HpHypA* samples, respectively. After running (120 - 160 V for 50 - 65 minutes), gels were stained using the *SimplyBlue Safestain (Invitrogen)*. *PageRuler™ Unstained Broad Range* or *Spectra™ Multicolor Broad Range* protein ladders (*Fermentas*) were used to estimate the correct molecular mass under denaturing conditions.

#### **2.4.2. Determination of protein identity by tryptic digestion and mass spectrometry**

The identity of purified *HpUreE* and *HpUreG*, recovered from an SDS-PAGE gel, was demonstrated by tryptic digestion, and the similarity between the expected and the experimental protein sequence was confirmed by ESI Q-TOF tandem mass spectrometry performed at *EMBL Proteomic Core Facility* (Heidelberg). The protein bands of interest, containing isolated proteins, were manually excised from the gel, and then treated according to the protocol of Rosenberg [313]. The spots were destained by immersion of gel pieces into acetonitrile followed by rehydration in 100 mM ammonium hydrogen carbonate. The proteins were reduced and alkylated by placing the gel pieces first into reducing solution, containing 100 mM ammonium hydrogen carbonate and 10 mM DTT, and then into alkylation solution, containing 100 mM ammonium hydrogen carbonate and 55 mM 2-iodoacetamide. For application of the trypsin the gel pieces were dehydrated in acetonitrile and submersed in digestion solution, containing 40 mM ammonium hydrogen carbonate, 4 mM calcium chloride and 1.5 ng  $\mu\text{L}^{-1}$  trypsin. The rehydration of the gel pieces with digestion solution was performed on ice for 45 min. Afterwards the remaining digestion solution was removed and replaced by 15  $\mu\text{L}$  of digestion buffer (40 mM ammonium hydrogen carbonate, 4 mM calcium chloride). The proteins were digested at 37°C over night and the resulting peptides were extracted in two steps. Acetonitrile (50  $\mu\text{L}$ ) was added and the gel pieces were sonicated for 15 min. The supernatant was transferred to a separate tube and the gel pieces were rehydrated with 30  $\mu\text{L}$  5% formic acid. For the second extraction, 60  $\mu\text{L}$  acetonitrile were added and sonicated again. The supernatant was removed and pooled with the peptides previously extracted. The latter were dried using a vacuum concentrator (*Eppendorf*) and then resuspended in 10  $\mu\text{l}$  of 5% formic acid.

Mass spectrometric analysis was performed using ESI (electrospray ionization) mode and a tandem MS/MS mass spectrometer with a hybrid quadrupole, orthogonal acceleration time-of-flight *Q-TOF2™* (*Waters*). The ESI-source was operated at a temperature of 100 °C with a nitrogen drying gas flow of 10 l h<sup>-1</sup>. To the nano flow needle a potential of 1500 V was applied. Full-scan mass spectra were acquired in the continuous data acquisition mode in the range *m/z* 500-1500 at a scan rate of 1 s and an inter-scan delay of 0.1 s per scan. In the MS/MS mode the quadrupole was used in the RF/DC (RF, radio frequency; DC, direct current) mode to select specific parent ions, which were subsequently fragmented in the hexapole collision cell using argon as collision gas (0.5 bar). The cone energy was always 40 V, whereas the collision energy was optimized on each precursor ion (28 - 40 V) for optimum fragmentation. The peptide mixture was dissolved in 5% formic acid, desalted with *C18 micro ZIP-TIP* columns (*Millipore*) and eluted with acetonitril/water/formic acid (49.5/49.5/1; *Merck*) directly into capillary nano flow glass needles (*Waters*).

### 2.4.3. Determination of protein concentration

The concentration of all protein samples was determined using a *BioPhotometer* spectrophotometer (*Eppendorf*) and dedicated plastic *UVette* (*Eppendorf*). For the evaluation theoretical molar extinction coefficients ( $\epsilon$ ) at 280 nm (see Table 15), calculated from the amino acid sequence by the *ProtParam* tool [312], were utilized.

Protein sample	$\epsilon_{280}$ ( $M^{-1} cm^{-1}$ )
<i>HpUreE</i> *	11.460
<i>HpHypA</i>	3.355
His <sub>6</sub> GB1- <i>HpHypA</i>	14.815
<i>HpUreG</i> *	10.095
His <sub>6</sub> -TEV protease	32.095

**Table 15 - Molar extinction coefficients for spectrophotometric evaluation of protein concentration**

\* For *HpUreE* and *HpUreG* mutated versions the same  $\epsilon_{280}$  of the wild-type protein was used, because a single amino acid substitution does not significantly affect the coefficient value.

For *HpUreE* and *HpUreG* concentration values were also confirmed by absolute protein quantitation, performed by measuring the amount of sulfur in a protein sample with known absorbance at 280 nm by ICP-ES (inductively coupled plasma emission spectroscopy). For instance, a sample with an estimated protein concentration of 25.5  $\mu M$ , a sulfur concentration of 284  $\mu M$  was measured for wild-type *HpUreG*. Because the amino acid composition of *HpUreG* contains twelve sulfur atoms, coming from three cysteine and nine methionine residues, per protein monomer, a concentration of 23.7  $\mu M$  was inferred for *HpUreG*, a value in good agreement with the theoretical estimation [251].

For *HpUreE* the protein concentration is always expressed by referring to the dimer (38.815 Da), whereas for *HpHypA* and *HpUreG* the concentration values refer to the monomeric form of both proteins (13.202 Da and 21.955 Da, respectively).

### 2.4.4. Determination of metal content in protein samples

The absence of any metal bound to the purified *HpUreE* and *HpUreG* protein samples was confirmed by ICP-ES, using a procedure as described in [260]. Metal analysis was performed using a *Spectro Ciros CCD ICP* optical emission spectrometer (*Spectro Analytical Instruments*) in combination with a *Lichte* nebulizer and a peristaltic pump for sample introduction. The ICP-ES system was calibrated by serial dilutions of appropriate single and multi-element standards (*CPI International*). The standardization curve was made using standard solutions in the range 0 - 500 mM of Ni and Zn in Tris HCl 2.5 mM, pH 8, and NaCl 7.5 mM with a linear fitting. An R<sub>f</sub> power of 1400 W, a nebulizer gas flow of 0.8 l min<sup>-1</sup> and a plasma gas flow of 14 l min<sup>-1</sup> were used. The sample uptake was set at 2 ml min<sup>-1</sup> for 24 s, and a wash time of 15 s at 4 ml min<sup>-1</sup> plus 45 s at 2 ml min<sup>-1</sup>, for each sample. Quality control was established by evaluation of buffer containing

standards. In order to estimate the total metal added to every protein sample, 200 mL of every metal solution were mixed in a 1:1 ratio with the blank buffer, diluted to 8 mL with milliQ water, and measured as the filtered samples. The 221.648 and 231.604 nm lines for Ni and the 202.548, 206.191, and 213.604 nm lines for Zn were used for analysis. The measured metal ion concentrations were corrected with the value obtained for the filtered solution of the protein incubated with the buffer blank, without metal ion. The experimental points were fitted using the *MacCurveFit* software, and the fit optimized using a *Quasi-Newton* algorithm.

## **2.5. HpUreE biochemical and structural characterization**

### **2.5.1. Circular dichroism spectroscopy**

The secondary structure of *HpUreE* was evaluated by CD (circular dichroism) spectroscopy, performed on the protein (5  $\mu$ M) diluted in 20 mM phosphate buffer, pH 8.0, using a *JASCO 810* spectropolarimeter flushed with N<sub>2</sub>, and a cuvette with 0.1 cm path-length. Ten spectra were accumulated from in the far-UV spectral region (190 - 240 nm) at 0.2 nm intervals, at 20°C and averaged to achieve an appropriate signal-to-noise ratio. The spectrum of the buffer was always subtracted. The secondary structure composition of *HpUreE* was evaluated using the *CDSSTR* tool [314] available on the *Dichroweb* server [315], with the reference sets 3, 4, 6, and 7 (<http://dichroweb.cryst.bbk.ac.uk/html>).

### **2.5.2. NMR spectroscopy experiments**

Uniformly <sup>15</sup>N-labeled *HpUreE* was produced with the same purification procedure used for the unlabeled proteins, using M9 auto-induction medium containing <sup>15</sup>NH<sub>4</sub>Cl as sole nitrogen source. <sup>1</sup>H-<sup>15</sup>N HSQC (heteronuclear single quantum correlation) experiments [316] and TROSY (transverse relaxation optimized spectroscopy)-HSQC experiments [317] were carried out at 298 K on a 800 MHz spectrometer (*Bruker*) equipped with a *TXI* cryoprobe, using an apo-*HpUreE* (200  $\mu$ M) solution. The same experimental setup was repeated for *HpUreE* samples pre-incubated with 200  $\mu$ M ZnSO<sub>4</sub> or NiSO<sub>4</sub>. The acquisition parameters for these NMR spectra are reported in Table 16 (see Section 2.6.4). Spectra were processed and analyzed using *TopSpin* (*Bruker*) and *iNMR* (<http://www.inmr.net>).

### **2.5.3. Light scattering measurements**

The oligomeric state and the hydrodynamic radius of *HpUreE* were determined using a combination of SEC (size exclusion chromatography), MALS (multiple angle light scattering) and QELS (quasi-elastic light scattering). In a typical experiment, a protein sample (100  $\mu$ L, 50  $\mu$ M)

was loaded onto a size-exclusion Superdex 200 HR 10/30 column (*GE Healthcare*), pre-equilibrated using 20 mM Tris HCl, pH 8.0, 150 mM NaCl, and eluted at a flow rate of 0.6 mL min<sup>-1</sup>. The same experiment was carried out in the presence of stoichiometric amount of zinc or nickel ions. The column was connected downstream to a multi-angle laser light (690.0 nm) scattering *DAWN EOS* photometer and to a *WyattQELS* device (*Wyatt Technology*). The concentration of the eluted protein was determined using a *Optilab DSP* refractive index detector (*Wyatt Technology*). The specific refractive index increment (dn/dc) for the proteins was taken as 0.185 ml g<sup>-1</sup> [32]. The value of 1.321 was used for the solvent refractive index. Molecular weights were determined from a *Zimm plot*. Data were recorded and processed using the *Astra 5.1.9* software (*Wyatt Technology*), following the manufacturer's indications. When the measurements were carried out in the presence of metal ions, stoichiometric amounts of ZnSO<sub>4</sub> or NiSO<sub>4</sub> were added to the protein samples before loading it onto the size-exclusion column, and the protein was eluted using the same buffer containing 20 μM ZnSO<sub>4</sub> or NiSO<sub>4</sub>. Analogous experimental set up was adopted for the characterization of H152A *HpUreE* mutant.

#### 2.5.4. ITC microcalorimetry experiments

In order to investigate the metal binding properties of *HpUreE*, the wild-type protein (10 μM) or its H102K and H152A mutants (10 μM) were titrated with 30 injections (10 μL each) of a solution containing 100 μM NiSO<sub>4</sub> or ZnSO<sub>4</sub>. Titration experiments were performed at 25 °C using a high-sensitivity *VP-ITC* microcalorimeter (*MicroCal LLC*, Northampton, MA). The proteins and the metal ions (from 100 mM stock solutions) were diluted using the same buffer (20 mM Tris HCl, pH 7.0, 150 mM NaCl) eluted from a size exclusion column utilized immediately before the ITC measurement to freshly purify the protein. The measuring cell contained 1.4093 ml of protein solution, and the reference cell was filled with deionized water. Before starting each experiment, the baseline stability was verified. A spacing of 400 - 600 sec 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 solution into the buffer alone, under identical conditions. Heats of dilution were negligible.

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 chi-square parameter  $\chi_v^2$  ( $\chi_v^2 = \chi^2/N$ , where N is the degrees of freedom,  $N = N_{idp} - N_{par}$ ;  $N_{idp}$  = number of points,  $N_{par}$  = 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 ( $\Delta H$ ), 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  $\Delta G = -RT \ln K_b$  ( $R = 1.9872 \text{ cal mol}^{-1} \text{ K}^{-1}$ ,  $T = 298 \text{ K}$ ) and  $\Delta G = \Delta H - T\Delta S$ . The dissociation constants and thermodynamic parameters provided in the article do not take into account possible events of

proton transfer linked to metal binding, or the presence, in solution, of complexes between the metal ions and the buffer. However, the values of the measured equilibrium constants compare well with those reported in the literature and determined using ITC or other methodologies such as equilibrium dialysis coupled to metal analysis, which, in principle, should also take into account similar effects. The values determined by ITC are therefore only used for comparison purposes.

## 2.6. Characterization of the *HpUreE-HpUreG* protein-protein interaction

### 2.6.1. Native PAGE electrophoresis

The formation of protein-protein complex between *HpUreE* and *HpUreG* was checked by native PAGE adapting the method of Laemmli [318], by using a *Mini-Protean II* apparatus (*BioRad*). Isolated *HpUreE* (5  $\mu$ M) and *HpUreG* (10  $\mu$ M) protein samples, as well as a solution containing equal amount of *HpUreE* (5  $\mu$ M) and *HpUreG* (10  $\mu$ M), were separated on 15% (w/v) acrylamide-bisacrylamide separating gels at 120 V for 45 minutes, using 25 mM Tris pH 8.3, containing 200 mM glycine as running buffer. Gel staining was carried out using the *SimplyBlue Safestain* (*Invitrogen*). The experiment was performed in the absence of any metal ions, as well as pre-incubating all protein samples with stoichiometric amount of ZnSO<sub>4</sub> and adding 1mM ZnSO<sub>4</sub> to running buffer. The composition of acrylamide gels is reported in Table 16.

Stacking gel (3%)		Separating gel (15%)	
Reagents	Volume	Reagents	Volume
Acrylamide-bisacrylamide 30:0.8	0.75 ml	Acrylamide-bisacrylamide 30:0.8	4 ml
Tris HCl 0.5 M, pH 6.8	1.25 ml	Tris HCl 1.5 M, pH 8.8	2 ml
ddH <sub>2</sub> O	3 ml	ddH <sub>2</sub> O	2 ml
Ammonium persulphate 10% (w/v)	30 $\mu$ l	Ammonium persulphate 10% (w/v)	45 $\mu$ l
TEMED	5 $\mu$ l	TEMED	10 $\mu$ l

**Table 16 - Composition of separating and stacking acrylamide gels**

### 2.6.2. Light scattering measurements

In order to explore the formation of a *HpUreE-HpUreG* protein complex, a solution containing *HpUreE* (50  $\mu$ M dimer) and *HpUreG* (100  $\mu$ M monomer) was analyzed in the absence and in the presence of 200  $\mu$ M NiSO<sub>4</sub> or 200  $\mu$ M ZnSO<sub>4</sub> under the same experimental conditions described in Section 2.5.3.

### 2.6.3. Thermal denaturation scanning

The thermal stability of dimeric *HpUreE* (5  $\mu$ M) and monomeric *HpUreG* (10  $\mu$ M) was monitored by applying circular dichroism in the absence of metal ions and in the presence of 10  $\mu$ M ZnSO<sub>4</sub>. The same experiment was conducted on a solution containing *HpUreE* (5  $\mu$ M dimer) and *HpUreG*

(10  $\mu\text{M}$  monomer). All proteins were buffered in 20 mM Tris HCl, pH 8.0, 150 mM NaCl. Three CD spectra were collected for each protein sample from 200 to 250 nm, at increasing values of temperature from 25°C to 90°C. For each temperature an equilibration time of 3 minutes was applied before collecting the data. Ellipticity at 218 and 222 nm were graphed as function on temperature and were normalized to fraction of unfolded protein using the equation (9):

$$\alpha = \frac{\theta_{obs} - \theta_N}{\theta_D - \theta_N} \quad (9)$$

where  $\alpha$  is the fraction of the protein denatured at each temperature,  $\theta_{obs}$  represents the measured ellipticity,  $\theta_N$  and  $\theta_D$  represent the ellipticity values for the fully folded and fully unfolded species at each temperature as calculated from the linear regression of the baselines preceding and following the transition region. Data were fitted using a two-state model, using the equation (10): [319]

$$\alpha = 1 - \frac{e^{\frac{\Delta H}{Rx}(\frac{T}{T_m} - 1)}}{1 + e^{\frac{\Delta H}{Rx}(\frac{T}{T_m} - 1)}} \quad (10)$$

where  $\Delta H$  is the enthalpy at the unfolding transition,  $T_m$  is the observed midpoint of the thermal transition,  $T$  is the denaturing temperature and  $R$  is the universal gas constant (1.987 cal K<sup>-1</sup> mol<sup>-1</sup>).

#### 2.6.4. NMR spectroscopy experiments

Uniformly <sup>15</sup>N-labeled *HpUreE* and *HpUreG* were produced with the same protocol used for the unlabeled proteins, using M9 auto-induction medium containing <sup>15</sup>NH<sub>4</sub>Cl as sole nitrogen source. Solutions containing <sup>15</sup>N-labeled *HpUreE* (200  $\mu\text{M}$ ) and unlabeled *HpUreG*, or <sup>15</sup>N-labeled *HpUreG* (300  $\mu\text{M}$ ) and unlabeled *HpUreE*, were prepared mixing one equivalent of the *HpUreE* dimer with two equivalents of the *HpUreG* monomer. <sup>1</sup>H-<sup>15</sup>NHSQC experiments and TROSY-HSQC experiments and spectra processing were performed as described in Section 2.5.2. The acquisition parameters for these NMR spectra are described in Table 17.

Experiment	Dimension of acquired data		Spectral width (p.p.m.)		Number of scans
	<sup>1</sup> H	<sup>15</sup> N	<sup>1</sup> H	<sup>15</sup> N	
<b><sup>15</sup>N <i>HpUreG</i></b>					
HSQC	1k	180	13	38	32
<b><sup>15</sup>N <i>HpUreG</i> + <i>HpUreE</i></b>					
HSQC	1k	128	13	38	128
TROSY-HSQC	1k	180	13	38	512
<b><sup>15</sup>N <i>HpUreE</i></b>					
HSQC	1k	128	14	40	32
TROSY-HSQC	1k	128	14	40	32
<b><sup>15</sup>N <i>HpUreE</i> + <i>HpUreG</i></b>					
TROSY-HSQC	1k	128	14	40	800

**Table 17 - Acquisition parameters for the NMR experiments**

### **2.6.5. ITC microcalorimetry experiments**

In order to determine the binding parameters of *HpUreE* to *HpUreG*, the latter protein (50  $\mu\text{M}$  monomer) was titrated with 30 injections (10  $\mu\text{L}$  each) of a solution containing 160  $\mu\text{M}$  *HpUreE* dimer in the same buffer. The effect on zinc on the protein complex stabilization was investigated performing  $\text{Zn}^{2+}$  titration onto the *HpUreE-HpUreG* complex (5  $\mu\text{M}$ ) generated in situ by mixing 5  $\mu\text{M}$  *HpUreE* and 10  $\mu\text{M}$  *HpUreG* monomer. In particular, 30 aliquots (10  $\mu\text{L}$  each) of a solution containing 70  $\mu\text{M}$   $\text{ZnSO}_4$  were injected into the protein complex solution. Identical set up was used for the related titration involving *HpUreE* (H102K, H152A) and *HpUreG* (C66A/H68A) mutants. Titration of  $\text{NiSO}_4$  onto the wild-type *HpUreE-HpUreG* complex was also performed, using the same experimental setup applied for  $\text{Zn}^{2+}$ -titration. All experiments were carried out using the same conditions, as well as data analysis and fitting procedures, described in Section 2.5.4.

### **2.6.6. Measurement of GTPase activity for *HpUreG* in complex with *HpUreE***

GTP-hydrolyzing activity of *HpUreG*, as well as *HpUreG* in complex with *HpUreE*, was measured using the colorimetric *Sensolyte MG phosphate assay Kit (Anaspec)*. The protein solution (10  $\mu\text{M}$ ) buffered in 20 mM Tris HCl, pH 7.0, containing 150 mM NaCl, was incubated at 37°C with 5 mM  $\text{MgCl}_2$ , 2 mM GTP, in the absence and in the presence of 10  $\mu\text{M}$   $\text{ZnSO}_4$ . Aliquots (30  $\mu\text{L}$ ) were removed at different incubation times (0 - 4 h) and added to 30  $\mu\text{L}$  of a 17.5% trichloroacetic acid/water solution and stored in ice until the measurement of the phosphate concentration determined colorimetrically. The experiment was also conducted using isolated *HpUreE*.

## **2.7. Characterization of the *HpUreE-HpHypA* protein-protein interaction**

### **2.7.1. ITC microcalorimetry experiments**

The interaction between *HpUreE* and *HpHypA* has been investigated by titrating the latter *HpUreE* dimer (19.5  $\mu\text{M}$  monomer) with 30 injections (10  $\mu\text{L}$  each) of a solution containing 320  $\mu\text{M}$  *HpHypA* monomer in the same buffer. Titrations were carried out using the same experimental conditions, as well as data analysis and fitting procedures, described in Section 2.5.4.

## Section B2 - Biocomputing procedures

### 2.8. Structural modeling of the *HpUreE-HpUreG* complex

#### 2.8.1. Sequence search and alignment of UreE proteins

Sequences of UreE proteins were searched as previously described [264], using sequence similarity criteria and the primary structure of *BpUreE* as template. The program *BLAST* (*Basic Local Alignment Search Tool*) [320, 321] available at [www.ncbi.nlm.nih.gov/BLAST](http://www.ncbi.nlm.nih.gov/BLAST) was utilized for the search. UreE sequences were retrieved from a non-redundant sum of different databases (*SwissProt*, *TrEMBL*, *TrEMBLNew*, *GenBank CDS*, *PDB*, *PIR*, *PRF*). Multiple sequence alignments were performed using the *ClustalW* program [322], available at [www.ebi.ac.uk/clustalw](http://www.ebi.ac.uk/clustalw), and alignment optimization was carried out using information deriving from secondary structure predictions provided by the program *JPRED* [323], available at [www.compbio.dundee.ac.uk/www-jpred](http://www.compbio.dundee.ac.uk/www-jpred).

#### 2.8.2. *HpUreE* structure prediction and homology modeling

The previously reported alignment of *BpUreE* and *HpUreE* [264] was used to calculate, using the *Modeller9v5* software [324], 50 structural models of the *HpUreE* dimer, imposing the structural identity of the two monomers. The structures of the *BpUreE* [262] and H144\**KaUreE* [263] dimers (PDB entries 1ear and 1gmw, respectively) were used as templates. The calculated model does not include the region containing the last 23 residues of *HpUreE*, due to the lack of suitable templates for the homology modelling. Indeed, both in the case of *BpUreE* and H144\**KaUreE*, this region was not observed in the crystal structures because of conformational disorder. The best model was selected on the basis of the lowest value of the *Modeller* objective function. The results of the *ProCheck* analysis [325] for the final model were fully satisfactory.

#### 2.8.3. *HpUreG* homology modeling

The structural model for dimeric *HpUreG* was calculated as described in [251].

#### 2.8.4. *HpUreE-HpUreG* molecular docking

The *RosettaDock* software [326] was used to calculate an initial complex between the model structure of dimeric *HpUreG* [251], and the central C-terminal domains of dimeric *HpUreE*. A search of 1.000 complexes was carried out by randomly translating and rotating the initial positions

of the interacting proteins. The complex with the best *RosettaDock* score was selected among all generated models for the subsequent refining run, carried out by applying 1000 times a perturbation to the starting structure. The C $\alpha$  trace of this complex was used, together with the crystal structures of *M. jannaschii* HypB, (PDB entry 1hf8) [253], *BpUreE* (PDB entry 1ear) [262] and H144\**KaUreE* (PDB entry 1gmw) [263] as templates to build a model of the *HpUreE-HpUreG* complex using the *Modeller9v5* software [324]. The alignment included a combination of the sequences of *HpUreE* with *BpUreE* [264], and of *HpUreG* with *MjHypB* [251]. The calculation, carried out imposing a structural identity of the two monomers of UreE and UreG, produced 200 structural models. The best model was selected on the basis of the lowest value of the *Modeller* objective function. The results of the *ProCheck* analysis [9] for the final model were fully satisfactory.

### **2.8.5. Calculation of electrostatic potential for the he *HpUreE-HpUreG* structural model**

The molecular (solvent-excluded) surfaces of dimeric *HpUreE* and *HpUreG* were calculated using the *UCSF Chimera* package [327]. All histidine residues were considered neutral. The electrostatic color-coding was generated using the *Delphi* software [328]. This program solves the linearized Poisson-Boltzmann equation to obtain the electrostatic potential in and around the protein, while taking the presence of solvent into account as a high dielectric continuum. The protein internal dielectric constant was set to 4 in all calculations, and the solvent dielectric constant was 80. The salt concentration was set to 150 mM NaCl, which corresponds to the physiological ionic strength.

## **2.9. Structural modeling of the *HpUreE-HpHypA* complex**

### **2.9.1. Sequence search and alignment of HypA proteins**

Sequences of HypA proteins were searched and aligned following the same procedure illustrated in Section 2.1.1.

### **2.9.2. *HpHypA* dimeric structure prediction and homology modeling**

The calculated sequence alignment of *TkHypA* and *HpHypA* was used to obtain using the *Modeller9v5* software [324], 50 structural models of the *HpHypA* dimer, imposing the structural identity of the two monomers. The dimeric crystal structure of the *TkHypA* [298] and the monomeric NMR structure of *HpHypA* [297] (PDB entries 3a44 and 2kdx, respectively) were used as templates. The best model was selected on the basis of the lowest value of the *Modeller* objective function. The results of the *ProCheck* analysis [325] for the final model were fully satisfactory.

### 2.9.3. *HpUreE*-*HpHypA* molecular docking

The *RosettaDock* protein-protein docking server [329], available on line at <http://rosettadock.graylab.jhu.edu/>, was used to calculate the two different *HpUreE*-*HpHypA* complexes between the model structure of dimeric *HpUreE*, and the experimental monomeric solution structure of *HpHypA* (PDB entry 2kdx) [297] or the calculated model structure of dimeric *HpHypA*, respectively. In the first case, two sequential docking process was performed, using the resulting [dimeric *HpUreE*-monomeric *HpHypA*] structure as input, together with the monomeric NMR structure of *HpHypA*. In this manner, a [dimeric *HpUreE*-dimeric *HpHypA*] structural model was generated, in order to respect the experimentally evaluated stoichiometry for the protein-protein assembly (see results on Section 3.4). The complexes with the best *RosettaDock* score was selected among all generated models. The results of the *ProCheck* analysis [9] for the finals model were fully satisfactory.

### 2.9.4. Calculation of electrostatic potential for the he *HpUreE*-*HpUreG* structural model

The molecular (solvent-excluded) surfaces of dimeric *HpUreE* and *HpHypA* were calculated as described in Section 2.8.5.

---

## **C - Results and Discussion**

---

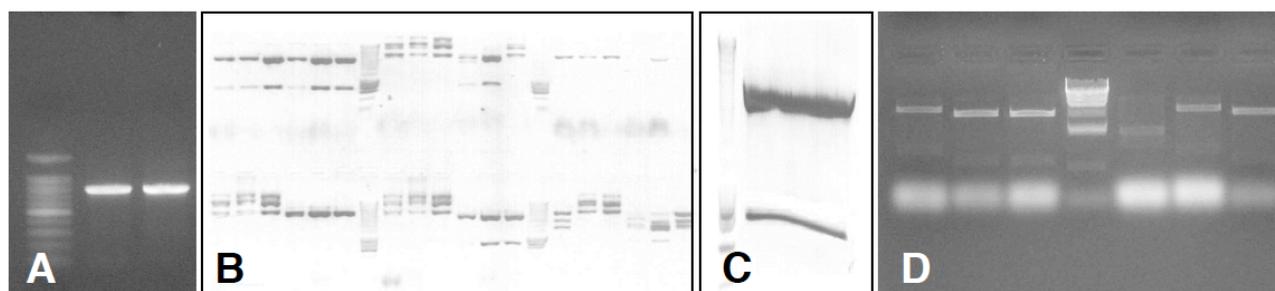


## Section C1 - Experimental results

### 3.1. Protein cloning, expression and purification

#### 3.1.1. Cloning and site-directed mutagenesis of *HpUreE*

In order to obtain a native purified full-length protein, two oligonucleotides were designed, with the required restriction sequences, and used to amplify by PCR reaction the *HpureE* gene from *H. pylori* strain G27 chromosomal DNA (Figure 45A). The PCR fragment was inserted into the cloning site of the *pGEM-T* vector by a T-A ligase reaction. The positive clones were identified by multiple restrictions of the purified ligase products (Figure 45B), and the plasmid DNA of one of them was purified in large amount. The construct was digested with *Bsp*HI and *Xho*I restriction enzymes, the first generating compatible ends with *Nco*I, to obtain the insert with the right overhang ends for the ligase (Figure 45C). This fragment was inserted between *Nco*I and *Xho*I sites of a *pET15b* plasmid, previously digested. In this manner, the His<sub>6</sub>-tag codifying sequence that precedes the multi-cloning site of *pET15b* is excised, allowing the transcription of the native *HpUreE* protein. Positive clones were identified by restriction screening (Figure 45D) and the sequence of *pET15b-HpureE* was confirmed by double-strand DNA sequencing. The sequence of the *HpureE* gene has been subsequently deposited in the NCBI database (code ABM16833).



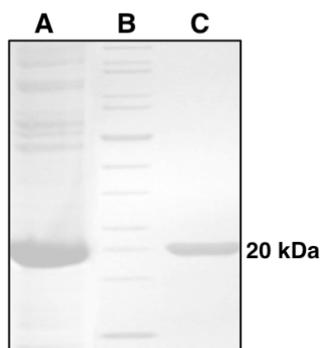
**Figure 45 - *HpureE* cloning into the *pET15b* vector**

(A) PCR products of *HpureE* with *Bsp*HI and *Xho*I sites inserted at the 5' and 3' ends. (B) Screening for the right *HpureE* insertion into the *pGEM-T* subcloning vector with multiple restrictions. (C) Digestion of *HpureE* gene from the *pGEM-T* vector, using *Bsp*HI/*Xho*I enzymes. (D) Screening for the right *HpureE* insertion into the *pET15b* expression vector through restrictions with *Bam*HI/*Pst*I, *Eco*RI/*Pst*I, *Bam*HI/*Xho*I enzymes.

The *HpUreE* mutants of crucial histidine residues involved in metal binding were produced by a PCR-based site-direct mutagenesis approach, using the *pET15b-HpureE* as DNA template (see Section 2.1.6). Positive clones were identified by restriction analysis (Figure 45D) and their sequences were confirmed by double-strand DNA sequencing.

### 3.1.2. *HpUreE* heterologous expression and purification

Based on the T7 expression system [330], large-scale expression of *HpUreE* was obtained transforming the *E. coli* BL21(DE3) strain with the *pET15b-HpureE* plasmid. Fractionation of the cellular extract showed that the over-produced protein accumulated in the soluble fraction. The observed molecular mass of the predominant polypeptide (20 kDa) is in complete agreement with the theoretical value calculated for the *HpUreE* monomer (19.408 kDa) (Figure 46, lane A).

