

## **THÈSE**

Pour obtenir le grade de

### **DOCTEUR DE LA COMMUNAUTE UNIVERSITE GRENOBLE ALPES**

Spécialité : **Chimie inorganique et bio inorganique**

Arrêté ministériel : 25 mai 2016

Présentée par

**Marila ALFANO**

Thèse dirigée par **Christine CAVAZZA**, Chercheur CEA, CEA  
**Grenoble**

préparée au sein du **Laboratoire de Chimie et Biologie des  
Métaux, Grenoble**  
dans l'**École Doctorale de Chimie et Science du vivant**

## **Etude de la biosynthèse de la monoxyde de carbone déshydrogénase, une enzyme-clé de la réaction du gaz a l'eau**

Thèse soutenue publiquement le **17 Octobre 2019**,  
devant le jury composé de :

**Dr. Carole Duboc**

Directrice de recherche CNRS Grenoble, Présidente

**Dr. Christelle Hureau**

Directrice de recherche CNRS Toulouse, Rapportrice

**Pr. Deborah Zamble**

Professeur des universités, Université de Toronto, Rapportrice

**Dr. Anne Volbeda**

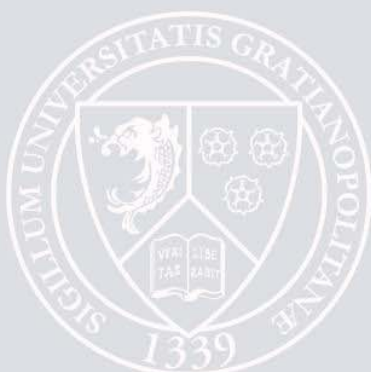
Chercheur CEA, CEA Grenoble, Examineur

**Dr. Laurent Terradot**

Directeur de recherche CNRS Lyon, Examineur

**Dr. Christine Cavazza**

Chercheur CEA, CEA Grenoble, Directrice de thèse





## Acknowledgments

After three years of working on my PhD project, the day of writing the final acknowledgments pages has arrived. It was a period of learning and maturation, from both a personal and scientific point of view. Writing this manuscript has had a strong impact on my personality and on how I currently perceive our society. I would like to say a few words of thanks to all the people who have supported and helped me during this period.

I would like to start thanking my supervisor Dr. Christine Cavazza for her continuous support during my PhD, starting from the first days with the easiest protocols to the last ones with the most hypothetical discussions about the CODH maturation. Your patience, motivation, knowledge and consistent guidance helped me reach the scientific maturity I have now. The results I achieved during this PhD are the clear example of the quality of your scientific approach and I will always be thankful for this and all the knowledges you transmitted me. I also really appreciated the scientific and non-scientific discussions we had during these years and the fact that you have always tried to recognize my efforts when I deserved it. Thanks also for all the opportunities you offered to present my data and increase my scientific knowledge. I could not imagine a better supervisor, and combined to such a successful PhD path, you showed me the best side of science.

Besides Christine, I would like to thank the rest of my thesis committee: Prof. Deborah Zamble, Dr. Christelle Hureau, Dr. Laurent Terradot, Dr. Anne Volbeda, and Dr. Carole Duboc, for their thoughtful comments on this manuscript and most importantly the discussion we had during the defense, which was my proof that science can be done in a constructive and cooperative way.

I thank the entire LCBM laboratory and the BIOCAT team for the welcome I received, for the stimulating discussions, for the advices endowed during the presentations and for forcing me into a perpetual intensive French class! Thanks to the chef Stephane Menage for the job he does maintaining the boat in the good direction, the always-targeted questions and the great humanity that characterizes him. Thanks to Julien, for the constant help in the lab and the bio-mol you tried to teach me; Christian, for the endless explanations on the purification-systems; Giulia, for the discussions on amide-ne in CooT; Philippe, for the crystal and non-crystal related chats and for being my office mate; Jaione, for all the chocolate we bought in the last years; Lucile, for the support during the last months and Umberto, to whom I wish good luck with the CODH.

A special thanks to my CSI committee, Isabelle and Irina, for the discussions which incited me to look at my research from various perspectives.

Thanks also to my external collaborators, Stefano Ciurli, Barbara Zambelli, Francesco Musiani, Luca Signor, to always be active and interested about this project.

An acknowledgment also to CEA for the grant that allowed this thesis to happen.

I also want to thank my Grenoble family (Raquel, Tomas, Jorge, Arianna, Marie Laure, Alvaro, Jaione, Xenia, Ana<sup>2</sup>, Javi, Dani, Andrea, Isa, Giorgos, Eugenia, etc.) for always being there for a laugh and the shoulder to depend on during these three years.

A big thanks to my family, my parents and my sisters, that are always there when we need each other. They are not completely sure about what I am doing in France, but they support me anyhow! Marta, Martina e Silvia, for always being there in the important moments, and Annalisa that has been my first 'PhD student' model giving me a preview on the PhD-life style.

Last but not least, I thank Vanni. Your constant support is my strength in the daily life, if I am the person and the scientist I am today, it is also because of you. Your point of view has always been my final confirmation and I am really lucky to share my life with you.

*" The saddest aspect of life right now is that science gathers knowledge faster than society gathers wisdom."*

*I. Asimov*



# Table of Contents

|   |        |
|---|--------|
| Preface .....   | 5      |
| Préambule .....   | 9      |
| <br><b>CHAPTER 1</b>  |        |
| Introduction .....  | 15     |
| <b>1.1 Nickel enzymes</b> .....                                       | 19     |
| 1.1.1 Urease.....   | 19     |
| 1.1.2 Glyoxalase I.....   | 21     |
| 1.1.3 Acireductone Dioxygenase (ARD).....                             | 23     |
| 1.1.4 Lactate racemase (LarA) .....                                   | 24     |
| 1.1.5 Superoxide Dismutase (NiSOD) .....                              | 26     |
| 1.1.6 Methyl Coenzyme M Reductase (MCR).....                          | 28     |
| 1.1.7 [Ni-Fe] hydrogenase .....                                       | 30     |
| 1.1.8 Acetyl-coenzyme A synthase/ Carbon Monoxide Dehydrogenase ..... | 33     |
| 1.1.9 Monofunctional Carbon Monoxide Dehydrogenase (CODH).....        | 36     |
| <b>1.2 Nickel in the active site: where is it coming from?</b> .....  | 43     |
| 1.2.1 Urease maturation .....   | 43     |
| 1.2.2 [Ni-Fe] hydrogenase maturation .....                            | 50     |
| 1.2.3 [NiFe]-Carbon Monoxide Dehydrogenase maturation.....            | 61     |
| <b>1.3 Heterologous production of CODH in <i>E. coli</i></b> .....    | 77     |
| <b>1.4 Our study model: <i>Rhodospirillum rubrum</i></b> .....        | 83     |
| <br>Statement of purpose .....  | <br>89 |
| <br><b>CHAPTER 2</b>  |        |
| CooT .....  | 91     |
| 2.1 CooT from <i>Rhodospirillum rubrum</i> .....                      | 93     |

|   |            |
|---|------------|
| 2.1.1 <i>RrCooT</i> paper .....   | 97         |
| 2.1.2 Unpublished data .....  | 125        |
| 2.1.2 Circular Dichroism characterization of CooT WT, CooT-Cys1 and CooT-Gly1 .....                   | 125        |
| 2.1.3 pH dependence .....   | 132        |
| 2.1.4 Crystallization assays for holo- <i>RrCooT</i> .....  | 134        |
| 2.1.5 N-terminal amine transamination reaction .....  | 137        |
| <b>2.2 <i>C. hydrogenoformans</i> CooT .....</b>  | <b>141</b> |
| 2.2.1 <i>ChCooT</i> paper .....   | 143        |
| <b>2.3 Comparison of CooT with HypC .....</b>   | <b>159</b> |
| <br><b>CHAPTER 3</b>  |            |
| <b><i>RrCooJ</i> .....</b>  | <b>163</b> |
| 3.1 Crystal structure of <i>RrCooJ</i> -Δ .....   | 165        |
| 3.2 The mystery of the 315 nm absorption peak.....  | 171        |
| 3.3 <i>RrCooJ</i> papers.....   | 173        |
| 3.3.1 Paper I.....  | 173        |
| 3.3.2 Paper II.....   | 203        |
| <br><b>CHAPTER 4</b>  |            |
| Heterologous production of CODH in <i>E. coli</i> .....   | 215        |
| <b>4.1 RESULTS AND DISCUSSIONS .....</b>  | <b>216</b> |
| 4.1.1 Heterologous expression and purification .....  | 216        |
| 4.1.2 <i>In cellulo</i> activity of <i>RrCODH</i> vs <i>RrCODH-CooJ</i> vs <i>RrCODH-CooJ</i> Δ ..... | 220        |
| <br><b>CHAPTER 5</b>  |            |
| Conclusions and perspectives for further research.....  | 225        |
| <br><b>CHAPTER 6</b>  |            |
| <b>Methods.....</b>   | <b>233</b> |
| 6.1 X-Ray Crystallography.....  | 233        |

|   |            |
|---|------------|
| 6.1.1 Crystallization .....                               | 233        |
| 6.1.2 Synchrotron radiation .....                         | 235        |
| 6.1.3 Data Analysis .....                                 | 235        |
| <b>6.2 Circular Dichroism.....</b>                        | <b>239</b> |
| <b>6.3 Isothermal titration calorimetry .....</b>         | <b>241</b> |
| <b>6.4 Multi angle laser light scattering.....</b>        | <b>243</b> |
| <b>6.5 Nuclear magnetic resonance .....</b>               | <b>245</b> |
| <b>6.6 X-ray absorption spectroscopy (XAS).....</b>       | <b>247</b> |
| <b>6.7 Biological small-angle scattering (SAXS) .....</b> | <b>249</b> |
| <br><b>Annex</b>  |            |
| Materials and methods from chapter 4 .....                | <b>253</b> |
| <br><b>Table 1 Nickel enzymes comparison .....</b>        | <b>259</b> |
| <br><b>Abbreviations .....</b>                            | <b>259</b> |
| <br><b>Review .....</b>                                   | <b>261</b> |
| <br><b>Reference .....</b>                                | <b>281</b> |



## Preface

In the current society, the severe exploitation of fossil fuels to match the constant improvements of living standards is creating major problems on our eco-system. The development of new sustainable strategies to meet the evolving demand of green-energy sources and eco-compatible commodity products is, currently, one of the biggest scientific and social concern. Among the possible renewable sources, biomass conversion has been appointed as one of the feasible alternatives to fossil fuels. Organic wastes are the biomass starting feedstock and can be gasified, producing a synthetic gas called “syngas”, mainly composed in  $H_2$ , CO and  $CO_2$  in a variable ratio. The accurate control of this ratio is the crucial step to create viable industrial products, and one of the most promising ways to achieve such control is by using the direct conversion of CO according to the reaction  $CO + H_2O \rightarrow H_2 + CO_2$ , known as the water gas shift (WGS) reaction. Nowadays, the WGS reaction is commonly employed in the industries but it requires the use of expensive inorganic catalysts, working at high temperatures and pressures, not compatible with the “green-energy” potential of this feedstock.

A possible renewable alternative comes from nature. In fact, a group of anaerobic microorganisms, survivors of the early anoxic life conditions, is capable to grow using CO as substrate, employing a perfect biological WGS reaction catalyzed at room temperature and pressure. In this process, the conversion of CO into  $H_2$  and  $CO_2$  requires two key enzymes, the [NiFe]-CO dehydrogenase (CODH), which oxidizes CO to  $CO_2$ , coupled to a [Ni-Fe]-hydrogenase, for the production of  $H_2$ . The possibility to draw inspiration directly from nature to solve our issues is becoming more and more a reality. However, in order to use microorganisms as a cost effective and environmentally friendly technology, we first need to acquire a deep understanding of how their enzymatic process occurs.

The key enzymes of the biological WGS reaction, [Ni-Fe]-CODH and [Ni-Fe]-hydrogenase, belong to the family of the metalloenzymes, known to be the most efficient catalysts in term of activity, substrates selectivity and product conversion. In particular, these enzymes harbor peculiar metallocluster-centers constituted by a heteronuclear Fe/S cluster with an additional Ni-ion, which constitute the catalytic site capable of promoting our desired WGS reaction. Given the complexity of such clusters, the need for specific machineries to build, assemble

and deliver these centers into the enzymes is not surprising and it requires elaborated maturation pathways. Understanding how the biosynthesis of these complex metal active sites occurs is one of the most challenging topics in the field of bio-inorganic chemistry. Regarding our two enzymes of interest, intense studies have been devoted to decode the reaction mechanism and the maturation pathway of the [Ni-Fe]-hydrogenase, leading to the synthesis of different bio-inspired complexes for bio-technological applications. On the other hand, there is an important lack of knowledge about [Ni-Fe]-CODH, particularly concerning its maturation mechanism. For this reason, in our project we are trying to provide a tangible contribution to unravel this knot.

In the context of my PhD project, titled "*Deciphering the biosynthetic pathway of Carbon Monoxide Dehydrogenase, a key enzyme of the water gas shift reaction*", I worked on the step concerning the enzyme activation via nickel insertion. While the structure of the enzyme is known and has been deeply characterized to unveil its reaction mechanism, little information is available on the nickel insertion and the activation of its active site.

Historically, the role of the chaperone proteins (CooC, CooT, CooJ) has been poorly investigated, mainly due to the difficulties encountered during their purification from *Rhodospirillum rubrum*. In this manuscript, I present the results we accomplished in the characterization of two out of three of these chaperones, CooT and CooJ, in the hope of providing one of the missing puzzle pieces needed to complete the main picture.

This manuscript is divided in 6 main chapters. The bibliographic introduction provides a general overview of the impact of nickel on the biological systems, to then enter in a brief description of the nine currently discovered nickel enzymes, comprising for each their function, structure and active site. A second section will open with a detailed description of the maturation mechanisms, in particular the nickel insertion step, for the urease, hydrogenase and mono-functional CODH. In this context, I place under the magnifying glass the currently known accessory proteins responsible for the nickel insertion into these enzymes active sites. Their structure, nickel coordination, interaction with other partners and the enzyme itself is described to the best of the present knowledge. The final section is devoted to the description of our study model, *R. rubrum*. Chapters 2, 3 and 4 are devoted to the presentation of the results obtained during these three years. Chapter 2 presents the results

accomplished on the characterization of CooT, Chapter 3 that ofCooJ and Chapter 4 the preliminary study accomplished on the recombinant *RrCODH* in *E. coli*. The conclusions and the further perspectives for the project are presented in Chapter 5. The final Chapter 6, is a quick overview of the physics behind the methods I used during the characterizations of these proteins.





## Préambule

Dans la société actuelle, l'exploitation intensive des combustibles fossiles pour faire face au développement continu du niveau de vie crée des problèmes majeurs dans notre écosystème. L'élaboration de nouvelles stratégies durables pour répondre à l'évolution de la demande de sources d'énergie verte et de produits de commodité éco-compatibles est, à l'heure actuelle, l'une des plus grandes préoccupations scientifiques et sociales. Parmi les sources renouvelables possibles, la conversion de la biomasse a été désignée comme l'une des alternatives possibles aux combustibles fossiles. Les déchets organiques sont la matière première de départ de la biomasse et peuvent être gazéifiés, produisant un gaz synthétique appelé "gaz de synthèse", principalement composé de CO<sub>2</sub>, H<sub>2</sub> et de CO dans un rapport variable. Le contrôle précis de ce rapport est l'étape cruciale pour créer des matières premières nobles et l'un des moyens les plus prometteurs d'y parvenir est d'utiliser la conversion directe du CO selon la réaction  $\text{CO} + \text{H}_2\text{O} \rightarrow \text{H}_2 + \text{CO}_2$  connue sous le nom de réaction du gaz à l'eau (WGS). De nos jours, la réaction WGS est couramment utilisée en industries mais elle nécessite l'utilisation de catalyseurs inorganiques coûteux, fonctionnant à haute température et pression, incompatibles avec la portée "énergie verte" de cette matière première.

Une alternative renouvelable possible vient de la nature. En fait, un groupe de micro-organismes anaérobies est capable de croître en utilisant le CO comme substrat, via une réaction WGS biologique catalysée à température et pression ambiante. Dans ce processus, la conversion du CO en H<sub>2</sub> et CO<sub>2</sub> nécessite deux enzymes clés, la [NiFe]-CO déshydrogénase (CODH), qui oxyde le CO en CO<sub>2</sub>, couplée à une [NiFe]-hydrogénase, pour produire du H<sub>2</sub>. La possibilité de s'inspirer directement de la nature pour résoudre nos problèmes devient de plus en plus une réalité. Cependant, l'utilisation des micro-organismes en tant que technologie rentable et respectueuse de l'environnement est fondamentale et nécessite souvent la compréhension de leur processus métaboliques.

Les enzymes clés de la réaction WGS biologique, la [NiFe]-CODH et la [NiFe]-hydrogénase, appartiennent à la famille des métalloenzymes, connus pour être les catalyseurs les plus efficaces en termes d'activité, de sélectivité des substrats et de conversion des produits. Les deux enzymes d'intérêt possèdent des centres multi métalliques particuliers, constitués d'un

cluster Fe/S hétéronucléaire avec un ion Ni supplémentaire et éventuellement des ligands diatomiques tel que du CO. En raison de leur complexité, le besoin de machineries spécifiques pour assembler et délivrer ces centres métalliques au cœur des enzymes n'est pas surprenant et cela exige des voies de maturation élaborées. Comprendre comment se produit la biosynthèse de ces sites actifs multi-métalliques est l'un des sujets les plus complexes en métallo-enzymologie. En ce qui concerne la [NiFe]-hydrogénase et la CODH, de nombreuses études ont été consacrées au décryptage du mécanisme réactionnel et de la maturation de la l'hydrogénase, qui ont conduit à la synthèse d'une collection des complexes bio-inspirés pour des applications biotechnologiques. En comparaison, beaucoup moins d'études ont été consacrées à la [NiFe]-CODH, en particulier concernant son mécanisme de maturation, et c'est pour cette raison que dans ce projet, nous essayons d'apporter notre contribution pour élargir le champ de connaissance dans ce domaine.

Dans le cadre de mon projet de thèse intitulé "*Décrypter la voie de biosynthèse de la monoxyde de carbone déshydrogénase, une enzyme clé de la réaction de déplacement du gaz de l'eau*", j'ai travaillé sur l'étape d'activation de l'enzyme par insertion du nickel. Bien que la structure de l'enzyme soit connue et que son mécanisme réactionnel ait été étudié en détails, peu d'informations sont disponibles sur la biogénèse et l'activation de son site actif. Historiquement, le rôle des protéines chaperons (CooC, CooT et CooJ) a été peu étudié, principalement en raison des difficultés rencontrées lors de leur purification à partir de la bactérie *Rhodospirillum rubrum*. Dans ce manuscrit, je présenterai les études que nous avons accomplies dans la caractérisation de deux de ces chaperons, CooT et CooJ, en espérant qu'un jour ils seront l'une des pièces manquantes du puzzle nécessaire pour compléter le schéma de maturation global.