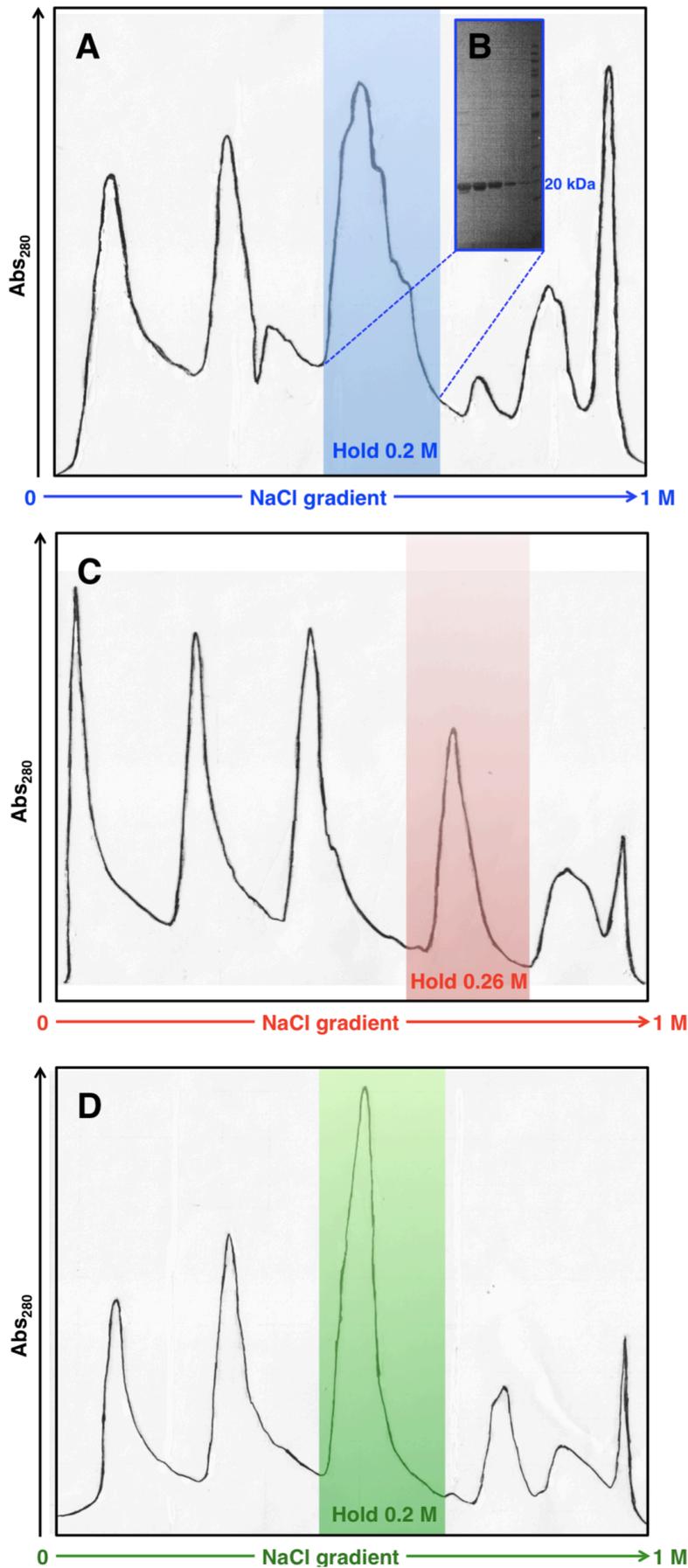


**Figure 46 - Expression and purification of *HpUreE***

SDS-PAGE of cell extracts of BL21(DE3) *E. coli* cells harboring *pET15b-HpureE* after growth in M9 auto-induction media (lane A). Molecular mass marker (lane B), and purified *HpUreE* (lane C) are also shown. The molecular weight of *HpUreE* (20 kDa) was verified by denaturing gel electrophoresis.

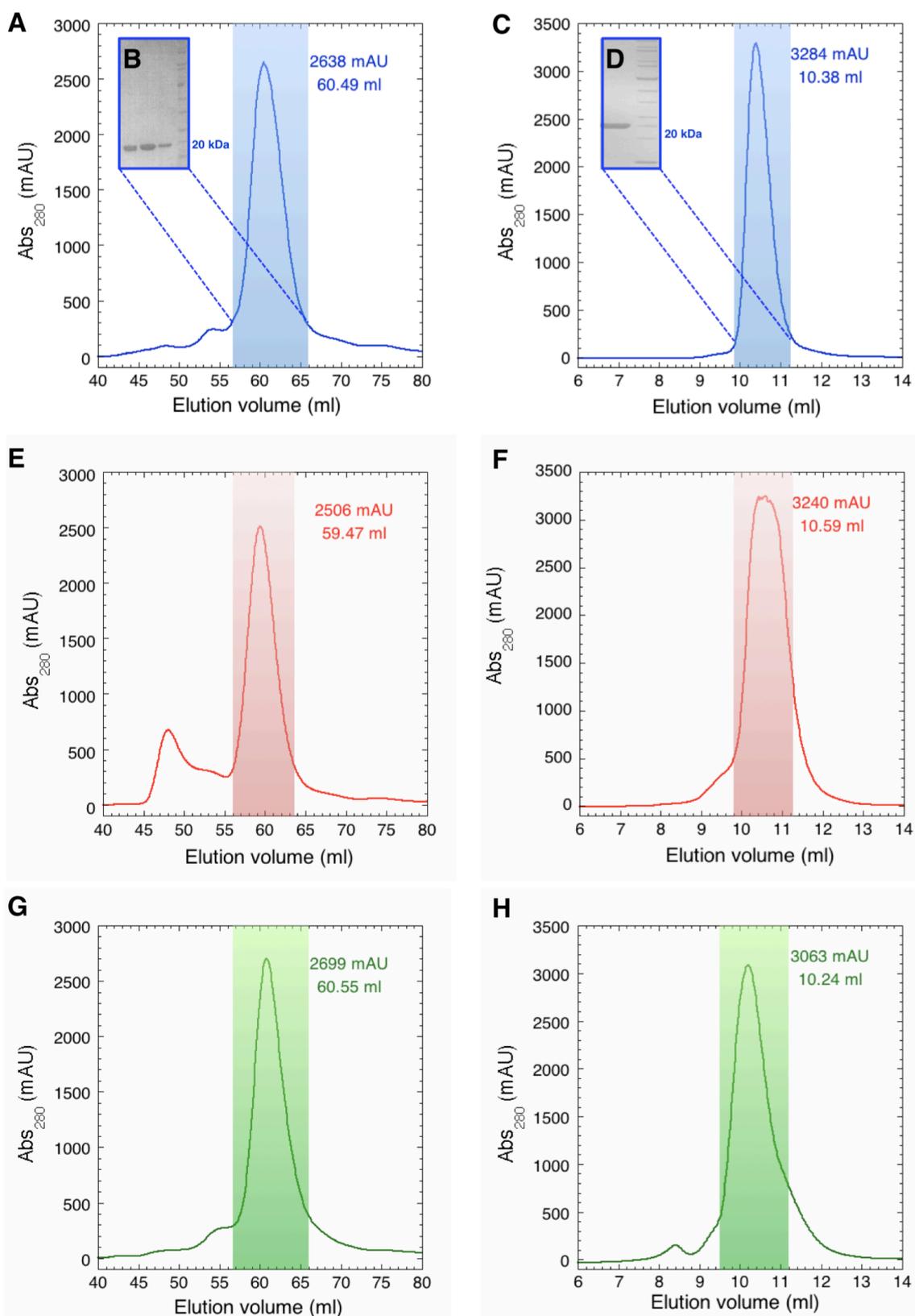
According with the predicted isoelectric point of *HpUreE* (8.56), calculated using the *ProtParam* tool [312], the solubilized cellular extract was purified by cation exchange chromatography using a linear ionic strength gradient ranging from 0 to 1M NaCl (Figure 47A). The isolated fractions eluted at about 0.2 M NaCl contained *HpUreE*, as verified by SDS-PAGE (Figure 47B). The soluble protein obtained was further purified by two size-exclusion chromatography steps (Figure 48A-D), achieving ~15 mg of native *HpUreE* per liter of culture, whose high purity grade (> 95%) has been checked by SDS-PAGE (Figure 46, lane D). The identity of the purified protein was demonstrated by tryptic digestion and ESI Q-TOF tandem mass spectrometry, which confirm the correspondence between the expected and the experimental protein sequence, through the sequencing of five digested fragments (DLNPLDFNVVDHVDLEWFETR, QGKDIAIR, LGLSQGDILFKEEK, HAALYYGESQFEFKTPFEKPTLALLEK, LTVSMPHSEPNFK) in complete agreement with the theoretical sequence of *HpUreE*. In addition, ICP-ES analysis confirmed the absence of any metal ions in all *HpUreE* preparations, as required in the following characterization steps.

All *HpUreE* mutants were purified analogously to the wild-type protein. However, the H102A *HpUreE* mutant could not be purified due to the formation of soluble aggregates that were not retained by ion exchange or size exclusion columns. Indeed, purification of H102Y *HpUreE* provided lower protein yield compared to the H102K mutant, so only the latter was isolated for characterization analysis. The elution profiles of the H102K and H152A *HpUreE* mutated versions were comparable with those obtained for the wild-type protein (Figure 47C,D and Figure 48E-G). The only exception is represented by the cation exchange elution profile of H102K *HpUreE* (Figure 48C), where appreciable differences were found compared to the wild-type. This protein mutant elutes at a higher ionic strength (0.26 M NaCl), probably due to the more positively charged surface generated by the substitution of the naturally present histidine His<sup>102</sup> with a lysine residue.



**Figure 47 - Cation exchange chromatography of *HpUreE***

The cation exchange chromatography was performed in native conditions using a *SP-Sepharose XK 16/10* column. Elution profiles monitored by the absorbance at 280 nm are displayed. The peaks corresponding to the almost pure wild-type (A) and mutated (C,D) *HpUreE* proteins are highlighted with colored boxes (wt *HpUreE*, blue; H102K *HpUreE*, red; H152A *HpUreE*, green). The presence of *HpUreE* in the selected peaks was revealed by SDS-PAGE, as reported in the insert (B). The salt concentration at which the *HpUreE* proteins eluted is reported.

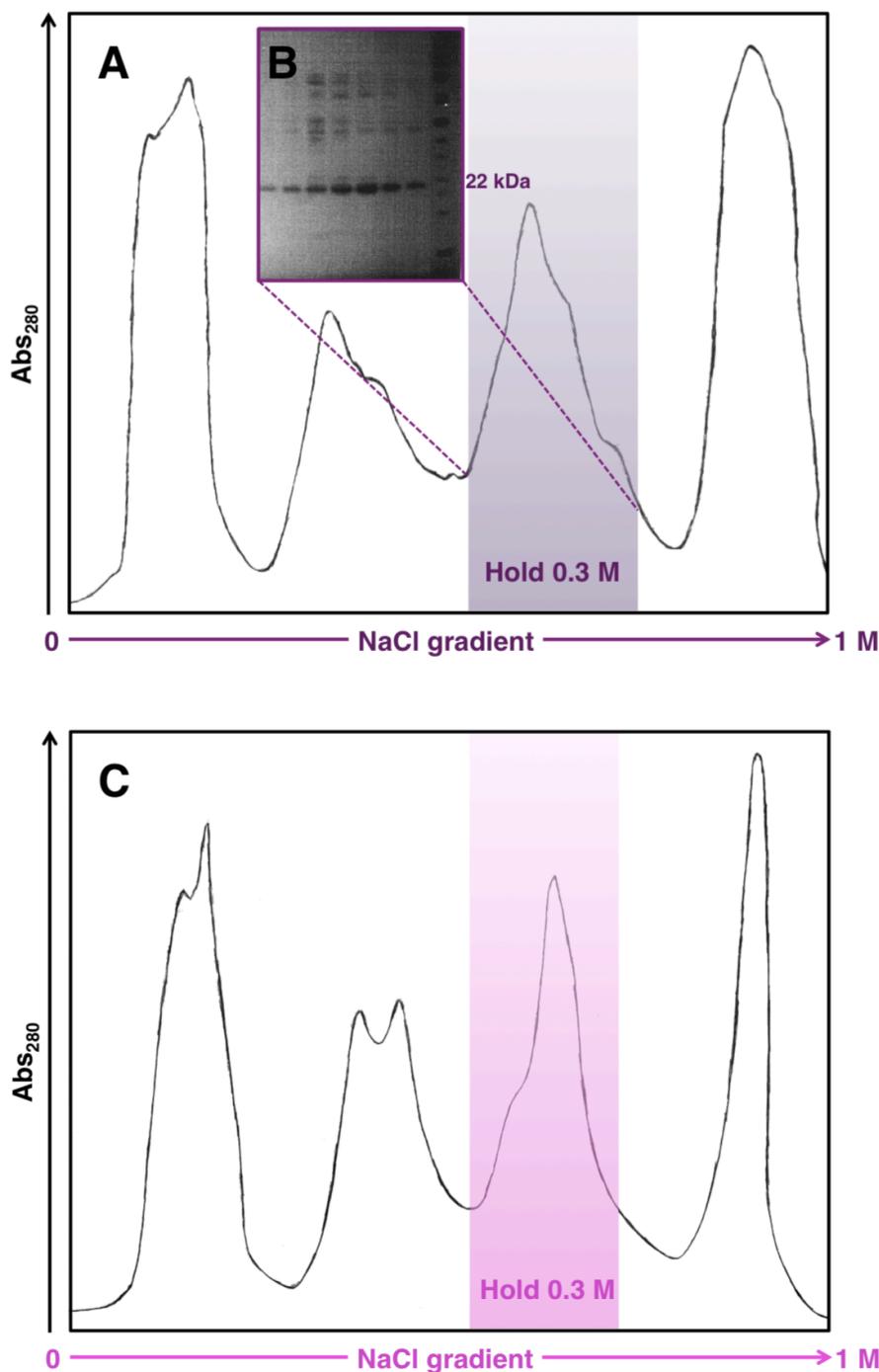


**Figure 48 - Size-exclusion chromatography of *HpUreE* proteins**

The separations were performed using *Superdex 75 XK 16/60* (left panels) and *Superdex 75 HR 10/30* (right panels) columns. Elution profiles monitored by the absorbance at 280 nm are displayed. The peaks corresponding to the pure wild-type (A,C) and mutated (E-H) *HpUreE* proteins are highlighted with colored boxes (wt *HpUreE*, blue; H102K *HpUreE*, red; H152A *HpUreE*, green). The presence of *HpUreE* in the selected peaks was revealed by SDS-PAGE, as reported in the inserts (B,D). For each peak corresponding to *HpUreE* proteins, elution volume and maximum absorbance value are reported.

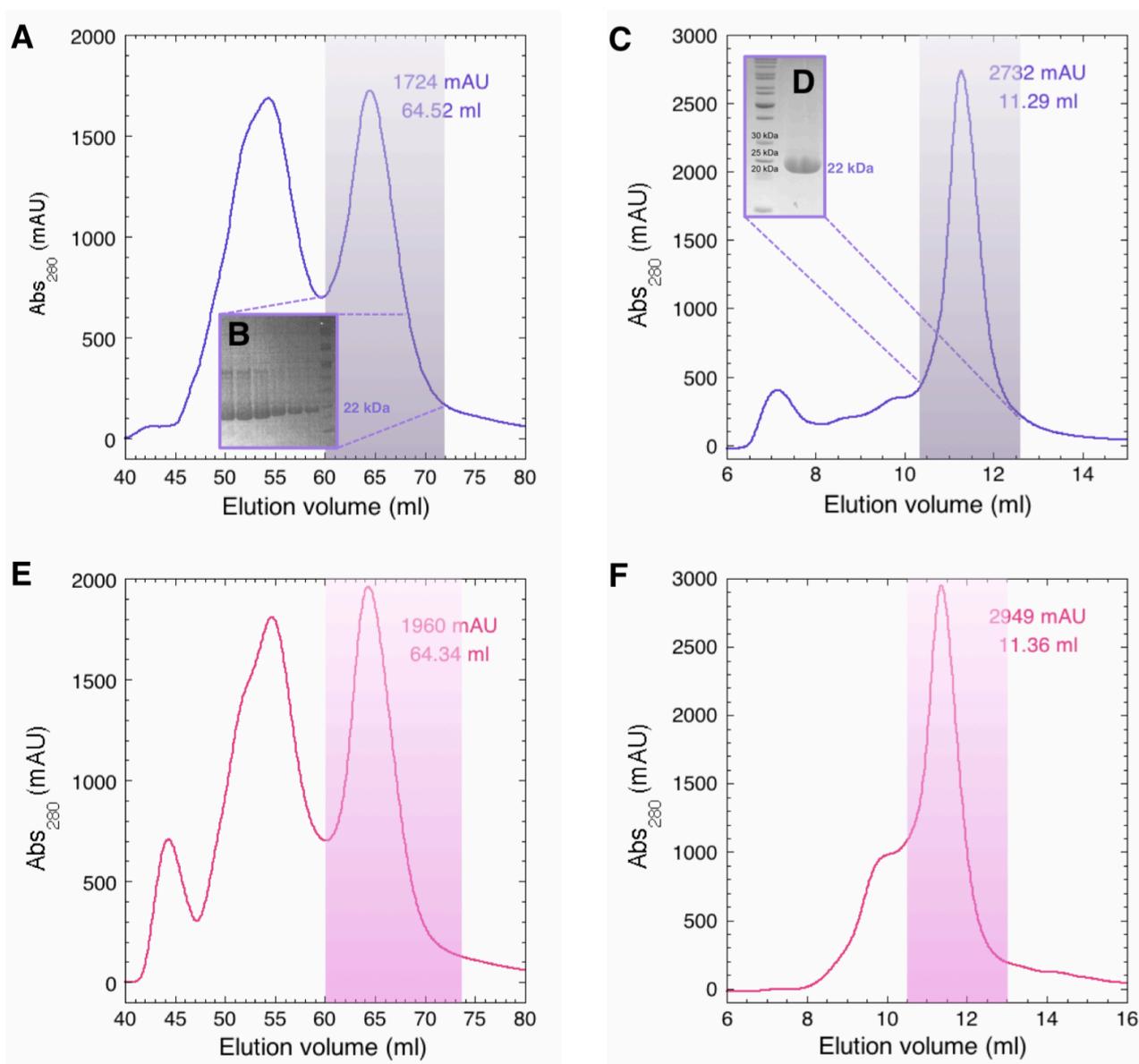
### 3.1.3. Heterologous expression and purification of *HpUreG*

Purification of wild-type *HpUreG* and C66A/H68A *HpUreG* mutant was achieved coupling an anion exchange chromatography (Figure 49) to two sequential size-exclusion passages (Figure 50), as recently described [251]. No differences between the elution profile of wild-type and mutated *HpUreG* were observed for all the purification steps.



**Figure 49 - Anion exchange chromatography of *HpUreG***

The anion exchange chromatography was performed in native conditions, using a *Q-Sepharose XK 26/10* column. Elution profiles monitored by the absorbance at 280 nm are displayed. The peaks corresponding to the fractions containing wild-type (A) and mutated (C) *HpUreG* proteins are highlighted with colored boxes (wt *HpUreG*, violet; C66A/H68A *HpUreG*). The presence of *HpUreG* in the selected peaks was revealed by SDS-PAGE, as reported in the insert (B). The salt concentration at which the *HpUreG* proteins eluted is reported.



**Figure 50 - Size-exclusion chromatography of *HpUreE* proteins**

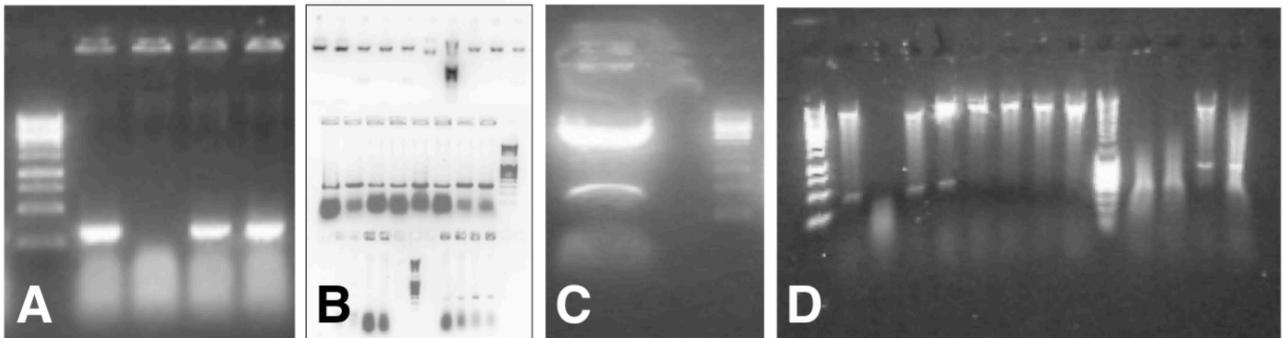
The separations were performed using *Superdex 75 XK 16/60* (left panels) and *Superdex 75 HR 10/30* (right panels) columns. Elution profiles monitored by the absorbance at 280 nm are displayed. The peaks corresponding to the pure or almost pure wild-type (A,C) and mutated (E,F) *HpUreG* proteins are highlighted with colored boxes (wt *HpUreG*, purple; C66A/H68A *HpUreG*, magenta). The presence of *HpUreG* in the selected peaks was revealed by SDS-PAGE, as reported in the inserts (B,D). For each peak corresponding to *HpUreG* proteins, elution volume and maximum absorbance value are reported.

### 3.1.4. Cloning of native, His<sub>6</sub>-tagged and His<sub>6</sub>-GB1-tagged *HpHypA*

Two oligonucleotides were designed, with the required restriction sequences (*Bam*HI and *Nde*I), in order to amplify the *HphypA* gene, using the *H. pylori* strain G27 genomic DNA as a template. The PCR fragment was inserted into the *pGEM-T* subcloning vector by a T-A ligase reaction. The positive clones were identified by multiple restrictions of the purified ligase products, and the plasmid DNA of one of them was purified in large amount. The PCR product was digested using

*Bam*HI and *Nde*I restriction enzymes and the gel-purified fragment was inserted into the *pET3a* and *pET15b* vectors, digested with the same endonucleases, in order to express the native and the His<sub>6</sub>-tagged *HpHypA*, respectively. The positive clones for both plasmid constructs were identified by multiple restrictions of the purified ligase products. For each of the two confirmed vectors, the plasmid DNA was purified and the sequence of the *HphypA* gene was confirmed by double-strand DNA sequencing.

The His<sub>6</sub>-GB1-tagged *HpHypA* protein was instead obtained by PCR amplification of the *HphypA* gene from the *H. pylori* G27 chromosomal DNA using another oligonucleotide pairs, containing the *Nco*I and *Bam*HI restriction sites (Figure 51A). The obtained fragment was ligated into the *pGEM-T* vector and the positive clones were identified by multiple restrictions of the purified ligase products (Figure 51B). The plasmid DNA of one positive clone was purified in large amount and digested with *Nco*I and *Bam*HI endonucleases. (Figure 51C). This fragment was then inserted in the *pETM-GB1* vector, previously digested with *Nco*I and *Bam*HI restriction enzymes. In this manner, the His<sub>6</sub>-GB1 codifying sequence of *pETM-GB1* is ligated in frame with the *HphypA* gene, allowing the transcription of the His<sub>6</sub>-GB1-*HpHypA* fusion protein. Positive clones were identified by restriction screening (Figure 51D) and the sequence of *pETM-GB1-HphypA* was confirmed by double-strand DNA sequencing.



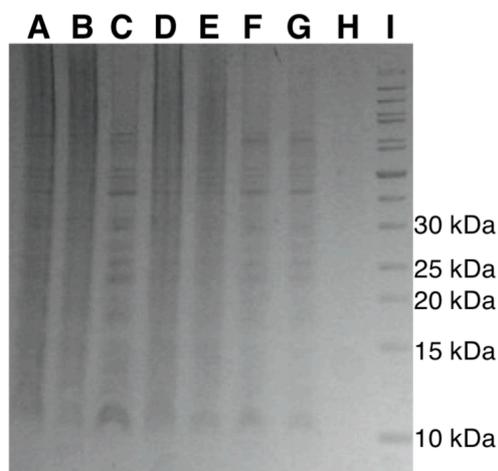
**Figure 51 - *HphypA* cloning into the *pETM-GB1* vector**

(A) PCR products of *HphypA* with *Nco*I and *Bam*HI sites inserted at the 5' and 3' ends. (B) Screening for the right *HphypA* insertion into the *pGEM-T* subcloning vector with multiple restrictions. (C) Digestion of *HphypA* gene from the *pGEM-T* vector, using *Nco*I/*Bam*HI enzymes. (D) Screening for the correct *HphypA* insertion into the *pETM-GB1* expression vector through restrictions with *Nco*I/*Bam*HI, *Xho*I/*Bam*HI, *Bgl*II/*Bam*HI enzymes.

### 3.1.5. Heterologous expression and purification of *HpHypA*

*E. coli* BL21(DE3) strain was transformed with the *pET3a-HphypA* and *pET15b-HphypA* plasmid, in order to over-produce the native and the His<sub>6</sub>-tagged *HpHypA* (13.2 kDa), respectively. The induction with IPTG does not provided the expression of the two expected polypeptides, as verified by SDS-PAGE comparing the total cell extract from the non-induced (Figure 52, lanes A,D) and the induced (Figure 52, lanes B,E) cells. Analogously, the soluble fractions of the cellular

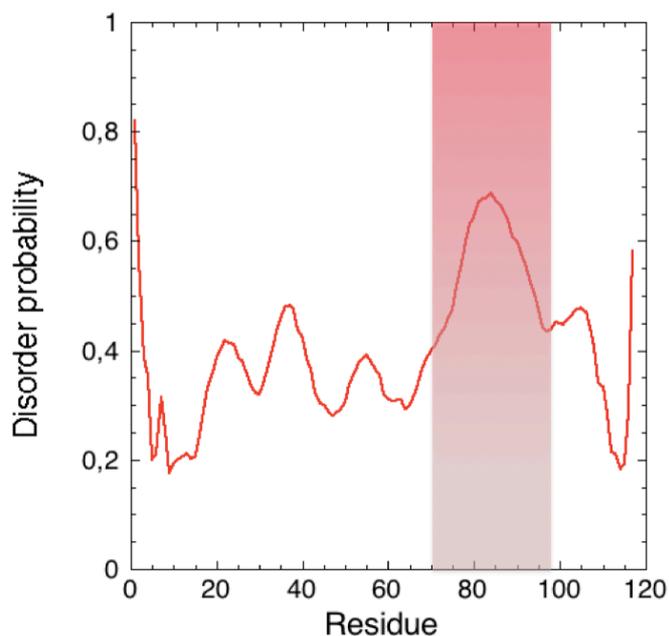
extract showed that *HpHypA* and His<sub>6</sub>-*HpHypA* were not detectable in the stained SDS-PAGE (Figure 52, lanes C,F). For His<sub>6</sub>-*HpHypA*, the soluble cell extract was also loaded on a Ni-affinity column but any polypeptide corresponding to the attended molecular weight of *HpHypA* were observed, either in the unbound (Figure 52, lane G) and retained fraction (Figure 52, lane H).



**Figure 52 - Expression of native and His<sub>6</sub>-tagged *HpHypA* proteins**

SDS-PAGE of *BL21(DE3)* *E. coli* cells harboring *pET3a-HphypA* (lane A-C) or *pET15b-HphypA* (lane D-H): total cell extracts prior to induction (lane A,D); total cell extracts after 3 h of induction with IPTG (lane B,E); soluble cell extract (lane C,F). Flow-through of the Ni-affinity column (lane G), retained fraction (lane H) and molecular mass marker (lane I) are also shown.

The lack of detectable *HpHypA* expression is probably due to its instability: the protein is classified as unstable on the basis of its instability index, calculated from the amino acid sequence using the *Protparam* tool [312]. Moreover, a C-terminal loop region (70-95) in sequence of *HpHypA* is predicted to be highly disordered (see Figure 53); this flexible portion involves all the four cysteine residues (Cys<sup>74</sup>, Cys<sup>77</sup>, Cys<sup>91</sup>, Cys<sup>94</sup>) required for the formation of a structural Zn<sup>2+</sup>-binding site, that is essential for correct protein folding [297].

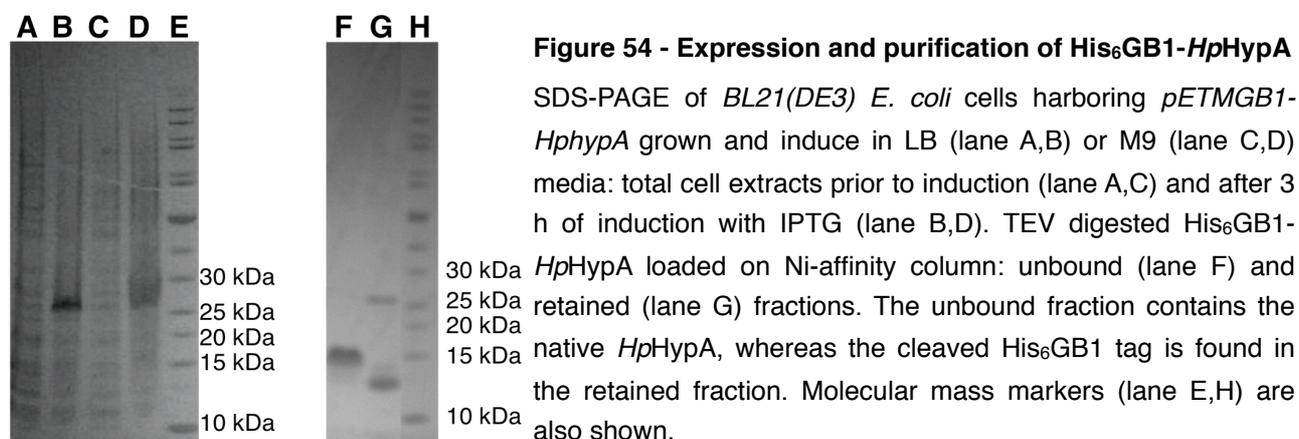


**Figure 53 - Disorder prediction of *HpHypA* sequence**

The prediction was performed by the *DisEMBL* intrinsic protein disorder prediction tool, available on line at <http://dis.embl.de/cgiDict.py>, using the amino acid sequence of *HpHypA* as input. The highly disorder region is indicated by a red box.

In order to improve expression levels and protein stability, the overproduction of *HpHypA* fused to a N-terminal His<sub>6</sub>GB1 domain (9 kDa) was achieved by transforming *E. coli BL21(DE3)* host with the *pETM-GB1-HphypA* plasmid. The induction with IPTG provided abundant polypeptide, with an apparent molecular mass of 25 kDa, as main result of the cellular protein synthesis (Figure 54,

lanes A). This protein was absent from the non-induced cells (Figure 54, lanes B). Interestingly, differences in the expression level were observed when cells were grown and induced in minimal M9 media (Figure 54, lanes C,D). This could be explained considering the increased protein instability due to the lack of zinc ions in the culture medium, supporting the physiological structural role of this metal in HypA proteins.



The His<sub>6</sub>GB1-tagged protein was firstly purified using a Ni<sup>2+</sup>-based affinity chromatography. This passage was highly specific for the His<sub>6</sub>GB1-*HpHypA*, which was completely retained in the column, while the majority of other proteins did not. The reversible Ni<sup>2+</sup>-protein bond was broken during the elution step, when the protein was recovered at very good level of purity. The eluted protein was then digested with TEV protease, in order to cleave the His<sub>6</sub>GB1. The product of proteolytic cleavage was further purified with a second passage on Ni<sup>2+</sup>-affinity column. In this way the obtained *HpHypA* protein (14 kDa) is not retained (Figure 54, lane F), whereas the His<sub>6</sub>GB1 tag and the uncleaved His<sub>6</sub>GB1-*HpHypA* are bound to the nickel resin (Figure 54, lane G).

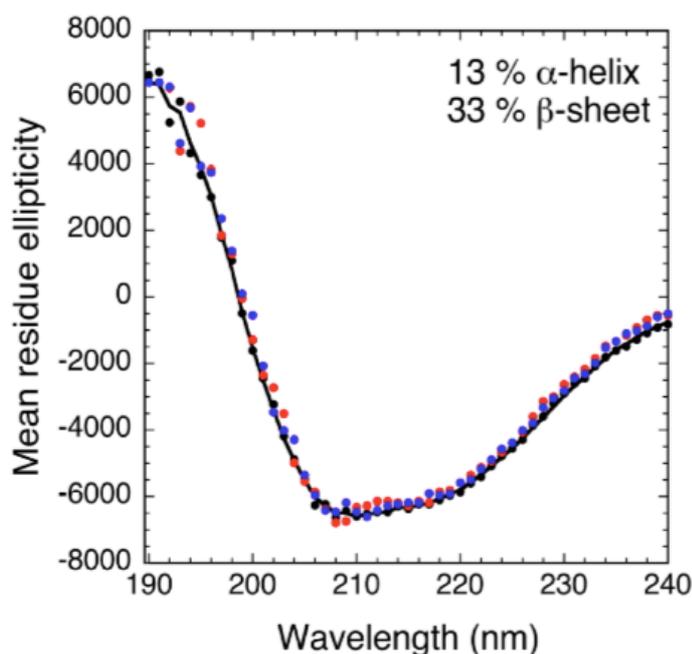
From a first examination, the use of the GB1 domain appeared a suitable method to solve the expression and instability problems of *HpHypA*, because of the possibility of obtaining high protein yield in a soluble stable form. Furthermore, the high affinity and specificity of the His<sub>6</sub>-tag for the Ni-affinity resin, gave also the opportunity to easily isolate and concentrate the protein.

## 3.2. Biochemical and structural characterization of *HpUreE*

### 3.2.1. Determination of *HpUreE* structural properties

In order to evaluate the secondary structure composition of the apo-protein in solution, CD (circular dichroism) spectroscopy was performed. The far-UV CD spectrum of apo-*HpUreE* (Figure 55, black dots) clearly shows the presence of both  $\alpha$ -helices and  $\beta$ -strands, as suggest by negative deflections around 218 nm and 208 nm and a positive peak at 190 nm. The CD spectrum was quantitatively analyzed and the best fit (NRMSD = 0.029) estimated a secondary structure

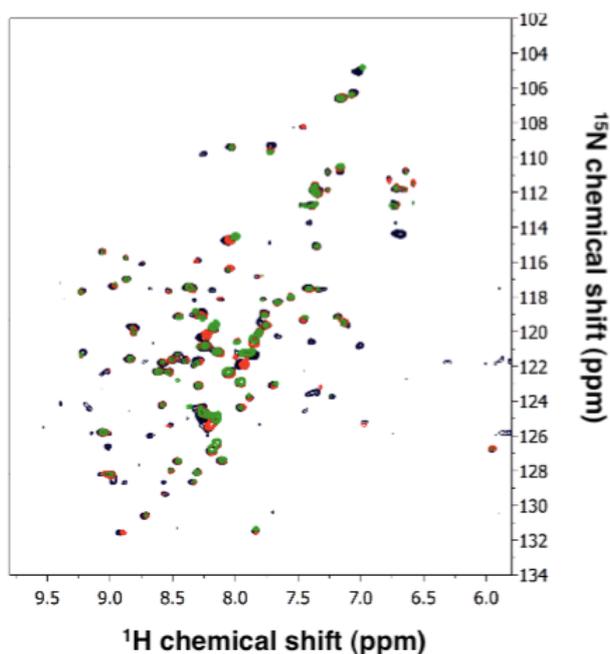
composition of 13%  $\alpha$ -helix, 33%  $\beta$ -strand, 23% turns and 30% random coil for *HpUreE*. This data slightly differs from a prediction of the secondary structure elements of *HpUreE* based on the *JPRED* algorithm [323], which indicates 21%  $\alpha$ -helix and 23%  $\beta$ -strand content, with the remaining 56% constituting turns or random coil conformations. However, the experimentally observed composition is similar to that calculated using the DSSP program [331] for the crystallographic structures of  $Zn^{2+}$ -bound *BpUreE* [262] (18%  $\alpha$ -helix, 37%  $\beta$ -sheet) and of the  $Cu^{2+}$ -bound H144\**KaUreE* [263] (13%  $\alpha$ -helix, 25%  $\beta$ -sheet). These evidences suggest that a similar fold is attained by several different UreE proteins, as previously proposed on the basis of modeling studies [264].



**Figure 55 - Far UV CD spectra of *HpUreE***

The experimental data of the apo- (black),  $Ni^{2+}$ -bound (blue) and  $Zn^{2+}$ -bound (red) protein are shown as dots. The solid line represents the best fit calculated for apo-*HpUreE* using the *CDSSTR* program available at the *Dichroweb* server. The calculated structural composition is reported.

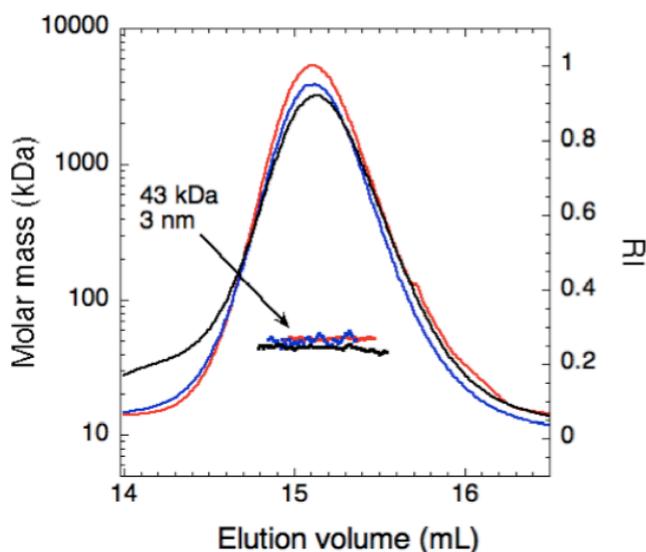
In order to investigate the conformational properties of the protein in solution, NMR spectroscopy was also applied. The  $^1H$  chemical shift spreading observed in the TROSY-HSQC NMR spectrum (Figure 56) ranges from 6.5 to 10 ppm, as observed in the  $^1H$ - $^{15}N$  HSQC NMR spectrum of *BpUreE* [332], suggesting similar extent of fold for the two proteins. The better quality of the  $^1H$ - $^{15}N$  TROSY-HSQC (Figure 56) with respect to the simple  $^1H$ - $^{15}N$  HSQC experiment further supports the presence of a dimeric form at 0.1 - 0.3 mM concentration. The symmetric architectural arrangement of the two monomers is revealed by the number of observed peptide NH peaks in the NMR spectrum: about 120 unique peaks, out of the expected 170 residues per monomer, can be observed, with missing signals likely including the C-terminal 30 residues predicted to be unstructured using *JPRED* tool [323]. In particular, in the case of glycine residues, seven Gly are present in the sequence of *HpUreE* and the same number of peaks is observed in the  $^{15}N$  100 - 110 ppm range typical for Gly NH signals (Figure 56).



**Figure 56 -  $^1\text{H}$ - $^{15}\text{N}$  TROSY-HSQC spectra of *HpUreE***

The spectra of the apo- (black),  $\text{Ni}^{2+}$ -bound (green) and  $\text{Zn}^{2+}$ -bound (red) protein were acquired at 800 MHz and 298 K.

Furthermore, the molecular mass ( $M_r$ ) and the hydrodynamic radius ( $R_h$ ) of apo-*HpUreE* in solution were determined using a combination of size exclusion chromatography and light scattering MALS/QELS (Figure 57, black line). The elution profile and the light scattering data show that *HpUreE* is a dimer in solution with  $M_r = 43.1 \pm 4.8$  kDa, in agreement with the calculated theoretical mass (39 kDa), and  $R_h = 3.0 \pm 1.4$  nm.



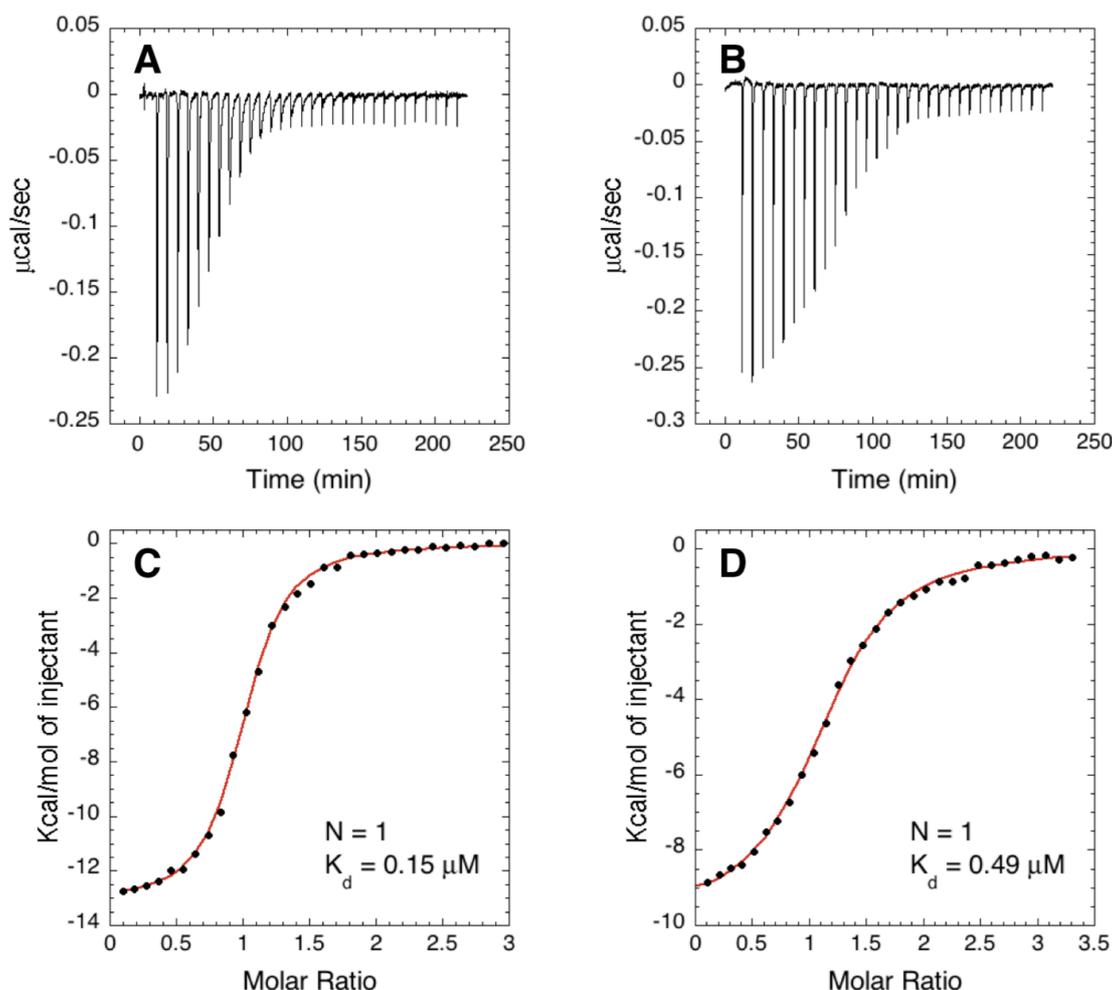
**Figure 57 - MALS/QELS analysis of *HpUreE***

Plot of the molar mass distribution for *HpUreE*. The solid lines indicate the *Superdex S-200* size-exclusion elution profile monitored by the refractive index detector, and the dots are the weight-averaged molecular masses for each slice, measured every second. Data obtained in the absence of metal ions (black) and in the presence of  $\text{Ni}^{2+}$  (blue) or  $\text{Zn}^{2+}$  (red) are reported. The average molecular mass and the hydrodynamic radius of apo-*HpUreE* are indicated.

This is consistent with all available crystallographic structural information on UreE proteins [262, 263] as well as with previous evidence collected on *HpUreE* based on size-exclusion chromatography criteria [257]. Moreover, the light scattering measurements exclude the possibility that oligomers of the apo-protein are formed in solution for concentrations lower than 50  $\mu\text{M}$ , as instead previously proposed for H144\**KaUreE* [259].

### 3.2.2. *HpUreE* metal binding properties

$\text{Ni}^{2+}$  is generally considered to be the physiological cofactor of UreE, and the understanding of the structural features of  $\text{Ni}^{2+}$  binding is therefore important to clarify the role of this chaperone *in vivo*. Moreover, in several recent instances, interplay between  $\text{Ni}^{2+}$  and  $\text{Zn}^{2+}$  has been observed and proposed to be functionally important in regulating cell trafficking of metal ions [248, 249, 251]. In particular,  $\text{Zn}^{2+}$  is involved in the dimerization of *HpUreG*, a process that plays a potential regulatory role in the urease active site assembly [251]. In the past, a dissociation constant of ca. 1  $\mu\text{M}$  and a 1:1 stoichiometry for the  $\text{Ni}^{2+}$  binding to the *HpUreE* dimer have been established by equilibrium dialysis experiments [257]. However, these measurements were carried out at pH 8.25 in an apparently non-buffered solution containing only NaCl, no thermodynamic parameters for the metal binding event were determined, nor the binding affinity for  $\text{Zn}^{2+}$  was measured [257]. For this reasons, the  $\text{Ni}^{2+}$  binding to *HpUreE* was investigated using microcalorimetry and a comparison between  $\text{Ni}^{2+}$  and  $\text{Zn}^{2+}$  binding was performed.

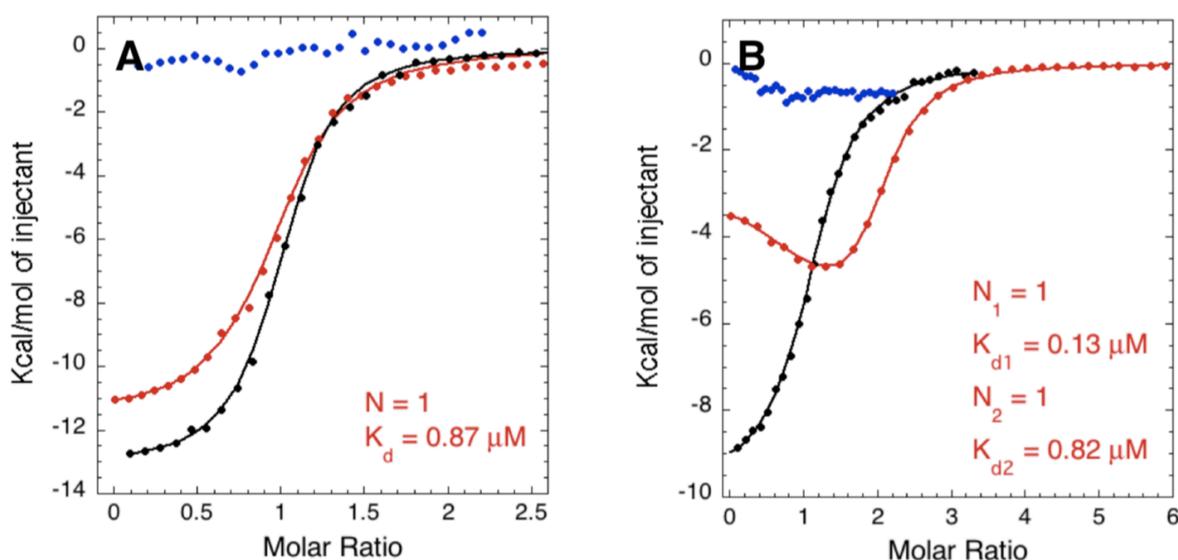


**Figure 58 - ITC data of  $\text{NiSO}_4$  and  $\text{ZnSO}_4$  binding to *HpUreE***

Upper panels display the plots of titration data corresponding to the thermal effect of  $30 \times 10 \mu\text{L}$  injections of  $\text{Ni}^{2+}$  (A) and  $\text{Zn}^{2+}$  (B) respectively, onto the *HpUreE* protein solution. Bottom panels show the best fits of the integrated data, represented as solid red lines, obtained by a non-linear least squares procedure of  $\text{Ni}^{2+}$  (C) and  $\text{Zn}^{2+}$  (D) titrations. The calculated numbers of sites and dissociation constants are indicated.

The ITC measurements were carried out by adding  $\text{Ni}^{2+}$  or  $\text{Zn}^{2+}$  to the apo-protein in a buffered solution at pH 7.0, and the occurrence of a binding event was revealed by the presence of exothermic peaks that followed each addition (Figure 58A,B). Fits of the integrated heat data (Figure 58C,D) were carried out using the simplest model, which entails a single binding event, and yielded a stoichiometry of one equivalent of  $\text{Ni}^{2+}$  or  $\text{Zn}^{2+}$  bound to the *HpUreE* dimer. Dissociation constants  $K_d(\text{Ni}) = 0.15 \pm 0.01 \mu\text{M}$  and  $K_d(\text{Zn}) = 0.49 \pm 0.01 \mu\text{M}$  were calculated for  $\text{Ni}^{2+}$  and  $\text{Zn}^{2+}$  binding, respectively. In both cases, these processes are driven by favorable enthalpic factors ( $\Delta H(\text{Ni}) = -13 \pm 1 \text{ kcal mol}^{-1}$ ,  $\Delta H(\text{Zn}) = -10 \pm 1 \text{ kcal mol}^{-1}$ ) that compensate the negative entropic values ( $\Delta S(\text{Ni}) = -13 \text{ cal mol}^{-1} \text{ K}^{-1}$ ,  $\Delta S(\text{Zn}) = -4 \text{ cal mol}^{-1} \text{ K}^{-1}$ ) calculated from the fit. The values of the dissociation constants measured for *HpUreE* are comparable to those established by ITC for the binding of  $\text{Ni}^{2+}$  and  $\text{Zn}^{2+}$  to *BpUreE* and H144\**KaUreE* [259], and by equilibrium dialysis for the binding of  $\text{Ni}^{2+}$  to *BpUreE* [260], of  $\text{Ni}^{2+}$  and  $\text{Zn}^{2+}$  to H144\**KaUreE* [258, 261] and of  $\text{Ni}^{2+}$  to *HpUreE* [257]. All these values are consistent with a role of intracellular metal ion transport associated with UreE proteins.

In the structure of  $\text{Zn}^{2+}$ -*BpUreE* [262] and  $\text{Cu}^{2+}$ -H144\**KaUreE* [263] the metal ions are bound to the surface of the protein using the conserved His<sup>100</sup> and His<sup>96</sup> residues, respectively. In order to firmly establish the role of the corresponding His<sup>102</sup> in the binding of  $\text{Ni}^{2+}$  and  $\text{Zn}^{2+}$  to *HpUreE*, the H102K mutant was obtained by site-directed mutagenesis. ITC titrations of H102K *HpUreE* with  $\text{Ni}^{2+}$  and  $\text{Zn}^{2+}$ , performed under identical conditions as for the wild-type protein, proved the absence of a binding event (Figure 59A,B, blue dots), confirming the key role of this residue in metal binding to *HpUreE*.



**Figure 59 - ITC data of  $\text{NiSO}_4$  and  $\text{ZnSO}_4$  binding to the *HpUreE* mutants**

Best fits of the integrated data, represented as solid lines, obtained by a non-linear least squares procedure of  $\text{Ni}^{2+}$  (A) and  $\text{Zn}^{2+}$  (A) titrations to *HpUreE* proteins. In each panel data obtained for the H120K (blue) and H152A (red) mutants are compared to those obtained for wild-type *HpUreE*. The calculated numbers of sites and dissociation constants of H152A *HpUreE* are reported.

However, the 1:1 metal ion binding stoichiometry established for *HpUreE* differs from previous data obtained for *KaUreE* and *BpUreE*, which indicated a 2:1 stoichiometry. *KaUreE* additionally binds three Ni<sup>2+</sup> ions to a His-rich tail containing 10 histidines among the last fifteen residues, absent both in *HpUreE* and in *BpUreE* [259]. In *BpUreE* and *KaUreE* a conserved HXH motif is present in this protein region: in particular, in *BpUreE* the HQH motif is located at the end of the sequence, while in *KaUreE* several possible HXH concatenated motifs constitute the His-rich tail (HGHHHAHHDHHAHSH). On the other hand, in *HpUreE* a single histidine (His<sup>152</sup>) is observed in the C-terminal tail. On these basis, a possible reason for the differences in the stoichiometry of the Ni<sup>2+</sup> binding among UreE proteins, might reside in the different sequence motifs of histidine residues found at their C-terminal tails, also suggesting a specialized role in metal ion storage and/or delivery for the C-terminal portion.

These observations prompted to investigate whether His<sup>152</sup> is involved in metal ion binding to *HpUreE*, performing ITC titrations on H152A *HpUreE* mutant. In the case of Ni<sup>2+</sup> binding to H152A mutant, the integrated heat data were fitted using a single site model (Figure 59A, red line) and yielded values for the dissociation constant ( $K_d(\text{Ni}) = 0.87 \pm 0.01 \mu\text{M}$ ), reaction enthalpy ( $\Delta H(\text{Ni}) = -12 \pm 1 \text{ kcal mol}^{-1}$ ) and reaction entropy ( $\Delta S(\text{Ni}) = -11 \text{ cal mol}^{-1} \text{ K}^{-1}$ ) that are consistent with those obtained for the wild-type protein (Figure 58C,D). These data indicate that the His<sup>152</sup> residue is not involved in binding the Ni<sup>2+</sup> ion in *HpUreE*, suggesting that this residue is not essential for the nickel delivery function.

On the other hand, the titration of H152A *HpUreE* with Zn<sup>2+</sup> showed clear differences in the binding mode as compared to the wild-type protein (Figure 59B). Best fits of the integrated heat data could be obtained using a model involving not one, as in the case of wild-type *HpUreE*, but two independent binding events, yielding dissociation constants  $K_{d1}(\text{Zn}) = 0.13 \pm 0.02 \mu\text{M}$  and  $K_{d2}(\text{Zn}) = 0.82 \pm 0.01 \mu\text{M}$ . Both events are driven by favorable enthalpic ( $\Delta H_1(\text{Zn}) = -4 \pm 1 \text{ kcal mol}^{-1}$ ,  $\Delta H_2(\text{Zn}) = -7 \pm 1 \text{ kcal mol}^{-1}$ ) and entropic ( $\Delta S_1(\text{Zn}) = 19 \text{ cal mol}^{-1} \text{ K}^{-1}$ ,  $\Delta S_2(\text{Zn}) = 6 \text{ cal mol}^{-1} \text{ K}^{-1}$ ) factors. The observation of an additional Zn<sup>2+</sup> binding site upon replacement of a histidine with a non-coordinating residue like alanine, could be, at first sight, bewildering. This apparent incongruity, resulting from our experimental data, can be explained by taking into consideration the peculiarity of the protein region where the mutation is carried out. This is a long extended and flexible stretch whose conformation or relative orientation with respect to the rest of the protein could change as a consequence of point mutations. Therefore, it is possible that, while His<sup>152</sup> is involved in Zn<sup>2+</sup> binding by isolated wild type *HpUreE*, the resulting conformation of the flexible C-terminal arm masks an additional binding site, which becomes accessible upon mutation of this residue.

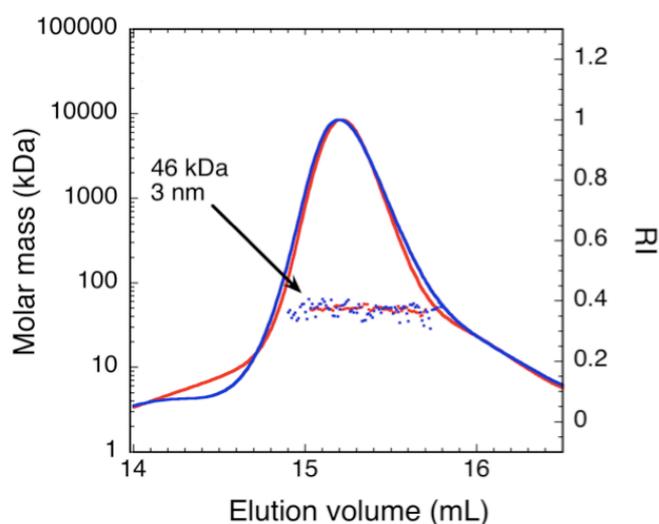
A different binding mode for the two metal ions to the wild type protein is also supported by a comparison of the TROSY-HSQC spectra of *HpUreE* in the apo-form with the same spectra of the Zn<sup>2+</sup>- and Ni<sup>2+</sup>-bound forms (see Figure 56). Residues changing their chemical shifts upon nickel

addition are also affected (and at the same extent) by the presence of zinc. In addition, a few more peaks change their position in the presence of  $Zn^{2+}$ , consistently with the involvement of a larger number of residues in the zinc binding event.

### 3.2.3. Influence of metal binding on *HpUreE* structure

The influence of metal binding on the quaternary structure of *HpUreE*, in the range of concentrations used in the microcalorimetric metal binding studies (10-20  $\mu$ M), was investigated using a combination of MALS/QELS light scattering. The values measured for  $Ni^{2+}$ -*HpUreE* ( $M_r = 45.7 \pm 5.1$  kDa,  $R_h = 3.4 \pm 1.5$  nm) and  $Zn^{2+}$ -*HpUreE* ( $M_r = 46.4 \pm 5.2$  kDa,  $R_h = 3.3 \pm 1.5$  nm) are similar to those established for the apo-protein (Figure 57), demonstrating that the metal-bound protein is a dimer, independently of the presence of bound metal ions. This is consistent also with CD spectroscopy measurements carried out in the presence of  $NiSO_4$  or  $ZnSO_4$  (Figure 55). No appreciable shifts from data acquired in the absence of metals were detected for both  $Ni^{2+}$ - and  $Zn^{2+}$ -bound samples, excluding a metal-driven conformational rearrangement of *HpUreE*.

Furthermore, the occurrence of oligomerization events for the H152A mutant in the presence of  $Zn^{2+}$  or  $Ni^{2+}$  was excluded, as demonstrated using light scattering experiments (Figure 60).



**Figure 60 - MALS/QELS analysis of H152A *HpUreE***

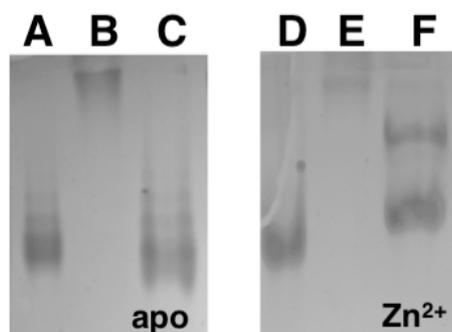
Plot of the molar mass distribution for H152A *HpUreE* in the presence of  $Ni^{2+}$  (red) or  $Zn^{2+}$  (blue). The solid lines indicate the *Superdex S-200* size-exclusion elution profile monitored by the refractive index detector, and the dots are the weight-averaged molecular masses for each slice, measured every second. The average molecular mass and the hydrodynamic radius of the two sample are indicated.

## 3.3. Characterization of the *HpUreE*-*HpUreG* interaction

### 3.3.1. Preliminary characterization of the *HpUreE*-*HpUreG* interaction

The available experimental evidences indicate that UreE and UreG form a functional complex *in vivo* [242, 245, 283]. In order to observe and characterize at preliminary stage this interaction *in vitro*, a native gel electrophoresis experiment was carried out running a solution containing *HpUreE* and *HpUreG*, as well as the two isolated proteins as controls. No distinguishable differences

respect to control lanes were observed in the absence of metal ions (Figure 61, lanes A-C), suggesting that no PPIs occur in apo- conditions. However, the same experiment performed in the presence of an excess of  $Zn^{2+}$  ions leads to the observation of a band with an intermediate molecular mass (Figure 61, lanes D-F) that could be ascribed to the formation of  $Zn^{2+}$ -mediated *HpUreE-HpUreG* assembly. In this regard, it is also possible to notice that *HpUreE* displays a peculiar migration profile, as a consequence of its intrinsically positive charge that obstructs the permeation in the gel; this phenomenon was not observed in the  $Zn^{2+}$ -*HpUreE-HpUreG* sample (Figure 61, lane F), supporting the hypothesis of a complex formation.

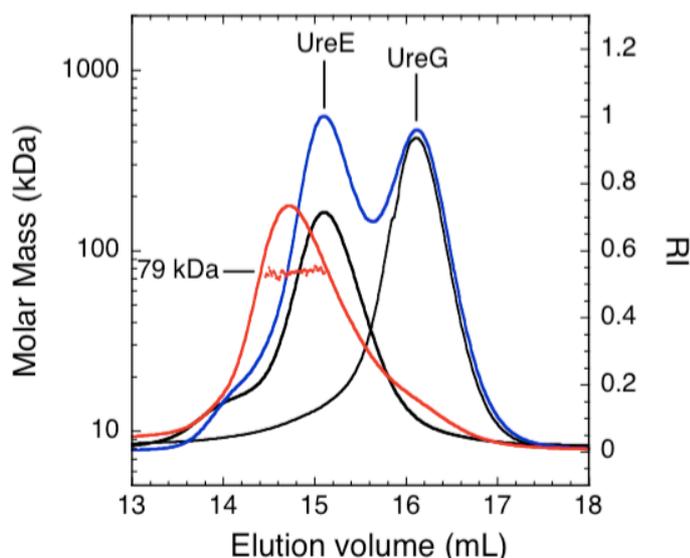


**Figure 61 - Preliminary characterization of the *HpUreE-HpUreG* interaction**

Native PAGE, in the absence (left panel) or in the presence (right panel) of  $Zn^{2+}$  ions, of *HpUreG* (lanes A,D), *HpUreE* (lanes B,E) and a solution containing the two proteins in equal amount (lanes C,F). The asterisk indicates the band ascribed to the formation of the *HpUreE-HpUreG* complex.