Ce manuscrit est divisé en 6 chapitres principaux. L'introduction bibliographique donne un aperçu général de l'impact du nickel dans les systèmes biologiques, poursuivie par une brève description des neuf enzymes à nickel actuellement connues, décrivant pour chacune leur fonction, structure et site actif. Une deuxième partie débutera par une description détaillée des mécanisme de maturation, en particulier l'étape d'insertion du nickel, de l'uréase, de l'hydrogénase et de la CODH. Dans ce contexte, le focus sera fait sur les protéines accessoires actuellement reconnues comme responsables de l'insertion du nickel dans les sites actifs de

ces enzymes. Leur structure, le mode de coordination du nickel, l'interaction avec d'autres partenaires et l'enzyme elle-même sont décrites au mieux des connaissances actuelles. La partie suivante est consacrée à la description de notre modèle d'étude, *Rhodospirillum rubrum*. Les chapitres 2, 3 et 4 sont consacrés à la présentation des résultats obtenus au cours de ces trois années. Le chapitre 2 présentera les résultats obtenus sur la caractérisation de CooT, le chapitre 3 sur celle de CooJ et le chapitre 4 sur l'étude préliminaire réalisée sur la CODH recombinante produite chez *E. coli*. Les conclusions et les perspectives du projet sont présentées au cours du chapitre 5. Le dernier chapitre 6, est un bref aperçu des méthodes que j'ai utilisées lors de la caractérisation de ces protéines.



# Chapter 1



# CHAPTER 1

## Introduction

At the beginning of life, the environment is assumed to be electron rich, with an atmosphere abundant in gasses such as  $H_2$ , CO and  $CO_2$  and hot oceans well-supplied in transition metals as  $Fe^{2+}$  and  $Ni^{2+}$ .<sup>1</sup> In order to handle these special early living conditions, nickel most likely featured as a key player in the catalyst development.<sup>2</sup> The appearance of oxygen in the atmosphere 2.4 billion years ago had a dramatic impact on metabolism evolution.<sup>3</sup> As oxygen conquered the ecosystem, many enzymatic reactions pathways, used by anoxic metabolism under reductive conditions, were replaced by aerobic ones. As possible evidence for this change, to date, nickel does not feature in any enzyme active site of higher eukaryotes, with the exception of plants.<sup>4,5</sup>

The reason why oxygen level increased in the atmosphere, namely the Great Oxidation Event (GOE), remains a mystery and different theories have been suggested. An interesting explanation focuses the attention on nickel and in its role as a metal co-factor in the metabolism of methanogenic archaea. The “nickel famine” theory has been proposed as a possible cause of methane collapse. It correlates the rise of oxygen in the atmosphere with a drastic reduction of nickel flux to archaean oceans, a consequence of the cooling of upper-mantle temperatures and decreased eruption of nickel-rich ultramafic rocks, which would have starved the major oceanic microbial community: the methanogens. Their falling left the place to the proliferation of other microorganisms, especially oceanic cyanobacteria, leading to an increased production of  $O_2$  via photosynthesis. Such unbalance in the global methanogen productivity would have caused a reduction of the methane atmospheric levels, contributing to the chain reaction that brought life as we know it. (Figure 1)<sup>1,6</sup>

No matter what was the cause for oxygen takeover, the result was a segregation of the microorganisms based on  $H_2$  and/or CO oxidation and  $CO_2$  reduction pathways to anaerobic niches, such as ocean’s bottom, animal’s digestive systems and volcanoic mud. In these habitats, they kept their intricate metal-based active sites, the core of the enzyme catalyst, which we are now trying to decrypt.<sup>7</sup>

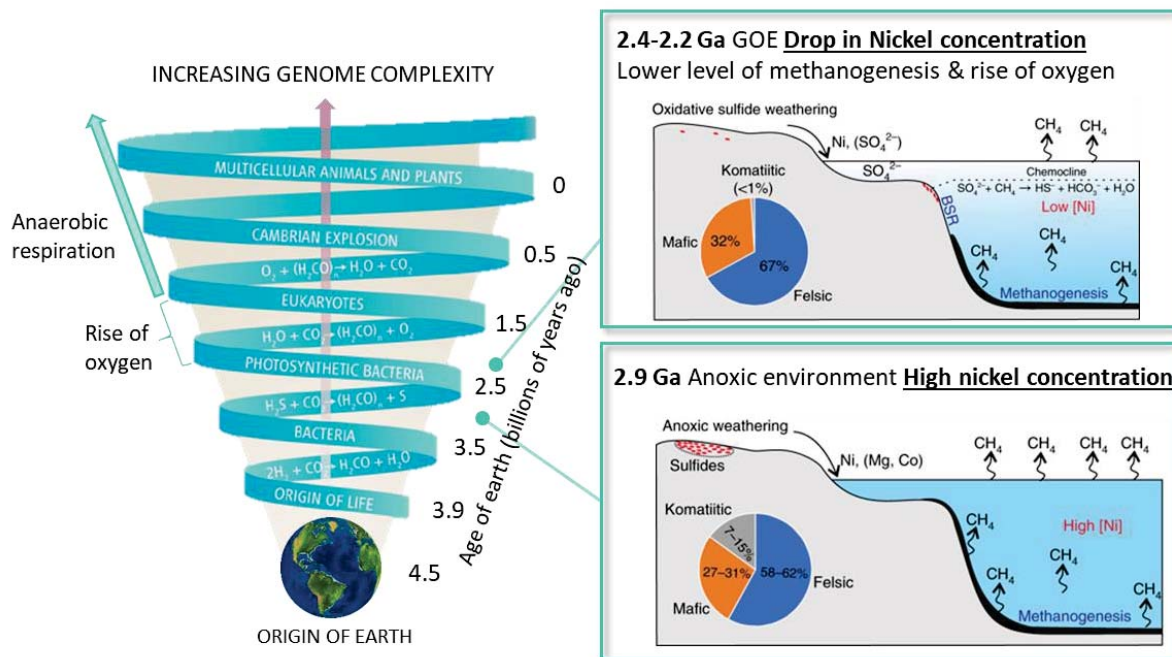


Figure 1. How the genomic and the metabolic pathways evolved during Earth's history. How life started is still unknown but there is evidence of microbial life in the mid-Proterozoic. The exact moment of photosynthesis evolution is not of easy determination, however geochemical data shown that between 2.3-2.2 billion years ago the oxygen amount in the atmosphere was sufficient to form the ozone layer. The GOE event looks like the cause of the increased complexity in both genomes and metabolic pathways. The nickel famine theory is pictured in the boxes, showing a decrease of nickel and consequently methane in the atmosphere during the GOE. Adapted from ref <sup>56</sup>.

To date, nine types of enzymes have been identified as reliant on nickel as a catalytic cofactor: [NiFe]-hydrogenase, urease, Ni-superoxide dismutase (NiSOD), [NiFe]-carbon monoxide dehydrogenase (CODH), acetyl-CoA decarbonylase/synthase (ACS), methyl-CoM reductase (MCR), glyoxylase I, acireductone dioxygenase and lactate racemase.<sup>8,9</sup> These enzymes, divided between hydrolases and redox ones, use nickel to catalyze essential reactions for the survival of the organisms, implying the need of a tight regulation of its intracellular concentration. In fact, as for other transition metals, nickel accumulation can be toxic for the cells, causing oxidative stress or replacing essential metals in metalloproteins. At the same time, the natural levels of nickel are generally low in the environment, generally in the nanomolar range with the exception of specific niches, and a constant and sufficient supply is



necessary to avoid any deficiency.<sup>3,10,11</sup> In responding to this urge, specific proteins are produced to control nickel homeostasis by transporting and/or storing the metal in the cell, assembling and delivering the metal-centers, modulating its import-efflux or regulating the expression of the other involved factors.<sup>12</sup>

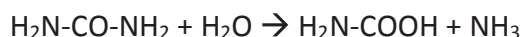
In the following pages, the nine nickel-dependent enzymes are presented, with a special attention to the anaerobic gas-processing ones. The focus will then switch to the CODH activation mechanism and to the role played by its chaperone proteins in the nickel insertion.



## 1.1 Nickel enzymes

### 1.1.1 Urease

Urease, an enzyme found in a plethora of organisms including plants, fungi, algae, archaea and bacteria, catalyzes the hydrolysis of urea to produce ammonia and carbamate.<sup>12,11,13</sup> The latter decomposes spontaneously into ammonia and carbonic acid following the reaction:



Considering the profusion of the urea substrate, urease is involved in different biological processes and is an active player in the global nitrogen cycle.

The production of ammonia leads to an overall increase in pH, which can affect both agriculture and human health.<sup>13</sup> In fact, urease is used as a virulence factor by many human pathogens such as *Helicobacter pylori*, which exploit the increase in pH to colonize the otherwise too acidic stomach environment. In agriculture, the rise of pH causes a decrease of the soil fertility but it also plays a role in the global cycling of nitrogen compounds, used as fertilizers, implying that a right balance is essential to guarantee its positive effect.

Notably, it was the first crystallized enzyme, leading to Sumner being awarded 1946 the Nobel Prize in chemistry, and it was the first evidence that nickel could be used as an enzymatic co-factor (Dixon, 1975).<sup>14,15</sup> Therefore, many years have passed since its discovery and significant steps have been made in order to understand its structure (49 different urease structures are deposited in the Protein Data Bank), catalytic mechanism, and the activation of this enzyme.<sup>16</sup>

From the structural point of view, all ureases possess several subunits that share similar tertiary structures. Most of the prokaryotic ureases contain three subunits  $\alpha$ ,  $\beta$  and  $\gamma$ , respectively corresponding to UreA, UreB and UreC, while they differ in their quaternary structures.<sup>11</sup> Considering as a model the urease from *Sporosarcina pasteurii*, its quaternary structure is built on a trimer of trimers  $(\alpha, \beta, \gamma)_3$  in which the catalytic center is contained in the  $\alpha$  subunit, making the total count of three active sites per enzyme.<sup>17</sup> The  $\alpha$  subunit possesses a flexible helix-turn-helix motif, highly conserved among all ureases, highlighting its suspected importance for the catalytic process. The active site is a di-nuclear center, with the

two nickel ions spatially separated by 3.5-3.7 Å, bridged by a carbamylated lysine residue and coordinated by two histidines each. Water molecules fill the remaining coordination positions, yielding one pentacoordinated distorted square-pyramidal based geometry (Ni1), while the second nickel (Ni2) is additionally bound to an aspartate resulting in a hexacoordinated distorted octahedral geometry.<sup>18</sup> (Figure 2)

The assembly of this active site requires different partners and, in most of the bacterial systems, four accessory proteins play this role: UreD, UreF, UreG and UreE. The simplified version of urease maturation sees UreDFG creating a complex that binds to urease apoenzyme to drive the lysine carbamylation, while UreE is the metallo-chaperone involved in nickel delivery.<sup>12</sup> A focus on these accessory proteins will be present in the following section.

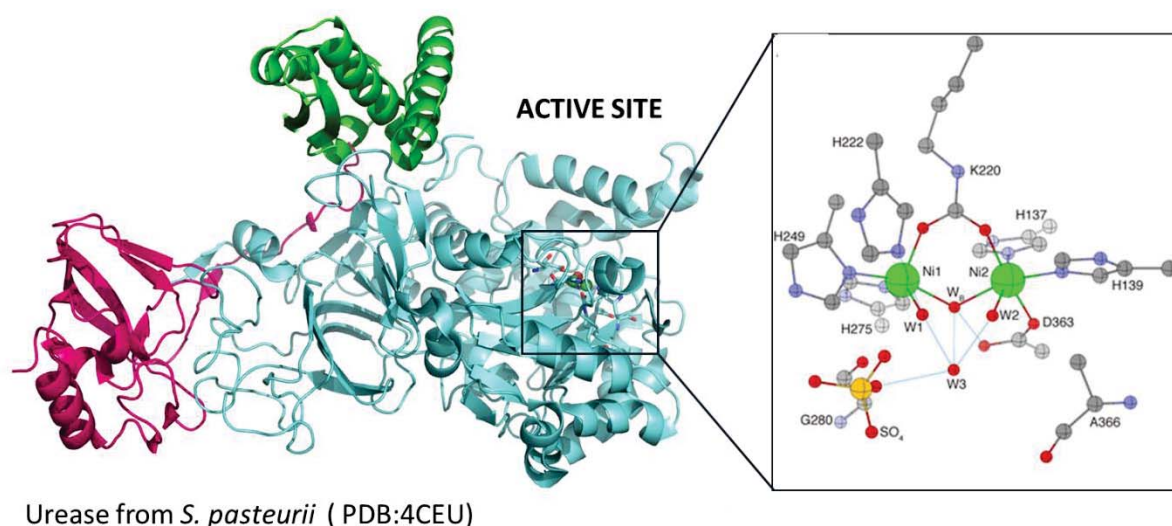


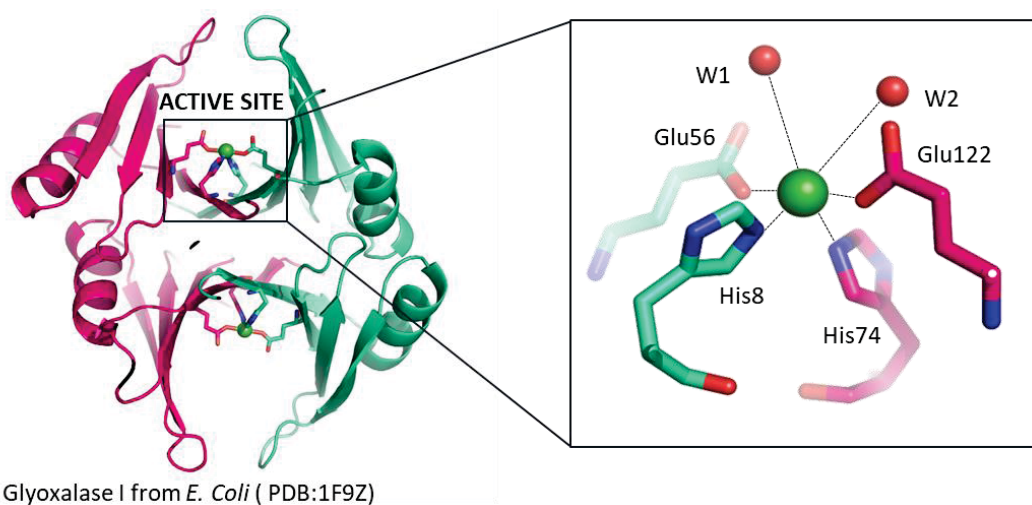
Figure 2. X-ray structure of *S. pasteurii* urease. The three different subunits are highlighted in green for the  $\beta$  subunit, in magenta for the  $\gamma$  subunit and in cyan for the  $\alpha$  subunit which contains the active site (PDB:4CEU). The zoom shows the di-nuclear active site, the two-nickel ions are bridged by a carbamylated lysine (K220) and a water molecule, and three additional water molecules (W1, W2 and W3) are present.<sup>18</sup>

### 1.1.2 Glyoxalase I

Glyoxalase I (Glx I), working in complex with glyoxalase II, is needed to transform the highly cytotoxic metabolic by-product methylglyoxal ( $\text{CH}_3\text{-CO-CHO}$ ) into non-toxic final ones. It converts the hemimercaptal substrate ( $\text{CH}_3\text{-CO-C(OH)-SG}$ ) formed by the reaction of methylglyoxal with reduced glutathione (GSH), to S-D-lactoylglutathione ( $\text{CH}_3\text{-CH(OH)-CO-SG}$ ), used in turn by Glx II as a substrate to re-generate reduced glutathione and non-toxic D-lactate ( $\text{CH}_3\text{-CH(OH)-COOH}$ ).<sup>11</sup>



Glx I from different organisms contains different metals and in *Escherichia coli* nickel features as the best performing one, maximizing its activity.<sup>19</sup> *E. coli* Glx I can also use  $\text{Co}^{2+}$ ,  $\text{Cd}^{2+}$  and  $\text{Mn}^{2+}$  as metal center, presenting a reduced reactivity, whereas  $\text{Zn}^{2+}$  substitution results in enzyme inactivation.<sup>19</sup> This change in reactivity is strictly related to the metal site coordination, which moves from an octahedral geometry for Ni, Co, Cd and Mn to a trigonal bi-pyramidal geometry for Zn.<sup>19</sup> The structure of *E. coli* Glx I is a homodimer, constituted by two subunits, each of which contributes to the formation of the two active sites. Nickel coordination is composed of a histidine and a glutamic acid from each monomer and is completed by two water molecules. (Figure 3) The potential metallochaperone proteins involved in *E. coli* Glx I nickel insertion have not yet been identified and it is unclear if they are needed.

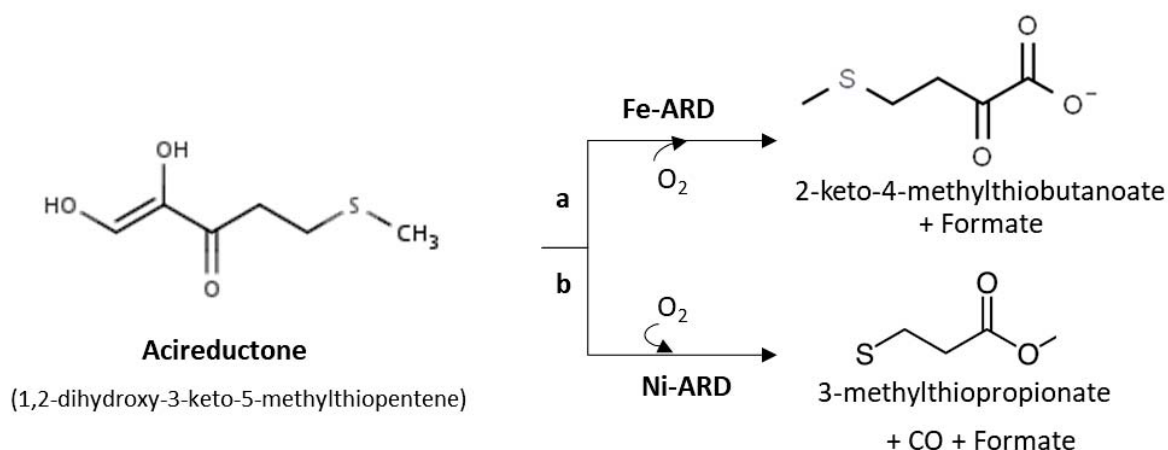


Glyoxalase I from *E. Coli* (PDB:1F9Z)

Figure 3. Structure of *E. coli* glyoxalase. The monomers constituting the homodimer are highlighted in green and magenta (PDB:1F9Z). The zoom shows the active site: the nickel ion (green) is coordinated by histidines (H8-56) and glutamic acids (E 56-112) while additional water ligands (red W1-W2) complete the geometry.<sup>19</sup>

### 1.1.3 Acireductone Dioxygenase (ARD)

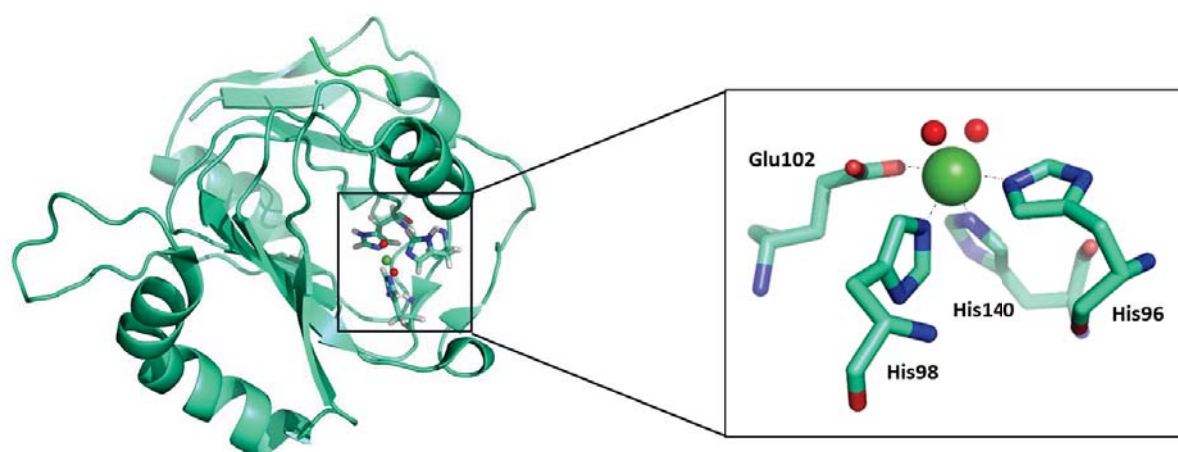
The methionine salvage pathway can be used to generate methionine from 5'-methylthioadenosine (MTA), produced from the S-adenosyl methionine during polyamine biosynthesis. During this pathway, 5'-methylthioadenosine is converted to an acireductone (1,2-dihydroxy-3-keto-5-methylthiopentene) which can be used as a substrate by the ARD enzyme to generate two different products with two different enzymatic reactions.<sup>9</sup> One reaction sees the conversion of 1,2-dihydroxy-3-keto-5-methylthiopentene into formate and 2-keto-4-methylthiobutanoate, then transaminated to methionine (reaction a). The other catalyzes an off-pathway shunt producing CO, 3-methylthiopropionate and formic acid (reaction b).



From the study of the *Klebsiella oxytoca* (formerly *K. pneumoniae*) system, it was discovered that the same enzyme catalyzes the two reactions, differing only in the nature of the coordinating metal.<sup>20</sup> When iron is present in the binding site, the ARD produce a ketoacid precursor that can be recycled back to methionine (reaction a). In contrast, when nickel or cobalt is present, ARD catalyzes a non-productive pathway of oxidation, which converts the acireductone into formate, carbon monoxide and methylthiobutyric acid (reaction b). The ARD from this organism is the only currently known enzyme, in the as-isolated state from the native organism, able to perform these two reactions depending on the nature of the metal ion bound to the protein active site.<sup>21</sup> The two forms of KoARD have the same sequence, but they can be separated chromatographically, indicating that the two structures should present

different conformations.<sup>20</sup> The structure of the Ni-KoARD was solved by NMR<sup>22</sup> (Figure 4) and the coordination of its active site is proposed via a homology model with the X-ray structure of ARD from mammalian *Mus musculus* (PDB:1VR3) (*MmARD*).<sup>21</sup> The ligand characterization, by spectroscopic and mutagenesis studies, suggested an octahedral coordination geometry for both metals involving three histidines and one glutamic acid as possible donors supplemented by two additional waters. Interestingly, from NMR spectroscopy studies on KoARD, it was shown that iron binding to the protein provokes an increase in order at the N-terminus and disorder at the C-terminus, while nickel binding generates the opposite effect. This re-arrangement in the structure, induced by the different metals, can lead to the active site changing its conformation, triggering a different selectivity to catalyze the two distinct reactions.<sup>23</sup>

Advances have been recently reported on understanding the reaction mechanism of the Human Fe-ARD, also thanks to its crystal structure which is supposed to be the native form under normal conditions. In Eukaryotes it is unclear if Ni-ARD has any biological role.<sup>24</sup> However, structural information regarding the substrate-ARD complex is poor and it is not evident how microorganisms can choose to insert different metal ions into the same catalytic site. To date, the nickel insertion pathway is unknown.

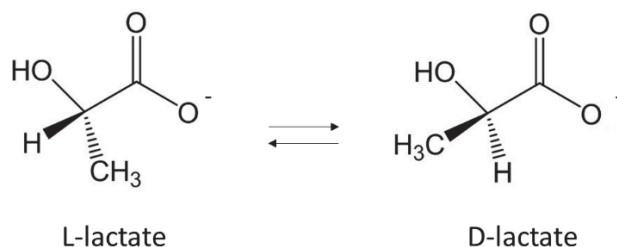


Acireductone dioxygenase from *Klebsiella oxytoca* ( PDB:1ZRR)

Figure 4. Structure of *K. oxytoca* Ni-ADR. The NMR structure is depicted in green (PDB:1ZRR). The zoom shows the active site, the 6-coordinate nickel ion (green) is proposed to be coordinated by three histidines (H98-140-96), one glutamic acid (E102) and two waters.<sup>22</sup>

### 1.1.4 Lactate racemase (LarA)

Lactate racemase (LarA), the last characterized nickel enzyme, catalyzes the interconversion between L- and D-lactic acid isomers, as depicted below.



The role of this enzyme is still under discussion and it most likely depends on the nature of the species where it is located. The latest information about LarA comes from *Lactobacillus plantarum*, which uses L-lactate as a substrate to produce D-lactate.<sup>25</sup> This bacteria does not need lactate to grow, thus the role of LarA has been linked to the rescue pathway for D-lactate production in case of stress (as in the presence of antibiotics).

In *L. plantarum*, two different operons (*larABCDE* and *larR(MN)QO*) have been identified as necessary for the enzyme activity. The enzyme itself is encoded by LarA, LarBCE play a role in the metal trafficking and co-factor biosynthesis, LarD is an aquaporin-like protein and Lar(MN)QO are related to the ABC-transporter for high-affinity nickel uptake. LarR is the transcriptional regulator. L-lactate positively regulates the system, while D-lactate represses the gene expression.<sup>26</sup>

The structure of this enzyme has been solved, disclosing the peculiar environment of its nickel coordination. In the active site is present a nicotinic acid mononucleotide derivative with two added thiocarboxylate groups, one of which forms a thioamide with lysine 184, namely nickel-pincer nucleotide (NPN) coenzyme (Figure 5). Pincer complexes are not new in chemistry but notably this is the first example of such a molecule in a biological system. The nickel is bound in an almost planar arrangement to the NPN pincer co-factor (using the thioamide sulfur atom, the pyridinium C<sub>4</sub> carbon and the thioacid sulfur atom) and the coordination is completed by a histidine. The stable Ni-C bond in the resting/stable state of this enzyme is surprising; in other Ni-containing enzymes Ni-C bonds have only been observed as reaction intermediates.<sup>13</sup> This coordination is essential to guarantee the stability of the metal in the complex. In fact, it appears that the enzyme can switch between an open and closed



conformations due to a distortion of the NPN pincer coordination and displacement of the His200. From the X-ray structure, it was possible to identify the presence of these two different states in the same asymmetric unit. The closed conformation shields the nickel from the solvent thanks to the close proximity between the N- and C- termini (Figure 5). The open conformation results in a loss of the nickel ion due to the easy accessibility of the active site to the solvent. This implies that: the access of the substrate to the catalytic center has to be rapid to avoid enzyme deactivation and that LarA is most likely able to dynamically switch between open and closed conformations.<sup>25</sup> The NPN biosynthesis is attributed to the proteins LarBCE. LarB is responsible for the production of the pre co-factor (P2CMN) from NAD, while LarE catalyzes a two sulfur transfer reactions forming the P2TMN version. LarC is the nickel insertase, which provides the nickel to generate the active form of the pincer cofactor that will bind to LarA (Figure 5).<sup>27,28 23</sup>

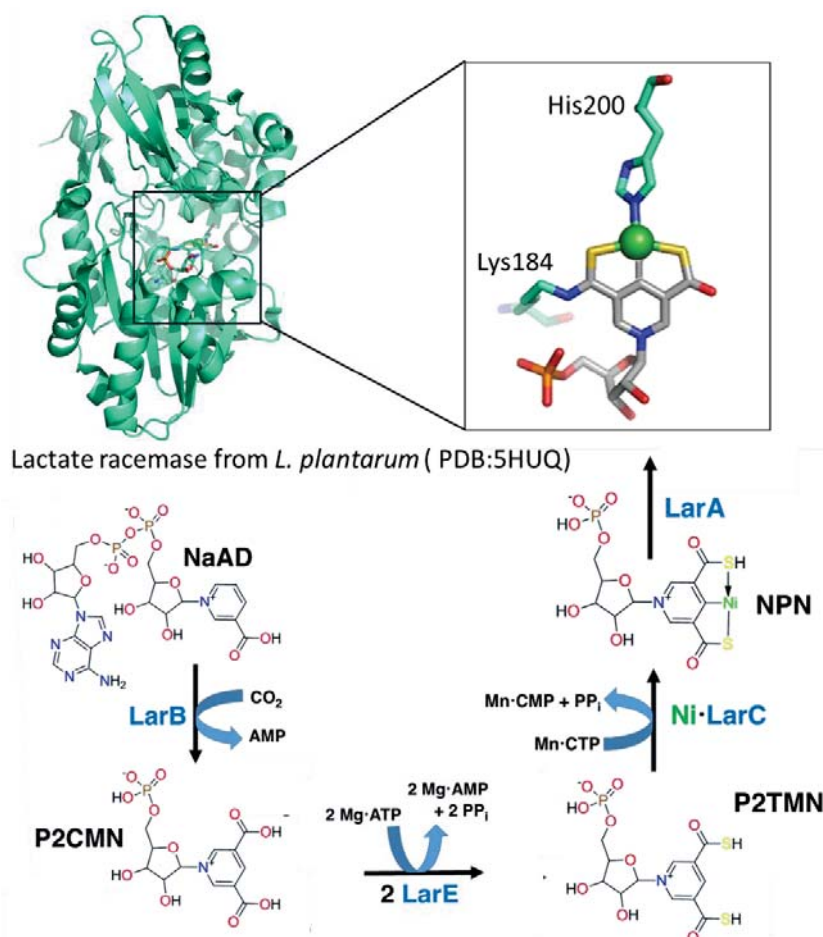
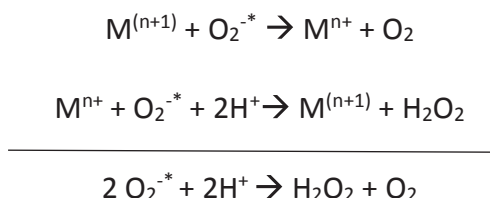


Figure 5. Structure of *L. plantarum* LarA. The nickel containing subunit is highlighted in green (PDB:5HUQ). The zoom shows the active site, the coordinate nickel ion (green) is coordinated by the NPN cofactor ( with C in gray, O in red, N in blue, S in yellow, and P in orange) and His200. The biosynthetic pathway for the NPN cofactor production starts with LarB forming P2CMN from NaAD. LarE catalyzes a sacrificial sulfur transfer reaction to synthesize P2TMN from P2CMN. LarC inserts nickel into P2TMN, forming the NPN cofactor. Adapted from <sup>23</sup>.

### 1.1.5 Superoxide Dismutase (NiSOD)

The superoxide dismutase (SOD) role is to protect biological systems from the oxidative damage caused by superoxide radical anions ( $O_2^{\cdot-}$ ), a byproduct of oxygen metabolism.<sup>29</sup> It catalyzes the dismutation of superoxide anions to hydrogen peroxide and oxygen following the reaction:



This enzyme can possess different metals in the active site. The most common versions are constituted either by a di-nuclear Cu-Zn center or by a mononuclear Fe or Mn, while the Ni enzyme has some unique characteristics. In fact, whereas the other transition metals can cycle between oxidized and reduced states using the natural midpoint-potential of  $O_2^{\cdot-}$  disproportionation ( $\sim 0.36$  V), Ni cannot catalyze superoxide disproportionation in aqueous solution. The reason why is related to the calculated reduction potential for the species  $Ni^{3+/+2}$ , which is +2.26 V, outside the range required for SOD catalysis.<sup>29</sup> It was, therefore, a surprise when NiSOD was detected and isolated in some *Streptomyces* bacteria and cyanobacterium. The adaptations required to use this metal as catalytic center (lowering the potential of the  $Ni^{3+/+2}$  couple by over 2 V) resulted in a unique evolution of this enzyme, which has no sequence homology with the other SODs. Nevertheless, NiSOD redox potential is similar to those of the other SODs and it catalyzes the reaction with the predicted  $Ni^{3+/+2}$  couple, as confirmed by EPR studies. The answer to how this enzyme is able to lower the  $Ni^{3+/+2}$  potential must reside within its structure.<sup>13</sup> The quaternary structure of NiSOD is exclusive in the SOD family. It forms an homo-hexamer where all the nickel sites are independent from each other (Figure 6). Interestingly, a posttranslational modification is needed to remove 14 amino acids at the N-terminus in order to promote nickel binding.<sup>30</sup> The metal binding site, is in fact, located by the N-terminus, known as a “Ni-hook”, which coordinates the Ni using the cysteines in position 2 (thiolate and amidate) and 6 (thiolate) and the histidine in position 1 (amine and imidazole). The nickel coordination changes depending on the oxidation state of the metal, switching from a pyramidal geometry for the oxidized state ( $Ni^{3+}$ ) to a square planar geometry for the reduced one ( $Ni^{2+}$ ), losing the apical imidazole donor (Figure 6).<sup>31</sup> While the thiolate

and imidazole ligands are critical for the reduction in potential, the role of the mixed amine-amidate coordination is the subject of different studies and it has been suggested as a possible player in maintaining the right redox potential for the nickel and to protect the active site towards oxygen.<sup>32</sup> In fact, in order to access  $\text{Ni}^{3+}$ , the only way discovered so far in nature is the use of cysteine thiolate coordination. However, this group can easily be oxidized by the products of the SOD reaction,  $\text{H}_2\text{O}_2$  and  $\text{O}_2$ , so cysteines are excluded from other metal SODs active sites. Interestingly, studies showed that the Cys ligands of NiSOD are unreactive with  $\text{H}_2\text{O}_2$  as long as the rate of catalysis is maintained, that the enzyme can be reduced by  $\text{H}_2\text{O}_2$  and that the excess peroxide can be removed without loss of activity.<sup>33</sup>

Concerning the possible accessory proteins involved in the nickel insertion, a putative nickel chelatase (CbiXhp) has been identified. The behavior of this enzyme is fascinating and many efforts have been devoted to understanding its chemical stability, also thanks to peptide-based models.<sup>30</sup> Nevertheless, further studies need to be done to understand the maturation and nickel insertion in the NiSOD production pathway.

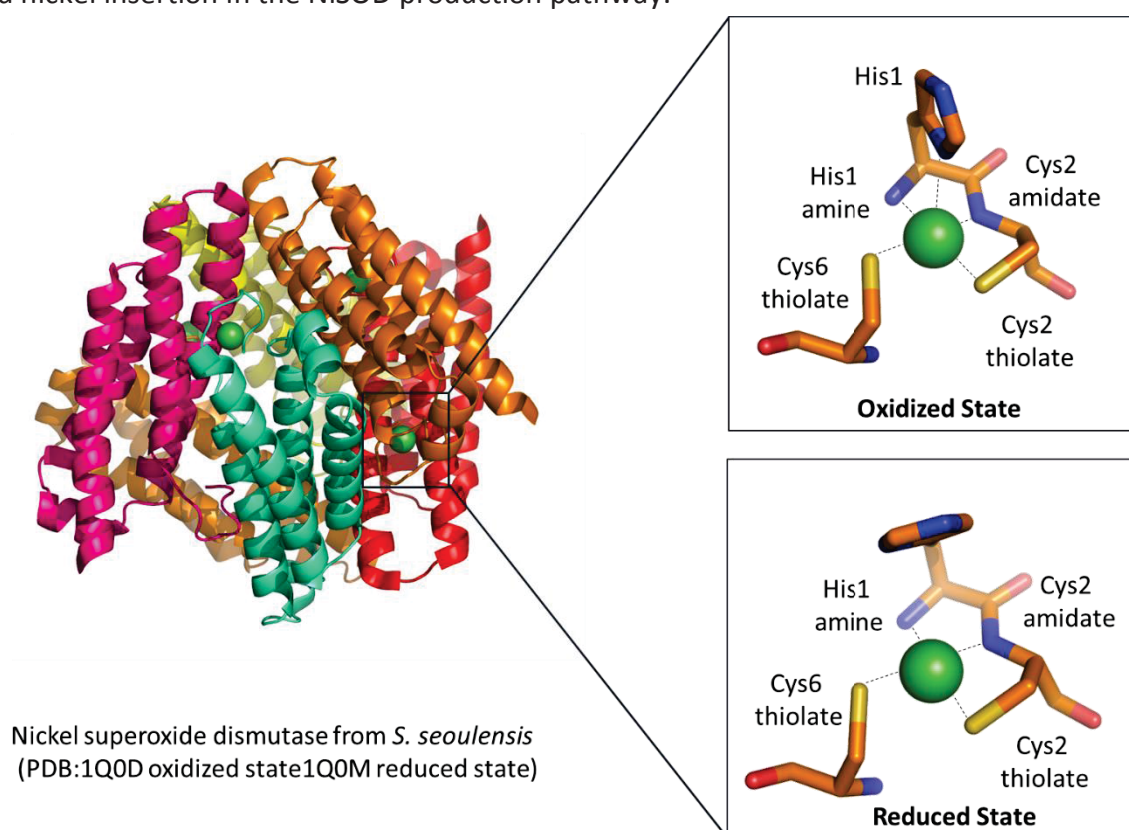
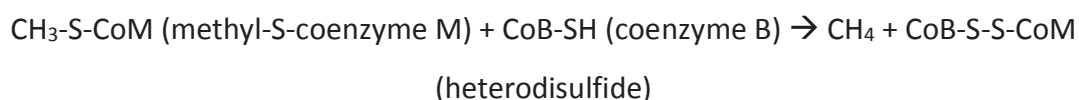


Figure 6. Structure of *S. seoulensis* NiSOD. The homo-hexameric structure shows each subunit in a different color. The active site describes the oxidized state ( $\text{Ni}^{3+}$ ) coordinated via the amino terminal amine, a backbone amide, two Cys, and an axial His (stick view). In the reduced state ( $\text{Ni}^{2+}$ ) the His is displaced.