### 3.3.2. Elucidation of the *HpUreE-HpUreG* complex formation

A more rigorous characterization of the interaction between *HpUreE* and *HpUreG* was achieved using a MALS/QELS approach: a solution containing equimolar amounts of the two purified apo-proteins was analyzed by size exclusion chromatography and light scattering (Figure 62, blue line). The result of this experiment indicates that the dimeric *HpUreE* elutes as a species separated from *HpUreG*, the latter being present in the monomeric state, as recently reported [251].

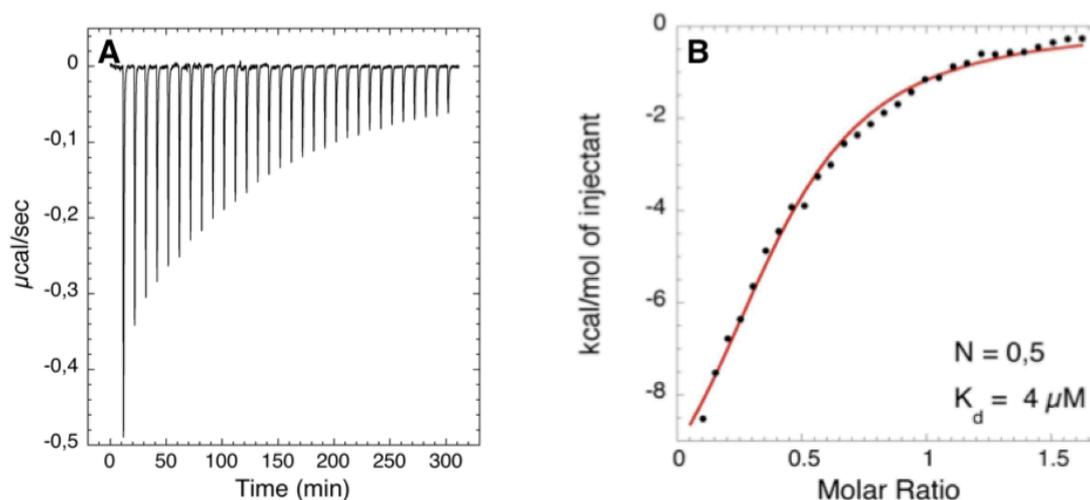


**Figure 62 - MALS/QELS data of the *HpUreE-HpUreG* complex**

The solid lines indicate the *Superdex S-200* size-exclusion elution profile monitored by the refractive index detector of a solution containing *HpUreE* (one equivalent of dimer) and *HpUreG* (two equivalents of monomer), in the absence (blue line) or in the presence (red line) of two equivalents of  $Zn^{2+}$ . Data of apo-*HpUreE* (black thin line) and apo-*HpUreG* (black thick line) are shown as references. The red dots are the weight-averaged molecular masses for each slice, measured every second.

Considering that the experimental setup used for the SEC-MALS-QELS measurement represents non-equilibrium conditions, the interaction was monitored more quantitatively using ITC.

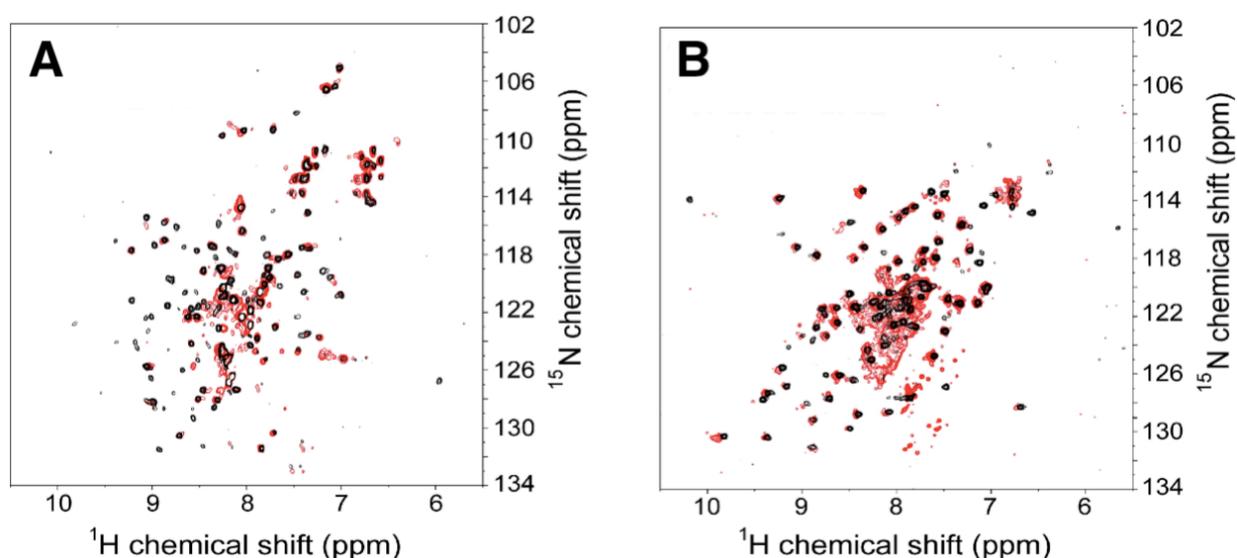
When a solution of *HpUreG* was titrated with a solution of *HpUreE* in the same buffer, clear exothermic peaks were observed (Figure 63A) which, after integration, revealed a curve (Figure 63B) that could be fitted using a single binding event model. The stoichiometry of the interaction suggests that two monomers of *HpUreG* bind to a single dimer of *HpUreE*, forming a *HpUreE-HpUreG* complex having a dissociation constant  $K_d = 4.0 \pm 0.3 \mu\text{M}$ ,  $\Delta H = -12.5 \pm 0.9 \text{ kcal mol}^{-1}$ , and  $\Delta S = -17.5 \text{ cal mol}^{-1} \text{ K}^{-1}$ .



**Figure 63 - ITC titration of dimeric *HpUreE* on monomeric *HpUreG***

Panel A displays the raw ITC data showing the thermal effect of 30 x 10  $\mu\text{L}$  injections of 160  $\mu\text{M}$  *HpUreE* dimer binding to 50  $\mu\text{M}$  *HpUreG* monomer. Panel B displays the best fit of the integrated data, represented as a solid line, obtained by a non-linear least squares procedure. The calculated numbers of sites and dissociation constant are indicated.

The formation of the *HpUreE-HpUreG* complex was also monitored using NMR spectroscopy. The TROSY-HSQC spectrum of the solution containing one equivalent of  $^{15}\text{N}$ -*HpUreE* dimer and two equivalents of unlabeled *HpUreG* monomer differs from that of *HpUreE* in the absence of *HpUreG* (Figure 64A). A general broadening of the peaks is observed, while a number of them experience small chemical shift changes ( $\Delta\delta_{\text{av}} \leq 0.1 \text{ ppm}$ ) consistent with the formation of a complex between the two proteins, as observed by ITC. The number of Gly resonances does not increase upon complex formation, suggesting maintenance of the *UreE* homo-dimeric symmetry. Addition of one equivalent of unlabeled *HpUreE* dimer to a solution of  $^{15}\text{N}$ -*HpUreG* causes broadening beyond detection for most of the backbone amide signals observed in a regular  $^1\text{H}$ - $^{15}\text{N}$  HSQC spectrum as compared to the same spectrum of *HpUreG* [251]. Resonances could be recovered by recording a  $^1\text{H}$ - $^{15}\text{N}$  TROSY-HSQC spectrum (Figure 64B), an effect that can be explained with the large molecular mass (ca. 80 kDa) of the *HpUreE-HpUreG* complex. Small chemical shift changes on selected resonances of backbone amides are observed ( $\Delta\delta_{\text{av}} \leq 0.05 \text{ ppm}$ ) upon complex formation. An assignment of the resonances, beyond the scope of the present study, would provide information on the surface contact areas.



**Figure 64 - NMR spectral changes induced by the interaction between *HpUreE* and *HpUreG***

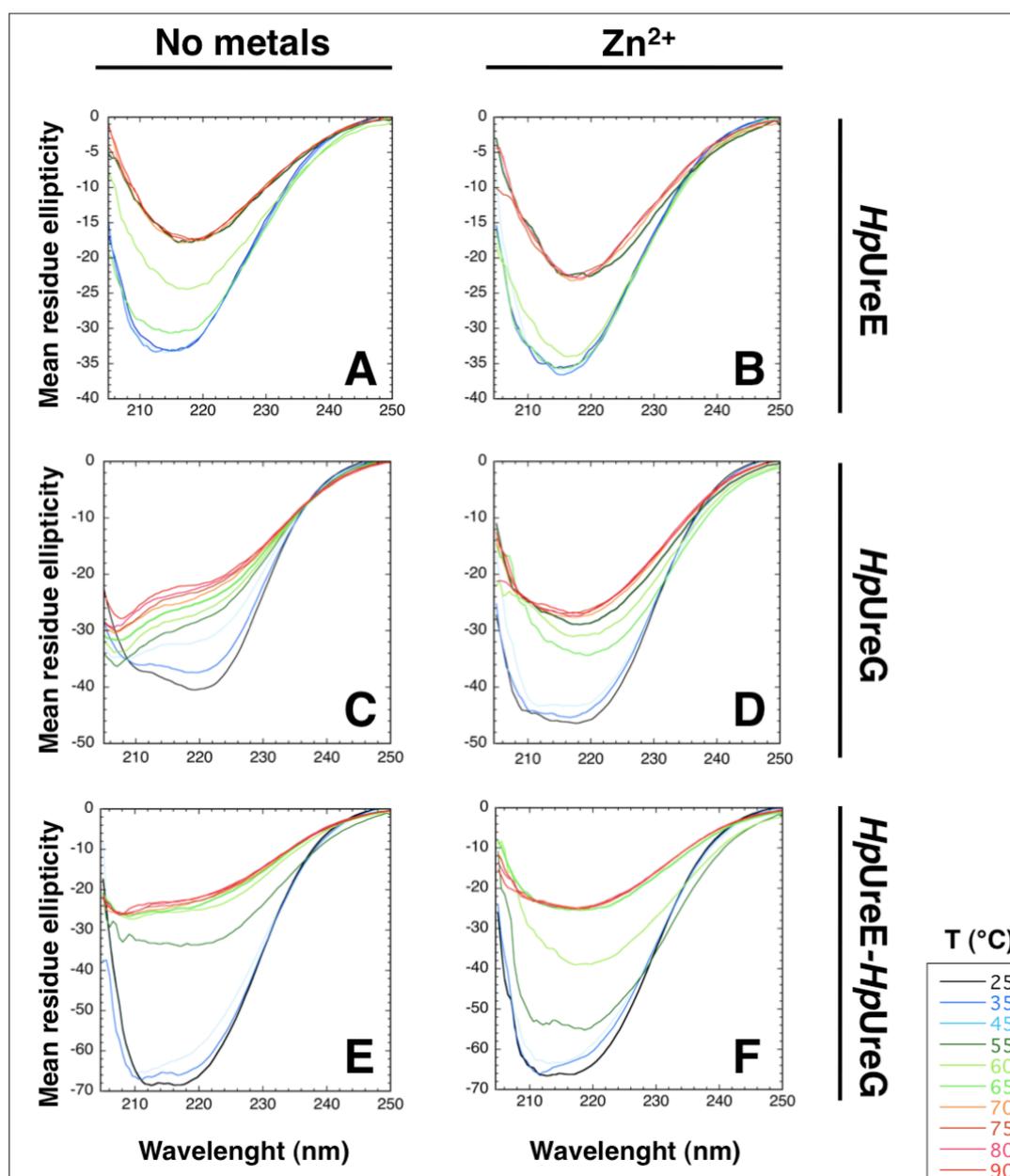
(A)  $^1\text{H}$ - $^{15}\text{N}$  TROSY-HSQC spectrum of the  $^{15}\text{N}$ -*HpUreE* dimer (black) superimposed with the  $^1\text{H}$ - $^{15}\text{N}$  TROSY-HSQC spectrum (red) recorded after addition of unlabeled *HpUreG*. (B)  $^1\text{H}$ - $^{15}\text{N}$  HSQC spectrum of  $^{15}\text{N}$ -*HpUreG* (black) superimposed with the  $^1\text{H}$ - $^{15}\text{N}$  TROSY-HSQC spectrum (red) recorded after addition of unlabeled *HpUreE*. Solutions of the two proteins were prepared mixing one equivalent of *HpUreE* dimer with two equivalents of *HpUreG* monomer. All spectra were acquired at 800 MHz and 298 K.

### 3.3.3. The role of $\text{Zn}^{2+}$ in the stabilization of the *HpUreE*-*HpUreG* complex

The observed 1:2 stoichiometry of the *HpUreE*-*HpUreG* complex, coupled to the previously reported dimerization of *HpUreG* selectively induced by the binding of one equivalent of  $\text{Zn}^{2+}$  per protein dimer [251], leads to the investigation of the role of this metal ions in the stabilization of the protein assembly by using MALS/QELS light scattering. The experiment was carried out as previously described, analyzing a solution containing dimeric *HpUreE* and monomeric *HpUreG* in a 1:2 ratio, but in the presence of  $\text{Zn}^{2+}$ . In this case, the elution peaks corresponding to the separate *HpUreE* and *HpUreG* proteins completely disappeared, while a new unique peak, with  $M_r = 79.4 \pm 2$  kDa and  $R_h = 5.8 \pm 0.1$  nm, was concomitantly observed (Figure 62, red line). This result indicates that the interaction between *HpUreE* and *HpUreG* is specifically stabilized by  $\text{Zn}^{2+}$ . On the basis of the theoretical masses of the *HpUreE* dimer (39 kDa) and of the *HpUreG* monomer (22 kDa), the mass of the new species formed in the presence of  $\text{Zn}^{2+}$  is fully consistent with the 1:2 stoichiometry established by ITC. The same experiment carried out in the presence of  $\text{Ni}^{2+}$  does not lead to the formation of a *HpUreE*-*HpUreG* complex: dimeric *HpUreE* elutes as a species separated from *HpUreG*, as observed in the absence of metal ions (Figure 62, blue line).

The role of zinc ions in the stabilization of the *HpUreE*-*HpUreG* was further demonstrated by thermal denaturation scanning, monitoring the protein unfolding by CD spectroscopy. CD spectra of isolated *HpUreE* and *HpUreG* proteins, as well as a *HpUreE*-*HpUreG* complex solution, were collected at different temperatures (from 25°C to 90°C), in the absence (Figure 65A,C,E) or in the

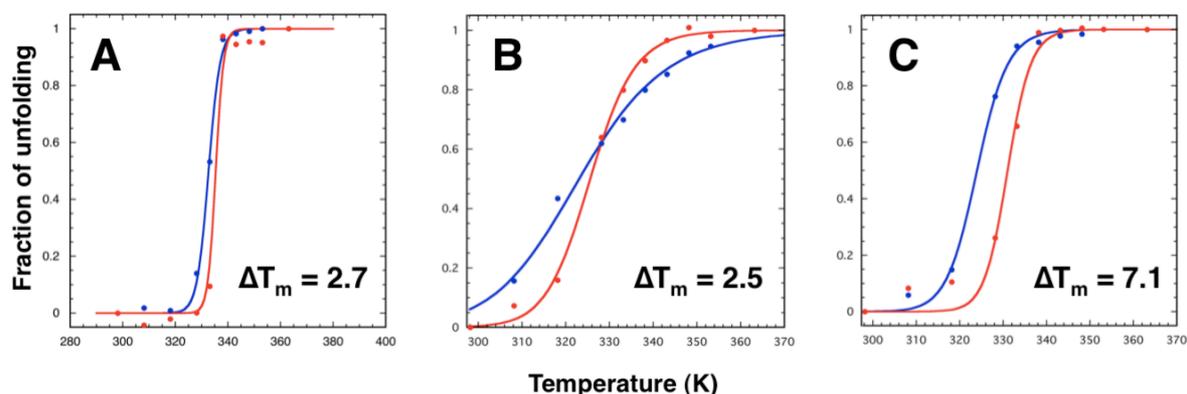
presence (Figure 65B,D,F) of stoichiometric amount of  $Zn^{2+}$ . The fraction of protein unfolding was then calculated from the ellipticity values at 222 nm.



**Figure 65 - Thermal denaturation scanning monitored by CD spectroscopy**

The CD spectra acquired at increasing temperatures (25–90°C, see the color code for the temperature in the side box) in the absence (left panels) and in the presence (right panels) of zinc ions, are shown for each protein samples (*HpUreE*, upper plots; *HpUreG*, central plots; *HpUreE-HpUreG* complex, bottom plots).

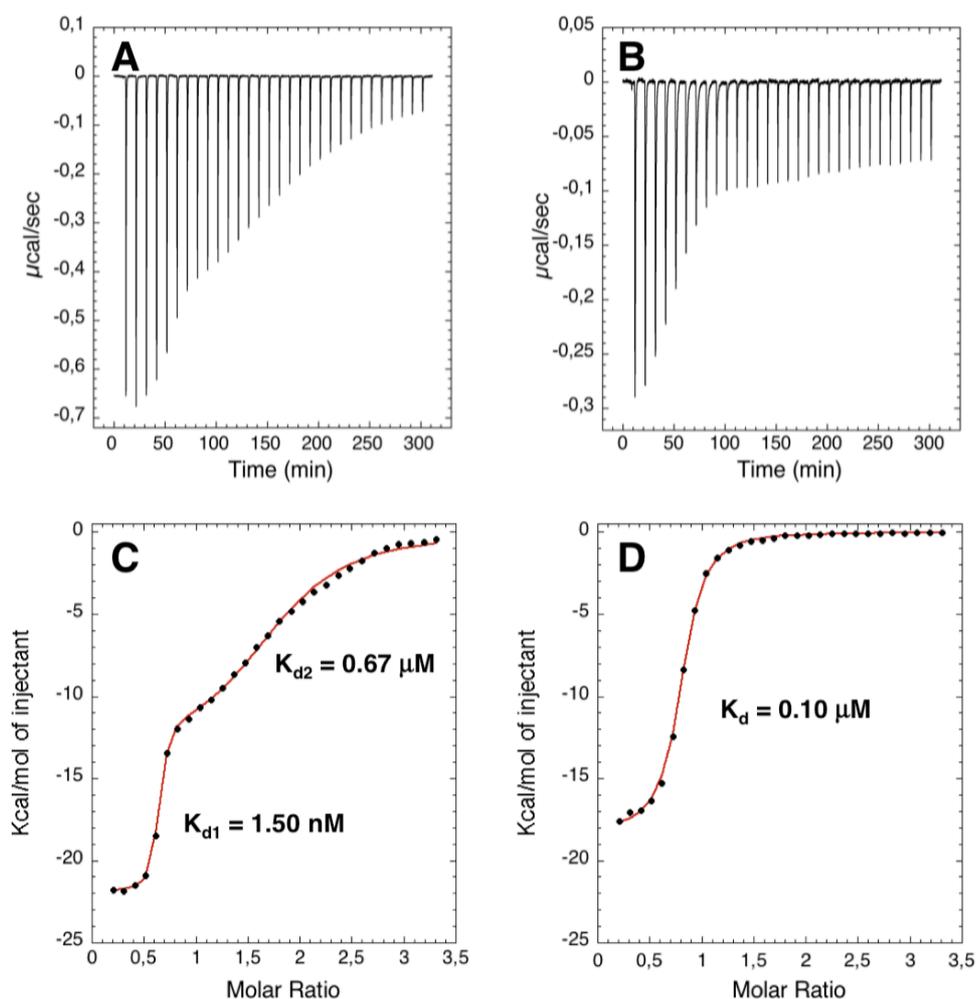
The  $Zn^{2+}$ -mediated stabilization of the *HpUreE-HpUreG* interaction is supported by the increased value of melting temperature ( $T_m$ ) for the protein assembly in the presence of zinc, compared to those observed in the absence of metal ions (Figure 66C). On the other hand, this effect is missing, or at least not appreciable, for the two isolated proteins (Figure 66A,B). This suggest that the  $Zn^{2+}$ -binding event, required for stabilizing the protein-protein interaction, would involve both proteins, probably occurring at the interface of the *HpUreE-HpUreG* complex.



**Figure 66 - Thermal denaturation scanning of *HpUreE-HpUreG* complex**

Plots of fraction of unfolded *HpUreE* (A), *HpUreG* (B) and *HpUreE-HpUreG* (C), in the absence (blue) and in the presence (red) of zinc ions, calculated from CD values at 222 nm acquired at different temperatures. Values of  $\Delta T_m$  (difference of melting temperatures between the apo- and the  $Zn^{2+}$ -bound state) are reported.

In order to experimentally verify the presence of this site, calorimetric titrations of  $Zn^{2+}$  onto a solution containing a preformed *HpUreE-HpUreG* complex (obtained in situ by mixing the two proteins with a 1:2 stoichiometry) were performed (Figure 66A,C).

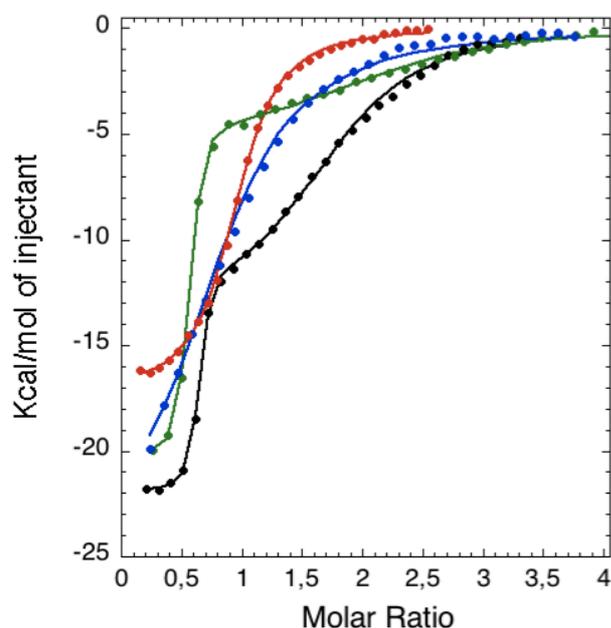


**Figure 67 - ITC data of  $ZnSO_4$  and  $NiSO_4$  binding to the *HpUreE-HpUreG* preformed complex**

Raw ITC data (upper panels) and best fit of the integrated data (bottom panels) of the titration of *HpUreE-HpUreG* complex with  $ZnSO_4$  (A,C) and  $NiSO_4$  (B,D). Dissociation constant values are indicated.

The curve obtained using the wild type proteins reveals indeed an event of binding, characterized by  $K_d = 1.5 \pm 0.3$  nM, which is ca. 2 - 3 orders of magnitude tighter than those observed for isolated *HpUreE* or *HpUreG*. This event is distinct from an additional following binding step with  $K_d = 0.67 \pm 0.05$   $\mu$ M. Therefore, the *HpUreE-HpUreG* complex binds two  $Zn^{2+}$  ions, in a high-affinity and a low-affinity site. Analogously,  $Ni^{2+}$  was titrated on a preformed *HpUreE-HpUreG* complex. In this case only one binding event was observed (Figure 68B) which, after integration, revealed a curve (Figure 63D) that could be fitted using a single binding event model. The observed stoichiometry ( $N = 1$ ), dissociation constant ( $K_d = 0.10 \pm 0.01$   $\mu$ M) and thermodynamics parameters ( $\Delta H = -18.2 \pm 0.2$  kcal mol<sup>-1</sup>, and  $\Delta S = -28.7$  cal mol<sup>-1</sup> K<sup>-1</sup>) are comparable with those previously measured for the binding of  $Ni^{2+}$  to isolated *HpUreE*. This result is in complete agreement with the lack of complex formation observed by MALS/QELS, suggesting that the binding of nickel is not sufficient to stabilize the *HpUreE-HpUreG* interaction, probably because it involves only *HpUreE* and not *HpUreG*.

The identity of the residues involved in the two  $Zn^{2+}$ -binding events was investigated by repeating the same experiment using, instead of the wild type proteins, the mutants H102K *HpUreE*, H152A *HpUreE*, or C66A/H68A *HpUreG* (Figure 68). In fact, the key importance of the conserved residues (Cys<sup>66</sup>, His<sup>68</sup>) in *HpUreG* for the binding of zinc has been recently established [251]. Moreover, the previously shown ITC experiments indicates a role for His<sup>102</sup> and His<sup>152</sup> in  $Zn^{2+}$ -binding to *HpUreE*.



**Figure 68 - Thermal denaturation scanning of *HpUreE-HpUreG* complex**

Best fit of the integrated raw ITC data of the titration of 5  $\mu$ M *HpUreE-HpUreG* complex, and its related mutants, with 70  $\mu$ M  $ZnSO_4$ , represented as a solid line, obtained by a non-linear least squares procedure (wild type, black; H102K *HpUreE - HpUreG*, red; H152A *HpUreE - HpUreG*, blue; *HpUreE - C66A/H68A HpUreG*, green). The calorimetric parameters derived from all fits are given in Table 18.

Protein complex	High affinity binding event			Low affinity binding event		
	$K_{d1}$ (nM)	$\Delta H_1$ (kcal mol <sup>-1</sup> )	$\Delta S_1$ (cal mol <sup>-1</sup> K <sup>-1</sup> )	$K_{d2}$ ( $\mu$ M)	$\Delta H_2$ (kcal mol <sup>-1</sup> )	$\Delta S_2$ (cal mol <sup>-1</sup> K <sup>-1</sup> )
<i>HpUreE - HpUreG</i>	1.5 $\pm$ 0.3	- 22.4 $\pm$ 0.2	- 34.9	0.67 $\pm$ 0.05	- 15.2 $\pm$ 0.4	- 23.4
H102K <i>HpUreE - HpUreG</i>	-	-	-	0.47 $\pm$ 0.03	- 17.2 $\pm$ 0.1	- 28.8
H152A <i>HpUreE - HpUreG</i>	-	-	-	0.80 $\pm$ 0.10	- 25.5 $\pm$ 1.0	- 57.5
<i>HpUreE - C66A/H68A HpUreG</i>	2.0 $\pm$ 0.4	- 20.1 $\pm$ 0.1	- 27.5	0.93 $\pm$ 0.10	- 5.0 $\pm$ 0.2	- 11.0

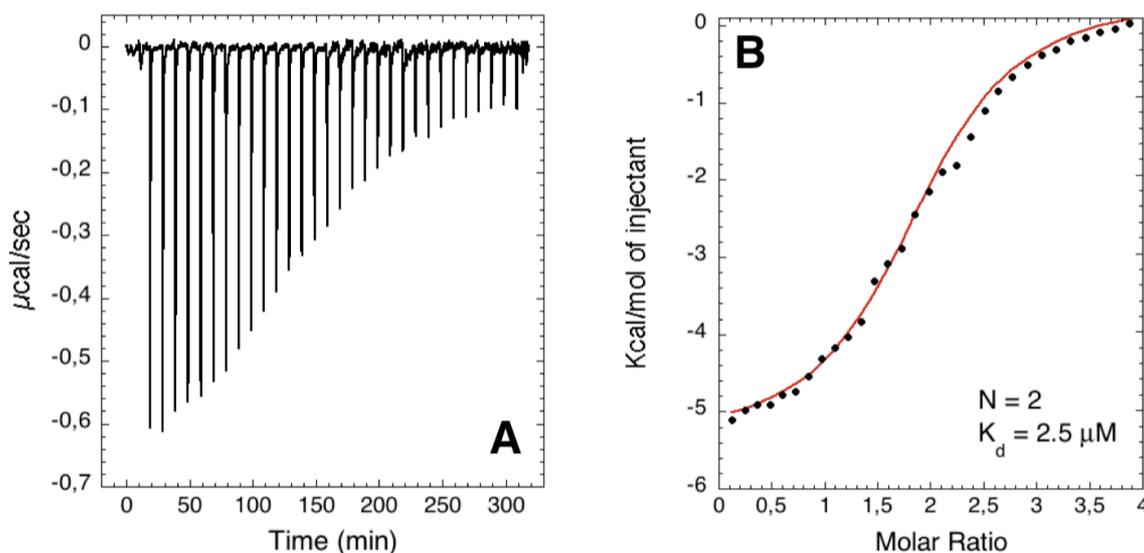
**Table 18** - Thermodynamic parameters of  $ZnSO_4$  binding to the *HpUreE-HpUreG* complex and its mutant

In the case of each of the *HpUreE* mutants, the tight binding event is not observed, while it is maintained when the mutant of *HpUreG* is used. On the other hand, the low affinity site is still present when the two mutants of *HpUreE* are utilized, but is disrupted in the case of C66A/H68A *HpUreG*. The residual binding of two  $Zn^{2+}$  ions in the latter case reproduces what previously observed for the *HpUreG* double mutant alone [251]. Overall, these data suggest that His<sup>102</sup> and His<sup>152</sup> contribute to the building of the high affinity site in the complex, while *HpUreG* Cys<sup>66</sup> and His<sup>68</sup> residues are responsible for the low affinity binding event. However, the exact topology of these two metal binding sites cannot be determined at the present stage of the study given the limited structural information on the flexible C-terminal arms containing His<sup>152</sup>.

Finally a GTPase activity evaluation was performed on *HpUreE*-*HpUreG* samples in the presence, or in the absence, of equimolar amount of  $Zn^{2+}$ , in order to estimate a possible enhancement of *HpUreG* capability after association with *HpUreE*. Samples containing identically amount of isolated proteins were analyzed as negative controls. No detectable increment in GTPase activity was measured for both apo- and  $Zn^{2+}$ -*HpUreE*-*HpUreG* species.

### 3.4. Preliminary characterization of the *HpUreE*-*HpHypA* interaction

Previous experimental data suggested that *HpHypA* can interact with the urease accessory system in *H. pylori* [307]. In particular, using a cross-linking approach the interaction between purified *HpHypA* and *HpUreE* was identified, leading to the formation of a 34 kDa hetero-dimeric complex [306]. However, the occurrence of an *HpHypA*-*HpUreE* assembly has not been confirmed experimentally. For this purpose, a solution of monomeric *HpHypA* was titrated with a solution dimeric of *HpUreE* in the same buffer.



**Figure 69 - ITC titration of monomeric *HpHypA* on dimeric *HpUreE***

(A) Raw ITC data showing the thermal effect of 30 x 10  $\mu\text{L}$  injections of 320  $\mu\text{M}$  *HpHypA* monomer binding to 19,5  $\mu\text{M}$  *HpUreE* dimer. (B) Best fit of the integrated data, represented as a solid line, obtained by a non-linear least squares procedure. The calculated numbers of sites and dissociation constant are indicated.

After addition, clear exothermic peaks were observed (Figure 69A) which, after integration, revealed a curve (Figure 69B) that could be fitted using a single binding event model. The stoichiometry of the interaction suggests that two monomers of *HpHypA* bind to a single dimer of *HpUreE*, forming a *HpUreE-HpHypA* complex having a dissociation constant  $K_d = 2.5 \pm 0.4 \mu\text{M}$ ,  $\Delta H = -5.7 \pm 0.2 \text{ kcal mol}^{-1}$ , and  $\Delta S = 6.46 \text{ cal mol}^{-1} \text{ K}^{-1}$ .

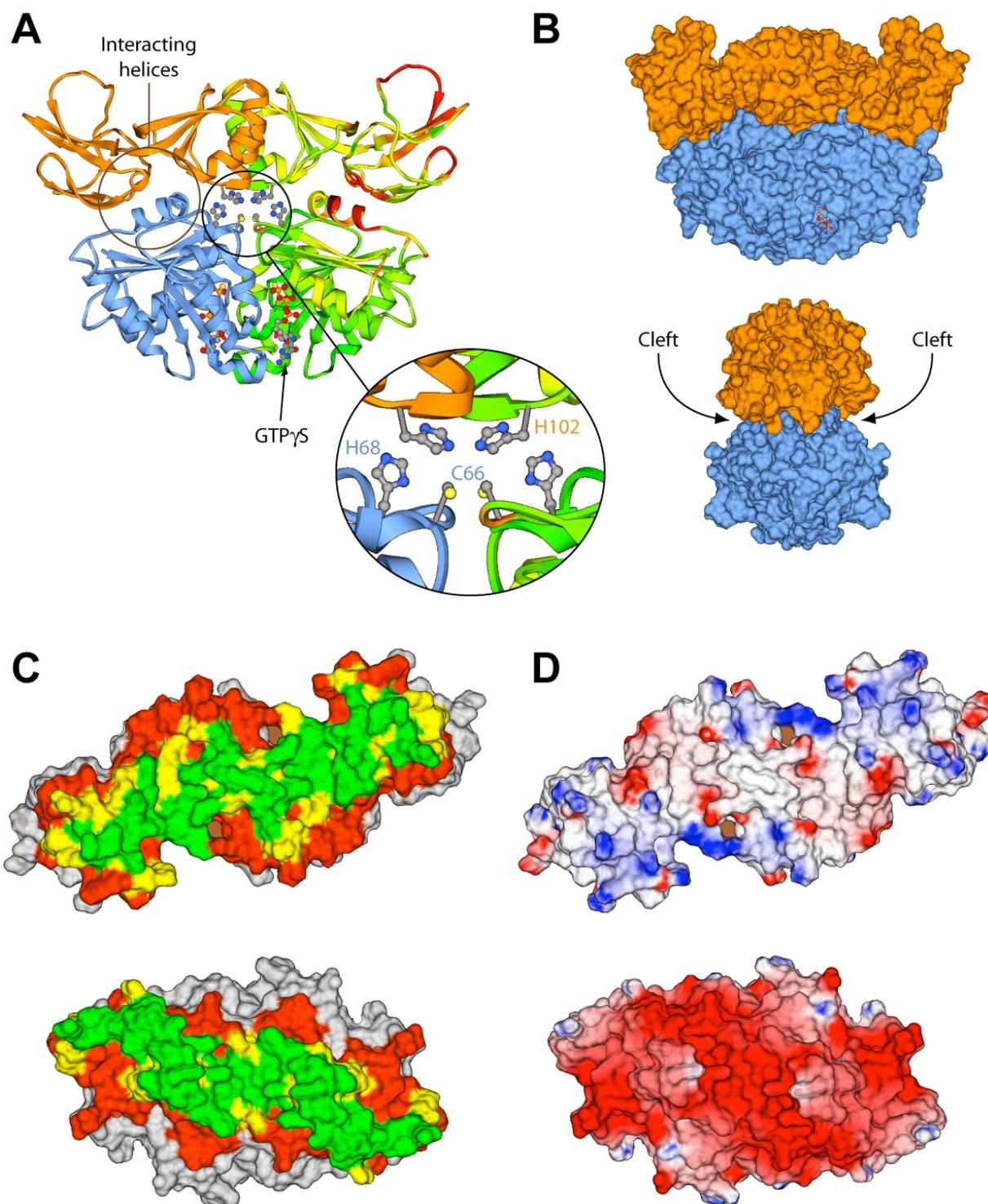
## Section C2 - Biocomputing results

### 3.5. Molecular modeling of the *HpUreE-HpUreG* complex

The viability of the *HpUreE-HpUreG* complex formation was investigated from a structural modeling point of view. The model structure of the *HpUreE* dimer was docked onto the model of the dimeric form of *HpUreG*, with optimization of protein backbone and side chains at the interface between the two homo-dimers. In the resulting structure, the two proteins face each other along their extended axes (Figure 70A), and only limited modifications of the proteins backbone, restricted both in extent and in topology distribution, were necessary in order to optimize the docking procedure.

Overall, a full size, shape, and electric charge complementarity between the surfaces of the two proteins is observed (Figure 70B-D). The central pocket formed on the *HpUreG* surface around the conserved Cys<sup>66</sup> and His<sup>68</sup> residues matches the shape and volume of the protruding region around the pair of conserved His<sup>102</sup> residues on the surface of *HpUreE*. Interestingly, the structure of the *HpUreE-HpUreG* complex features Cys<sup>66</sup>, His<sup>68</sup> of *HpUreG*, and His<sup>102</sup> of *HpUreE*, in neighbouring positions in the central core of the assembly. This suggests the building up of a novel metal binding site at the interface between the two protein partners (see close-up in Figure 70A). Furthermore, the shallow crevice formed between the central C-terminal domain and the peripheral N-terminal domain of *HpUreE* is filled with the bulge found on the surface of *HpUreG* around the rim of the protein dimerization interface (Figure 70B).

The formation of the complex results in a large total surface area (6378 Å<sup>2</sup>, 40.4% of *HpUreE* and 36.3% of *HpUreG*) that is buried by the two interacting homo-dimers. Most of the atomic contacts (defined as involving atoms positioned at distances  $\leq 0.4$  Å smaller than the sum of their van der Waals radii) between the two proteins involve hydrophobic interactions. These contacts are localized between the helix that separates the two parallel  $\beta$ -sheets in the N-terminal domain of *HpUreE* and the long central helix of *HpUreG* (Figure 70A). However, the interaction is additionally favored by electrostatic forces between a positively charged patch located on the N-terminal domains of *HpUreE* and a negatively charged region of the *HpUreG* surface (Figure 70D). Finally, H-bonds are formed between the side chains of *HpUreE* Arg<sup>31</sup> and *HpUreG* Ser<sup>76</sup>, and between the carboxylic group of Glu<sup>83</sup> on the *HpUreG* side and backbone N and O atoms of Lys<sup>32</sup>, Lys<sup>33</sup>, Ile<sup>34</sup>, and Ala<sup>35</sup> on the *HpUreE* side.



**Figure 70 - Model structure of the *HpUreE-HpUreG* assembly**

Ribbon diagram (A) and solvent excluded surface (B) of the complex between *HpUreE* (orange) and *HpUreG* (light blue). In the right side of panel A the ribbons are colored according to the backbone root mean square deviation with respect to the separated protein model structures, ranging from 0.0 Å (green) to 0.75 Å (yellow) to greater than 1.5 Å (red). Residues involved in metal binding, GTP $\gamma$ S molecules and Mg<sup>2+</sup> ions are shown as *ball and stick* and colored as follow: Mg, dark green; C, grey; H, white; N, blue; and O, red. The position of the surface clefts are indicated in panel B. Panels C and D report the solvent excluded surfaces of *HpUreE* and *HpUreG*, oriented in order to expose the interaction surfaces. In panel C the surface is colored according to the distance between the docked proteins: gray: > 10 Å; red: 5 - 10 Å; yellow: 2.5 - 5 Å; green: < 2.5 Å. In panel D the surface is colored according to the surface electrostatic potential.

### 3.6. Molecular modeling of the *HpUreE-HpHypA* complex

#### 3.6.1. HypA multiple sequence alignment construction and analysis

A HypA sequence search resulted in 112 hits, from different bacterial sources (Table 19). Identical amino acid sequences, as well as isoforms of HypA, were excluded from further analysis. These sequences are highlighted in grey in Table 19. Notably, the sequences from *H. pylori* strains 26695 and G27 are identical. The alignment of the remaining 37 HypA proteins was optimized using information derived from secondary structure predictions. In this optimized multiple alignment, reported in Figure 71, the sequences were found to feature an identity with respect to the sequence of *HpHypA* ranging between 53.4% and 20.3%.

From such alignment it is possible to extract a whole set of considerations. The structural zinc binding site, composed by two sequential CXXC motifs (Cys<sup>74</sup>, Cys<sup>77</sup>, Cys<sup>91</sup>, Cys<sup>94</sup> in *HpHypA*), is fully conserved. On the other hand, only three amino acids suggested to be involved in Ni<sup>2+</sup> binding (Met<sup>1</sup>, His<sup>2</sup>, Glu<sup>3</sup> in *HpHypA*) show a complete conservation. The only other fully conserved residue is a glycine (Gly<sup>34</sup> in *HpHypA*), found at the end of the first  $\beta$ -strand. Furthermore, three amino acids are conservatively mutated: Lys<sup>25</sup> and Ile<sup>31</sup>, from the first  $\beta$ -strand, and Ile<sup>63</sup>, found in the second  $\beta$ -strand. Some HypA proteins present large differences, as in the case of two sequences featuring an additional N-terminal portion (Figure 71, sequences n.15, 23). More interestingly, eleven sequences (Figure 71) show a long insertion between the  $\beta$ 3 strand and the  $\beta$ 4 strand, that in the structure of dimeric *TkHypA* fold as two  $\alpha$ -helices and mediate the domain-swapping protein dimerization (see Section 1.11.6).

Bacterial source	Genome	Gene	Length (aa)	Annotation
<i>Acaryochloris marina</i>	NC_009925	<i>hypA</i>	126	hydrogenase nickel insertion protein HypA
<i>Acaryochloris marina</i>	NC_009929	<i>hypA</i>	113	hydrogenase nickel insertion protein HypA
<i>Acidithiobacillus ferrooxidans</i>	NC_011761	<i>hypA-1</i>	109	hydrogenase nickel insertion protein HypA
<i>Acidithiobacillus ferrooxidans</i>	NC_011761	<i>hypA-2</i>	113	hydrogenase nickel insertion protein HypA
<i>Aeromonas hydrophila</i>	NC_008570	<i>hypA</i>	113	hydrogenase nickel insertion protein HypA
<i>Aeromonas salmonicida</i>	NC_009348	<i>hypA</i>	113	hydrogenase nickel insertion protein HypA
<i>Aquifex aeolicus</i>	NC_000918	<i>hypA</i>	115	hydrogenase accessory protein HypA
<i>Archaeoglobus fulgidus</i>	NC_000917	<i>hypA</i>	110	hydrogenase expression/formation protein ( <i>hypA</i> )
<i>Arcobacter butzleri</i>	NC_009850	<i>hypA</i>	123	hydrogenase expression/formation protein HypA
<i>Azoarcus</i> BH72	NC_008702	<i>hypA</i>	114	hydrogenase nickel incorporation protein <i>hypA</i>
<i>Azorhizobium caulinodans</i>	NC_009937	<i>hypA</i>	113	hydrogenase expression/synthesis protein
<i>Azotobacter vinelandii</i>	NC_012560	<i>hypA</i>	110	hydrogenase nickel incorporation protein HypA
<i>Bradyrhizobium BTAi1</i>	NC_009485	<i>hypA</i>	112	hydrogenase formation/expression
<i>Bradyrhizobium BTAi1</i>	NC_009485	<i>hypA</i>	110	hydrogenase expression/formation
<i>Bradyrhizobium japonicum</i>	NC_004463	<i>hypA</i>	113	hydrogenase nickel incorporation protein
<i>Bradyrhizobium japonicum</i>	NC_004463	<i>hypA</i>	113	HypA protein
<i>Bradyrhizobium</i> ORS278	NC_009445	<i>hypA</i>	113	hydrogenase expression/formation protein HypA
<i>Campylobacter concisus</i>	NC_009802	<i>hypA</i>	113	hydrogenase nickel insertion protein HypA
<i>Campylobacter curvus</i>	NC_009715	<i>hypA</i>	113	hydrogenase nickel insertion protein HypA

Bacterial source	Genome	Gene	Length (aa)	Annotation
<i>Campylobacter jejuni</i>	NC_009839	<i>hypA</i>	114	hydrogenase nickel insertion protein HypA
<i>Campylobacter jejuni</i>	NC_008787	<i>hypA</i>	114	hydrogenase nickel insertion protein HypA
<i>Campylobacter jejuni doylei</i>	NC_009707	<i>hypA</i>	114	hydrogenase nickel insertion protein HypA
<i>Campylobacter jejuni</i>	NC_002163	<i>hypA</i>	114	hydrogenase expression/formation protein
<i>Campylobacter jejuni</i>	NC_003912	<i>hypA</i>	114	hydrogenase nickel insertion protein HypA
<i>Campylobacter lari</i>	NC_012039	<i>hypA</i>	121	hydrogenase expression/formation protein HypA
<i>Carboxydotherrmus hydrogenoformans</i>	NC_007503	<i>hypA1</i>	113	hydrogenase nickel insertion protein HypA
<i>Carboxydotherrmus hydrogenoformans</i>	NC_007503	<i>hypA2</i>	114	hydrogenase nickel insertion protein HypA
<i>Citrobacter koseri</i>	NC_009792	<i>hypA</i>	129	hydrogenase nickel incorporation protein
<i>Clostridium botulinum</i> A2 Kyoto	NC_012563	<i>hypA</i>	113	hydrogenase nickel insertion protein HypA
<i>Clostridium botulinum</i> A2 Kyoto	NC_012563	<i>hypA</i>	113	hydrogenase nickel insertion protein HypA
<i>Clostridium botulinum</i> A3 Loch Maree	NC_010520	<i>hypA</i>	113	hydrogenase nickel insertion protein HypA
<i>Clostridium botulinum</i> A ATCC 19397	NC_009697	<i>hypA-2</i>	113	hydrogenase nickel insertion protein HypA
<i>Clostridium botulinum</i> A	NC_009495	<i>hypA</i>	114	hydrogenase nickel insertion protein HypA
<i>Clostridium botulinum</i> B1 Okra	NC_010516	<i>hypA</i>	113	hydrogenase nickel insertion protein HypA
<i>Clostridium botulinum</i> F Langeland	NC_009699	<i>hypA-1</i>	113	hydrogenase nickel insertion protein HypA
<i>Clostridium botulinum</i> F Langeland	NC_009699	<i>hypA-2</i>	113	hydrogenase nickel insertion protein HypA
<i>Cyanobacteria bacterium</i> Yellowstone	NC_007776	<i>hypA</i>	132	hydrogenase nickel insertion protein HypA
<i>Cyanothece</i> ATCC 51142	NC_010546	<i>hypA</i>	114	hydrogenase expression/synthesis
<i>Dehalococcoides</i> CBDB1	NC_007356	<i>hypA</i>	119	hydrogenase nickel insertion protein HypA
<i>Dehalococcoides ethenogenes</i>	NC_002936	<i>hypA</i>	119	hydrogenase nickel insertion protein HypA
<i>Desulfobacterium autotrophicum</i>	NC_012108	<i>hypA</i>	115	HypA
<i>Desulfotalea psychrophila</i>	NC_006138	<i>hypA</i>	115	hydrogenase accessory protein HypA
<i>Desulfurococcus kamchatkensis</i>	NC_011766	<i>hypA</i>	129	hydrogenase nickel incorporation protein
<i>Dictyoglomus thermophilum</i>	NC_011297	<i>hypA</i>	132	hydrogenase nickel incorporation protein
<i>Dictyoglomus turgidum</i>	NC_011661	<i>hypA</i>	132	hydrogenase nickel incorporation protein
<i>Enterobacter</i> 638	NC_009436	<i>hypA</i>	116	hydrogenase nickel incorporation protein
<i>Erwinia carotovora atroseptica</i>	NC_004547	<i>hypA</i>	117	hydrogenase nickel incorporation protein
<i>Escherichia coli</i> O127:H6	NC_011601	<i>hypA</i>	116	hydrogenase nickel incorporation protein
<i>Escherichia coli</i> 536	NC_008253	<i>hypA</i>	116	hydrogenase nickel incorporation protein
<i>Escherichia coli</i> APEC O1	NC_008253	<i>hypA</i>	120	hydrogenase nickel incorporation protein
<i>Escherichia coli</i> CFT073	NC_004431	<i>hypA</i>	120	hydrogenase nickel incorporation protein
<i>Escherichia coli</i> SMS 3 5	NC_004431	<i>hypA</i>	120	hydrogenase nickel incorporation protein
<i>Frankia alni</i> ACN14a	NC_008278	<i>hypA2</i>	120	hydrogenase nickel incorporation protein
<i>Frankia alni</i> ACN14a	NC_008278	<i>hypA1</i>	127	hydrogenase nickel incorporation protein hypA
<i>Geobacter sulfurreducens</i>	NC_002939	<i>hypA</i>	110	hydrogenase expression/formation protein hupa
<i>Hahella chejuensis</i> KCTC 2396	NC_007645	<i>hypA</i>	107	hydrogenase formation/expression protein HypA
<i>Helicobacter acinonychis</i> Sheeba	NC_008229	<i>hypA</i>	117	hydrogenase nickel incorporation protein
<i>Helicobacter hepaticus</i>	NC_004917	<i>hypA</i>	141	hydrogenase nickel incorporation protein
<i>Helicobacter pylori</i> 26695	NC_004917	<i>hypA</i>	117	hydrogenase nickel incorporation protein
<i>Helicobacter pylori</i> G27	NC_011333	<i>hypA</i>	117	hydrogenase nickel incorporation protein
<i>Helicobacter pylori</i> P12	NC_011498	<i>hypA</i>	117	hydrogenase nickel incorporation protein
<i>Heliobacterium modesticaldum</i>	NC_010337	<i>hypA</i>	115	hydrogenase nickel insertion protein hypa
<i>Klebsiella pneumoniae</i> 342	NC_011283	<i>hypA</i>	114	hydrogenase nickel incorporation protein
<i>Klebsiella pneumoniae</i> MGH 78578	NC_009648	<i>hypA</i>	114	hydrogenase nickel incorporation protein
<i>Legionella pneumophila</i> Corby	NC_009494	<i>hypA</i>	113	hydrogenase nickel incorporation protein HypA
<i>Legionella pneumophila</i> Lens	NC_006369	<i>hypA</i>	113	hydrogenase nickel incorporation protein HypA
<i>Methanobacterium thermoautotrophicum</i>	NC_000916	<i>hypA</i>	125	hydrogenase nickel incorporation protein
<i>Methanobrevibacter smithii</i>	NC_009515	<i>hypA</i>	125	hydrogenase nickel incorporation protein
<i>Methanococcus maripaludis</i>	NC_005791	<i>hypA</i>	126	hydrogenase nickel insertion protein HypA
<i>Methanosarcina acetivorans</i>	NC_003552	<i>hypA</i>	133	hydrogenase nickel incorporation protein
<i>Methanosarcina barkeri</i>	NC_007355	<i>hypA</i>	133	hydrogenase nickel incorporation protein
<i>Methanosarcina mazei</i>	NC_003901	<i>hypA</i>	137	hydrogenase nickel incorporation protein

Bacterial source	Genome	Gene	Length (aa)	Annotation
<i>Methanospaera stadthanae</i>	NC_007681	<i>hypA</i>	127	hydrogenase nickel incorporation protein
<i>Methylococcus capsulatus</i> Bath	NC_002977	<i>hypA</i>	117	hydrogenase expression/formation protein HypA
<i>Mycobacterium marinum</i>	NC_010612	<i>hypA</i>	130	hydrogenase expression/synthesis protein, HypA
<i>Mycobacterium smegmatis</i> MC2	NC_008596	<i>hypA</i>	111	hydrogenase nickel insertion protein HypA
<i>Mycobacterium smegmatis</i> MC2	NC_008596	<i>hypA</i>	109	hydrogenase nickel insertion protein HypA
<i>Nautilia profundicola</i>	NC_012115	<i>hypA</i>	119	hydrogenase nickel insertion protein HypA
<i>Pelotomaculum thermopropionicum</i>	NC_009454	<i>hypA</i>	137	hydrogenase nickel incorporation protein
<i>Persephonella marina</i>	NC_012440	<i>hypA</i>	114	hydrogenase nickel insertion protein HypA
<i>Pyrococcus abyssi</i>	NC_000868	<i>hypA</i>	139	hydrogenase nickel incorporation protein
<i>Pyrococcus furiosus</i>	NC_003413	<i>hypA</i>	139	hydrogenase nickel incorporation protein
<i>Pyrococcus horikoshii</i>	NC_000961	<i>hypA</i>	145	hydrogenase nickel incorporation protein
<i>Ralstonia eutropha</i>	NC_005241	<i>hypA1</i>	119	HypA1
<i>Ralstonia eutropha</i>	NC_005241	<i>hypA3</i>	109	HypA3
<i>Ralstonia eutropha</i>	NC_005241	<i>hypA2</i>	113	HypA2
<i>Rhodobacter sphaeroides</i>	NC_007493	<i>hypA</i>	113	hydrogenase expression/synthesis, HypA family
<i>Rhodococcus erythropolis</i>	NC_012490	<i>hypA</i>	109	hydrogenase nickel incorporation protein HypA
<i>Rhodococcus opacus</i>	NC_012522	<i>hypA</i>	109	hydrogenase nickel incorporation protein HypA
<i>Rhodopseudomonas palustris</i>	NC_005296	<i>hypA</i>	113	hydrogenase formation/expression protein hypA
<i>Saccharopolyspora erythraea</i>	NC_009142	<i>hypA1</i>	109	hydrogenase nickel incorporation protein HypA
<i>Salmonella enterica arizonae</i> serovar	NC_010067	<i>hypA</i>	123	hydrogenase nickel incorporation protein
<i>Salmonella enterica</i> Choleraesuis	NC_006905	<i>hypA</i>	118	hydrogenase nickel incorporation protein
<i>Salmonella enterica</i> serovar Agona	NC_011149	<i>hypA</i>	113	hydrogenase nickel insertion protein HypA
<i>Salmonella enterica</i> serovar Newport	NC_011080	<i>hypA</i>	113	hydrogenase nickel insertion protein HypA
<i>Serratia proteamaculans</i>	NC_009832	<i>hypA</i>	114	hydrogenase nickel incorporation protein
<i>Shewanella oneidensis</i>	NC_004347	<i>hypA</i>	118	hydrogenase expression/formation protein HypA
<i>Shigella boydii</i>	NC_007613	<i>hypA</i>	116	hydrogenase nickel incorporation protein
<i>Shigella dysenteriae</i>	NC_007606	<i>hypA</i>	116	hydrogenase nickel incorporation protein
<i>Shigella flexneri</i>	NC_008258	<i>hypA</i>	116	hydrogenase nickel incorporation protein
<i>Staphylothermus marinus</i>	NC_009033	<i>hypA</i>	136	hydrogenase nickel incorporation protein
<i>Streptomyces avermitilis</i>	NC_003155	<i>hypA</i>	131	[NiFe]-hydrogenase expression/formation protein
<i>Synechococcus elongatus</i>	NC_006576	<i>hypA</i>	112	hydrogenase expression/formation protein HypA
<i>Synechococcus</i> PCC 7002	NC_010475	<i>hypA</i>	113	hydrogenase nickel insertion protein HypA
<i>Synechococcus</i> PCC 7002	NC_010475	<i>hypA</i>	121	hydrogenase nickel insertion protein
<i>Synechocystis</i> PCC6803	NC_000911	<i>hypA</i>	113	hydrogenase expression/formation protein; HypA
<i>Thermococcus kodakaraensis</i> KOD1	NC_006624	<i>hypA</i>	139	hydrogenase nickel incorporation protein
<i>Thermococcus onnurineus</i>	NC_011529	<i>hypA</i>	139	hydrogenase nickel incorporation protein
<i>Thermodesulfovibrio yellowstonii</i>	NC_011296	<i>hypA</i>	116	hydrogenase nickel insertion protein HypA
<i>Treponema denticola</i>	NC_002967	<i>hypA</i>	113	hydrogenase nickel insertion protein HypA
<i>Wolinella succinogenes</i>	NC_005090	<i>hypA</i>	119	hydrogenase nickel incorporation protein
<i>Yersinia enterocolitica</i>	NC_008800	<i>hypA</i>	115	hydrogenase nickel incorporation protein

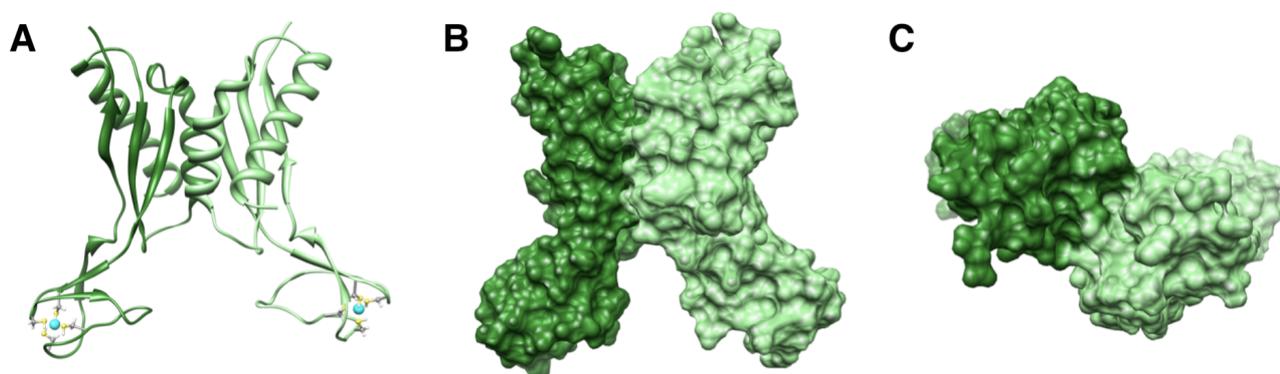
**Table 18 - HypA sequences retrieved for analysis and comparison**

For each sequence, the bacterial source, genomic NCBI code, gene locus, amino acidic length and annotation are reported. Excluded sequences are colored in grey.



### 3.6.2. Structural model of dimeric *HpHypA*

Recently, the crystal structure of dimeric HypA from *Thermococcus kodakaraensis* KOD1 (*TkHypA*), has been elucidated [298]. This, coupled to the experimentally observed 2:1 stoichiometry of the association between monomeric *HpHypA* and dimeric *HpUreE*, motivates the calculation of a feasible homo-dimeric structural architecture for *HpHypA*. At this purpose, the sequence alignment between *HpHypA* and *TkHypA* (Figure 71, sequences n. 16 and 32, respectively), together with the corresponding monomeric and dimeric structures, was used as input to achieve the modeling process (Figure 72). In the calculated structure, the head-head dimerization involved the N-terminal domains, that pack the two chains. On the contrary, the C-terminal domains, containing the structural Zn<sup>2+</sup>-binding sites, are forced out in opposite directions.

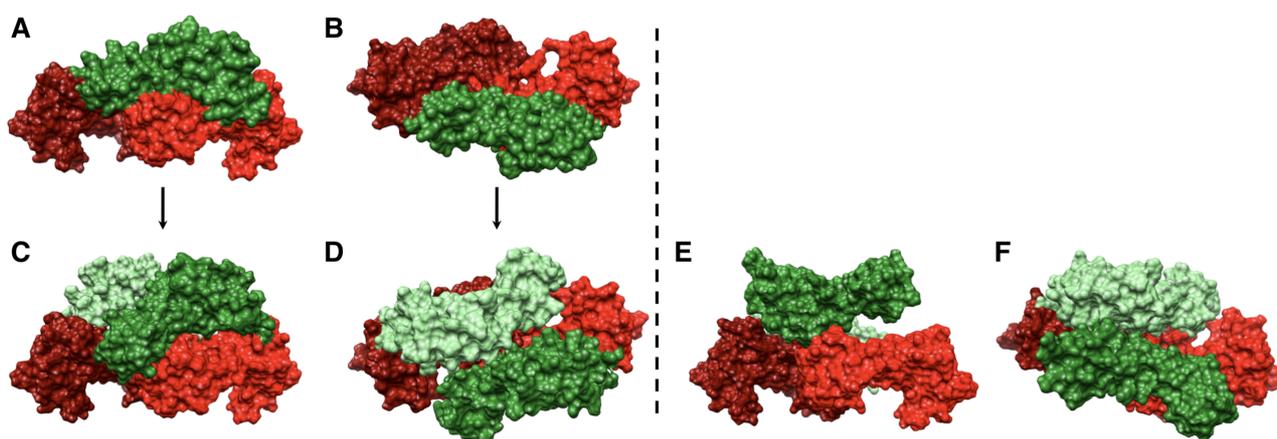


**Figure 72 - Structural model of dimeric *HpHypA***

Ribbon diagram (A) and solvent excluded surfaces (B,C) of the dimeric *HpHypA* model. The two chains are colored differently. In panel A residues that form the structural Zn<sup>2+</sup>-binding sites are shown as *ball and stick* and colored according to the following atomic scheme: carbon, grey; hydrogen, white; sulfur, yellow; zinc, cyan. In panel C the structure is rotated by 90° around the horizontal axis.

### 3.6.3. Molecular models of the *HpUreE-HpHypA* interaction

The *HpUreE-HpHypA* interaction was investigated from a structural modeling point of view using two distinct approaches. In fact, the experimental monomeric structure of *HpHypA* on one side, and the calculated dimeric model of *HpHypA* on the other, were separately docked onto the model of dimeric *HpUreE*. The first model reflects a mechanism in which *HpHypA* interacts with *HpUreE* undergoing dimerization on the surface of the protein partner (Figure 73A-D), whereas the second model assumes that a preformed *HpHypA* dimer associates with the dimeric architecture of *HpUreE* (Figure 73E,F). The model structure of *HpUreE-HpHypA* complex obtained starting from a preformed *HpHypA* dimer is asymmetrical, showing only one *HpHypA* that prevalently contacts *HpUreE*. On the other hand, the model obtained by sequential docking of two *HpHypA* monomers on the *HpUreE* displays a compact structure with an extended interacting surface, due to the symmetrical head-tail dimerization of *HpHypA* on *HpUreE*.



**Figure 73 - Model structures of the *HpUreE-HpHypA* assembly**

Solvent excluded surfaces of *HpUreE-HpHypA* complexes obtained by sequential docking of two *HpHypA* monomers on *HpUreE* (A-D) or direct docking of preformed *HpHypA* dimer on *HpUreE* (E,F). The chains of *HpUreE* (red and dark red) and *HpHypA* (light green and dark green) are differently colored. In panels B,D,F the structures are rotated by 90° around the horizontal axis.

The more favorable values of *RosettaDock* score for the sequential docking of monomeric *HpHypA* on *HpUreE* seems to confirm what previously observed from a topological point of view. Furthermore, the [*HpHypA* monomer-*HpUreE* dimer] assembly isolated by cross-linking experiments [306] supports the mechanism of a sequential binding of two *HpHypA* monomers, dictated by the homo-dimeric architecture of *HpUreE*.