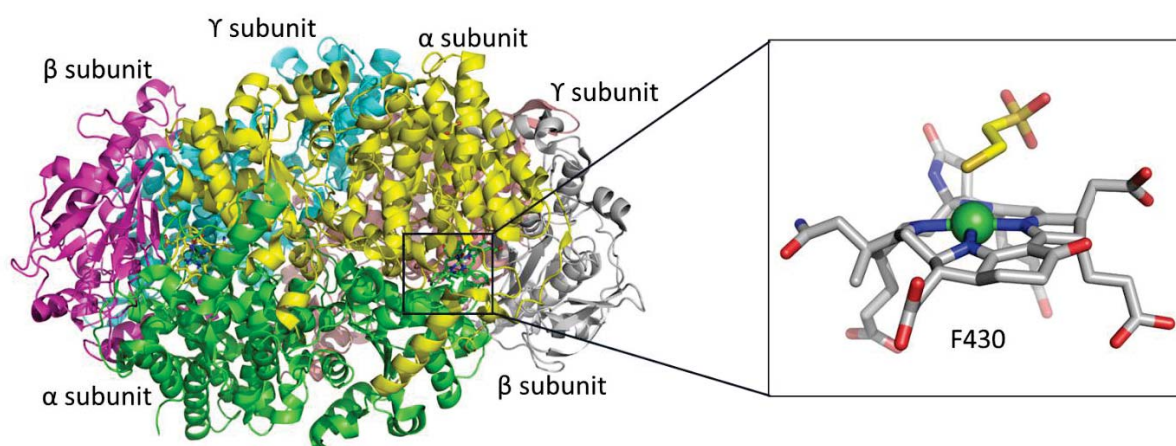
### 1.1.6 Methyl Coenzyme M Reductase (MCR)

Methyl Coenzyme M Reductase (MCR), catalyzing the reduction of methyl-S-coenzyme M by coenzyme B to methane and disulfide bridging CoB and CoM, plays an important role in the global carbon cycle.<sup>13</sup>



The methane formation is the last step of methanogenic archaea metabolism. These strictly anaerobic microorganisms, able to grow on acetate, methanol, formate, CO<sub>2</sub> and H<sub>2</sub>, are responsible for more than 90 % of the methane present on earth. The only other known source of biological methane comes from the methylphosphonates metabolism (4%),<sup>34</sup> so MCR methanogenic archaea catalysis is its main producer.<sup>35</sup>

The crystal structure of MCR from *Methanothermobacter marburgensis* reveals a heterohexamer constituted of three different subunits ( $\alpha_2\beta_2\gamma_2$ ) and a Ni-containing co-factor, namely tetrapyrrole F<sub>430</sub>, in each  $\alpha$  subunit (Figure 7).<sup>36</sup>



Methyl-coenzyme M from *M. marburgensis* (PDB:3POT)

Figure 7. Structure of *M. marburgensis* MCR. The dimer is shown with each subunit ( $\alpha\beta\gamma$ )<sub>2</sub> in different color. Each active site (Ni in green) contains coenzyme F430 with an axial Gln147 ligand, and an additional axially coordinated CoM-SH in this particular structure.<sup>36</sup>

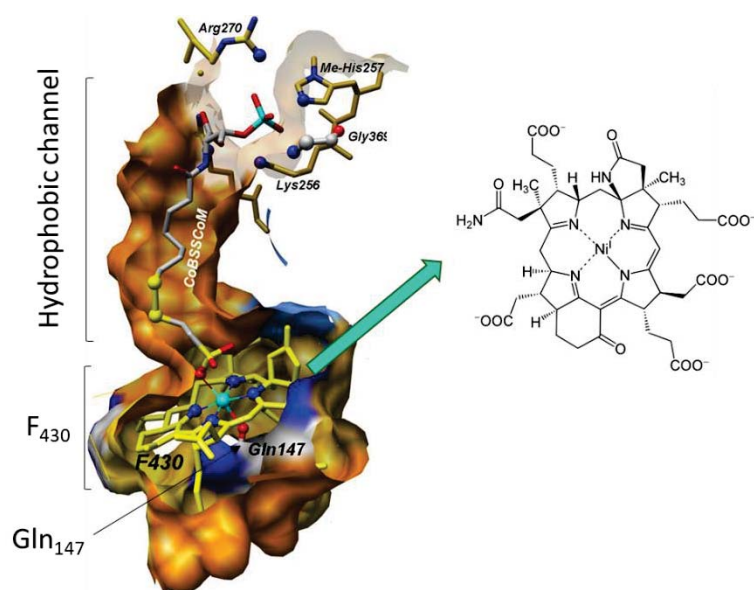


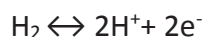
Figure 8. On the left is the representation of the hydrophobic channel in MCR. The F<sub>430</sub> cofactor is present at the bottom of it and the Ni is coordinated by the tetrapyrrole ring and Gln147. On the right is the structure of the F<sub>430</sub> cofactor. Adapted from <sup>13</sup>.

The name of the co-factor derives from its absorption band, precisely at 430 nm, due to the Ni<sup>+2</sup> oxidation state. The enzyme also needs a Ni<sup>+1</sup> state in order to be active, implying that a tight control of the metal redox state is required for its functionality. The F<sub>430</sub> is not covalently bound in the active site, but it is deeply buried at the end of a hydrophobic channel that shelters the substrates/products. This arrangement is to prevent solvent accessibility, which could cause a change in the oxidation states. The substrate methyl-S-coenzyme M must enter in this channel before the second substrate, coenzyme B, whose arrival completely impedes the access to the binding site. The ring of the F<sub>430</sub> co-factor chelates the nickel ion and its coordination is completed via an axial glutamine 147 on the opposite side of the ring from the channel.<sup>36</sup> The conformational arrangement is shown in Figure 8.

Nickel is present as an inorganic metal complex. The presence of a specific nickel chelatase (CfbA) to promote the insertion of Ni in the sirohydrochlorin has been proven both *in vitro* and *in vivo*.<sup>37</sup> From this starting intermediate, other Cfb enzymes (CfbEDCB) are needed to complete the cofactor biosynthesis, which in any case, is still not clearly understood.<sup>37,38,39</sup>

### 1.1.7 [Ni-Fe] hydrogenase

Hydrogenase catalyzes the oxidation of hydrogen into protons and electrons in a reversible way, following the reaction:



Hydrogen can have a double role, either as an energy source for the system or as a final product to remove a possible excess of reducing equivalents.<sup>7</sup>

Different kinds of hydrogenases are found in nature, present in diverse organisms such as archaea, bacteria and some specific eukaryotes. It is possible to divide them, depending on the metal ions used in their active site, into three not phylogenetically related groups: [Fe-Fe] hydrogenase, [Fe] hydrogenase and [Ni-Fe] hydrogenase.<sup>40</sup> [NiFe]-hydrogenase, found in bacteria and archaea, are divided into four main groups depending on their function:

Group 1- Membrane-associated H<sub>2</sub> uptake hydrogenases.

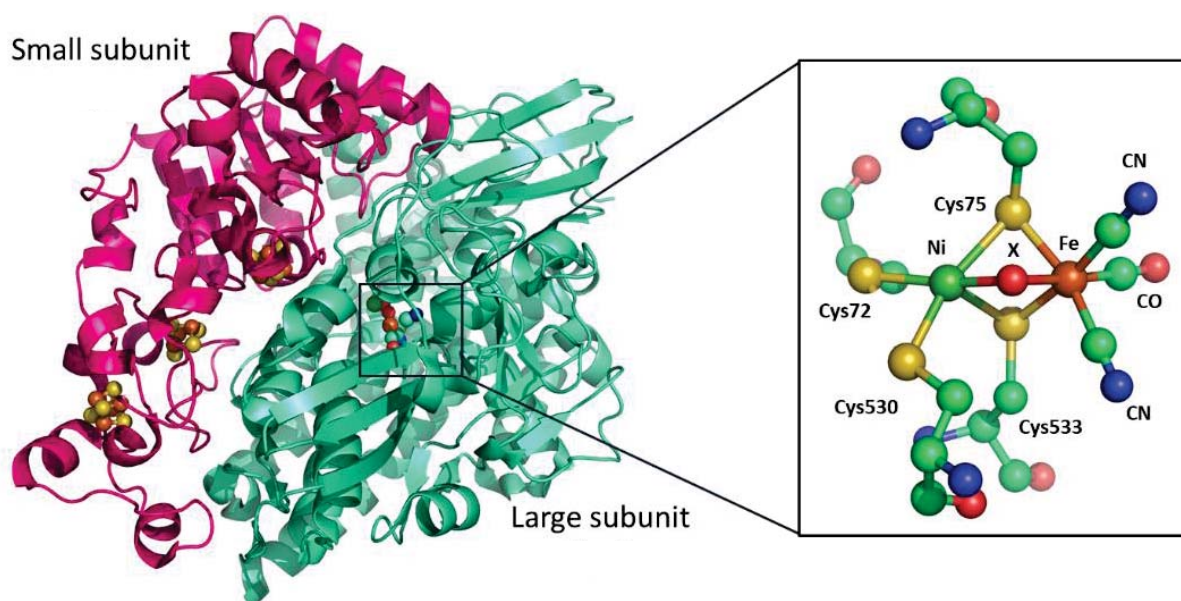
Group 2- Soluble uptake hydrogenases and sensory hydrogenases.

Group 3- Heteromultimeric cytoplasmic hydrogenases harboring a reducible cofactor (F420 or NAD(P)).

Group 4- Energy-converting hydrogenases. This group includes the CO-induced hydrogenases involved in our research scope reaction (*the WGS reaction see session 1.4*). Each hydrogenase from *Methanosarcina barkeri* and hydrogenase 3 from *E. coli* can be considered as the most studied examples of this class.

A common feature of all [Ni-Fe] hydrogenases is their general assembly in at least two subunits (Figure 9).<sup>9</sup> A large catalytic subunit hosts a Ni-Fe active site and a small subunit hosts the Fe-S clusters needed for electron transfer. The number of Fe-S clusters differs for each organism. Various structures of this enzyme have been solved, unveiling the presence of a di-nuclear active site composed of a nickel ion coordinated to cysteines of two CxxC motifs, of which one of each pair is used to bridge to the iron ion.<sup>12</sup> The Fe coordination is completed by one CO and two CN<sup>-</sup> ligands. Between Ni and Fe, a third labile binding position is present for substrate binding.<sup>41,42</sup>





Ni-Fe Hydrogenase from *Desulfovibrio fructosivorans* (PDB:1YQW)

Figure 9. Structure of *D. fructosivorans* [Ni-Fe]-hydrogenase. The large subunit is shown in green, the small one in magenta. The Fe-S clusters in the small subunit are shown in spheres and the [Ni-Fe]-cluster is highlighted in the active site. (x=labile position for substrate binding)

A common feature among all the structures is the highly buried position of the active site, which suggests the presence of a hydrophobic tunnel for dihydrogen access, as well as proton pathways.<sup>43</sup> The biosynthesis of [Ni-Fe] hydrogenase is a well-studied process and different proteins have been identified as responsible for the biogenesis of such a complex enzyme.<sup>44</sup> Their role is described in section 1.2.2.

In *E. coli*, the *hycBCDEFG* operon encodes for the hydrogenase 3, responsible for hydrogen production from formate thanks to its coupling with formate dehydrogenase-H (fdhF), in the formate hydrogenlyase complex (FHL).<sup>45</sup> The *hycABCDEFGH* operon stands for the large (HycE) and small (HycG) subunits, characteristic of [NiFe]-hydrogenases, plus two additional hydrophilic subunits (HycB and HycF) and two inner membrane subunits (HycC and HycD).<sup>44</sup> The HycA is the negative transcriptional regulator of the formate regulon, while HycH is predicted to have an accessory function. (Figure 10)



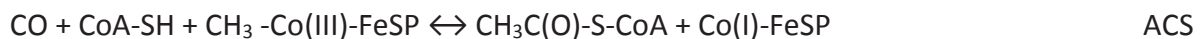
Figure 10. Comparison of the subunits encoded by the *hyc* operon from *E. coli* and the *cooMKLXUH* operon from *R. rubrum*. Proteins without sequence similarity are shown in white. Adapted from <sup>43</sup>.

In hydrogenogenic carboxydrotrophs, such as our study model *Rhodospirillum rubrum*, the hydrogenase is enclosed by the CO-induced operon *cooMKLXUH*, present upstream of the CODH one.<sup>46</sup> CooM and CooK (corresponding to the previously described HycC and HycD in *E. coli*) are two integral membrane-bound proteins, CooX (corresponding to *E. coli* HycF) is described as a Fe<sub>4</sub>S<sub>4</sub> ferredoxin-like protein and CooU has no homologue in *E. coli* hydrogenase 3. CooH and CooL encode for the large and small subunit, respectively, and they corresponds to HycE and HycG in *E. coli* hydrogenase 3 (Figure 10).



### 1.1.8 Acetyl-coenzyme A synthase/ Carbon Monoxide Dehydrogenase

Carbon monoxide dehydrogenase (CODH) can be present either as a monofunctional enzyme, as in the case for *R. rubrum* (discussed in the 1.1.9 session), or as a bifunctional complex with the acetyl-S-CoA synthase (ACS, ACDS in methanogens). The ACS/CODH complex catalyses the synthesis of acetyl-S-CoA from coenzyme A (CoA-SH) condensed with CO (derived from CO<sub>2</sub> reduction catalyzed by CODH) and a methyl group (derived from a corrinoid/ Fe-S protein Co(III)-FeSP) as described in the reaction:



This bi-functional enzyme plays an important role in the Wood-Ljungdahl pathway, one of the six carbon fixation pathways known on Earth. It consists of two branches, defined as the methyl and the carbonyl pathways. The *methyl branch* differs between acetogenic bacteria and methanogenic archaea to convert the CO<sub>2</sub> into the methyl group. The *carbonyl branch* is common in acetogens and methanogens and consists of the conversion of carbon dioxide into carbon monoxide, catalyzed by the CODH. In a second step, the CO coming from the *carbonyl branch* produced by the CODH is combined to the methyl coming from the *methyl branch* and CoA by the Ni-ACS enzyme to form acetyl-CoA, the most common carbon source in biology (Figure 11).<sup>47</sup>

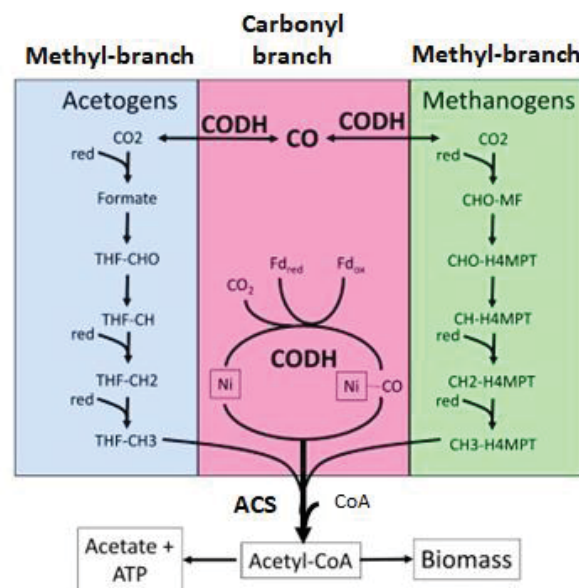
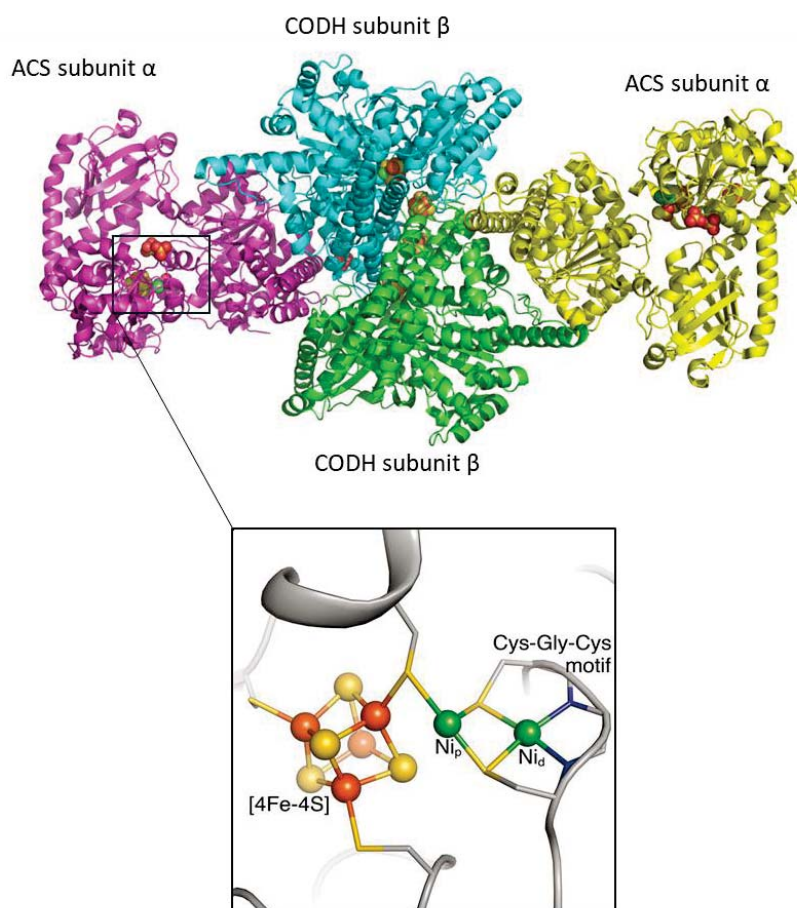


Figure 11. Wood-Ljungdahl carbon fixation pathway. The carbonyl branch is in pink, the generalized methyl branch used in acetogenic bacteria is in light blue and the generalized methyl branch used in methanogenic archaea is in green. Adapted from <sup>48</sup>.

As expected, bifunctional CODH/ACS contains two distinct active sites, one present in the CODH subunit and one in the ACS subunit. The two active sites are connected by hydrophobic channels to allow the CO produced from the CODH to reach the ACS active site, regulating in this way the import-export of the substrates.<sup>49</sup>

In the model organism *Moorella thermoacetica* the ACS/CODH is an  $\alpha_2\beta_2$  tetrameric protein formed by two CODH  $\beta$ -subunits in the center and an ACS  $\alpha$ -subunit on each extremity (Figure 12). The ACS subunit contains three different domains: the N-terminal domain, responsible for the interaction of ACS with CODH, the central domain and the C-terminal domain, which contains the active site, called the A-cluster. This site consists of a [4Fe-4S] cluster bridged to a proximal Ni ( $Ni_p$ ) via a cysteine thiolate (Cys 509).  $Ni_p$  is in turn bridged to a distal Ni ( $Ni_d$ ) via two additional cysteine thiolates (Cys 595 and 597) and its coordination is completed in a square planar geometry by an unidentified exogenous ligand, in labile position for substrate and/or intermediates binding. Historically, it was supposed that the  $Ni_p$  center could be physiologically substituted by other metals, resulting in a Cu-Ni and Zn-Ni forms. However, this hypothesis has been ruled out and the scientific community now agrees that the A-cluster is formed by a [4Fe-4S]-  $Ni_d$ -  $Ni_p$  center (Figure 12).<sup>49,13</sup> Concerning the  $Ni_d$ , its coordination is held by the two thiolates and the two backbone amide groups (one from Gly and one from Cys) of a Cys-Gly-Cys motif, in a stable  $S_2N_2$  square planar geometry (Figure 12).  $Ni_d$  does not have redox activity and it will likely remain a diamagnetic  $Ni^{2+}$  through catalysis.  $Ni_p$ , on the other hand, presents different oxidation states during the catalysis, indicating its catalytic substrate-converting role. Even though the structure of the A-cluster has been solved, the catalytic mechanism is still under debate.<sup>49,50,51</sup>

Consensus has been reached on the role of  $Ni_p$  as responsible for binding the CO and the methyl group, followed by methyl migration/CO insertion to result in a  $Ni_p$ -acetyl bound group, to finish with CoA substitution to produce acetyl-CoA. The role of the distant nickel  $Ni_d^{+2}$  has been proposed either as a supporting ligand, to stabilize the proximal  $Ni_p$  in a low-valence redox state, or as a hemilabile site responsible for modulating of the active conformation, facilitating its catalytic cycle.<sup>49,52</sup> The inorganic models made for this center all point towards a neither reducible nor catalytic center  $Ni_d$ .<sup>53</sup>



Carbon monoxide dehydrogenase/acetyl-CoA synthase(CODH/ACS) from *Moorella thermoacetica* (PDB:1OAO)

Figure 12. Structure of *M. thermoacetica* CODH/ACS. The tetramer  $\alpha_2\beta_2$  subunits are shown in different colors, the CODH homodimer  $\beta_2$  is in green and cyan, the ACS  $\alpha_2$  subunits are in pink and yellow. The metalloclusters are shown in spheres. The active site, A-cluster, is in the zoom region where the  $[4\text{Fe-4S}]$ -  $\text{Ni}_d$ -  $\text{Ni}_p$  center in ACS subunit are shown in spheres greens for nickel, orange for iron and yellow for sulfur. Adapted from <sup>13</sup>.