To sum up, the two different topologies calculated for the *HpHypA-HpUreE* complex clearly depend on the dimerization mode of *HpHypA*. In particular, the head-tail dimerization originates a dimeric *HpHypA* structure that is more complementary in shape with those shown by *HpUreE*. On the contrary, the head-head dimerization leads to the formation of a more stable association between the two polypeptide chains of *HpHypA*, that is less compatible with the overall architecture of *HpUreE*. This dimeric assembly is observed in *TkHypA*, that undergoes a head-head dimerization, associated to a domain swapping between the two chains, through the presence of two additional helices, that are missing in *HpHypA* (see sequence alignment in Figure 71). Interestingly, these additional helices are found only in *Archaea* microorganisms that do not possess either the urease enzymes and all the urease accessory proteins, including *UreE* (see Table 19). Moreover, the majority of these do not possess also *HypB*, which interacts with *HypA* to correctly deliver  $\text{Ni}^{2+}$  to the [NiFe]-hydrogenase active site. All these observations allow to speculate that the head-head dimerization of *HypA* proteins, occurring through additional not-conserved  $\alpha$ -helices, is a typical feature of *Archaea* species evolutionally conserved in order to permit the stabilization of *HypA* proteins in the absence of its interacting partners *UreE* and/or *HypB*.

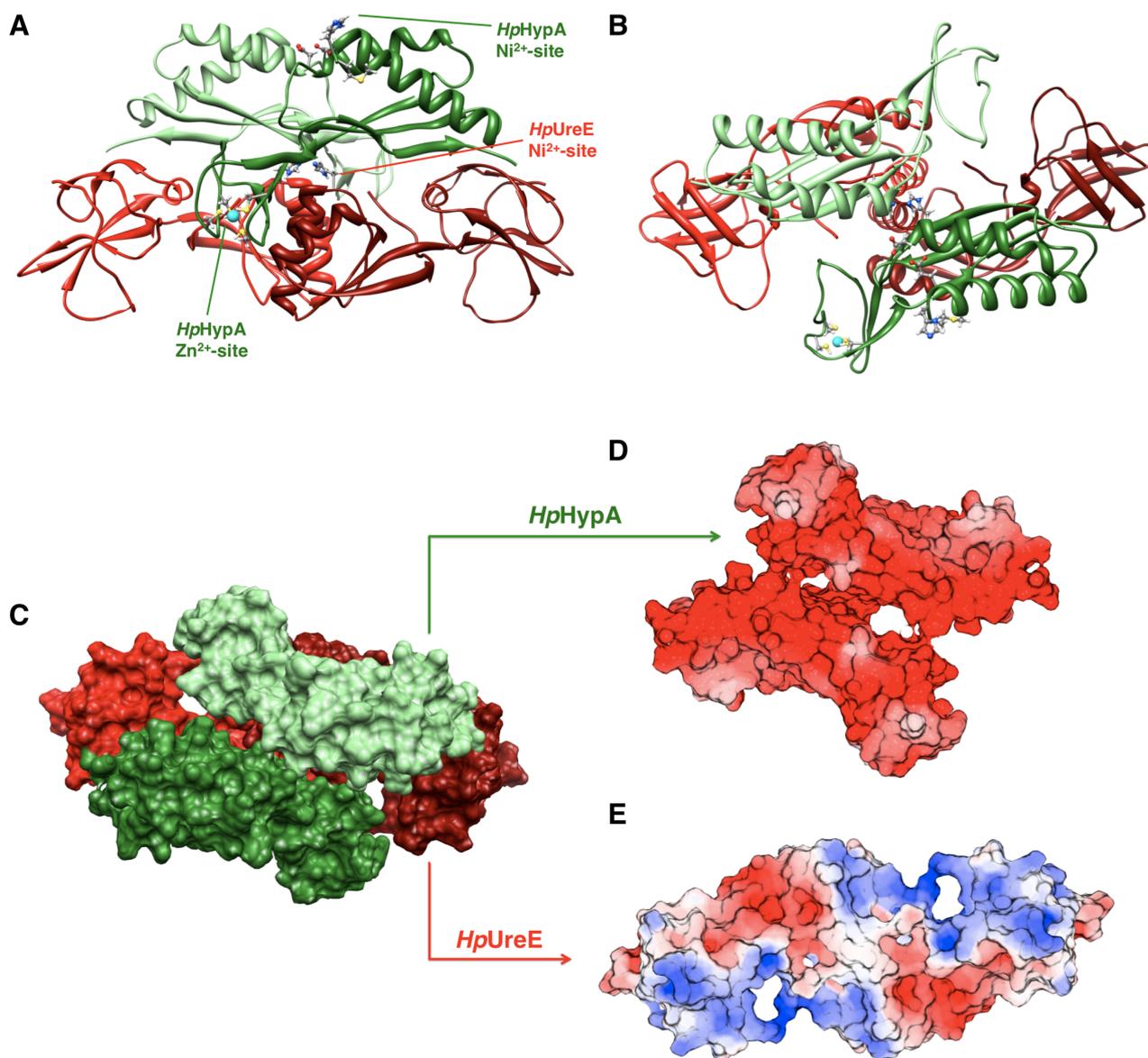
ID	Organism	Genome NCBI code	Urease	Ure accessory	Hydrogenase	Hyp accessory
01	<i>Acaryochloris marina</i>	NC_009925 NC_009929	Yes	D,E,F,G	Yes	A,B,C,D,E,F
02	<i>Aquifex aeolicus</i>	NC_000918	No	-	Yes	A,B,D,E,F
03	<i>Bradyrhizobium japonicum</i>	NC_009485	Yes	D,E,F,G	Yes	A,B,C,D,E,F
04	<i>Campylobacter jejuni</i>	NC_009839	No	-	Yes	A,B,C,D,E,F
05	<i>Campylobacter lari</i>	NC_012039	No	-	Yes	A,B,C,D,E,F
06	<i>Clostridium botulinum</i>	NC_012563	No	-	Yes	A,B,C,D,E,F
07	<i>Synechococcus</i> sp. JA-2-3B'a	NC_007776	Yes	D,E,F,G	Yes	A,B
08	<i>Dehalococcoides ethenogenes</i>	NC_002936	No	-	Yes	A,B,C,D,E,F
09	<i>Desulfurococcus kamchatkensis</i>	NC_011766	No	-	Yes	A,C,D,E,F
10	<i>Dictyoglomus thermophilum</i>	NC_011297	No	-	Yes	A,C,D,E,F
11	<i>Escherichia coli</i> O127:H6	NC_011601	No	-	Yes	A,B,C,D,E
12	<i>Escherichia coli</i> APEC O1	NC_008563	No	-	Yes	A,B,C,D,E,F
13	<i>Geobacter sulfurreducens</i>	NC_002939	No	-	Yes	A,B,C,D,E,F
14	<i>Helicobacter acinonychis</i>	NC_008229	Yes	D,E,F,G	Yes	A,B,C,D,E,F
15	<i>Helicobacter hepaticus</i>	NC_004917	Yes	D,E,F,G	Yes	A,B,C,D,E,F
16	<i>Helicobacter pylori</i> 26695	NC_000915	Yes	D,E,F,G	Yes	A,B,C,D,E,F
17	<i>Klebsiella pneumoniae</i>	NC_011283	Yes	D,E,F,G	Yes	A,B,C,D,E,F
18	<i>Methanobacterium thermoautotrophicum</i>	NC_000916	No	-	Yes	A,B,C,D,E,F
19	<i>Methanococcus maripaludis</i>	NC_005791	No	-	Yes	A,B,C,D,E,F
20	<i>Methanosarcina acetivorans</i>	NC_003552	No	-	Yes	A,B,C,D,E,F
21	<i>Methanosphaera stadtmanae</i>	NC_007681	No	-	Yes	A,B,E,F
22	<i>Methylococcus capsulatus</i>	NC_002977	No	-	Yes	A,B,C,D,E,F
23	<i>Mycobacterium marinum</i>	NC_010612	Yes	D,F,G	Yes	A,B,C,D,E,F
24	<i>Pelotomaculum thermopropionicum</i>	NC_009454	No	-	Yes	A,B,C,D,E,F
25	<i>Pyrococcus furiosus</i>	NC_003413	No	-	Yes	A
26	<i>Pyrococcus horikoshii</i>	NC_000961	No	-	Yes	A,C,D,F
27	<i>Ralstonia eutropha</i>	NC_008313 NC_005241	Yes	D,E,F,G	Yes	A,B,C,D,E,F
28	<i>Rhodobacter sphaeroides</i>	NC_007493	Yes	D,E,F,G	Yes	A,B,C,D,E,F
29	<i>Salmonella enterica</i> serovar	NC_011149	No	-	Yes	A,B,C,D,E,F
30	<i>Staphylothermus marinus</i>	NC_009033	No	-	Yes	A,C,D,E,F
31	<i>Streptomyces avermitilis</i>	NC_003155	Yes	D,F,G	Yes	A,B,C,D,E,F
32	<i>Thermococcus kodakaraensis</i> KOD1	NC_006624	No	-	Yes	A,C,D,E,F
33	<i>Thermodesulfovibrio yellowstonii</i>	NC_011296	No	-	Yes	A,B,C,D,E,F
34	<i>Wolinella succinogenes</i>	NC_005090	No	-	Yes	A,B,C,D,E,F
35	<i>Yersinia enterocolitica</i>	NC_008800	Yes	D,E,F,G	Yes	A,B,C,D,E
36	<i>Dictyoglomus turgidum</i>	NC_011661	No	-	Yes	A,C,D,E,F
37	<i>Pyrococcus abyssi</i>	NC_000868	No	-	Yes	A,C,D,E,F

**Table 19 - Archaeal and bacterial HypA sequences: analysis and comparison**

The sequence numeration reflects the one used for the alignment in Figure 71. For each sequence, the bacterial source and the genomic NCBI code are indicated. The presence or the absence in the corresponding genome of the two Ni<sup>2+</sup>-enzymes urease and [NiFe]-hydrogenase, and the proteins constituting their accessory systems, is reported. Sequences containing the additional not-conserved helices mediating the head-head dimerization are colored in grey.

In the final *HpUreE-HpHypA* model structure, the two *HpHypA* monomers are differently oriented achieving a head-tail dimerization on the surface of the *HpUreE* dimer (Figure 74A-C). Overall, a good size, shape, and electric charge complementarity between the surfaces of the two

proteins is observed (Figure 74D,F). In particular the C-terminal domain of the *HpHypA* monomer, containing the structural  $Zn^{2+}$ -binding domain, fits the shape and volume of the protruding region around the pair of conserved His<sup>102</sup> residues on the surface of *HpUreE*. Moreover, the surface between the central C-terminal domain and the peripheral N-terminal domain of *HpUreE* is filled with the N-terminal domain of *HpHypA*, where are found the residues involved in  $Ni^{2+}$  coordination (Figure 74A). Interestingly, the  $Ni^{2+}$ -binding domains of the two proteins, although not so close in space, are aligned, suggesting a possible mobilization of nickel ions between the two chaperones.



**Figure 74 - Model structure of the *HpUreE*-*HpHypA* assembly**

Ribbon diagram (A,B) and solvent excluded surface (C) of the *HpUreE*-*HpHypA* complex. The chains of *HpUreE* (red and dark red) and *HpHypA* (light green and dark green) are differently colored. In panel A the metal binding sites of the two proteins are indicated. Residues involved in metal binding and  $Zn^{2+}$  ions are shown as *ball and stick* and colored as follow: Zn, cyan; C, grey; H, white; N, blue; S, yellow; and O, red. Panels D and E report the solvent excluded surfaces of *HpHypA* and *HpUreE*, oriented in order to expose the interaction surfaces. Here, the surface is colored according to the surface electrostatic potential.

The formation of the complex results in a mean surface area of 490 Å<sup>2</sup>, that is buried by the two interacting homo-dimers. Most of the atomic contacts between the two proteins involve hydrophobic interactions, located the two parallel β-sheets in the N-terminal domain of *HpUreE* and the C-terminal strands of *HpHypA* (Figure 74A,B). However, the interaction is additionally favored by electrostatic forces between a positively charged patch located on the N-terminal domains of *HpUreE* and the extended negatively charged region of the *HpHypA* surface (Figure 74C,D). Notably, H-bonds are formed between the side chains of *HpUreE* Lys<sup>32</sup>, Arg<sup>36</sup> and *HpHypA* Glu<sup>113</sup>, Glu<sup>117</sup>, and between the carboxylic group of Glu<sup>68</sup>, on the *HpHypA* side ,and side chain of Arg<sup>101</sup> and backbone of Asn<sup>100</sup>, on the *HpUreE* side.

---

## **D - Conclusions**

---



## Section D - Conclusions

The present study focused on the accessory systems of two Ni<sup>2+</sup>-dependent enzymes, urease and [NiFe]-hydrogenase, from *H. pylori*, and specifically on the required metal-mediated PPIs that allow the nickel delivery and incorporation into the enzymes' active sites. In particular, the aim of this investigation was the elucidation of the interaction and crosstalk mechanisms between the urease and hydrogenase accessory proteins at the molecular level.

The results presented here allow us to envisage a mechanism for the urease assembly that entails a specific role for both Ni<sup>2+</sup> and Zn<sup>2+</sup> ions. An exchange of Zn<sup>2+</sup> for Ni<sup>2+</sup> binding to *HpUreE* could be the initial switch that modulates the interaction between *HpUreE* and *HpUreG*. In this view, Ni<sup>2+</sup> released from *HpUreE* could be incorporated into the apo-urease active site, concomitantly with the Zn<sup>2+</sup>-induced *HpUreE-HpUreG* complex formation and consequent stimulation of GTPase activity catalyzed by *HpUreG*. This step would lead to the carbamylation of the lysine residue in the urease active site, thus finalizing the activation of the enzyme.

The influence of Zn<sup>2+</sup> in the Ni<sup>2+</sup>-dependent urease system principally relies on the different Ni<sup>2+</sup>- and Zn<sup>2+</sup>-binding modes observed experimentally for *HpUreE*. The higher availability of intracellular Zn<sup>2+</sup> as compared to Ni<sup>2+</sup>, together with the similar affinity of *HpUreE* for these two metal ions, suggests that the specificity of binding different metals must rely on changes in ligand environment. In fact, while Ni<sup>2+</sup> is bound to the conserved His<sup>102</sup> on the surface of the dimer, without any involvement of His<sup>152</sup>, Zn<sup>2+</sup>-binding not only requires the His<sup>102</sup> pair, but is also modulated by the two His<sup>152</sup> residues at the C-terminal position.

The capability of *HpUreE* to distinguish between Ni<sup>2+</sup> and Zn<sup>2+</sup> is probably dictated by the occurrence of another functional role for this metallo-chaperone, apart of its function as nickel transporter. Here, a role for Zn<sup>2+</sup>, and not for Ni<sup>2+</sup>, in the stabilization of the *HpUreE-HpUreG* interaction was experimentally demonstrated. In particular, titration of Zn<sup>2+</sup> on the *HpUreE-HpUreG* complex results in two distinct binding event, the first with high affinity in the nano-molar range and the second with lower affinity in the micro-molar range. This stabilizing effect, occurring specifically only in the presence of Zn<sup>2+</sup>, is significant within the framework of the known role of *HpUreG* *in vivo*: this protein is an enzyme that catalyzes GTP hydrolysis necessary to the urease activation process [251]. *HpUreG* belongs to a class of homo-dimeric GTPases (or ATPases) that use GTP (or ATP) hydrolysis as a conserved molecular switch to regulate a large number of cellular processes [52]. The activity of these hydrolases is, in general, tightly controlled by different factors, such as protein dimerization and subsequent interaction with GAP (GTPase activator protein) and GEF (guanine nucleotide exchange factor) proteins. These GTPase regulators are stable functional dimers, as observed for *UreE*. However, the interaction with *HpUreE* stabilized by Zn<sup>2+</sup> is

not sufficient to promote any detectable GTPase activity. This result is not surprising: it is known that the GTP-dependent process of nickel incorporation into apo-urease occurs only in the presence of a UreDFG complex, implying that UreD and UreF must also play an essential role in UreG activation.

The chaperones involved in the maturation of [NiFe]-hydrogenase HypA and HypB, also display a specific Zn<sup>2+</sup>-binding capability. Of these, *HpHypA*, responsible for the activation of both urease and [NiFe]-hydrogenase enzymes, also shows Ni<sup>2+</sup>-binding capability *in vitro*. The occurrence of a specific interaction between *HpHypA* and *HpUreE*, previously hypothesized on the basis of cross-linking and immuno-blotting experiments [306], was here demonstrated to occur *in vitro* by microcalorimetric titrations. Interestingly, the calculated dissociation constant for the *HpHypA*-*HpUreE* complex ( $K_d = 2.5 \mu\text{M}$ ) is very similar to that observed for the *HpUreE*-*HpUreG* assembly ( $K_d = 4 \mu\text{M}$ ), suggesting that this interaction would be functional to mediate Ni<sup>2+</sup> transfer from *HpHypA*, or *HpHypA*-*HpHypB* complex, to *HpUreE*, or *HpUreE*-*HpUreG* complex, *in vivo*. Indeed, specific PPIs were observed between *HpHypA* and *HpHypB* during cross-linking experiments, as well as between *HpHypB* and *HpUreG* from tandem affinity purification [333]. These data, coupled to the observation of specific *HpUreE*-*HpUreG* and *HpUreE*-*HpHypA* interactions, indicate the possible presence of cross-talk mechanisms *in vivo*, involving *HpHypA*, *HpHypB*, *HpUreG* and *HpUreE*. It is interesting to notice that all these proteins possess Ni<sup>2+</sup> and/or Zn<sup>2+</sup> binding capability. Therefore, the Zn<sup>2+</sup>-dependent interaction between *HpUreE* and *HpUreG*, as well as the interdependence between Ni<sup>2+</sup> and Zn<sup>2+</sup>, emerging in this study for the *H. pylori* urease system, suggests a functional role for metal binding to these accessory proteins, modulating the formation of the PPIs necessary for enzyme maturation. In particular, a role for Ni<sup>2+</sup> and/or Zn<sup>2+</sup> in regulating the *HpHypA*-*HpUreE* interaction could be hypothesized, but further analysis are needed to confirm it experimentally, as in the case of the *HpUreE*-*HpUreG* association.

## Bibliography

1. Szilagy, A., et al., *Prediction of physical protein-protein interactions*. Phys Biol, 2005. **2**(2): p. S1-16.
2. Whitman, W.B., D.C. Coleman, and W.J. Wiebe, *Prokaryotes: the unseen majority*. Proc Natl Acad Sci U S A, 1998. **95**(12): p. 6578-83.
3. Caetano-Anolles, G., et al., *The origin, evolution and structure of the protein world*. Biochem J, 2009. **417**(3): p. 621-37.
4. Bull, A.T., M. Goodfellow, and J.H. Slater, *Biodiversity as a source of innovation in biotechnology*. Annu Rev Microbiol, 1992. **46**: p. 219-52.
5. Drake, J.W., et al., *Rates of spontaneous mutation*. Genetics, 1998. **148**(4): p. 1667-86.
6. Linderstrom-Lang, K. and J.A. Schellman, *Protein structure and enzymatic activity*, in *The Enzymes 2nd ed.*, H. Lardy and K. Myrback, Editors. 1959, Academic Press. : New York. p. 443-510.
7. Salem, G.M., et al., *Correlation of observed fold frequency with the occurrence of local structural motifs*. J Mol Biol, 1999. **287**(5): p. 969-81.
8. Govindarajan, S. and R.A. Goldstein, *Why are some proteins structures so common?* Proc Natl Acad Sci U S A, 1996. **93**(8): p. 3341-5.
9. Soding, J. and A.N. Lupas, *More than the sum of their parts: on the evolution of proteins from peptides*. Bioessays, 2003. **25**(9): p. 837-46.
10. Alberts, B., *The cell as a collection of protein machines: preparing the next generation of molecular biologists*. Cell, 1998. **92**(3): p. 291-4.
11. Li, R., et al., *Alpha A-crystallin and alpha B-crystallin, newly identified interaction proteins of protease-activated receptor-2, rescue astrocytes from C2-ceramide- and staurosporine-induced cell death*. J Neurochem, 2009. **110**(5): p. 1433-44.
12. Ryan, J.G., et al., *Clinical features and functional significance of the P369S/ R408Q variant in pyrin, the familial Mediterranean fever protein*. Ann Rheum Dis, 2009.
13. Ausio, J., et al., *Syndromes of disordered chromatin remodeling*. Clin Genet, 2003. **64**(2): p. 83-95.
14. Belperio, J.A., et al., *Critical role for CXCR2 and CXCR2 ligands during the pathogenesis of ventilator-induced lung injury*. J Clin Invest, 2002. **110**(11): p. 1703-16.
15. Lomas, D.A., *New insights into the structural basis of alpha 1-antitrypsin deficiency*. QJM, 1996. **89**(11): p. 807-12.
16. Goh, C.S., et al., *Co-evolution of proteins with their interaction partners*. J Mol Biol, 2000. **299**(2): p. 283-93.
17. Bahadur, R.P. and M. Zacharias, *The interface of protein-protein complexes: analysis of contacts and prediction of interactions*. Cell Mol Life Sci, 2008. **65**(7-8): p. 1059-72.
18. Nooren, I.M. and J.M. Thornton, *Diversity of protein-protein interactions*. EMBO J, 2003. **22**(14): p. 3486-92.
19. Nooren, I.M., et al., *The solution structure and dynamics of an Arc repressor mutant reveal premelting conformational changes related to DNA binding*. Biochemistry, 1999. **38**(19): p. 6035-42.
20. Keskin, O., et al., *Principles of protein-protein interactions: what are the preferred ways for proteins to interact?* Chem Rev, 2008. **108**(4): p. 1225-44.
21. Faust, P.L., S. Kornfeld, and J.M. Chirgwin, *Cloning and sequence analysis of cDNA for human cathepsin D*. Proc Natl Acad Sci U S A, 1985. **82**(15): p. 4910-4.
22. van de Locht, A., et al., *Two heads are better than one: crystal structure of the insect derived double domain Kazal inhibitor rhodniin in complex with thrombin*. EMBO J, 1995. **14**(21): p. 5149-57.

23. Oldham, W.M. and H.E. Hamm, *Structural basis of function in heterotrimeric G proteins*. Q Rev Biophys, 2006. **39**(2): p. 117-66.
24. Janin, J., S. Miller, and C. Chothia, *Surface, subunit interfaces and interior of oligomeric proteins*. J Mol Biol, 1988. **204**(1): p. 155-64.
25. Bahadur, R.P., et al., *A dissection of specific and non-specific protein-protein interfaces*. J Mol Biol, 2004. **336**(4): p. 943-55.
26. Kleanthous, C., *Protein-protein recognition*. Frontiers in molecular biology 31. 2000, Oxford ; New York: Oxford University Press. xix, 314 p.
27. Jones, S. and J.M. Thornton, *Principles of protein-protein interactions*. Proc Natl Acad Sci U S A, 1996. **93**(1): p. 13-20.
28. Chandler, D., *Interfaces and the driving force of hydrophobic assembly*. Nature, 2005. **437**(7059): p. 640-7.
29. Lo Conte, L., C. Chothia, and J. Janin, *The atomic structure of protein-protein recognition sites*. J Mol Biol, 1999. **285**(5): p. 2177-98.
30. O'Neil, K.T. and W.F. DeGrado, *How calmodulin binds its targets: sequence independent recognition of amphiphilic alpha-helices*. Trends Biochem Sci, 1990. **15**(2): p. 59-64.
31. Gellman, S.H., *On the role of methionine residues in the sequence-independent recognition of nonpolar protein surfaces*. Biochemistry, 1991. **30**(27): p. 6633-6.
32. Stites, W.E., *Protein-Protein Interactions: Interface Structure, Binding Thermodynamics, and Mutational Analysis*. Chem Rev, 1997. **97**(5): p. 1233-1250.
33. Smith-Gill, S.J., *Protein-protein interactions: structural motifs and molecular recognition*. Curr Opin Biotechnol, 1991. **2**(4): p. 568-75.
34. Remaut, H. and G. Waksman, *Protein-protein interaction through beta-strand addition*. Trends Biochem Sci, 2006. **31**(8): p. 436-44.
35. Janin, J. and C. Chothia, *The structure of protein-protein recognition sites*. J Biol Chem, 1990. **265**(27): p. 16027-30.
36. Smith, G.R., M.J. Sternberg, and P.A. Bates, *The relationship between the flexibility of proteins and their conformational states on forming protein-protein complexes with an application to protein-protein docking*. J Mol Biol, 2005. **347**(5): p. 1077-101.
37. Dyson, H.J. and P.E. Wright, *Intrinsically unstructured proteins and their functions*. Nat Rev Mol Cell Biol, 2005. **6**(3): p. 197-208.
38. Dyson, H.J. and P.E. Wright, *According to current textbooks, a well-defined three-dimensional structure is a prerequisite for the function of a protein. Is this correct?* IUBMB Life, 2006. **58**(2): p. 107-9.
39. Dosztanyi, Z., et al., *Disorder and sequence repeats in hub proteins and their implications for network evolution*. J Proteome Res, 2006. **5**(11): p. 2985-95.
40. Tompa, P., Z. Dosztanyi, and I. Simon, *Prevalent structural disorder in E. coli and S. cerevisiae proteomes*. J Proteome Res, 2006. **5**(8): p. 1996-2000.
41. Neduva, V. and R.B. Russell, *Peptides mediating interaction networks: new leads at last*. Curr Opin Biotechnol, 2006. **17**(5): p. 465-71.
42. Stein, A. and P. Aloy, *Contextual specificity in peptide-mediated protein interactions*. PLoS One, 2008. **3**(7): p. e2524.
43. Neduva, V., et al., *Systematic discovery of new recognition peptides mediating protein interaction networks*. PLoS Biol, 2005. **3**(12): p. e405.
44. Smoluchowski, M.V., *Versuch einer mathematischen Theorie der Koagulationskinetik kolloidaler Lösungen* Z Phys Chem, 1917. **92**(2): p. 129-68.

45. Northrup, S.H. and H.P. Erickson, *Kinetics of protein-protein association explained by Brownian dynamics computer simulation*. Proc Natl Acad Sci U S A, 1992. **89**(8): p. 3338-42.
46. Schreiber, G., *Kinetic studies of protein-protein interactions*. Curr Opin Struct Biol, 2002. **12**(1): p. 41-7.
47. Berg, O.G. and P.H. von Hippel, *Diffusion-controlled macromolecular interactions*. Annu Rev Biophys Chem, 1985. **14**: p. 131-60.
48. Sheinerman, F.B., R. Norel, and B. Honig, *Electrostatic aspects of protein-protein interactions*. Curr Opin Struct Biol, 2000. **10**(2): p. 153-9.
49. Selzer, T. and G. Schreiber, *Predicting the rate enhancement of protein complex formation from the electrostatic energy of interaction*. J Mol Biol, 1999. **287**(2): p. 409-19.
50. Selzer, T., S. Albeck, and G. Schreiber, *Rational design of faster associating and tighter binding protein complexes*. Nat Struct Biol, 2000. **7**(7): p. 537-41.
51. Piehler, J. and G. Schreiber, *Biophysical analysis of the interaction of human ifnar2 expressed in E. coli with IFNalpha2*. J Mol Biol, 1999. **289**(1): p. 57-67.
52. Vijayakumar, M., et al., *Electrostatic enhancement of diffusion-controlled protein-protein association: comparison of theory and experiment on barnase and barstar*. J Mol Biol, 1998. **278**(5): p. 1015-24.
53. Wendt, H., et al., *Very rapid, ionic strength-dependent association and folding of a heterodimeric leucine zipper*. Biochemistry, 1997. **36**(1): p. 204-13.
54. Sydor, J.R., et al., *Transient kinetic studies on the interaction of Ras and the Ras-binding domain of c-Raf-1 reveal rapid equilibration of the complex*. Biochemistry, 1998. **37**(40): p. 14292-9.
55. Estrada, S., et al., *The N-terminal region of cystatin A (stefin A) binds to papain subsequent to the two hairpin loops of the inhibitor. Demonstration of two-step binding by rapid-kinetic studies of cystatin A labeled at the N-terminus with a fluorescent reporter group*. Protein Sci, 2000. **9**(11): p. 2218-24.
56. Li, Y., et al., *Mutations of an epitope hot-spot residue alter rate limiting steps of antigen-antibody protein-protein associations*. Biochemistry, 2001. **40**(7): p. 2011-22.
57. Baugh, R.J. and C.G. Trowbridge, *Calorimetry of some trypsin-trypsin inhibitor reactions*. J Biol Chem, 1972. **247**(23): p. 7498-501.
58. Scheraga, H.A., G. Nemethy, and I.Z. Steinberg, *The contribution of hydrophobic bonds to the thermal stability of protein conformations*. J Biol Chem, 1962. **237**: p. 2506-8.
59. Ross, P.D. and M.V. Rekharsky, *Thermodynamics of hydrogen bond and hydrophobic interactions in cyclodextrin complexes*. Biophys J, 1996. **71**(4): p. 2144-54.
60. Prabhu, N.V. and K.A. Sharp, *Heat capacity in proteins*. Annu Rev Phys Chem, 2005. **56**: p. 521-48.
61. Jin, R., et al., *Botulinum neurotoxin B recognizes its protein receptor with high affinity and specificity*. Nature, 2006. **444**(7122): p. 1092-5.
62. Zhou, Y.L., et al., *Thermodynamics of the interaction of xanthine oxidase with superoxide dismutase studied by isothermal titration calorimetry and fluorescence spectroscopy*. Thermochim Acta, 2005. **426**(1.2): p. 173-78.
63. Liang, Y., *Applications of isothermal titration calorimetry in protein science*. Acta Biochim Biophys Sin (Shanghai), 2008. **40**(7): p. 565-76.
64. Velazquez Campoy, A. and E. Freire, *ITC in the post-genomic era...? Priceless*. Biophys Chem, 2005. **115**(2-3): p. 115-24.
65. Doyle, M.L., *Characterization of binding interactions by isothermal titration calorimetry*. Curr Opin Biotechnol, 1997. **8**(1): p. 31-5.
66. Kumar, M.D. and M.M. Gromiha, *PINT: Protein-protein Interactions Thermodynamic Database*. Nucleic Acids Res, 2006. **34**(Database issue): p. D195-8.
67. Bogan, A.A. and K.S. Thorn, *Anatomy of hot spots in protein interfaces*. J Mol Biol, 1998. **280**(1): p. 1-9.