### 1.1.9 Monofunctional Carbon Monoxide Dehydrogenase (CODH)

Carbon monoxide dehydrogenase catalyzes the reversible oxidation of carbon monoxide to carbon dioxide following the reaction:



Two primary classes of CODH have been identified in nature, differing from each other in their active sites metal composition. One consist in a Cu-Mo-Fe-S cluster, which catalyzes a irreversible reaction and is found in aerobic microorganisms. The other class possesses a Ni-containing active site and is recurrent in anaerobic bacteria and archaea. In both case, CODH plays an essential role in the global carbon cycle.<sup>47</sup> As already mentioned in the previous section, Ni-CODH can occur as monofunctional or bifunctional enzyme and the following part will focus on monofunctional Ni-containing CODHs.<sup>54</sup>

To decrypt this enzyme, the majority of the information in the hand of the scientific community comes mainly from two organisms: *R. rubrum* (RrCODH) and *Carboxydotherrmus hydrogenoformans* (ChCODH).

In *C. hydrogenoformans* five different genes encoding CODHs are present:

- ChCODH-I, which is coupled to a [Ni-Fe] hydrogenase to catalyze the so-called water gas shift (WGS) reaction ( $\text{CO} + \text{H}_2\text{O} \leftrightarrow \text{CO}_2 + 2\text{H}^+ + 2\text{e}^-$  (CODH) and  $2\text{e}^- + 2\text{H}^+ \rightarrow \text{H}_2$  (hydrogenase)). This process is vital for the organism considering that it is a strictly anaerobic thermophilic bacterium. In fact, it can grow using CO as a sole carbon and energy source thanks to this WGS reaction<sup>55</sup>;
- ChCODH-II, involved in NADH generation<sup>56</sup>;
- ChCODH-III, coupled with the ACS enzyme to perform the acetyl-CoA biosynthesis pathway;
- ChCODH-IV, probably involved in oxidative stress response as a mechanism of oxygen detoxification<sup>57</sup>;
- ChCODH-V, its role has not yet been identified. Recently, its structure has been solved\*, showing a  $[\text{Fe}_4\text{S}_3\text{O}_2]$  cluster, which can be converted into a  $[\text{Fe}_4\text{S}_3]$  cluster by reduction. Thus, it shares more similarities with hybrid cluster proteins (HCPs). HCPs

are responsible for nitrosative stress protection during anaerobic growth in the presence of nitrate/nitrite and they harbor a 4Fe-2S-2O iron-sulfur-oxygen cluster.<sup>58</sup> Considering that CODH-V cannot convert CO or CO<sub>2</sub>, it may represent the evolutionary link between HCPs and CODHs. (\*results published in 2015 in Fessler J. thesis from Dobbek group)

In *R. rubrum* only one CODH (*Rr*CODH) is present and, as for the *Ch*CODH-I, it is involved in the WGS reaction, allowing the bacteria to use CO as a sole energy source.<sup>46</sup>

Crystal structures of *Ch*CODH-I/II/IV/V and *Rr*CODH have been solved, displaying an overall similar structure.<sup>59,60</sup> They are homodimers, of which each monomer is composed of three different domains known as the N-terminal helical domain, the central domain and the C-terminal  $\alpha/\beta$  (Rossmann-like) domains (Figure 13). Each homodimer contains five metallocusters of three different natures: two unique [Ni-4Fe-4S] clusters (namely C-cluster), two [4Fe-4S] clusters (namely B-cluster) and one intermolecular [4Fe-4S] cluster (namely D-cluster). The C-cluster is a distorted Ni-3Fe-4S cubane coordinated to an additional fourth iron (Fe<sub>u</sub>), unique in biology.<sup>59</sup> The global arrangement places the clusters in a C-B-D-B\*-C\* position, where the bridging D-cluster sits between the two monomers (Figure 14). The reaction is catalyzed at the C-cluster, the buried active site in the protein, which oxidizes CO and releases CO<sub>2</sub>, protons and electrons. The electrons, in order to reach the CODH's partners, need to be transferred to the protein surface. The B-cluster is the most likely candidate to fulfil this role, due to their proximity to the active site and the protein surface.



Carbon monoxide dehydrogenase from *R. rubrum* (PDB:1JQK)

Figure 13. Structure of *R. rubrum* CODH. The monomers are shown in magenta and green. The metallocusters are shown as spheres.

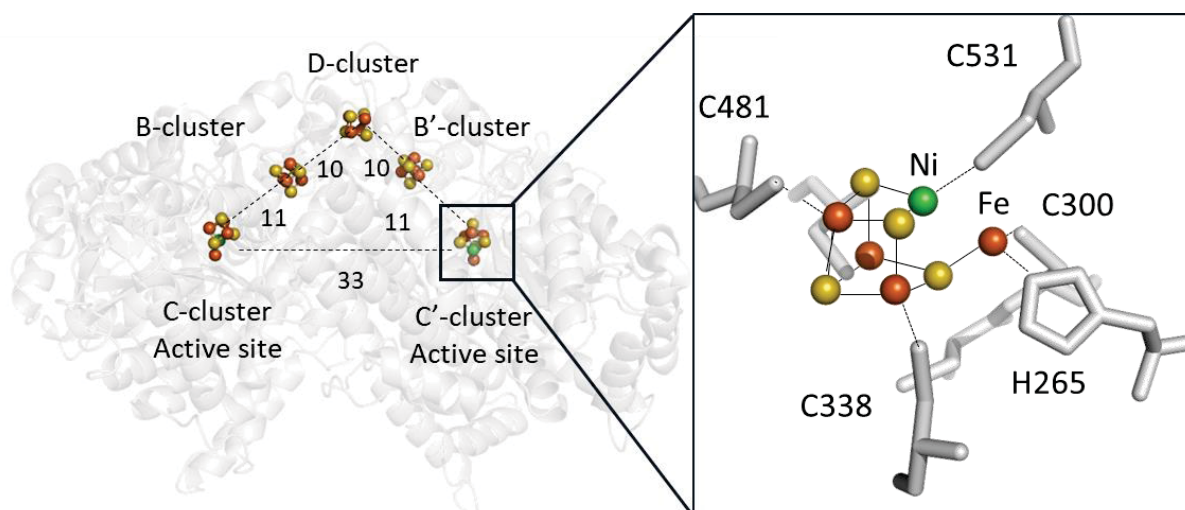


Figure 14. Structure of *R. rubrum* CODH. The cluster rearrangement is highlighted on the structure of the homodimer. The active site, C-cluster, is in the zoom region where the metals are shown in spheres (green for nickel, orange for iron and yellow for sulfur). The reported distances between Fe atoms of individual clusters are in Ångstrom.

Conversely, the role of the D-cluster remains to be elucidated. It might be involved in the electron transfer with the B- cluster, but its low reducing potential (-530 mV) could be an indication of a more 'oxygen-protective' role.<sup>61</sup>

### 1.9.1 Reaction Mechanism

Different approaches have been used to elucidate the CODH reaction mechanism, allowing the identification of the main steps involved and the postulation of the following working model.

Prior to the structural characterization of the C-cluster, four different redox states were proposed via spectroscopic studies:  $C_{ox}$ ,  $C_{red1}$ ,  $C_{int}$  and  $C_{red2}$ , all differing from their reduction/oxidation state by one electron.  $C_{ox}$  is catalytically inactive and EPR silent.  $C_{red1}$  binds CO and analogues such as cyanide. This state is paramagnetic ( $S=1/2$ ) and displays an EPR signal with  $g$ -values of 2.03, 1.88, and 1.71.  $C_{int}$  is the two-electron reduced state, also EPR silent.  $C_{red2}$  is the three-electron reduced state able to bind  $CO_2$  and analogues such as cyanate and presents an EPR signature with  $g$ -values of 1.97, 1.87 and 1.75.<sup>62</sup> All together these data had contributed to the hypothesis that CO would bind to the  $C_{red1}$ , the catalytic active state,



and when it is oxidized to  $\text{CO}_2$  the C-cluster gains two more electrons and is in the  $\text{C}_{\text{red2}}$  form. The cluster is then re-oxidized via electron transfer to the B-cluster passing through a  $\text{C}_{\text{int}}$  state (Figure 15). Where the two additional electrons are located in the  $\text{C}_{\text{red2}}$  state is still not clear, with the hypotheses of a possible  $\text{Ni}^0$  state, or hydride-bound  $\text{Ni(II)}$  or a dative Ni-Fe bond under debate.<sup>63,64,65</sup>

The difference in redox potential among these states is a key feature in order to activate correctly the enzyme. It has been shown that the  $\text{C}_{\text{ox}}/\text{C}_{\text{red1}}$  redox potential is -200 mV and  $\text{C}_{\text{red1}}/\text{C}_{\text{red2}}$  has a redox potential of -530 mV, matching almost perfectly with that of  $\text{CO}_2/\text{CO}$  which is -558 mV at pH 7.<sup>66</sup>

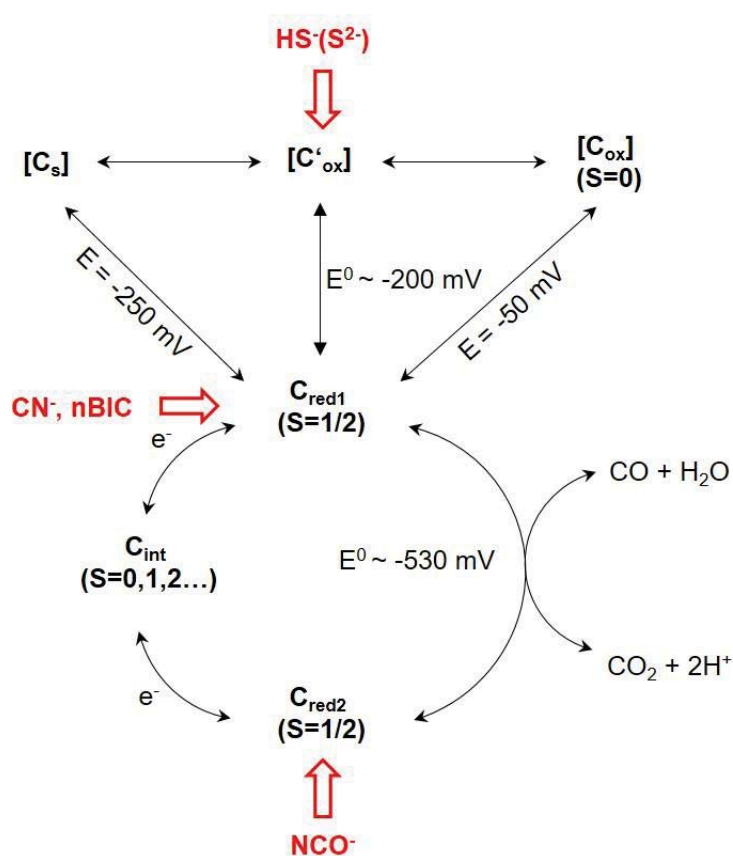


Figure 15. Redox states of CODH active site. The electronic states are reported with their relative spin states and mid-point potentials. The inhibitors of each state are also reported in red.  $\text{CN}^-$  inhibits CO oxidation by binding to the  $\text{C}_{\text{red1}}$  but it does not inhibit  $\text{CO}_2$  reduction.  $\text{NCO}^-$  inhibits  $\text{CO}_2$  reduction binding to the  $\text{C}_{\text{red2}}$  but it does not inhibit CO oxidation. Adapted from <sup>67</sup>. (nBIC : n-Butyl Isocyanide a CO analogue,  $\text{CN}^-$ : cyanide a CO analogue,  $\text{NCO}^-$  : cyanate a  $\text{CO}_2$  analogue)

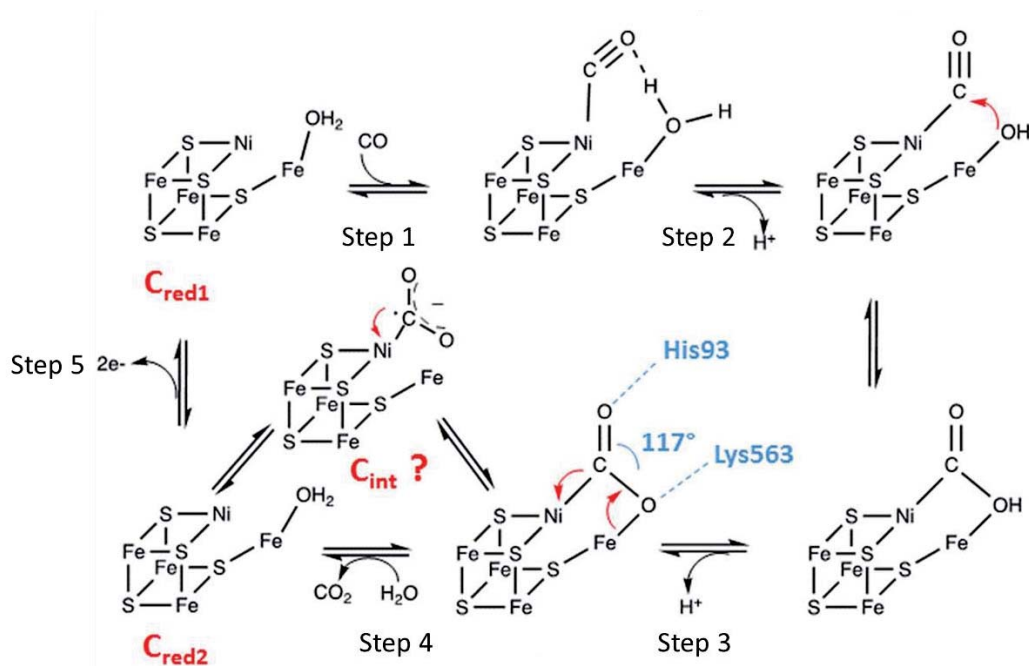


Figure 16. Proposed reaction mechanism of [NiFe]-CODH. Adapted from <sup>43</sup>.

The determination of diverse crystal structures, in complex with the CO-CO<sub>2</sub> substrates, inhibitors or analogues (CN<sup>-</sup> and NCO<sup>-</sup>), was also a great support to allow the proposition of the following mechanistic model (Figure 16):

Step 1. CO binds to Ni(II) and the H<sub>2</sub>O to the Fe(II) ion of the C-cluster (C<sub>red1</sub>) ;

Step 2. Fe-H<sub>2</sub>O deprotonates, resulting in a nucleophilic hydroxide species. OH<sup>-</sup> nucleophilic attacks of the CO results in a Ni-Fe bridged carboxylic group;

Step 3. Second deprotonation. The Ni-Fe now is bridged by a COO<sup>-</sup> intermediate;

Step 4. Release of the CO<sub>2</sub> and two electron cluster reduction (C<sub>red2</sub>);

Step 5. Electrons are transferred to the B-cluster and the C-cluster goes back to its C<sub>red1</sub> state.

Even though the mechanism appears clear, different points still need to be elucidated.

For example during the nucleophilic attack (Figure 16 step 2) a “carbon shift” should occur in order to place the carbon closer to the resulting hydroxide, which is otherwise too far away.<sup>68</sup>

Moreover, the C<sub>int</sub> state role during the CO<sub>2</sub> release process is still not clear. Electron transfer via the Ni center is too rapid to visualize reduced Ni states. To obtain more insights into this last step, structures of ChCODH-II bound to CO<sub>2</sub> and cyanate (NCO<sup>-</sup> a CO<sub>2</sub><sup>-</sup> analogue) have been recently determined to 1.03 and 1.06 Å resolution, using anaerobically poised crystals at -600 mV to mimic the C<sub>red2</sub> state.<sup>69</sup> Both structures showed a  $\mu_2, \eta^2$  bent bridging CO<sub>2</sub>/NCO<sup>-</sup>



coordination between the Ni and the Fe (step 2), with two unexpected identical C-O bonds lengths (1.32 and 1.3 Å). CO<sub>2</sub> is stabilized by H-bonding with His93 and Lys563. Interestingly, the certitude of the CO<sub>2</sub> binding geometry, ensured by the high atomic resolution, allowed the identification of the O-C-O angle as 117°, different from the previously reported value of 130°. This smaller value suggests that the CO<sub>2</sub> activation is mediated by a particularly nucleophilic Ni species that pushes electrons into CO<sub>2</sub>, making it appear as a two-electron-reduced carboxylate (Ni-COO<sup>-</sup>) waiting to be protonated, in favor of a 2e<sup>-</sup> transfer process. However, the existence of a transient CO<sub>2</sub><sup>·-</sup> intermediate radical cannot be completely excluded.

In addition, access of CO/CO<sub>2</sub> and proton transfer must be allowed between the buried active site and the solvent. Hydrophobic channels have been detected between the Ni ion and the protein surface as a possible gateway for the gas molecules.<sup>67</sup> A polar channel would be needed for the proton transfer and several water molecules linked by histidine residues were observed concerning the Ni to the solvent in all of the CODH structures (Figure 17). Moreover, a recent IR-X-ray-*in silico* combined study of the cyanide (CN<sup>-</sup>)/CODH complex showed that the proton channel between the C-cluster and the solution phase is strictly directional and dependent on the oxidation state of the cluster.<sup>70,71</sup>

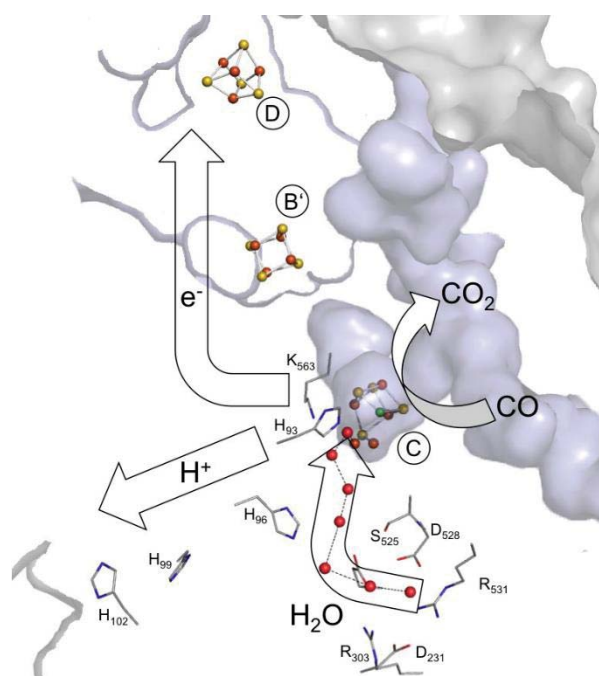


Figure 17. Proposed channels involved in the access and transfer of products and substrates. Adapted from <sup>67</sup>.



## 1.2 Nickel in the active site: where is it coming from?

Considering all the peculiar nickel active sites catalyzing the so far exposed reactions, the natural question that should arise from the reader should be: from where is this nickel coming from? How does it appear in the active site?

First, it is important to consider that this metal needs to be imported by the cell from the natural environment, passing through the cytoplasmic membrane. A tightly regulated import mechanism is essential and many organisms have developed high-affinity nickel transport systems to fulfil this role. In the case of gram-negative bacteria, nickel needs to first cross the outer membrane. Small molecules and ions can pass through this membrane by passive diffusion *via* non specific transmembrane porins but when their concentration is low, an energized import is required. Lately, the existence of dedicated transporters, namely TonB-dependent transporter (TBDT), has been postulated also in the outer membrane for nickel.<sup>72</sup> Once in the periplasm, two main classes of import systems are found in the cytoplasmic membrane: the ABC-type importers and the Ni/Co permeases.<sup>12</sup> (*NikABCDE transporter in E. coli is probably the best studied ABC-type importer, whereas HoxN in Cupriavidus necator is a good representative for nickel/cobalt permeases*).