68. Clackson, T. and J.A. Wells, *A hot spot of binding energy in a hormone-receptor interface*. Science, 1995. **267**(5196): p. 383-6.
69. Wells, J.A., *Systematic mutational analyses of protein-protein interfaces*. Methods Enzymol, 1991. **202**: p. 390-411.
70. Wells, J.A., *Additivity of mutational effects in proteins*. Biochemistry, 1990. **29**(37): p. 8509-17.
71. Morrison, K.L. and G.A. Weiss, *Combinatorial alanine-scanning*. Curr Opin Chem Biol, 2001. **5**(3): p. 302-7.
72. Jones, D.H., et al., *Reversible dissociation of dimeric tyrosyl-tRNA synthetase by mutagenesis at the subunit interface*. Biochemistry, 1985. **24**(21): p. 5852-7.
73. Ward, W.H., D.H. Jones, and A.R. Fersht, *Effects of engineering complementary charged residues into the hydrophobic subunit interface of tyrosyl-tRNA synthetase. Appendix: Kinetic analysis of dimeric enzymes that reversibly dissociate into inactive subunits*. Biochemistry, 1987. **26**(13): p. 4131-8.
74. Hartman, F.C., et al., *Function of Lys-166 of Rhodospirillum rubrum ribulosebisphosphate carboxylase/oxygenase as examined by site-directed mutagenesis*. J Biol Chem, 1987. **262**(8): p. 3496-501.
75. Brange, J., et al., *Monomeric insulins obtained by protein engineering and their medical implications*. Nature, 1988. **333**(6174): p. 679-82.
76. Borchert, T.V., et al., *The crystal structure of an engineered monomeric triosephosphate isomerase, monoTIM: the correct modelling of an eight-residue loop*. Structure, 1993. **1**(3): p. 205-13.
77. Beernink, P.T. and D.R. Tolan, *Disruption of the aldolase A tetramer into catalytically active monomers*. Proc Natl Acad Sci U S A, 1996. **93**(11): p. 5374-9.
78. Roberts, S., J.C. Cheetham, and A.R. Rees, *Generation of an antibody with enhanced affinity and specificity for its antigen by protein engineering*. Nature, 1987. **328**(6132): p. 731-4.
79. Thorn, K.S. and A.A. Bogan, *ASEdb: a database of alanine mutations and their effects on the free energy of binding in protein interactions*. Bioinformatics, 2001. **17**(3): p. 284-5.
80. Samanta, U., D. Pal, and P. Chakrabarti, *Environment of tryptophan side chains in proteins*. Proteins, 2000. **38**(3): p. 288-300.
81. DeLano, W.L., *Unraveling hot spots in binding interfaces: progress and challenges*. Curr Opin Struct Biol, 2002. **12**(1): p. 14-20.
82. Pal, G., et al., *Intramolecular cooperativity in a protein binding site assessed by combinatorial shotgun scanning mutagenesis*. J Mol Biol, 2005. **347**(3): p. 489-94.
83. Horovitz, A., *Double-mutant cycles: a powerful tool for analyzing protein structure and function*. Fold Des, 1996. **1**(6): p. R121-6.
84. Gregoret, L.M. and R.T. Sauer, *Additivity of mutant effects assessed by binomial mutagenesis*. Proc Natl Acad Sci U S A, 1993. **90**(9): p. 4246-50.
85. Bernat, B., et al., *Dissecting the binding energy epitope of a high-affinity variant of human growth hormone: cooperative and additive effects from combining mutations from independently selected phage display mutagenesis libraries*. Biochemistry, 2004. **43**(20): p. 6076-84.
86. Greenspan, N.S. and E. Di Cera, *Defining epitopes: It's not as easy as it seems*. Nat Biotechnol, 1999. **17**(10): p. 936-7.
87. Buczek, O., et al., *Analysis of serine proteinase-inhibitor interaction by alanine shaving*. Protein Sci, 2002. **11**(4): p. 806-19.
88. Kortemme, T., D.E. Kim, and D. Baker, *Computational alanine scanning of protein-protein interfaces*. Sci STKE, 2004. **2004**(219): p. pl2.
89. Kortemme, T. and D. Baker, *A simple physical model for binding energy hot spots in protein-protein complexes*. Proc Natl Acad Sci U S A, 2002. **99**(22): p. 14116-21.

90. McFarland, B.J., et al., *Symmetry recognizing asymmetry: analysis of the interactions between the C-type lectin-like immunoreceptor NKG2D and MHC class I-like ligands*. Structure, 2003. **11**(4): p. 411-22.
91. Boulanger, M.J., et al., *Convergent mechanisms for recognition of divergent cytokines by the shared signaling receptor gp130*. Mol Cell, 2003. **12**(3): p. 577-89.
92. Kollman, P.A., et al., *Calculating structures and free energies of complex molecules: combining molecular mechanics and continuum models*. Acc Chem Res, 2000. **33**(12): p. 889-97.
93. Knight, J.L. and C.L. Brooks, 3rd, *Lambda-dynamics free energy simulation methods*. J Comput Chem, 2009. **30**(11): p. 1692-700.
94. Fernandez-Recio, J., M. Totrov, and R. Abagyan, *Identification of protein-protein interaction sites from docking energy landscapes*. J Mol Biol, 2004. **335**(3): p. 843-65.
95. Moreira, I.S., P.A. Fernandes, and M.J. Ramos, *Unraveling the importance of protein-protein interaction: application of a computational alanine-scanning mutagenesis to the study of the IgG1 streptococcal protein G (C2 fragment) complex*. J Phys Chem B, 2006. **110**(22): p. 10962-9.
96. Gouda, H., et al., *Free energy calculations for theophylline binding to an RNA aptamer: Comparison of MM-PBSA and thermodynamic integration methods*. Biopolymers, 2003. **68**(1): p. 16-34.
97. Salgado, E.N., J. Faraone-Mennella, and F.A. Tezcan, *Controlling protein-protein interactions through metal coordination: assembly of a 16-helix bundle protein*. J Am Chem Soc, 2007. **129**(44): p. 13374-5.
98. Berridge, M.J., P. Lipp, and M.D. Bootman, *The versatility and universality of calcium signalling*. Nat Rev Mol Cell Biol, 2000. **1**(1): p. 11-21.
99. Norris, V., et al., *Calcium signalling in bacteria*. J Bacteriol, 1996. **178**(13): p. 3677-82.
100. Gangola, P. and B.P. Rosen, *Maintenance of intracellular calcium in Escherichia coli*. J Biol Chem, 1987. **262**(26): p. 12570-4.
101. Tisa, L.S. and J. Adler, *Cytoplasmic free-Ca<sup>2+</sup> level rises with repellents and falls with attractants in Escherichia coli chemotaxis*. Proc Natl Acad Sci U S A, 1995. **92**(23): p. 10777-81.
102. Swulius, M.T. and M.N. Waxham, *Ca(2+)/calmodulin-dependent protein kinases*. Cell Mol Life Sci, 2008. **65**(17): p. 2637-57.
103. Chin, D. and A.R. Means, *Calmodulin: a prototypical calcium sensor*. Trends Cell Biol, 2000. **10**(8): p. 322-8.
104. Tanaka, T., *Calmodulin-dependent calcium signal transduction*. Jpn J Pharmacol, 1988. **46**(2): p. 101-7.
105. Yazawa, M., F. Matsuzawa, and K. Yagi, *Inter-domain interaction and the structural flexibility of calmodulin in the connecting region of the terminal two domains*. J Biochem, 1990. **107**(2): p. 287-91.
106. Schumacher, M.A., et al., *Structure of the gating domain of a Ca<sup>2+</sup>-activated K<sup>+</sup> channel complexed with Ca<sup>2+</sup>/calmodulin*. Nature, 2001. **410**(6832): p. 1120-4.
107. Wall, M.E., J.B. Clarage, and G.N. Phillips, *Motions of calmodulin characterized using both Bragg and diffuse X-ray scattering*. Structure, 1997. **5**(12): p. 1599-612.
108. Osawa, M., et al., *A novel target recognition revealed by calmodulin in complex with Ca<sup>2+</sup>-calmodulin-dependent kinase kinase*. Nat Struct Biol, 1999. **6**(9): p. 819-24.
109. Drum, C.L., et al., *Structural basis for the activation of anthrax adenyl cyclase exotoxin by calmodulin*. Nature, 2002. **415**(6870): p. 396-402.
110. Gumbiner, B.M., *Regulation of cadherin-mediated adhesion in morphogenesis*. Nat Rev Mol Cell Biol, 2005. **6**(8): p. 622-34.
111. Nagar, B., et al., *Structural basis of calcium-induced E-cadherin rigidification and dimerization*. Nature, 1996. **380**(6572): p. 360-4.

112. Rae, T.D., et al., *Undetectable intracellular free copper: the requirement of a copper chaperone for superoxide dismutase*. Science, 1999. **284**(5415): p. 805-8.
113. Malmstrom, B.G. and J. Leckner, *The chemical biology of copper*. Curr Opin Chem Biol, 1998. **2**(2): p. 286-92.
114. Banci, L., et al., *The Atx1-Ccc2 complex is a metal-mediated protein-protein interaction*. Nat Chem Biol, 2006. **2**(7): p. 367-8.
115. Dancis, A., et al., *Molecular characterization of a copper transport protein in S. cerevisiae: an unexpected role for copper in iron transport*. Cell, 1994. **76**(2): p. 393-402.
116. Askwith, C., et al., *The FET3 gene of S. cerevisiae encodes a multicopper oxidase required for ferrous iron uptake*. Cell, 1994. **76**(2): p. 403-10.
117. Huffman, D.L. and T.V. O'Halloran, *Energetics of copper trafficking between the Atx1 metallochaperone and the intracellular copper transporter, Ccc2*. J Biol Chem, 2000. **275**(25): p. 18611-4.
118. Pufahl, R.A., et al., *Metal ion chaperone function of the soluble Cu(I) receptor Atx1*. Science, 1997. **278**(5339): p. 853-6.
119. Arnesano, F., et al., *Solution structure of the Cu(I) and apo forms of the yeast metallochaperone, Atx1*. Biochemistry, 2001. **40**(6): p. 1528-39.
120. Banci, L., et al., *Solution structure of the yeast copper transporter domain Ccc2a in the apo and Cu(I)-loaded states*. J Biol Chem, 2001. **276**(11): p. 8415-26.
121. Hung, I.H., et al., *HAH1 is a copper-binding protein with distinct amino acid residues mediating copper homeostasis and antioxidant defense*. J Biol Chem, 1998. **273**(3): p. 1749-54.
122. Lutsenko, S., et al., *Human copper-transporting ATPase ATP7B (the Wilson's disease protein): biochemical properties and regulation*. J Bioenerg Biomembr, 2002. **34**(5): p. 351-62.
123. Tanchou, V., et al., *Copper-mediated homo-dimerisation for the HAH1 metallochaperone*. Biochem Biophys Res Commun, 2004. **325**(2): p. 388-94.
124. Wernimont, A.K., et al., *Structural basis for copper transfer by the metallochaperone for the Menkes/Wilson disease proteins*. Nat Struct Biol, 2000. **7**(9): p. 766-71.
125. Banci, L., et al., *Understanding copper trafficking in bacteria: interaction between the copper transport protein CopZ and the N-terminal domain of the copper ATPase CopA from Bacillus subtilis*. Biochemistry, 2003. **42**(7): p. 1939-49.
126. Auld, D.S., *Zinc coordination sphere in biochemical zinc sites*. Biometals, 2001. **14**(3-4): p. 271-313.
127. Golovin, A., et al., *MSDsite: a database search and retrieval system for the analysis and viewing of bound ligands and active sites*. Proteins, 2005. **58**(1): p. 190-9.
128. Maret, W., *Zinc coordination environments in proteins determine zinc functions*. J Trace Elem Med Biol, 2005. **19**(1): p. 7-12.
129. Papageorgiou, A.C. and K.R. Acharya, *Microbial superantigens: from structure to function*. Trends Microbiol, 2000. **8**(8): p. 369-75.
130. Bohach, G.A., et al., *Staphylococcal and streptococcal pyrogenic toxins involved in toxic shock syndrome and related illnesses*. Crit Rev Microbiol, 1990. **17**(4): p. 251-72.
131. Chi, Y.I., et al., *Zinc-mediated dimerization and its effect on activity and conformation of staphylococcal enterotoxin type C*. J Biol Chem, 2002. **277**(25): p. 22839-46.
132. Papageorgiou, A.C., et al., *Crystal structure of the superantigen enterotoxin C2 from Staphylococcus aureus reveals a zinc-binding site*. Structure, 1995. **3**(8): p. 769-79.
133. Miller, J., A.D. McLachlan, and A. Klug, *Repetitive zinc-binding domains in the protein transcription factor IIIA from Xenopus oocytes*. EMBO J, 1985. **4**(6): p. 1609-14.
134. Mackay, J.P. and M. Crossley, *Zinc fingers are sticking together*. Trends Biochem Sci, 1998. **23**(1): p. 1-4.

135. Brayer, K.J., S. Kulshreshtha, and D.J. Segal, *The protein-binding potential of C2H2 zinc finger domains*. Cell Biochem Biophys, 2008. **51**(1): p. 9-19.
136. Brayer, K.J. and D.J. Segal, *Keep your fingers off my DNA: protein-protein interactions mediated by C2H2 zinc finger domains*. Cell Biochem Biophys, 2008. **50**(3): p. 111-31.
137. Gamsjaeger, R., et al., *Sticky fingers: zinc-fingers as protein-recognition motifs*. Trends Biochem Sci, 2007. **32**(2): p. 63-70.
138. Liew, C.K., et al., *Zinc fingers as protein recognition motifs: structural basis for the GATA-1/friend of GATA interaction*. Proc Natl Acad Sci U S A, 2005. **102**(3): p. 583-8.
139. Alam, S.L., et al., *Ubiquitin interactions of NZF zinc fingers*. EMBO J, 2004. **23**(7): p. 1411-21.
140. Reyes-Turcu, F.E., et al., *The ubiquitin binding domain ZnF UBP recognizes the C-terminal diglycine motif of unanchored ubiquitin*. Cell, 2006. **124**(6): p. 1197-208.
141. Opipari, A.W., Jr., M.S. Boguski, and V.M. Dixit, *The A20 cDNA induced by tumor necrosis factor alpha encodes a novel type of zinc finger protein*. J Biol Chem, 1990. **265**(25): p. 14705-8.
142. Freyd, G., S.K. Kim, and H.R. Horvitz, *Novel cysteine-rich motif and homeodomain in the product of the Caenorhabditis elegans cell lineage gene lin-11*. Nature, 1990. **344**(6269): p. 876-9.
143. Gross, C.T. and W. McGinnis, *DEAF-1, a novel protein that binds an essential region in a Deformed response element*. EMBO J, 1996. **15**(8): p. 1961-70.
144. Kadrmaz, J.L. and M.C. Beckerle, *The LIM domain: from the cytoskeleton to the nucleus*. Nat Rev Mol Cell Biol, 2004. **5**(11): p. 920-31.
145. Liu, Y., et al., *Structural basis for recognition of SMRT/N-CoR by the MYND domain and its contribution to AML1/ETO's activity*. Cancer Cell, 2007. **11**(6): p. 483-97.
146. Blobel, G.A., *CREB-binding protein and p300: molecular integrators of hematopoietic transcription*. Blood, 2000. **95**(3): p. 745-55.
147. Freedman, S.J., et al., *Structural basis for recruitment of CBP/p300 by hypoxia-inducible factor-1 alpha*. Proc Natl Acad Sci U S A, 2002. **99**(8): p. 5367-72.
148. Freedman, S.J., et al., *Structural basis for negative regulation of hypoxia-inducible factor-1alpha by CITED2*. Nat Struct Biol, 2003. **10**(7): p. 504-12.
149. Andreini, C., et al., *Metal ions in biological catalysis: from enzyme databases to general principles*. J Biol Inorg Chem, 2008. **13**(8): p. 1205-18.
150. Ma, Z., F.E. Jacobsen, and D.P. Giedroc, *Coordination chemistry of bacterial metal transport and sensing*. Chem Rev, 2009. **109**(10): p. 4644-81.
151. Rees, D.C., E. Johnson, and O. Lewinson, *ABC transporters: the power to change*. Nat Rev Mol Cell Biol, 2009. **10**(3): p. 218-27.
152. Kuhlbrandt, W., *Biology, structure and mechanism of P-type ATPases*. Nat Rev Mol Cell Biol, 2004. **5**(4): p. 282-95.
153. Nevo, Y. and N. Nelson, *The NRAMP family of metal-ion transporters*. Biochim Biophys Acta, 2006. **1763**(7): p. 609-20.
154. Montanini, B., et al., *Phylogenetic and functional analysis of the Cation Diffusion Facilitator (CDF) family: improved signature and prediction of substrate specificity*. BMC Genomics, 2007. **8**: p. 107.
155. Murakami, S., *Multidrug efflux transporter, AcrB--the pumping mechanism*. Curr Opin Struct Biol, 2008. **18**(4): p. 459-65.
156. Davidson, A.L., et al., *Structure, function, and evolution of bacterial ATP-binding cassette systems*. Microbiol Mol Biol Rev, 2008. **72**(2): p. 317-64, table of contents.
157. Lu, M. and D. Fu, *Structure of the zinc transporter YiiP*. Science, 2007. **317**(5845): p. 1746-8.
158. Toyoshima, C. and H. Nomura, *Structural changes in the calcium pump accompanying the dissociation of calcium*. Nature, 2002. **418**(6898): p. 605-11.

159. Wu, C.C., W.J. Rice, and D.L. Stokes, *Structure of a copper pump suggests a regulatory role for its metal-binding domain*. *Structure*, 2008. **16**(6): p. 976-85.
160. Chen, P.R. and C. He, *Selective recognition of metal ions by metalloregulatory proteins*. *Curr Opin Chem Biol*, 2008. **12**(2): p. 214-21.
161. San Francisco, M.J., et al., *Identification of the metalloregulatory element of the plasmid-encoded arsenical resistance operon*. *Nucleic Acids Res*, 1990. **18**(3): p. 619-24.
162. Morby, A.P., et al., *SmtB is a metal-dependent repressor of the cyanobacterial metallothionein gene smtA: identification of a Zn inhibited DNA-protein complex*. *Nucleic Acids Res*, 1993. **21**(4): p. 921-5.
163. Brown, N.L., et al., *The MerR family of transcriptional regulators*. *FEMS Microbiol Rev*, 2003. **27**(2-3): p. 145-63.
164. Liu, T., et al., *CsoR is a novel Mycobacterium tuberculosis copper-sensing transcriptional regulator*. *Nat Chem Biol*, 2007. **3**(1): p. 60-8.
165. Wunderli-Ye, H. and M. Solioz, *Copper homeostasis in Enterococcus hirae*. *Adv Exp Med Biol*, 1999. **448**: p. 255-64.
166. Orth, P., et al., *Structural basis of gene regulation by the tetracycline inducible Tet repressor-operator system*. *Nat Struct Biol*, 2000. **7**(3): p. 215-9.
167. Hantke, K., *Iron and metal regulation in bacteria*. *Curr Opin Microbiol*, 2001. **4**(2): p. 172-7.
168. Ding, X., et al., *Identification of the primary metal ion-activation sites of the diphtheria tox repressor by X-ray crystallography and site-directed mutational analysis*. *Nat Struct Biol*, 1996. **3**(4): p. 382-7.
169. Chivers, P.T. and R.T. Sauer, *NikR is a ribbon-helix-helix DNA-binding protein*. *Protein Sci*, 1999. **8** (11): p. 2494-500.
170. Saridakis, V., et al., *Structural insight on the mechanism of regulation of the MarR family of proteins: high-resolution crystal structure of a transcriptional repressor from Methanobacterium thermoautotrophicum*. *J Mol Biol*, 2008. **377**(3): p. 655-67.
171. Hall, D.R., et al., *The high-resolution crystal structure of the molybdate-dependent transcriptional regulator (ModE) from Escherichia coli: a novel combination of domain folds*. *EMBO J*, 1999. **18**(6): p. 1435-46.
172. Watanabe, S., et al., *Crystal structure of the [2Fe-2S] oxidative-stress sensor SoxR bound to DNA*. *Proc Natl Acad Sci U S A*, 2008. **105**(11): p. 4121-6.
173. Traore, D.A., et al., *Crystal structure of the apo-PerR-Zn protein from Bacillus subtilis*. *Mol Microbiol*, 2006. **61**(5): p. 1211-9.
174. O'Halloran, T.V. and V.C. Culotta, *Metallochaperones, an intracellular shuttle service for metal ions*. *J Biol Chem*, 2000. **275**(33): p. 25057-60.
175. Rosenzweig, A.C., *Metallochaperones: bind and deliver*. *Chem Biol*, 2002. **9**(6): p. 673-7.
176. Huffman, D.L. and T.V. O'Halloran, *Function, structure, and mechanism of intracellular copper trafficking proteins*. *Annu Rev Biochem*, 2001. **70**: p. 677-701.
177. Outten, C.E. and T.V. O'Halloran, *Femtomolar sensitivity of metalloregulatory proteins controlling zinc homeostasis*. *Science*, 2001. **292**(5526): p. 2488-92.
178. Andrews, N.C., *Metal transporters and disease*. *Curr Opin Chem Biol*, 2002. **6**(2): p. 181-6.
179. Luk, E., L.T. Jensen, and V.C. Culotta, *The many highways for intracellular trafficking of metals*. *J Biol Inorg Chem*, 2003. **8**(8): p. 803-9.
180. Bruce, M.G. and H.I. Maarsoos, *Epidemiology of Helicobacter pylori infection*. *Helicobacter*, 2008. **13 Suppl 1**: p. 1-6.
181. Ferreira, A.C., et al., *Helicobacter and gastric malignancies*. *Helicobacter*, 2008. **13 Suppl 1**: p. 28-34.
182. Marshall, B.J. and J.R. Warren, *Unidentified curved bacilli in the stomach of patients with gastritis and peptic ulceration*. *Lancet*, 1984. **1**(8390): p. 1311-5.
183. Marshall, B., *Helicobacter pylori: 20 years on*. *Clin Med*, 2002. **2**(2): p. 147-52.

184. Scott, D., et al., *The life and death of Helicobacter pylori*. Gut, 1998. **43 Suppl 1**: p. S56-60.
185. Eaton, K.A. and S. Krakowka, *Effect of gastric pH on urease-dependent colonization of gnotobiotic piglets by Helicobacter pylori*. Infect Immun, 1994. **62**(9): p. 3604-7.
186. Ha, N.C., et al., *Supramolecular assembly and acid resistance of Helicobacter pylori urease*. Nat Struct Biol, 2001. **8**(6): p. 505-9.
187. Olson, J.W. and R.J. Maier, *Molecular hydrogen as an energy source for Helicobacter pylori*. Science, 2002. **298**(5599): p. 1788-90.
188. Burne, R.A. and Y.Y. Chen, *Bacterial ureases in infectious diseases*. Microbes Infect, 2000. **2**(5): p. 533-42.
189. Wen, Y., et al., *Acid-adaptive genes of Helicobacter pylori*. Infect Immun, 2003. **71**(10): p. 5921-39.
190. Benoit, S.L. and R.J. Maier, *Hydrogen and nickel metabolism in helicobacter species*. Ann N Y Acad Sci, 2008. **1125**: p. 242-51.
191. Testerman, T.L., et al., *Nutritional requirements and antibiotic resistance patterns of Helicobacter species in chemically defined media*. J Clin Microbiol, 2006. **44**(5): p. 1650-8.
192. Wang, G., P. Alamuri, and R.J. Maier, *The diverse antioxidant systems of Helicobacter pylori*. Mol Microbiol, 2006. **61**(4): p. 847-60.
193. Hausinger, R.P. and P.A. Karplus, *Urease*, in *Handbook of Metalloproteins*, A. Messerschmidt, et al., Editors. 2001, John Wiley & Sons: Chichester, U.K. p. 867-879.
194. Ciurli, S., *Urease: recent insights on the role of nickel*, in *Nickel and Its Surprising Impact in Nature: Metal Ions in Life Sciences*, A. Sigel, H. Sigel, and R.K.O. Sigel, Editors. 2007, John Wiley & Sons Ltd., West Sussex, United Kingdom. p. 241-277.
195. Callahan, B.P., Y. Yuan, and R. Wolfenden, *The burden borne by urease*. J Am Chem Soc, 2005. **127**(31): p. 10828-9.
196. Ciurli, S. and S. Mangani, *Nickel-containing enzymes*, in *Handbook on Metalloproteins*, I. Bertini, A. Sigel, and S. H., Editors. 2001, Marcel Dekker: New York, USA. p. 669-708.
197. Sumner, J.B., *The isolation and crystallization of the enzyme urease*. J Biol Chem, 1926. **69**: p. 435-41.
198. Dixon, N.E., et al., *Letter: Jack bean urease (EC 3.5.1.5). A metalloenzyme. A simple biological role for nickel?* J Am Chem Soc, 1975. **97**(14): p. 4131-3.
199. Mobley, H.L., M.D. Island, and R.P. Hausinger, *Molecular biology of microbial ureases*. Microbiol Rev, 1995. **59**(3): p. 451-80.
200. Jabri, E., et al., *The crystal structure of urease from Klebsiella aerogenes*. Science, 1995. **268**(5213): p. 998-1004.
201. Benini, S., et al., *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, 1999. **7**(2): p. 205-16.
202. Moncrief, M.B., et al., *Urease activity in the crystalline state*. Protein Sci, 1995. **4**(10): p. 2234-6.
203. Benini, S., et al., *Molecular details of urease inhibition by boric acid: insights into the catalytic mechanism*. J Am Chem Soc, 2004. **126**(12): p. 3714-5.
204. Karplus, P.A., M.A. Pearson, and R.P. Hausinger, *70 years of crystalline urease: what have we learned?* Acc Chem Res, 1997. **30**: p. 330-7.
205. Musiani, F., et al., *Structure-based computational study of the catalytic and inhibition mechanisms of urease*. J Biol Inorg Chem, 2001. **6**(3): p. 300-14.
206. Lubitz, W., M.V. Gastel, and W. Gartner, *Nickel iron hydrogenases*, in *Nickel and Its Surprising Impact in Nature*, A. Sigel, H. Sigel, and R.K.O. Sigel, Editors. 2007, John Wiley & Sons Ltd.: West Sussex, UK. p. 279-322.

207. Vignais, P.M. and B. Billoud, *Occurrence, classification, and biological function of hydrogenases: an overview*. Chem Rev, 2007. **107**(10): p. 4206-72.
208. Fontecilla-Camps, J.C., et al., *Structure/function relationships of [NiFe]- and [FeFe]-hydrogenases*. Chem Rev, 2007. **107**(10): p. 4273-303.
209. Lyon, E.J., et al., *UV-A/blue-light inactivation of the 'metal-free' hydrogenase (Hmd) from methanogenic archaea*. Eur J Biochem, 2004. **271**(1): p. 195-204.
210. Shima, S. and R.K. Thauer, *A third type of hydrogenase catalyzing H<sub>2</sub> activation*. Chem Rec, 2007. **7**(1): p. 37-46.
211. Garcin, E., et al., *The crystal structure of a reduced [NiFeSe] hydrogenase provides an image of the activated catalytic center*. Structure, 1999. **7**(5): p. 557-66.
212. Volbeda, A., et al., *Crystal structure of the nickel-iron hydrogenase from Desulfovibrio gigas*. Nature, 1995. **373**(6515): p. 580-7.
213. Higuchi, Y., T. Yagi, and N. Yasuoka, *Unusual ligand structure in Ni-Fe active center and an additional Mg site in hydrogenase revealed by high resolution X-ray structure analysis*. Structure, 1997. **5**(12): p. 1671-80.
214. Montet, Y., et al., *Gas access to the active site of Ni-Fe hydrogenases probed by X-ray crystallography and molecular dynamics*. Nat Struct Biol, 1997. **4**(7): p. 523-6.
215. Matias, P.M., et al., *[NiFe] hydrogenase from Desulfovibrio desulfuricans ATCC 27774: gene sequencing, three-dimensional structure determination and refinement at 1.8 Å and modelling studies of its interaction with the tetrahaem cytochrome c<sub>3</sub>*. J Biol Inorg Chem, 2001. **6**(1): p. 63-81.
216. Lill, S.O. and P.E. Siegbahn, *An autocatalytic mechanism for NiFe-hydrogenase: reduction to Ni(I) followed by oxidative addition*. Biochemistry, 2009. **48**(5): p. 1056-66.
217. Lee, M.H., S.B. Mulrooney, and R.P. Hausinger, *Purification, characterization, and in vivo reconstitution of Klebsiella aerogenes urease apoenzyme*. J Bacteriol, 1990. **172**(8): p. 4427-31.
218. Park, I.S. and R.P. Hausinger, *Requirement of carbon dioxide for in vitro assembly of the urease nickel metallocenter*. Science, 1995. **267**(5201): p. 1156-8.
219. Mulrooney, S.B. and R.P. Hausinger, *Sequence of the Klebsiella aerogenes urease genes and evidence for accessory proteins facilitating nickel incorporation*. J Bacteriol, 1990. **172**(10): p. 5837-43.
220. de Koning-Ward, T.F., A.C. Ward, and R.M. Robins-Browne, *Characterisation of the urease-encoding gene complex of Yersinia enterocolitica*. Gene, 1994. **145**(1): p. 25-32.
221. Toffanin, A., et al., *Characterization of the urease gene cluster from Rhizobium leguminosarum bv. viciae*. Arch Microbiol, 2002. **177**(4): p. 290-8.
222. Bosse, J.T., H.D. Gilmour, and J.I. MacInnes, *Novel genes affecting urease activity in Actinobacillus pleuropneumoniae*. J Bacteriol, 2001. **183**(4): p. 1242-7.
223. Palinska, K.A., et al., *Prochlorococcus marinus strain PCC 9511, a picoplanktonic cyanobacterium, synthesizes the smallest urease*. Microbiology, 2000. **146 Pt 12**: p. 3099-107.
224. Labigne, A., V. Cussac, and P. Courcoux, *Shuttle cloning and nucleotide sequences of Helicobacter pylori genes responsible for urease activity*. J Bacteriol, 1991. **173**(6): p. 1920-31.
225. Weeks, D.L., et al., *A H<sup>+</sup>-gated urea channel: the link between Helicobacter pylori urease and gastric colonization*. Science, 2000. **287**(5452): p. 482-5.
226. Cussac, V., R.L. Ferrero, and A. Labigne, *Expression of Helicobacter pylori urease genes in Escherichia coli grown under nitrogen-limiting conditions*. J Bacteriol, 1992. **174**(8): p. 2466-73.
227. Stoof, J., et al., *Inverse nickel-responsive regulation of two urease enzymes in the gastric pathogen Helicobacter mustelae*. Environ Microbiol, 2008. **10**(10): p. 2586-97.
228. Reytrat, J.M., F.X. Berthet, and B. Gicquel, *The urease locus of Mycobacterium tuberculosis and its utilization for the demonstration of allelic exchange in Mycobacterium bovis bacillus Calmette-Guerin*. Proc Natl Acad Sci U S A, 1995. **92**(19): p. 8768-72.