Once nickel is present in the cell, its trafficking needs to be tightly controlled to avoid the presence of free toxic metal and it has to be correctly delivered and incorporated into the desired enzyme active sites. These operations often require an intricate team of accessory proteins, working in concert for the biosynthesis of these complex metalloclusters enzymes. Understanding the steps necessary for the cluster assembly and insertion into the enzyme, deciphering the maturation machinery involved, are some of the main challenges in the metallo-enzymology field. In the next session, the identified accessory proteins involved in the Ni-enzyme maturation pathways for urease, hydrogenase and CODH are discussed. A special attention will be devoted to nickel chaperones, a class of proteins able to reversibly bind the metal, transport it in the cell and deliver it to the apo-enzyme metallocenter.

### 1.2.1 Urease maturation

As already mentioned, urease is first produced in the apo-form which then needs to be activated through two main steps: lysine carbamylation and nickel insertion into the active site. Considering the buried position of this site in the apo-protein, the occurrence of a protein folding modification is predicted to allow nickel insertion.

To carry out these steps and fully complete urease biosynthesis, four accessory proteins are usually needed: UreD (UreH in *H. pylori*), UreF, UreG and UreE. From the studies of *Klebsiella aerogenes* urease, it was evidenced that the apo-enzyme can be partially activated upon nickel and CO<sub>2</sub> exposure but its activity is enhanced when it is co-expressed with UreD.<sup>13</sup> This suggests the need for the co-expression of structural and accessory proteins to fully activate the enzyme.

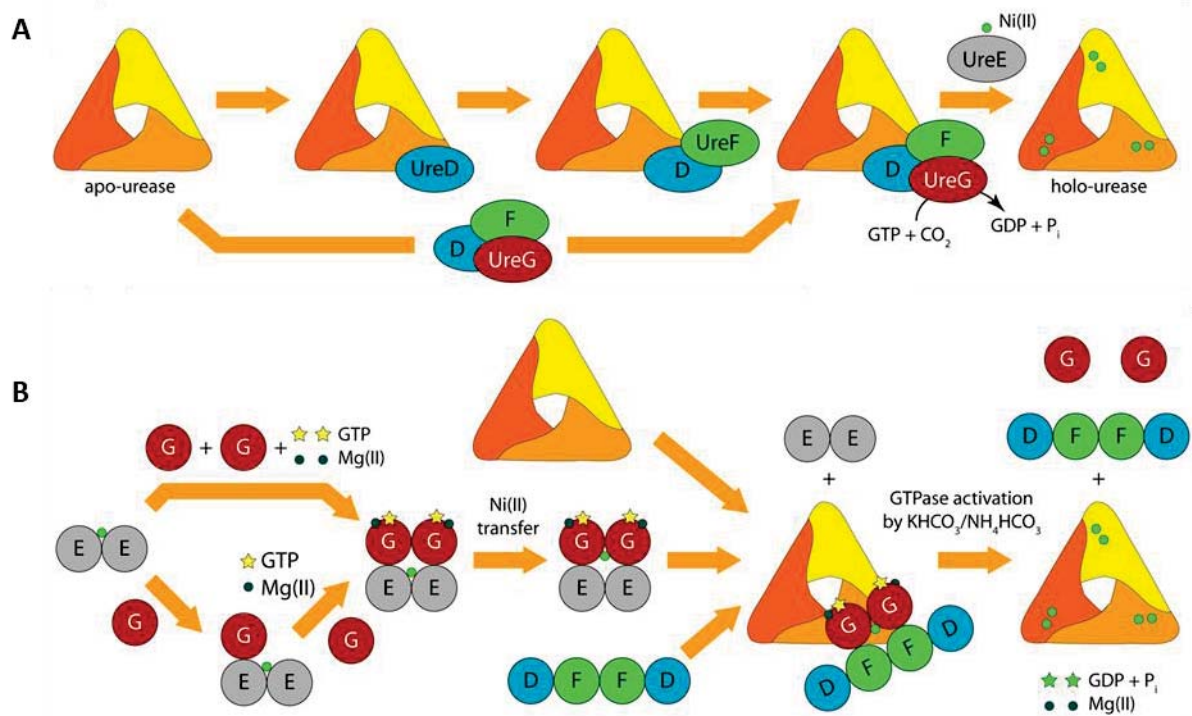


Figure 18. Proposed urease maturation process. A) Classical proposed mechanism B) the new proposed one. Adapted from <sup>13</sup>.

The classical model, currently proposed for the maturation mechanism even though the role of each single protein are not fully understood to date, is the following:

UreD, the first protein to interact with the apo-urease, binds to the enzyme and acts as a protein scaffold to promote the interactions of the other partners.<sup>73</sup> UreF interacts with UreD possibly provoking a conformational change of the enzyme around the active site and enhancing Ni accessibility.<sup>74</sup> At this point, UreG binds to UreF and promotes the lysine carbamylation, which relies upon GTP hydrolysis by UreG, defining this protein as the energy driving force. Finally, the urease-UreDFG complex interacts with UreE, responsible for the nickel insertion and completes the enzyme maturation. (Figure 18A) This model derives from the ability to isolate the apo-urease-UreDFG complex and for this reason it is proposed to be the functional unit.<sup>73</sup> However, no crystal structure is available for urease in complex with any of these accessory proteins. The complex UreHFG from *H. pylori* has been recently crystallized, showing a dimer of heterotrimers.<sup>75</sup> The interaction between UreG and UreE in *H. pylori* has been proven both *in vivo* (yeast two-hybrid analysis<sup>76,77</sup> and immunoprecipitation assay<sup>77</sup>) and *in vitro* (calorimetry and NMR spectroscopy<sup>78</sup>).

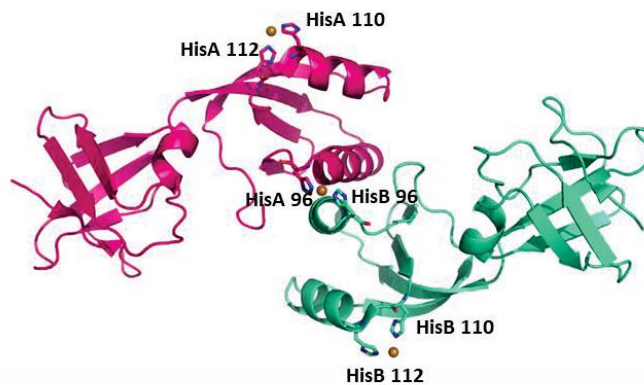
In view of the most recent results, a second model was postulated for the maturation of urease active site (Figure 18B).<sup>79</sup> In this scenario, a first interaction between Ni-UreE and UreG is proposed, forming a complex Ni-(UreEG)<sub>2</sub>. The complex (UreDF)<sub>2</sub> will then interact with Ni-(UreEG)<sub>2</sub>, competing for the Ni-UreG<sub>2</sub> binding. Nickel is delivered to the apo-urease by this last the complex Ni-(UreDFG)<sub>2</sub>.<sup>79</sup> A more detailed explanation about this last mechanism is reported in the next section, with a complete description of UreE, UreG and their complex.

## UreE

Historically, the role of this protein as nickel chaperone during urease maturation was suggested after examining the UreE sequences from *K. aerogenes* and *Proteus mirabilis*. Both sequences are, in fact, histidine rich in their C-terminal domains, well known as potential nickel binding motifs. However, the His-rich terminal region is not an essential prerequisite for the UreE family, considering that some members from different organisms have been identified lacking this, such as the one from *H. pylori* (HpUreE) (Figure 19). Moreover, their role in the enzyme maturation differs for each organism, from being essential for several *Helicobacter* species or substitutable by nickel addition in the growth medium for others.<sup>80</sup>

### UreE *Klebsiella aerogenes* 5-6 Ni per dimer (truncated version up to H144\* 1-2 Ni per dimer) (PDB:1GMW)

MLYLTQRLEI PAAATASVTI PIDVRVKSRV KVTI<sup>NDGRDA</sup> GLLLPRGLLL RGGDVLSNEE GTEFVQVIAA DEEVS<sup>VVRCD</sup>  
DPFMLAKACY HLGNRHVPLQ IMPGELRYHH DHVLD<sup>DDMLRQ</sup> FGLTVTFGQL PFEPEAGAYA SESH (GHHHAH HDHHAHSH)



### UreE *Helicobacter pylori* 1 Ni per dimer (PDB:3TJ8)

MIIERLVGNL RDLNPLDFS<sup>V</sup> DHV<sup>DLEWFET</sup> RKKIARFKTR QGKDIAIRLK DAPKLGLSQG DILFKEEKEI IAVNILDSEV  
IH<sup>I</sup>QAKSVAE VAKI<sup>CE</sup>Y<sup>EIGN</sup> RHAALY<sup>YGES</sup> QFEFKTPFEK PTLALLEKLG VQNRVLSSKL DSKERLTVM PHSEPNFKVS  
LASDFKVVVK

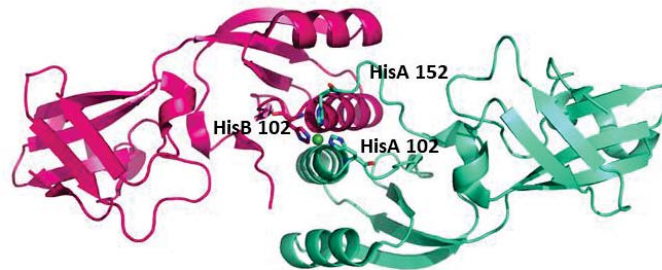


Figure 19. Top part: UreE from *K. aerogenes*. The full sequence is reported, the brackets show the truncated version in the case of the mutant H144\*. The crystal structure (PDB:1GMW) comes from this truncated version. Bottom part: UreE from *H. pylori*. The proteins are both in a dimeric conformation and present a nickel-binding site at the interface between the two monomers. In the sequences, the histidines are colored in red, the cysteines in green and the amino acids involved in nickel binding are underlined.

Due to the high variability of UreE sequences, the number of nickels bound to the protein varies depending on the organism, going from 1 Ni per dimer for *H. pylori* to 6 Ni per dimer for *K. aerogenes*. In order to understand why this histidine rich region is present only in some version of UreE and if its role is essential during urease maturation, different mutants have been produced. The result was that the truncation of the His-region in *KaUreE* (H144\*) does not completely hinder urease activation. This mutant can, in fact, still bind 1-2 Ni per dimer and activate the urease *in vivo*<sup>81</sup>, raising even more questions about the role of this double set of sites.<sup>82</sup> The crystal structure of this truncated version highlighted the presence of a second binding site in the N-terminal domain in addition to the histidine tail C-terminal domain, spatially separated by a flexible linker (similar to the N-terminal domain from *HpUreE* Figure 19). In the case of *H. pylori* UreE, the histidine rich domain was added to the protein sequence to check how the urease activation would differ. The constructs shown an increase in the urease activity level for high nickel concentrations, proving that a possible nickel sequestration mechanism via the his-tail could help the urease activation.<sup>83</sup>

From these results, the presence of two spatially separated sites suggests a possible double function of UreE. One site could be involved in the interaction with the other maturation proteins, while the other could be related to a metal transport/storage function.

*HpUreE* coordinates the Ni(II) ions at the dimer interface using two histidines from the well-structured region (one from each His102) and two histidines residues from the more flexible C-terminal part (His152). This mix between stable and flexible regions could be an indication of a transient conformation, needed to modulate the nickel binding and release.<sup>84</sup> Also in the case of *K. aerogenes*, the binding occurs at the dimer interface and involves the conserved histidine in the structural region (His96).<sup>81</sup> Considering its “open” conformation, it was suggested that the histidine in the C-terminal tail would probably closed up to coordinate metal (His 102-110), in agreement with the “close” structure found for *H. pylori* UreE. (see the comparison of the two structures in Figure 20)

The interaction between UreE and the other accessory proteins responsible for urease activation has been proven. Considering our interest in the nickel delivery process, only its interaction with UreG will be presented.



## UreG and nickel delivery

UreG is a GTPase, which couples GTP hydrolysis to allosterically control Ni binding.<sup>85</sup> The GTP-dependent conformational change of UreG is one of the key steps during the urease maturation. Interestingly, this protein has been identified as intrinsically disordered, belonging to the class of IDP (Intrinsically Disordered Protein), and with low GTPase activity if present alone. In order to achieve a good GTPase activity, UreG needs to form a complex with the other accessory proteins to promote its conformational change.

The UreHFG complex has been recently isolated and crystallized, showing a dimer of heterotrimers organization (Figure 20).<sup>75</sup> The metal-binding site of UreG is located at the dimer interface, binding one nickel per dimer using the Cys66 and the His68. This motif is essential to promote the nickel binding and its release is regulated via GTP hydrolysis, which induces a conformational change.<sup>86</sup> This nickel-binding site is interconnected by a tunnel to the UreH surface passing through UreF, as supported by structural model (Figure 20).<sup>74</sup>

Focusing on UreG, when UreDFG are associated together<sup>75</sup> the binding of GTP on UreG will cause its dissociation (Figure 21A step 1). Dimeric nickel bound UreE would bind two apo-UreG monomers, allowing the GTP uptake by UreG in the presence of Mg ions. In the (UreEG)<sub>2</sub> complex, the nickel will be transferred from UreE<sub>2</sub> to UreG<sub>2</sub> (Figure 21A step 2). At this point the complex (UreDF)<sub>2</sub> competes with UreE<sub>2</sub> for the Ni-UreG<sub>2</sub> complex in order to get to the main complex apo-urease-(UreDFG)<sub>2</sub>-Ni (Figure 21A step 3).

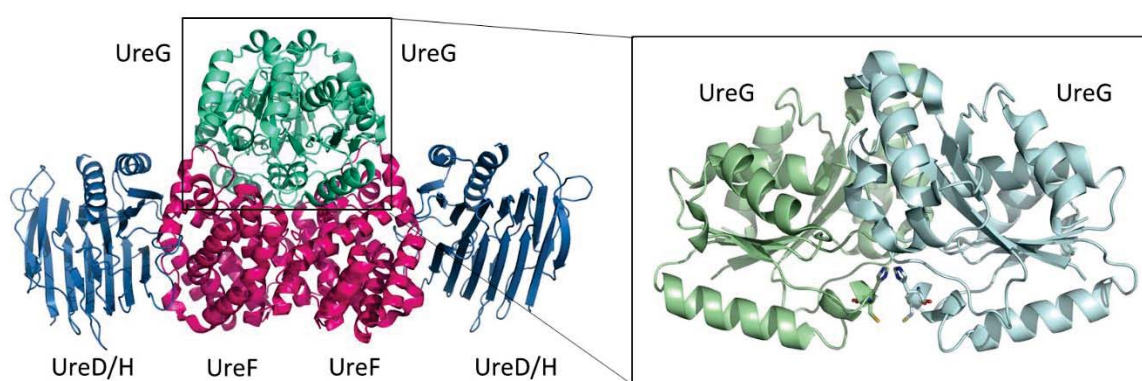


Figure 20. Overall structure of *Helicobacter pylori* UreGFH complex. UreG, UreF, and UreD/H are in green, magenta and light blue, respectively (PDB: 4HI0). Zoom in region of UreG homodimer structure. The nickel binding coordinating sidechains of two conserved metal binding residues from each monomer, Cys66 and His68, shown as sticks.



GTP hydrolysis will then induce a conformational change in the UreG unit that would cause the release of nickel to the enzyme, completing the insertion into apo-urease and reversing the UreG to its GDP-bound state. At this point, the UreDFG-GDP bound complex dissociates from the holo-enzyme and it is ready for the next activation round (Figure 21A step 4).<sup>86</sup>

Interestingly, *H. pylori* UreE was shown to interact with UreG to form a stable UreE<sub>2</sub>G<sub>2</sub> complex in the presence of Mg(II) and GTP, transferring the nickel ion from UreE to UreG.<sup>87</sup> In this scenario, the presence of the [NiFe]-hydrogenase accessory proteins is also postulated. In fact, apo-UreE would acquire the nickel ion from HypA (*nickel chaperone for the hydrogenase maturation 1.2.2*) and then transfer it to UreG (Figure 21B). Once again nature shows its variability, with maturation pathways of two different nickel-containing enzymes able to “cross-talk” between each other.

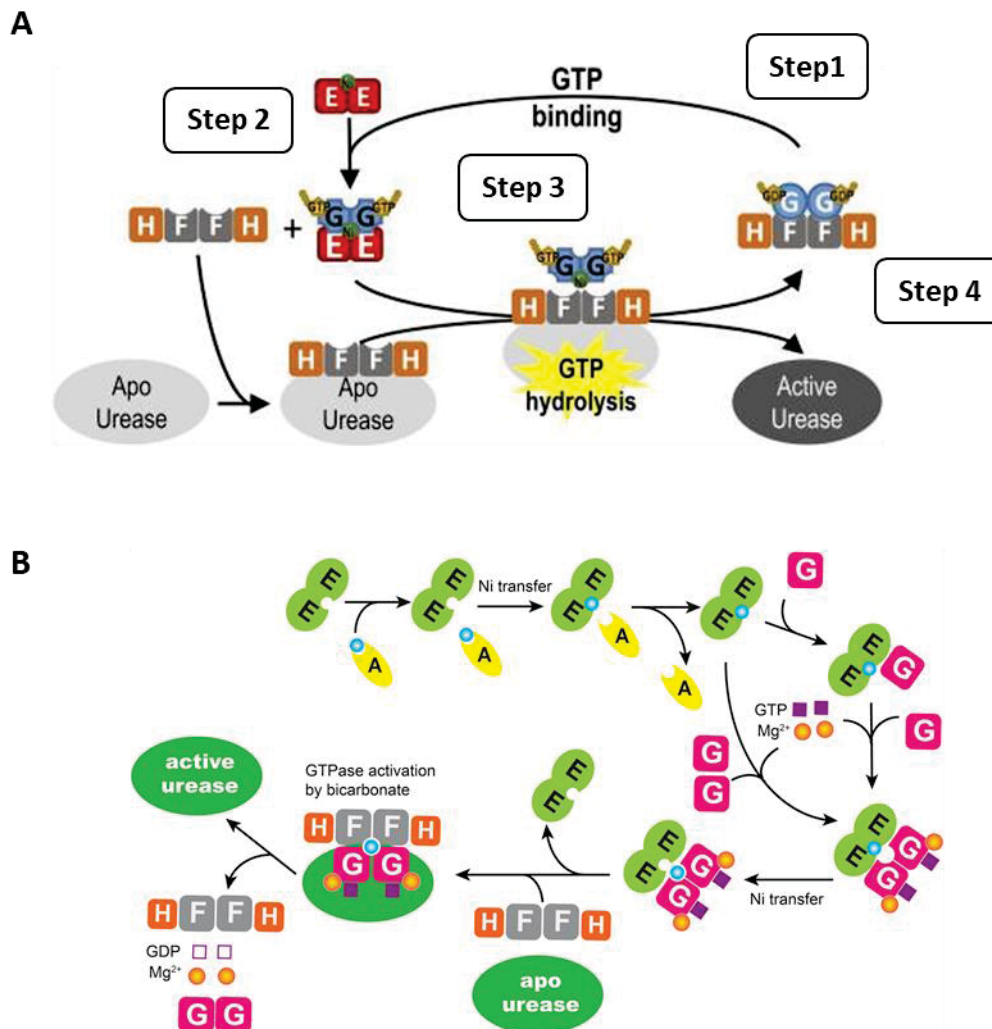


Figure 21. A) Proposed nickel insertion mechanism on apo-urease. B) Proposed nickel insertion mechanism on in the case of cross-talk for UreE and HypA in *H. pylori*. Adapted from <sup>86</sup> and <sup>87</sup> respectively.

### 1.2.2 [Ni-Fe] hydrogenase maturation

During the [Ni-Fe]-hydrogenase maturation process, the common thought is that  $\text{Fe}(\text{CN})_2\text{CO}$  unit is assembled and inserted prior to nickel. Thus, two consecutive events need to occur to promote the formation of this enzyme catalytic site: iron center biosynthesis and assembly and nickel delivery.<sup>88</sup>

In *E. coli* [Ni-Fe] hydrogenase 3, the large subunit (HycE) requires at least 8 accessory proteins for its synthesis encoded by the *hypABCDEF* operon, as well as *slyD* and *hycl* genes (Figure 22), and it is allegedly matured according to the following steps. At first, HypEF mediates the formation of the  $\text{Fe}(\text{CN})_2(\text{CO})$  unit in concert with HypDC. This step is followed by the insertion of this Fe site in the large subunit. The insertion of Ni occurs via a GTP-dependent mechanism mediated by HypAB + SlyD, followed by the proteolytic cleavage of the C-terminus of HycE by a specific protease (Hycl). It is presumably this last step that is responsible for the internalization of the catalytic center, essential to guarantee the enzyme activity (Figure 22).

Considering the scope of this thesis, I am going to focus on the nickel insertion step.

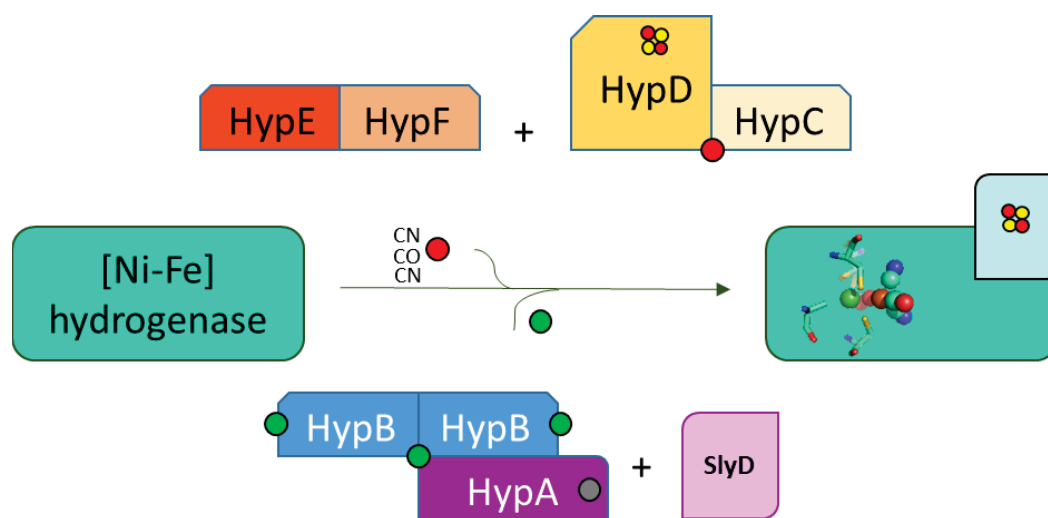


Figure 22. Over simplified allegedly maturation mechanism for [Ni-Fe]-hydrogenase. Adapted from <sup>88</sup>.

## Nickel delivery

Three different metallochaperones, HypA, HypB and SlyD, are thought to be involved in the hydrogenase nickel insertion. In *E. coli*, their cooperative role was suggested considering that they could be pulled down in a complex with the large subunit (HycE) (*see following section*). The impact of their knockout on the enzyme maturation showed that the hydrogenase activity could be partially restored via nickel addition in the culture media, reinforcing their need to obtain the maximal enzyme activity. Considering the impossibility to reach the same activity levels as in the wild type strain, it is evident that they play an important role in the delicate nickel equilibrium of the cell.<sup>89</sup>

For *H. pylori* it has been shown that both the presence of HypA and HypB are needed for the full activation of both the [Ni-Fe]-hydrogenase and urease *in vivo*, and *in vitro* studies also displayed an interaction between HypB and SlyD, even if the role of SlyD is really not clear.<sup>90,91,92</sup>

## HypA

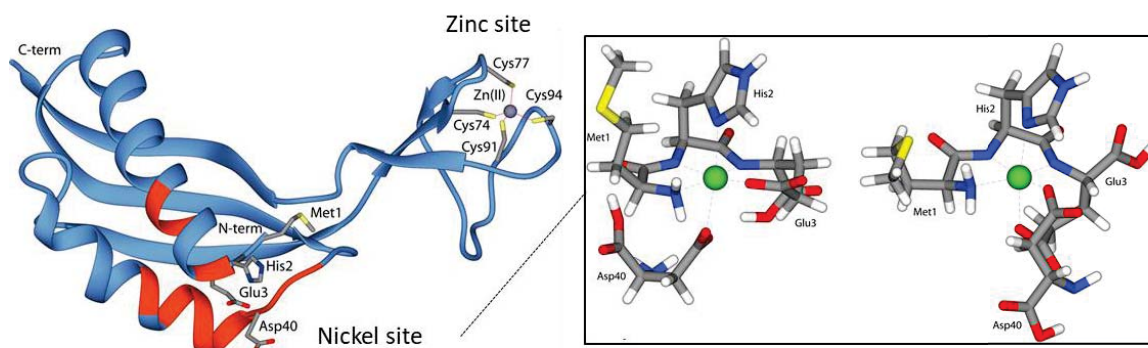
HypA is a ~13 kDa protein required for the activation of [Ni-Fe] hydrogenase 3 in *E. coli*.

In this organism, a homologous form of HypA is also present, HybF, responsible for the maturation of hydrogenase 1 and 2. Structures of HypA from *H. pylori* and from *Thermococcus kodakarensis* have been obtained via NMR and X-ray crystallography, respectively, both in a Ni-free form. These structures highlighted the elongated shape of the protein and the presence of two spatially separated domains connected via a linker: an N-terminal nickel-binding site and a C-terminal structural zinc site (Figure 23).

Concerning the **nickel-binding site**, deletion or mutation of the residues in the conserved N-terminal amino acid sequence (**MHE**) resulted in a loss of both hydrogenase and urease activities in *H. pylori*.<sup>91</sup> The coordination of this site is not clear and different hypotheses have been reported.

### HypA *Helicobacter pylori* (PDB:6G81)

**MHE**YSVVSSL IALCEEHAKK NQAHKIERVV VGIGERSAMD KSLFVSAFET FREESLVCKD  
AILDIVDEKV ELE**CKD**CSHV FKPNALDYGV **CEK**CHSKNVI ITQGNEMRLS SLEMLAE



### HypA *Escherichia coli*

**MHE**ITLCQRA LELIEQQAQK HGAKRVTGVW LKIGAFSCVE TSSLAFCFDL VCRGSVAEGC  
KLHLEEQEAE CW**CET**CQQYV TLLTQVRRC PQ**CH**GDMLOI VADDGLQIRR IEIDQE

### HypA *Thermococcus kodakarensis* (PDB:5AUN)

**MHE**WALADAI VRTVLDYAQR EGASRVKAVR VVLGELQDVA EDIVKFAMEQ LFAGTIAEGA  
EIEFVEEEAV FK**CRN**CNYEW KLKEVKDKFD ERIKEDI**HFI** PEVVHAFLAC PK**CG**SHDFEV  
VKGRGVYVAG IKIEKEGGS

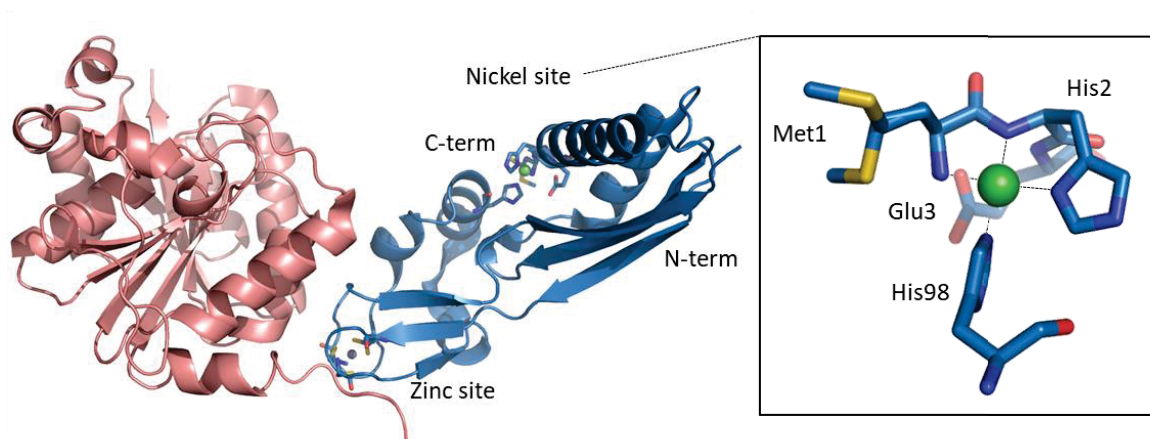


Figure 23. HypA from *H. pylori*, *E. coli* and *T. kodakarensis*. The sequences of the three proteins are reported highlighting the conserved N-terminal region (MHE) responsible for nickel binding and the CXXC (in orange) responsible for Zn binding. The NMR structure is reported for *H. pylori* with the latest proposed coordination motif for the Ni. The crystal structure of HypA-HypB complex from *T. kodakarensis* is shown (pink hypB, blue HypA) with its peculiar nickel binding motif in the zoom. As it is possible to observe, for both structures the Ni-site is spatially separated from the Zn-site via a flexible linker.

For *HpHypA*, a hexacoordinate nickel site has been recently suggested, formed by the N-terminal amine, His2 and Glu3 amide N atoms, His2 imidazole side chain and the carboxylate side chains of Glu3 and Asp40 (Figure 23).<sup>93</sup> In *TkHypA*, a square planar coordination was unveiled from the structure obtained in complex with *TkHypB*, which completes the His2-Glu3-amine coordination with an additional histidine (His98) ligand not conserved in *H. pylori* and *E. coli* (Figure 23). Interestingly, the interaction between *TkHypA* and *TkHypB* increases the affinity of the nickel site from micromolar to nanomolar range and it is a unique feature for this organism.<sup>94</sup> Comprehensively all these results indicates that the nickel transfer mechanism could be different among different organisms.

The **zinc site** is associated with two conserved **CXXC** sequences and is essential to guarantee the stability of the protein in solution. The affinity for this site is in the nanomolar range for *TkHypA*, and it contains most of the residues involved in the *TkHypA-TkHypB* interaction (Figure 23). For *HpHypA*, it was suggested that Zn coordination would change in response to nickel binding and a pH decrease, switching from a 4Cys coordination to a 2Cys-2His.<sup>95,96</sup> However, this hypothesis did not match the latest results obtained from NMR experiments on Ni binding to Zn-*HpHypA* according to the pH, which sees a stable Zn coordination unperturbed for nickel or pH changes.<sup>93</sup>

Considering that from the structural evidence the two sites are far apart, no information could be determined on how they could cooperate for delivering the metal. Recently, the new NMR study of *HpHypA* excludes intercommunication between the nickel and zinc sites and postulates that Ni, stabilizing the N-terminal region, would facilitate its interaction with the other partners.<sup>93</sup>

## HypB

HypB is a ~30 kDa metal-dependent NTPase<sup>97</sup>, presenting a high variability among different organisms, as we observed for HypA. They all possess at least one metal-binding site and NTPase activity, while a histidine rich region and a high-affinity N-terminal site are present in only some homologues. During the [Ni-Fe]-hydrogenase maturation, the NTPase activity is one of the needed features to insert correctly the nickel. The reason why is not totally understood, but it is probably related to a conformational change in the protein that results in a better nickel loading process. From the NTPase activity, the HypB homologues can be divided into either GTPases or ATPases. The GTP-dependent HypBs have low turnover rate in solution<sup>98,99</sup>, whereas the ATP-dependent ones behave differently according to the organism's origin. *TkHypB* accelerates its ATPase activity in the presence of HypA, while this is not the case for *E. coli* HypB.<sup>88</sup>

The X-ray structures of HypB from *Methanocaldococcus jannaschii*, *Archaeoglobus fulgidus*, *T. kodakarensis* and *H. pylori* have been determined. They all share a homodimeric conformation with the nucleotide-binding site close to the interface. The dimerization step is triggered by nickel and/or nucleotide binding on the protein, which accelerates the hydrolysis rate thanks to the docking of an invariant lysine, one from each monomer, to the bound nucleotide (Figure 24 *violet region highlighted on the sequences*). This dimer formation step is supposed to be essential to fully activate the enzyme. However, dimer-preventing mutants shown a partial hydrogenase activation ability, rising questions on the role of this process.<sup>88</sup>

Another way to divide the HypB homologues is by the number and the position of the metal-binding sites (Figure 24). All sequences present a metal binding-NTPase domain (Figure 24 *green region*), while the his-rich domain (Figure 24 *yellow region*) and/or the N-terminal domain (Figure 24 *red region*) is/are present only in some cases.



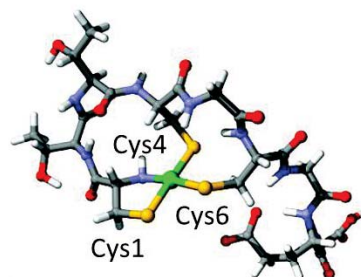
HypB *Bradyrhizobium japonicum* GTPase 3Ni sites (« high affinity site »; His rich; C-term « low affinity site »)

MCTVCGCSDG KASIEHAHDH HHDHGHDDH GHDGHHHHH GHDQDHHHHH DHAHGDAGLL DCGANPAGQK IAGMSSDRII  
QVERDILGKN DRLAADNRAR FRADEVLA FN LVSSPGAGKT SLLVRAVSEL KDSFAIGVIE GDQTSNDAE RIRATGVP AI  
QVNTGKGCHL DAAMVGEAYD RLPWLNGLL FIENVGNLVC PAAFDLGEAC KIVVFSTTEG EDKPLKYPDM FAASSMLLIN  
KIDLASVLDF DLARTIEYAR RVNPKIEVLT LSARTGEGFA AFYAWIRKRM AATTPAAMTA AE

HypB *Escherichia coli*- GTPase 2Ni sites (N-term « high affinity site »; C-term « low affinity site »)

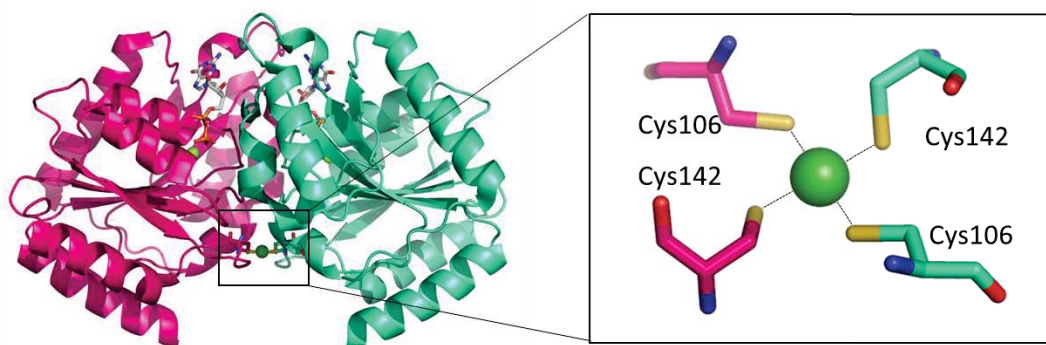
MCTTCGCGEG NLYIEGDEHN PHSAFRSAPF APAARPKMKI TGIIKAPEFTP SQTEEGDLHY GHGEAGTHAP GMSQRRMLEV  
EIDVLDKNNR LAERNRARA ARKQLVLNLV SSPGSGKTTL LTETLMRLKD SVP CAVIEGD QQTVNDAARI RATGTPAIQV  
NTGKGCHLDA QMIADAAPRL PLDDNGILFI ENVGNLCPA SFDLGEKHKV AVLSVTEGED KPLKYPHMF AASLMLLNKV  
DLLPYLNF DV EKC IACAREV NPEIEIILIS ATSGEGMDQW LNWLETQRCA

DFT high affinity site  
coordination model



HypB *Helicobacter pylori* (PDB:4LPS)- GTPase – 1Ni sites (C-term « low affinity site »)

MSEQRQESLQ NNPNL SKKDV KIVEKILSKN DIKAAEMKER YLKEGLYVLN FMSSPGSGKT TMLNLADEFK DFKFCVVEGD  
LQTNRDADRL RKKGVSAHQI TTGEACHLEA SMIEGAFDLL KDEGALEKSD FLIIENVGNL VCPSSYNLGA AMNIVLLSVP  
EGDDKVLKYP TMFMCADAVI ISKADMVEVF NFRVSQVKED MQKLKPEAPI FLMSKDPKS LEDFKNFLE KKRENYQSTH  
SF



## HypB *Methanocaldococcus jannaschii* (PDB:2HF8) – 1Ni sites (C-term « **low affinity site** »)

MHLVGVLDA KDILKANKRL ADKNRKLNLK HGVVAFDFMG AIGSGKTLLI EKLIDNLKDK YKIACIAGDV IAKFDAERME  
 KHGAKVVPLN TGKECHLDAH LVGHAEEDLN LDEIDLLFIE NVGNLICPAD FDLGTHKRIV VISTTEGDDT IEKHPGIMKT  
 ADLIVINKID LADAVGADIK KMENDAKRIN PDAEVVLLSL KTMGEFDKVL EFIEKSVKEV K

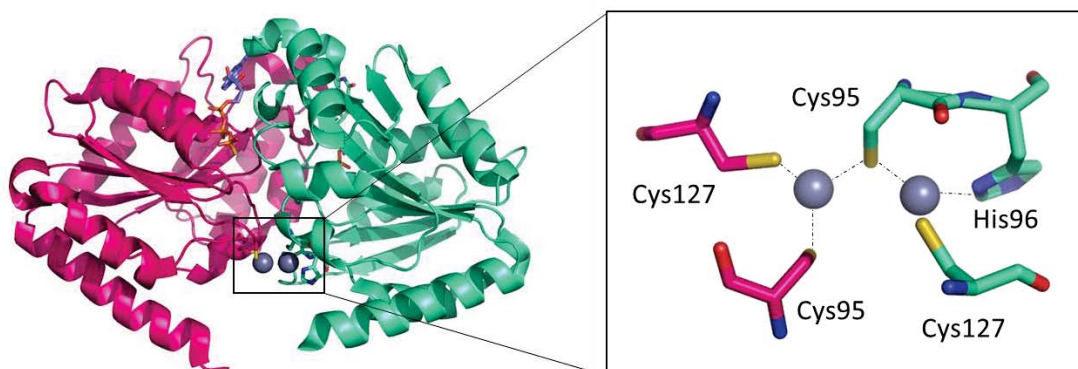


Figure 24. HypB from *B. japonicum*, *E. coli*, *H. pylori* and *M. jannaschii*. The sequences of the four proteins are reported highlighting (when present) in green the conserved “low affinity” site (CH-C), in red the N-term “high affinity” site, in yellow the histidine rich region and in violet the nucleotide-binding domain. For *E. coli* HypB the computational model of the N-term high affinity domain is reported. For *H. pylori* the crystal structure of the Ni-bound state shows the coordination of the nickel at the interface of the dimer. Only cysteines are involved in the coordination. For *M. jannaschii*, the crystal structure of the Zn-GTP-loaded HypB shows that the coordination includes the cysteines and one histidine. The histidines are colored in red, the cysteines in green and the amino acids involved in the nickel binding are underlined.