229. Kim, J.K., S.B. Mulrooney, and R.P. Hausinger, *Biosynthesis of active Bacillus subtilis urease in the absence of known urease accessory proteins*. J Bacteriol, 2005. **187**(20): p. 7150-4.
230. Lee, M.H., et al., *Klebsiella aerogenes urease gene cluster: sequence of ureD and demonstration that four accessory genes (ureD, ureE, ureF, and ureG) are involved in nickel metallocenter biosynthesis*. J Bacteriol, 1992. **174**(13): p. 4324-30.
231. Mulrooney, S.B., S.K. Ward, and R.P. Hausinger, *Purification and properties of the Klebsiella aerogenes UreE metal-binding domain, a functional metallochaperone of urease*. J Bacteriol, 2005. **187**(10): p. 3581-5.
232. Walz, S.E., et al., *Multiple proteins encoded within the urease gene complex of Proteus mirabilis*. J Bacteriol, 1988. **170**(3): p. 1027-33.
233. Gerlach, G.F., S. Clegg, and W.A. Nichols, *Characterization of the genes encoding urease activity of Klebsiella pneumoniae*. FEMS Microbiol Lett, 1988. **50**: p. 131-5.
234. Mulrooney, S.B., et al., *Purification, characterization, and genetic organization of recombinant Providencia stuartii urease expressed by Escherichia coli*. J Bacteriol, 1988. **170**(5): p. 2202-7.
235. Quiroz, S., et al., *Chaperones of nickel metabolism*, in *Nickel and its Surprising Impact in Nature*, A. Sigel, H. Sigel, and R.K.O. Sigel, Editors. 2007, John Wiley & Sons, Ltd.: Chichester. p. 519-44.
236. Park, I.S., M.B. Carr, and R.P. Hausinger, *In vitro activation of urease apoprotein and role of UreD as a chaperone required for nickel metallocenter assembly*. Proc Natl Acad Sci U S A, 1994. **91**(8): p. 3233-7.
237. Chang, Z., 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, 2004. **279**(15): p. 15305-13.
238. Park, I.S. and R.P. Hausinger, *Metal ion interaction with urease and UreD-urease apoproteins*. Biochemistry, 1996. **35**(16): p. 5345-52.
239. Quiroz-Valenzuela, S., et al., *The structure of urease activation complexes examined by flexibility analysis, mutagenesis, and small-angle X-ray scattering*. Arch Biochem Biophys, 2008. **480**(1): p. 51-7.
240. Moncrief, M.B. and R.P. Hausinger, *Purification and activation properties of UreD-UreF-urease apoprotein complexes*. J Bacteriol, 1996. **178**(18): p. 5417-21.
241. Heimer, S.R. and H.L. Mobley, *Interaction of Proteus mirabilis urease apoenzyme and accessory proteins identified with yeast two-hybrid technology*. J Bacteriol, 2001. **183**(4): p. 1423-33.
242. Voland, P., et al., *Interactions among the seven Helicobacter pylori proteins encoded by the urease gene cluster*. Am J Physiol Gastrointest Liver Physiol, 2003. **284**(1): p. G96-G106.
243. Salomone-Stagni, M., et al., *A model-based proposal for the role of UreF as a GTPase-activating protein in the urease active site biosynthesis*. Proteins, 2007. **68**(3): p. 749-61.
244. Park, I.S. and R.P. Hausinger, *Evidence for the presence of urease apoprotein complexes containing UreD, UreF, and UreG in cells that are competent for in vivo enzyme activation*. J Bacteriol, 1995. **177**(8): p. 1947-51.
245. Soriano, A. 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, 1999. **96**(20): p. 11140-4.
246. Colpas, G.J. and R.P. Hausinger, *In vivo and in vitro kinetics of metal transfer by the Klebsiella aerogenes urease nickel metallochaperone, UreE*. J Biol Chem, 2000. **275**(15): p. 10731-7.
247. Moncrief, M.B. 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, 1997. **179**(13): p. 4081-6.
248. Zambelli, B., et al., *UreG, a chaperone in the urease assembly process, is an intrinsically unstructured GTPase that specifically binds Zn<sup>2+</sup>*. J Biol Chem, 2005. **280**(6): p. 4684-95.

249. Zambelli, B., et al., *Biochemical studies on Mycobacterium tuberculosis UreG and comparative modeling reveal structural and functional conservation among the bacterial UreG family*. *Biochemistry*, 2007. **46**(11): p. 3171-82.
250. Mehta, N., S. Benoit, and R.J. Maier, *Roles of conserved nucleotide-binding domains in accessory proteins, HypB and UreG, in the maturation of nickel-enzymes required for efficient Helicobacter pylori colonization*. *Microb Pathog*, 2003. **35**(5): p. 229-34.
251. Zambelli, B., et al., *Zn<sup>2+</sup>-linked dimerization of UreG from Helicobacter pylori, a chaperone involved in nickel trafficking and urease activation*. *Proteins*, 2009. **74**(1): p. 222-39.
252. Olson, J.W., N.S. Mehta, and R.J. Maier, *Requirement of nickel metabolism proteins HypA and HypB for full activity of both hydrogenase and urease in Helicobacter pylori*. *Mol Microbiol*, 2001. **39**(1): p. 176-82.
253. Gasper, R., A. Scrima, and A. Wittinghofer, *Structural insights into HypB, a GTP-binding protein that regulates metal binding*. *J Biol Chem*, 2006. **281**(37): p. 27492-502.
254. Lee, M.H., et al., *Purification and characterization of Klebsiella aerogenes UreE protein: a nickel-binding protein that functions in urease metallocenter assembly*. *Protein Sci*, 1993. **2**(6): p. 1042-52.
255. Sriwanthana, B., et al., *Single-step purification of Proteus mirabilis urease accessory protein UreE, a protein with a naturally occurring histidine tail, by nickel chelate affinity chromatography*. *J Bacteriol*, 1994. **176**(22): p. 6836-41.
256. Ciurli, S., et al., *Molecular characterization of Bacillus pasteurii UreE, a metal-binding chaperone for the assembly of the urease active site*. *J Biol Inorg Chem*, 2002. **7**(6): p. 623-31.
257. Benoit, S. and R.J. Maier, *Dependence of Helicobacter pylori urease activity on the nickel-sequestering ability of the UreE accessory protein*. *J Bacteriol*, 2003. **185**(16): p. 4787-95.
258. Colpas, G.J., et al., *Identification of metal-binding residues in the Klebsiella aerogenes urease nickel metallochaperone, UreE*. *Biochemistry*, 1999. **38**(13): p. 4078-88.
259. Grosseohme, N.E., et al., *Thermodynamics of Ni<sup>2+</sup>, Cu<sup>2+</sup>, and Zn<sup>2+</sup> binding to the urease metallochaperone UreE*. *Biochemistry*, 2007. **46**(37): p. 10506-16.
260. Stola, M., et al., *The nickel site of Bacillus pasteurii UreE, a urease metallo-chaperone, as revealed by metal-binding studies and X-ray absorption spectroscopy*. *Biochemistry*, 2006. **45**(20): p. 6495-509.
261. Brayman, T.G. and R.P. Hausinger, *Purification, characterization, and functional analysis of a truncated Klebsiella aerogenes UreE urease accessory protein lacking the histidine-rich carboxyl terminus*. *J Bacteriol*, 1996. **178**(18): p. 5410-6.
262. Remaut, H., et al., *Structural basis for Ni(2+) transport and assembly of the urease active site by the metallochaperone UreE from Bacillus pasteurii*. *J Biol Chem*, 2001. **276**(52): p. 49365-70.
263. Song, H.K., et al., *Crystal structure of Klebsiella aerogenes UreE, a nickel-binding metallochaperone for urease activation*. *J Biol Chem*, 2001. **276**(52): p. 49359-64.
264. Musiani, F., et al., *Nickel trafficking: insights into the fold and function of UreE, a urease metallochaperone*. *J Inorg Biochem*, 2004. **98**(5): p. 803-13.
265. Kim, K.Y., C.H. Yang, and M.H. Lee, *Expression of the recombinant Klebsiella aerogenes UreF protein as a MalE fusion*. *Arch Pharm Res*, 1999. **22**(3): p. 274-8.
266. Kim, J.K., S.B. Mulrooney, and R.P. Hausinger, *The UreEF fusion protein provides a soluble and functional form of the UreF urease accessory protein*. *J Bacteriol*, 2006. **188**(24): p. 8413-20.
267. Lam, R., et al., *Crystal structure of the Helicobacter pylori urease accessory protein UreF*. 2010.
268. Deckert, G., et al., *The complete genome of the hyperthermophilic bacterium Aquifex aeolicus*. *Nature*, 1998. **392**(6674): p. 353-8.
269. Du, L., et al., *Sequences, organization and analysis of the hupZMNOQRTV genes from the Azotobacter chroococcum hydrogenase gene cluster*. *J Mol Biol*, 1994. **243**(4): p. 549-57.

270. Maier, T. and A. Böck, *Nickel incorporation into hydrogenases*, in *Mechanisms of Metallocenter Assembly*, R.P. Hausinger, G.L. Eichhorn, and L.G. Marzilli, Editors. 1996, VCH Publishers: New York. p. 173-92.
271. Eitinger, T. and B. Friedrich, *Microbial nickel transport and incorporation into hydrogenases in Transition Metals in Microbial Metabolism* G. Winkelmann and C.J. Carrano, Editors. 1997, Harwood Academic Publishers. p. 235-56.
272. Menon, N.K., et al., *Cloning, sequencing, and mutational analysis of the *hyb* operon encoding *Escherichia coli* hydrogenase 2*. J Bacteriol, 1994. **176**(14): p. 4416-23.
273. Bernhard, M., et al., *The *Alcaligenes eutrophus* membrane-bound hydrogenase gene locus encodes functions involved in maturation and electron transport coupling*. J Bacteriol, 1996. **178**(15): p. 4522-9.
274. Sauter, M., R. Bohm, and A. Bock, *Mutational analysis of the operon (*hyc*) determining hydrogenase 3 formation in *Escherichia coli**. Mol Microbiol, 1992. **6**(11): p. 1523-32.
275. Jacobi, A., R. Rossmann, and A. Bock, *The *hyp* operon gene products are required for the maturation of catalytically active hydrogenase isoenzymes in *Escherichia coli**. Arch Microbiol, 1992. **158**(6): p. 444-51.
276. Wolf, I., et al., *Duplication of *hyp* genes involved in maturation of [NiFe] hydrogenases in *Alcaligenes eutrophus* H16*. Arch Microbiol, 1998. **170**(6): p. 451-9.
277. Bock, A., et al., *Maturation of hydrogenases*. Adv Microb Physiol, 2006. **51**: p. 1-71.
278. Winter, G., et al., *A model system for [NiFe] hydrogenase maturation studies: Purification of an active site-containing hydrogenase large subunit without small subunit*. FEBS Lett, 2005. **579**(20): p. 4292-6.
279. Paschos, A., R.S. Glass, and A. Bock, *Carbamoylphosphate requirement for synthesis of the active center of [NiFe]-hydrogenases*. FEBS Lett, 2001. **488**(1-2): p. 9-12.
280. Li, C., et al., *X-ray crystal structure of aminoimidazole ribonucleotide synthetase (PurM), from the *Escherichia coli* purine biosynthetic pathway at 2.5 Å resolution*. Structure, 1999. **7**(9): p. 1155-66.
281. Reissmann, S., et al., *Taming of a poison: biosynthesis of the NiFe-hydrogenase cyanide ligands*. Science, 2003. **299**(5609): p. 1067-70.
282. Jones, A.K., et al., *NiFe hydrogenase active site biosynthesis: identification of Hyp protein complexes in *Ralstonia eutropha**. Biochemistry, 2004. **43**(42): p. 13467-77.
283. Rain, J.C., et al., *The protein-protein interaction map of *Helicobacter pylori**. Nature, 2001. **409**(6817): p. 211-5.
284. Magalon, A. and A. Bock, *Analysis of the HypC-hycE complex, a key intermediate in the assembly of the metal center of the *Escherichia coli* hydrogenase 3*. J Biol Chem, 2000. **275**(28): p. 21114-20.
285. Roseboom, W., et al., *The biosynthetic routes for carbon monoxide and cyanide in the Ni-Fe active site of hydrogenases are different*. FEBS Lett, 2005. **579**(2): p. 469-72.
286. Mehta, N., J.W. Olson, and R.J. Maier, *Characterization of *Helicobacter pylori* nickel metabolism accessory proteins needed for maturation of both urease and hydrogenase*. J Bacteriol, 2003. **185**(3): p. 726-34.
287. Atanassova, A. and D.B. Zamble, **Escherichia coli* HypA is a zinc metalloprotein with a weak affinity for nickel*. J Bacteriol, 2005. **187**(14): p. 4689-97.
288. Kennedy, D.C., et al., *A dynamic Zn site in *Helicobacter pylori* HypA: a potential mechanism for metal-specific protein activity*. J Am Chem Soc, 2007. **129**(1): p. 16-7.
289. Olson, J.W. and R.J. Maier, *Dual roles of *Bradyrhizobium japonicum* nickel protein in nickel storage and GTP-dependent Ni mobilization*. J Bacteriol, 2000. **182**(6): p. 1702-5.
290. Blokesch, M., et al., *Metal insertion into NiFe-hydrogenases*. Biochem Soc Trans, 2002. **30**(4): p. 674-80.
291. Leach, M.R., et al., *Metal binding activity of the *Escherichia coli* hydrogenase maturation factor HypB*. Biochemistry, 2005. **44**(36): p. 12229-38.

292. Zhang, J.W., et al., *A role for SlyD in the Escherichia coli hydrogenase biosynthetic pathway*. J Biol Chem, 2005. **280**(6): p. 4360-6.
293. Mulrooney, S.B. and R.P. Hausinger, *Nickel uptake and utilization by microorganisms*. FEMS Microbiol Rev, 2003. **27**(2-3): p. 239-61.
294. Theodoratou, E., R. Huber, and A. Bock, *[NiFe]-Hydrogenase maturation endopeptidase: structure and function*. Biochem Soc Trans, 2005. **33**(Pt 1): p. 108-11.
295. Vignais, P.M. and A. Colbeau, *Molecular biology of microbial hydrogenases*. Curr Issues Mol Biol, 2004. **6**(2): p. 159-88.
296. Blokesch, M., et al., *HybF, a zinc-containing protein involved in NiFe hydrogenase maturation*. J Bacteriol, 2004. **186**(9): p. 2603-11.
297. Xia, W., et al., *Structure of a nickel chaperone, HypA, from Helicobacter pylori reveals two distinct metal binding sites*. J Am Chem Soc, 2009. **131**(29): p. 10031-40.
298. Watanabe, S., et al., *Crystal structure of HypA, a nickel-binding metallochaperone for [NiFe] hydrogenase maturation*. J Mol Biol, 2009. **394**(3): p. 448-59.
299. Fu, C., J.W. Olson, and R.J. Maier, *HypB protein of Bradyrhizobium japonicum is a metal-binding GTPase capable of binding 18 divalent nickel ions per dimer*. Proc Natl Acad Sci U S A, 1995. **92**(6): p. 2333-7.
300. Maier, T., et al., *The product of the hypB gene, which is required for nickel incorporation into hydrogenases, is a novel guanine nucleotide-binding protein*. J Bacteriol, 1993. **175**(3): p. 630-5.
301. Kuchar, J. and R.P. Hausinger, *Biosynthesis of metal sites*. Chem Rev, 2004. **104**(2): p. 509-25.
302. Leipe, D.D., et al., *Classification and evolution of P-loop GTPases and related ATPases*. J Mol Biol, 2002. **317**(1): p. 41-72.
303. Hottenrott, S., et al., *The Escherichia coli SlyD is a metal ion-regulated peptidyl-prolyl cis/trans-isomerase*. J Biol Chem, 1997. **272**(25): p. 15697-701.
304. Li, M.F., et al., *Helicobacter pylori HspA Heat-shock Protein Gene Cloning, Expression and Immunogenicity*. Sheng Wu Hua Xue Yu Sheng Wu Wu Li Xue Bao (Shanghai), 1999. **31**(3): p. 264-268.
305. Gilbert, J.V., et al., *Protein Hpn: cloning and characterization of a histidine-rich metal-binding polypeptide in Helicobacter pylori and Helicobacter mustelae*. Infect Immun, 1995. **63**(7): p. 2682-8.
306. Benoit, S.L., et al., *Interaction between the Helicobacter pylori accessory proteins HypA and UreE is needed for urease maturation*. Microbiology, 2007. **153**(Pt 5): p. 1474-82.
307. Maier, R.J., S.L. Benoit, and S. Seshadri, *Nickel-binding and accessory proteins facilitating Ni-enzyme maturation in Helicobacter pylori*. Biometals, 2007. **20**(3-4): p. 655-64.
308. Bogomolovas, J., et al., *Screening of fusion partners for high yield expression and purification of bioactive viscotoxins*. Protein Expr Purif, 2009. **64**(1): p. 16-23.
309. Phan, J., et al., *Structural basis for the substrate specificity of tobacco etch virus protease*. J Biol Chem, 2002. **277**(52): p. 50564-72.
310. Cheng, Y. and D.J. Patel, *An efficient system for small protein expression and refolding*. Biochem Biophys Res Commun, 2004. **317**(2): p. 401-5.
311. Lucast, L.J., R.T. Batey, and J.A. Doudna, *Large-scale purification of a stable form of recombinant tobacco etch virus protease*. Biotechniques, 2001. **30**(3): p. 544-6, 548, 550 passim.
312. Wilkins, M.R., et al., *Protein identification and analysis tools in the ExPASy server*. Methods Mol Biol, 1999. **112**: p. 531-52.
313. Rosenfeld, J., et al., *In-gel digestion of proteins for internal sequence analysis after one- or two-dimensional gel electrophoresis*. Anal Biochem, 1992. **203**(1): p. 173-9.

314. Sreerama, N. and R.W. Woody, *Estimation of protein secondary structure from circular dichroism spectra: comparison of CONTIN, SELCON, and CDSSTR methods with an expanded reference set.* Anal Biochem, 2000. **287**(2): p. 252-60.
315. Whitmore, L. and B.A. Wallace, *DICHROWEB, an online server for protein secondary structure analyses from circular dichroism spectroscopic data.* Nucleic Acids Res, 2004. **32**(Web Server issue): p. W668-73.
316. Sklenar, V., et al., *Gradient-tailored water suppression for 1H-15N HSQC experiments optimized to retain full sensitivity.* J Magn Reson, 1993. **102** p. 241-5.
317. Pervushin, K., et al., *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, 1997. **94**(23): p. 12366-71.
318. Laemmli, U.K., *Cleavage of structural proteins during the assembly of the head of bacteriophage T4.* Nature, 1970. **227**(5259): p. 680-5.
319. Greenfield, N.J., *Applications of circular dichroism in protein and peptide analysis.* Trends Analytic Chem, 1999. **18**(4): p. 236-44.
320. Altschul, S.F., et al., *Basic local alignment search tool.* J Mol Biol, 1990. **215**(3): p. 403-10.
321. Altschul, S.F., et al., *Gapped BLAST and PSI-BLAST: a new generation of protein database search programs.* Nucleic Acids Res, 1997. **25**(17): p. 3389-402.
322. Thompson, J.D., D.G. Higgins, and T.J. Gibson, *CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice.* Nucleic Acids Res, 1994. **22**(22): p. 4673-80.
323. Cuff, J.A., et al., *JPred: a consensus secondary structure prediction server.* Bioinformatics, 1998. **14** (10): p. 892-3.
324. Marti-Renom, M.A., et al., *Comparative protein structure modeling of genes and genomes.* Annu Rev Biophys Biomol Struct, 2000. **29**: p. 291-325.
325. Laskowski, R.A., et al., *PROCHECK: a program to check the stereochemical quality of protein structures.* J Appl Crystallograph, 1993. **26**: p. 283-91.
326. Gray, J.J., et al., *Protein-protein docking with simultaneous optimization of rigid-body displacement and side-chain conformations.* J Mol Biol, 2003. **331**: p. 281-99.
327. Pettersen, E.F., et al., *UCSF Chimera - A visualization system for exploratory research and analysis.* J Comput Chem, 2004. **25**: p. 1605-12.
328. Honig, B. and A. Nicholls, *Classical electrostatics in biology and chemistry.* Science, 1995. **268**: p. 1144-9.
329. Lyskov, S. and J.J. Gray, *The RosettaDock server for local protein-protein docking.* Nucleic Acids Res, 2008. **36**(Web Server issue): p. W233-8.
330. Studier, F.W., et al., *Use of T7 RNA polymerase to direct expression of cloned genes.* Methods Enzymol, 1990. **185**: p. 60-89.
331. Kabsch, W. and C. Sander, *Dictionary of protein secondary structure: pattern recognition of hydrogen-bonded and geometrical features.* Biopolymers, 1983. **22**(12): p. 2577-637.
332. Won, H.S., et al., *Structural characterization of the nickel-binding properties of Bacillus pasteurii urease accessory protein (Ure)E in solution.* J Biol Chem, 2004. **279**(17): p. 17466-72.
333. Stingl, K., et al., *In vivo interactome of Helicobacter pylori urease revealed by tandem affinity purification.* Mol Cell Proteomics, 2008. **7**(12): p. 2429-41.
334. Fukada, H., K. Takahashi, and J.M. Sturtevant, *Thermodynamics of the binding of Streptomyces subtilisin inhibitor to alpha-chymotrypsin.* Biochemistry, 1985. **24**(19): p. 5109-15.
335. Milos, M., et al., *Microcalorimetric investigation of the interaction of calmodulin with seminalplasmin and myosin light chain kinase.* J Biol Chem, 1988. **263**(19): p. 9218-22.

336. Kelley, R.F., et al., *Antigen binding thermodynamics and antiproliferative effects of chimeric and humanized anti-p185HER2 antibody Fab fragments*. *Biochemistry*, 1992. **31**(24): p. 5434-41.
337. Tsumoto, K., et al., *Contribution to antibody-antigen interaction of structurally perturbed antigenic residues upon antibody binding*. *J Biol Chem*, 1994. **269**(46): p. 28777-82.
338. Schwarz, F.P., et al., *Thermodynamics of antigen-antibody binding using specific anti-lysozyme antibodies*. *Eur J Biochem*, 1995. **228**(2): p. 388-94.
339. Tello, D., et al., *Structural and physicochemical analysis of the reaction between the anti-lysozyme antibody D1.3 and the anti-idiotopic antibodies E225 and E5.2*. *J Mol Recognit*, 1994. **7**(1): p. 57-62.
340. Evans, L.J., A. Cooper, and J.H. Lakey, *Direct measurement of the association of a protein with a family of membrane receptors*. *J Mol Biol*, 1996. **255**(4): p. 559-63.
341. Kelley, R.F., et al., *Analysis of the factor VIIa binding site on human tissue factor: effects of tissue factor mutations on the kinetics and thermodynamics of binding*. *Biochemistry*, 1995. **34**(33): p. 10383-92.
342. McLean, M.A. and S.G. Sligar, *Thermodynamic characterization of the interaction between cytochrome b5 and cytochrome c*. *Biochem Biophys Res Commun*, 1995. **215**(1): p. 316-20.
343. Murphy, K.P., E. Freire, and Y. Paterson, *Configurational effects in antibody-antigen interactions studied by microcalorimetry*. *Proteins*, 1995. **21**(2): p. 83-90.
344. Raman, C.S., M.J. Allen, and B.T. Nall, *Enthalpy of antibody--cytochrome c binding*. *Biochemistry*, 1995. **34**(17): p. 5831-8.
345. Li, J., et al., *The response regulators CheB and CheY exhibit competitive binding to the kinase CheA*. *Biochemistry*, 1995. **34**(45): p. 14626-36.
346. Philo, J.S., et al., *Dimerization of the extracellular domain of the erythropoietin (EPO) receptor by EPO: one high-affinity and one low-affinity interaction*. *Biochemistry*, 1996. **35**(5): p. 1681-91.
347. Pearce, K.H., Jr., et al., *Structural and mutational analysis of affinity-inert contact residues at the growth hormone-receptor interface*. *Biochemistry*, 1996. **35**(32): p. 10300-7.
348. Philo, J.S., et al., *Human stem cell factor dimer forms a complex with two molecules of the extracellular domain of its receptor, Kit*. *J Biol Chem*, 1996. **271**(12): p. 6895-902.
349. Renzoni, D.A., et al., *Structural and thermodynamic characterization of the interaction of the SH3 domain from Fyn with the proline-rich binding site on the p85 subunit of PI3-kinase*. *Biochemistry*, 1996. **35**(49): p. 15646-53.
350. Murphy, K.P., et al., *Structural energetics of peptide recognition: angiotensin II/antibody binding*. *Proteins*, 1993. **15**(2): p. 113-20.
351. Page, J.D., et al., *Localization of the binding site on plasma kallikrein for high-molecular-weight kininogen to both apple 1 and apple 4 domains of the heavy chain*. *Arch Biochem Biophys*, 1994. **314**(1): p. 159-64.
352. Ladbury, J.E., et al., *Measurement of the binding of tyrosyl phosphopeptides to SH2 domains: a reappraisal*. *Proc Natl Acad Sci U S A*, 1995. **92**(8): p. 3199-203.
353. Ladbury, J.E., et al., *Alternative modes of tyrosyl phosphopeptide binding to a Src family SH2 domain: implications for regulation of tyrosine kinase activity*. *Biochemistry*, 1996. **35**(34): p. 11062-9.
354. Mandiyan, V., et al., *Thermodynamic studies of SHC phosphotyrosine interaction domain recognition of the NPXpY motif*. *J Biol Chem*, 1996. **271**(9): p. 4770-5.
355. Leder, L., et al., *Spectroscopic, calorimetric, and kinetic demonstration of conformational adaptation in peptide-antibody recognition*. *Biochemistry*, 1995. **34**(50): p. 16509-18.
356. Milos, M., et al., *Microcalorimetric investigation of the interactions in the ternary complex calmodulin-calcium-melittin*. *J Biol Chem*, 1987. **262**(6): p. 2746-9.
357. Weber, P.C., M.W. Pantoliano, and L.D. Thompson, *Crystal structure and ligand-binding studies of a screened peptide complexed with streptavidin*. *Biochemistry*, 1992. **31**(39): p. 9350-4.

358. Wu, J., et al., *The receptor binding site for the methyltransferase of bacterial chemotaxis is distinct from the sites of methylation*. *Biochemistry*, 1996. **35**(15): p. 4984-93.
359. Jelesarov, I. and H.R. Bosshard, *Thermodynamic characterization of the coupled folding and association of heterodimeric coiled coils (leucine zippers)*. *J Mol Biol*, 1996. **263**(2): p. 344-58.
360. Lin, Z., F.P. Schwartz, and E. Eisenstein, *The hydrophobic nature of GroEL-substrate binding*. *J Biol Chem*, 1995. **270**(3): p. 1011-4.
361. Lee, S., et al., *Structural basis for ubiquitin recognition and autoubiquitination by Rabex-5*. *Nat Struct Mol Biol*, 2006. **13**(3): p. 264-71.
362. Vaynberg, J., et al., *Structure of an ultraweak protein-protein complex and its crucial role in regulation of cell morphology and motility*. *Mol Cell*, 2005. **17**(4): p. 513-23.
363. Pena, P.V., et al., *Molecular mechanism of histone H3K4me3 recognition by plant homeodomain of ING2*. *Nature*, 2006. **442**(7098): p. 100-3.
364. Dames, S.A., et al., *Structural basis for Hif-1 alpha /CBP recognition in the cellular hypoxic response*. *Proc Natl Acad Sci U S A*, 2002. **99**(8): p. 5271-6.
365. Zheng, N., et al., *Structure of a c-Cbl-UbcH7 complex: RING domain function in ubiquitin-protein ligases*. *Cell*, 2000. **102**(4): p. 533-9.
366. Deane, J.E., et al., *Tandem LIM domains provide synergistic binding in the LMO4:Ldb1 complex*. *EMBO J*, 2004. **23**(18): p. 3589-98.
367. Petrella, E.C., et al., *Structural requirements and thermodynamics of the interaction of proline peptides with profilin*. *Biochemistry*, 1996. **35**(51): p. 16535-43.
368. Schmidt, T.G., et al., *Molecular interaction between the Strep-tag affinity peptide and its cognate target, streptavidin*. *J Mol Biol*, 1996. **255**(5): p. 753-66.
369. Varadarajan, R., et al., *Heat capacity changes for protein-peptide interactions in the ribonuclease S system*. *Biochemistry*, 1992. **31**(5): p. 1421-6.
370. Lemmon, M.A., et al., *Independent binding of peptide ligands to the SH2 and SH3 domains of Grb2*. *J Biol Chem*, 1994. **269**(50): p. 31653-8.
371. Gomez, J. and E. Freire, *Thermodynamic mapping of the inhibitor site of the aspartic protease endothiapepsin*. *J Mol Biol*, 1995. **252**(3): p. 337-50.
372. Baker, B.M. and K.P. Murphy, *Dissecting the energetics of a protein-protein interaction: the binding of ovomucoid third domain to elastase*. *J Mol Biol*, 1997. **268**(2): p. 557-69.
373. Lee, C.H., et al., *A single amino acid in the SH3 domain of Hck determines its high affinity and specificity in binding to HIV-1 Nef protein*. *EMBO J*, 1995. **14**(20): p. 5006-15.
374. Chauvin, F., et al., *The N-terminal domain of Escherichia coli enzyme I of the phosphoenolpyruvate/glycose phosphotransferase system: molecular cloning and characterization*. *Proc Natl Acad Sci U S A*, 1996. **93**(14): p. 7028-31.
375. Johanson, K., et al., *Binding interactions of human interleukin 5 with its receptor alpha subunit. Large scale production, structural, and functional studies of Drosophila-expressed recombinant proteins*. *J Biol Chem*, 1995. **270**(16): p. 9459-71.
376. Kresheck, G.C., L.B. Vitello, and J.E. Erman, *Calorimetric studies on the interaction of horse ferricytochrome c and yeast cytochrome c peroxidase*. *Biochemistry*, 1995. **34**(26): p. 8398-405.
377. Martinez, J.C., et al., *A calorimetric study of the thermal stability of barnase and its interaction with barnase*. *Biochemistry*, 1995. **34**(15): p. 5224-33.
378. Jelesarov, I. and H.R. Bosshard, *Thermodynamics of ferredoxin binding to ferredoxin:NADP+ reductase and the role of water at the complex interface*. *Biochemistry*, 1994. **33**(45): p. 13321-8.
379. Bhat, T.N., et al., *Bound water molecules and conformational stabilization help mediate an antigen-antibody association*. *Proc Natl Acad Sci U S A*, 1994. **91**(3): p. 1089-93.

380. Hibbits, K.A., D.S. Gill, and R.C. Willson, *Isothermal titration calorimetric study of the association of hen egg lysozyme and the anti-lysozyme antibody HyHEL-5*. *Biochemistry*, 1994. **33**(12): p. 3584-90.
381. Takahashi, K. and H. Fukada, *Calorimetric studies of the binding of Streptomyces subtilisin inhibitor to subtilisin of Bacillus subtilis strain N<sup>+</sup>*. *Biochemistry*, 1985. **24**(2): p. 297-300.
382. Castro, M.J. and S. Anderson, *Alanine point-mutations in the reactive region of bovine pancreatic trypsin inhibitor: effects on the kinetics and thermodynamics of binding to beta-trypsin and alpha-chymotrypsin*. *Biochemistry*, 1996. **35**(35): p. 11435-46.
383. Vignais, P.M., B. Billoud, and J. Meyer, *Classification and phylogeny of hydrogenases*. *FEMS Microbiol Rev*, 2001. **25**(4): p. 455-501.