In *Bradyrhizobium japonicum* and *Rhizobium leguminosarum*, HypB possesses a histidine rich N-terminal region. In *E. coli*, HypB possesses two distinct nickel-binding sites. The first one is formed by a **CXXCGCG** motif at the N-terminus, known as the “**high affinity site**” due to its picomolar affinity. XAS and DFT analysis of this site identified three cysteines and the amine of the first amino acid as potential nickel-binding ligands, coordinating Ni in a square planar geometry.<sup>97</sup> Recently, its ability to bind Zn in a 4-coordinate tetrahedral geometry completed by 3 cysteines and 1 glutamate oxygen was unveiled via computational modeling and XAS, thus differing from the nickel one.<sup>8</sup> Mutations of this high affinity site resulted in a deficient hydrogenase maturation *in vivo* and in the loss of nickel binding ability *in vitro*.

The second site is in the NTPase domain and it is the one conserved among all the homologues. It consists of **one histidine and two cysteines** forming the so called “**low affinity site**”. The nickel affinity for this site is in the micromolar range, while zinc binds an order of magnitude



tighter.<sup>8</sup> Most probably, the role of this site is linked to NTP hydrolysis modulation, which is reduced when the metal is bound in *E. coli*. In the case of *H. pylori*, it was shown that a GTP/GDP loaded HypB, which possesses only the “low-affinity” site, causes a decrease in the nickel ion affinity and no changes for the zinc one. *H. pylori* HypB is the only structure obtained with loaded Ni and it displays a square planar geometry coordination constituted by four cysteines, two from each monomer, at the interface of the HypB dimer, suggesting that the role of the histidine in the conserved motif is most probably related to zinc binding (Figure 24).<sup>100</sup> Interestingly, the Zn coordination on the nucleotide-bound state differs from the Ni one and solution studies revealed the His as ligand donor, confirming the previous hypothesis. This is also reinforced by the structure of *Methanocaldococcus jannaschii* HypB loaded in Zn-GTP, which shows two zinc ions coordinated by four cysteines, two from each monomer, and one histidine (Figure 24). This evidence suggests a metal recognition role for the histidine, possibly used for a regulatory function.

The variety of HypB protein sequences is emblematic and it seemingly reflects the diversity in nickel needs for organisms living in different environments and how their protein production is regulated to fulfil as best the nickel request.

## SlyD

SlyD belongs to the FKBP family of the peptide-prolyl isomerases (PPIase).<sup>101</sup> This protein has been deeply study in *E. coli*, where its knock out has a low impact on the hydrogenase activity, in contrast to the results obtained from either HypA or HypB deletion<sup>102</sup>, suggesting a possible role of SlyD in optimizing the metal transfer to other partners.

Three different domains are identified from its structure: a C-terminal domain (Figure 25 *green region*), a PPIase domain (Figure 25 *blue region*) and an IF domain (Figure 25 *red region*).<sup>103</sup>

The metal chaperone properties of SlyD are mainly attributed to its C-terminal domain, unusually rich in potential metal-binding residues, such as histidines and cysteines. The *E. coli* SlyD C-domain has 15 histidines and 6 cysteines able to coordinate multiple nickel ions. The C-domain is a transiently disordered region<sup>104</sup>, which is structured in the presence of metal. The metal binding affects the PPIase domain conformation, although PPIase activity has been shown to be not essential during hydrogenase maturation.

During the hydrogenase maturation pathway, three main roles are supposed for this protein: as a nickel source for HypB, as a promoter for the nickel release from HypB after its binding to this protein or as a direct nickel donor to the hydrogenase 3 in *E. coli*.<sup>105</sup>

**SlyD** *Escherichia coli* (PDB: 2KR7) (IF domain ; PPlase domain; C-domain)

```

MKVAKDLVVS LAYQVRTEDG VLVDESPVSA PLDYLHGHGS LISGLETALE GHEVGDKFDV
AVGANDAYGQ YDENLVQRVP KDVFMGVDEL QVGMRFLEET DQGPVPVEIT AVEDDHVVVD
GNHMLAGQNL KFNVEVVAIR EATEEELAHG HVHGAHDHHH DHDHDGCCGG HGHDHGHEHG
GEGCCGGKGN GCGCH

```



Figure 25. SlyD from *E. coli*. The sequence shows the high density of potential metal-binding residues and the organization of the three different domains, with the C-term (responsible for the nickel binding and that not present in the crystal structure) in a flexible conformation. The IF domain is highlighted in red, the PPlase domain in blue and the C-domain in green. In the sequence, the histidines are colored in red and the cysteines in green.

### HypA-HypB-SlyD interaction

As above described, HypA, HypB and SlyD have a substantial role in the nickel ion delivery during [Ni-Fe]-hydrogenase maturation. Their cooperation is proposed in different models and a recent study discards the idea of a multi-complex formation, re-orienting it towards a “pair-related” scenario with different relative affinities as HypB-SlyD > HypB-HypA > HypA-SlyD.<sup>106</sup> HypB-SlyD interaction results in a decrease of the nickel affinity of the “high affinity site” of HypB, promoting the metal release. HypB-HypA complex is formed to transfer nickel from the low affinity-domain site of HypB to HypA, a process that occurs faster in the presence of GDP than GTP (for *E.coli*). When HypA receives the metal, it dissociates from HypB. HypA-SlyD interacts poorly. In addition, the presence of SlyD prevents HypA binding to HypB, indicating that they most likely bind to a similar region of the protein and highlighting the central role of HypB during the nickel insertion in the enzyme maturation steps.

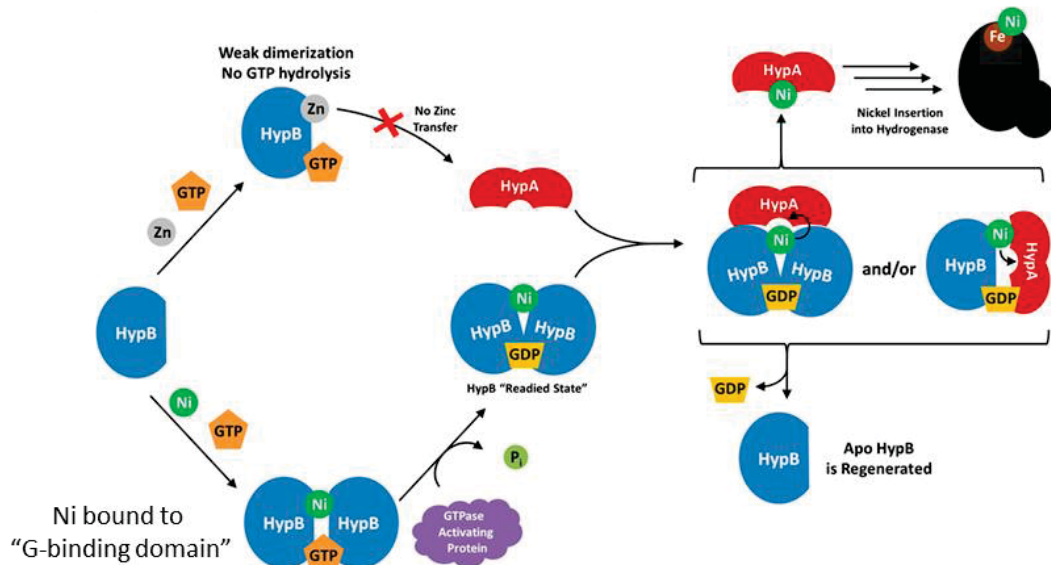
In *E. coli*, two different nickel transfer pathways are proposed, based on the nature of the two nickel-binding sites of HypB involved (“high-affinity site” vs “low affinity site”) (Figure 26).

**HypB-HypA nickel transfer model.** In this model, GTP bound HypB would transfer the nickel from its “low affinity domain” to apo-HypA upon GTP hydrolysis, a factor that induces the formation of the HypA-HypB complex and decreases the nickel-binding affinity of HypB. At this point, HypA-Ni loaded will bind to the hydrogenase large subunit and release the metal. The zinc would play in this context a possible regulatory role, because loaded Zn-HypB does not interact with HypA.<sup>107</sup>

**HypB-SlyD nickel transfer model.** This second model sees the N-terminal “high affinity” site of HypB as the nickel donor, in a process facilitated by SlyD. The interaction of these two proteins occurs between a proline-rich region of HypB, situated between the GTPase domain and the N-terminal site, and the IF domain of SlyD. The formation of the HypB-SlyD complex would promote the release of Ni from the high-affinity site towards other acceptors, considering that HypB does not interact with the hydrogenase large subunit. The role of SlyD would facilitate the nickel release that will be ultimately delivered to the hydrogenase. Also in

this case, the interaction between the two proteins blocks the release of zinc, possibly suggesting a regulatory role of the metal pathway.<sup>108 106</sup>

### HypB-HypA nickel transfer model



### HypB-SlyD nickel transfer model

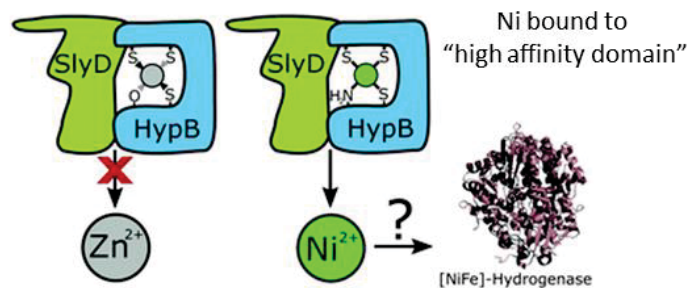


Figure 26. Two different nickel insertion models in the large subunit of hydrogenase 3 in *E. coli*. Adapted from <sup>107</sup> and <sup>8</sup>, respectively.

### 1.2.3 [NiFe]-Carbon Monoxide Dehydrogenase maturation

As for the hydrogenase, the maturation step of [NiFe]-CODH is thought to involve two consecutive steps: an initial Fe/S cluster insertion followed by the nickel delivery.

In our study model *R. rubrum* the operon *cooFSCTJ*, present downstream of the hydrogenase one, encodes for a ferredoxin CooF, the CODH (*cooS*) and three different nickel accessory proteins CooC, CooT and CooJ (Figure 27).<sup>109</sup> The expression of this operon is regulated by CooA, a CO-sensing transcriptional activator responsible also for the hydrogenase expression.<sup>110</sup> CooA possesses a DNA-binding domain and two heme domains. The latter are responsible for CO binding, which induces an allosteric conformational change on CooA that promotes the interaction with specific target regions, resulting in positive gene regulation.<sup>111</sup> CooF is responsible for the intermolecular electron transfer, from the CODH to its partner hydrogenase. CooC is the most recognized and distributed accessory protein in the genomes, while CooJ and CooT are trickier to identify.

As for HypA and HypB during hydrogenase maturation, CooC is essential for CODH maturation in the absence of high nickel concentrations. In fact, in nickel-depleted culture media, CODH is produced as a deficient-Ni form, suggesting that its accessory proteins are needed for Ni insertion.<sup>112</sup> As shown in *R. rubrum*, strain with a deletion or mutation of the *cooC* gene require a 1000-fold-higher nickel concentration than the wild type strain to be able to grow in presence of CO, while a 50-fold-higher concentration was needed to compensate for *cooJ* and/or *cooT* deletions. However, the enzyme has only 50-60% of the native activity, showing the requirement of the three proteins for the full activation of CODH.

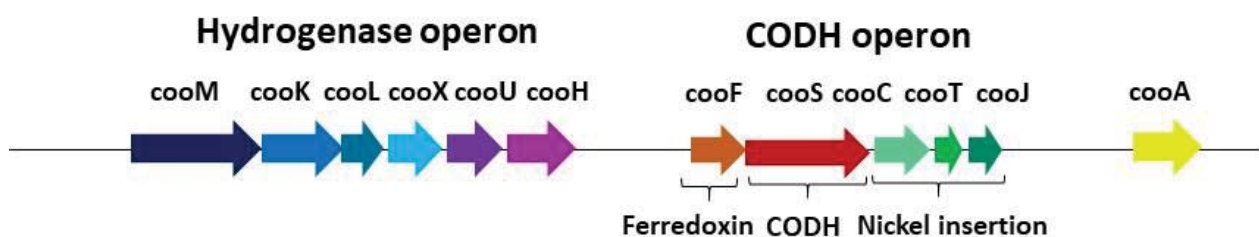


Figure 27. Hydrogenase (*cooMKLXUH*) and CODH (*cooFSCTJ*) operons regulated via transcriptional activator CooA in *R. rubrum*.

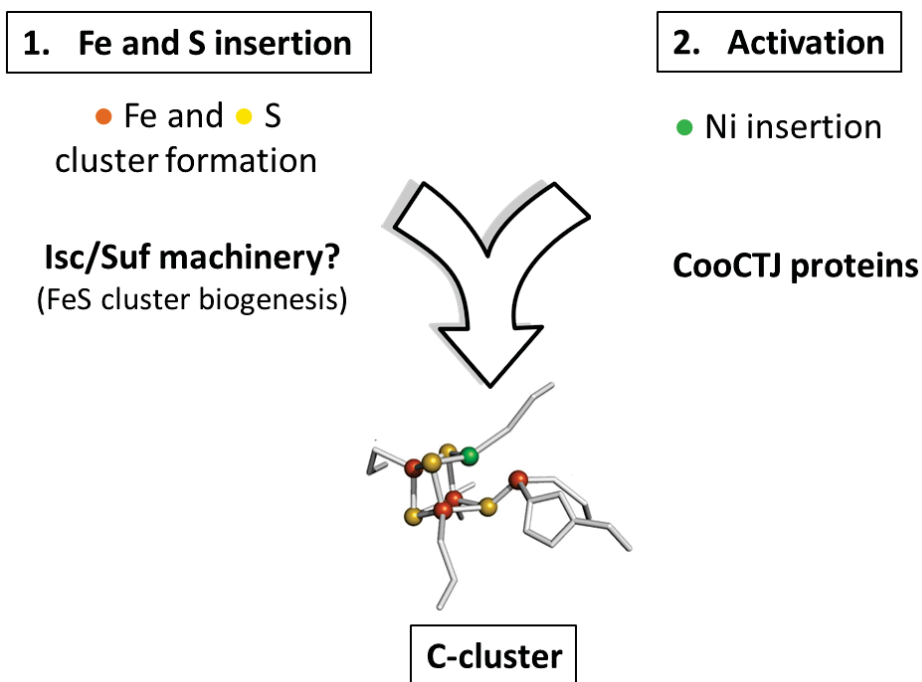


Figure 28. Metal insertion into CODH. (A) Schematic representation of Fe, S and Ni insertion and metallochaperones.

The precise role of these proteins and their interplay still needs to be elucidated but they are all required for the optimal biosynthesis of the nickel center (Figure 28).

Concerning the building of the FeS unit in the C-cluster, the involved accessory proteins have not been identified to date. It is clear, however, that it is not related to the presence of nickel during the culture growth. In fact, the FeS cluster is correctly inserted in the enzyme nickel-free form and can be partly re-activated *in vitro* by addition of nickel salt under reducing conditions.<sup>113</sup> Trickier is to identify what occurs concerning the mononuclear Fe<sub>u</sub> site. Mössbauer studies showed that, if nickel is not present, this site is not detectable and the spectroscopic properties match those of a classical 4Fe-4S cluster.<sup>114</sup> Under reductive conditions, the addition of nickel in the presence of CO is sufficient to activate a Ni-free C-cluster, indicating that this metal insertion might modify the coordination environment of one of the fourth iron atoms. From the crystal structure, it was evinced that the amino acids that play a role in the coordination of the C-cluster are His265, Cys300, Cys338, Cys 451, Cys481 and a Cys531 directly coordinating the nickel ion (Figure 29) . Surprisingly, single mutations of the Cys338, Cys 451, Cys481 and even Cys531 did not affect the nickel binding ability, which remains 1 nickel atom per cluster.<sup>115</sup> However the CODH activity of these mutants was strongly affected, exhibiting low or no activity, evidence that these amino acids are essential for the correct C-cluster coordination.

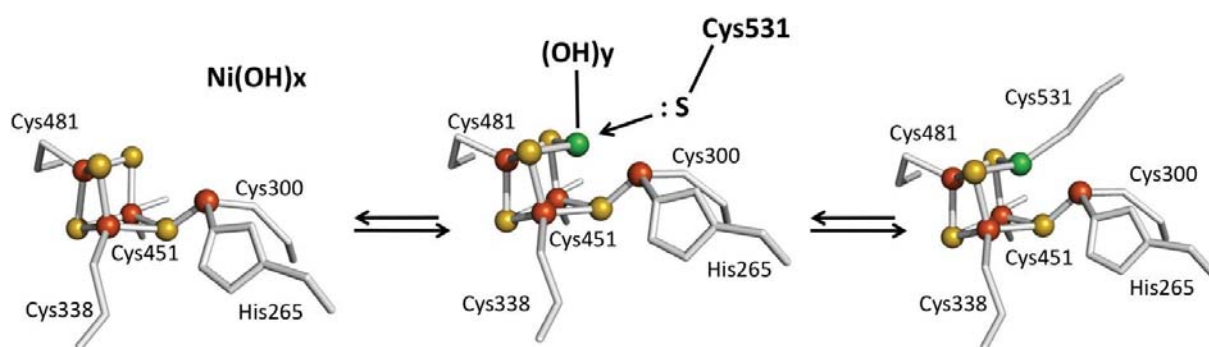


Figure 29. Proposed model for the mechanism of Ni insertion into Ni free-CODH in *R. rubrum* under reducing conditions, adapted from <sup>43</sup>.

Interestingly, the mutation of His265 that coordinates the mononuclear  $\text{Fe}_u$  ion, showed that this is the only essential amino acid for nickel binding, as a His265Ala mutant contains only 0.25 Ni atoms per monomer.<sup>114</sup> The mechanism by which this amino acid would influence the assembly of the cluster is still unknown but a model has been proposed based on these site directed mutagenesis studies. In this optic, the [3Fe-4S]-subcluster would bind the nickel and the thiolate group of Cys531 would convert the not active Ni-complex into an active one (Figure 29).<sup>115</sup> However, this model does not fit the general agreement of the scientific community, in particular on the holding coordination of the mononuclear iron site, which would indicate the presence of a distorted and open 4Fe-4S cluster coordination. If this supposition is incorrect and the cluster has a classical 4Fe-4S cubic coordination, as shown by spectroscopic studies, it would imply that the nickel insertion would cause a bond breaking. This could be an index of why the His265 has an impact on the nickel insertion even if it is not one of its ligands.

Recently a new study on CODH from *Desulfovibrio vulgaris* showed the plasticity of the C-cluster, from the capturing of a crystal in the oxidized state.<sup>116</sup> Their observations agree with the “bond breaking” theory due to the fact that they observe a Ni ion bound to the site formerly occupied by  $\text{Fe}_u$ , coordinated by the equivalent His265 and Cys300 (Figure 30). However, in their model, the  $\text{Fe}_u$  would not be in a prior [4Fe-4S] cluster but it will form an open bridge position between the nickel, a sulfur and the remaining [3Fe-3S] cluster and, such a conformation has never been observed during the spectroscopic studies.



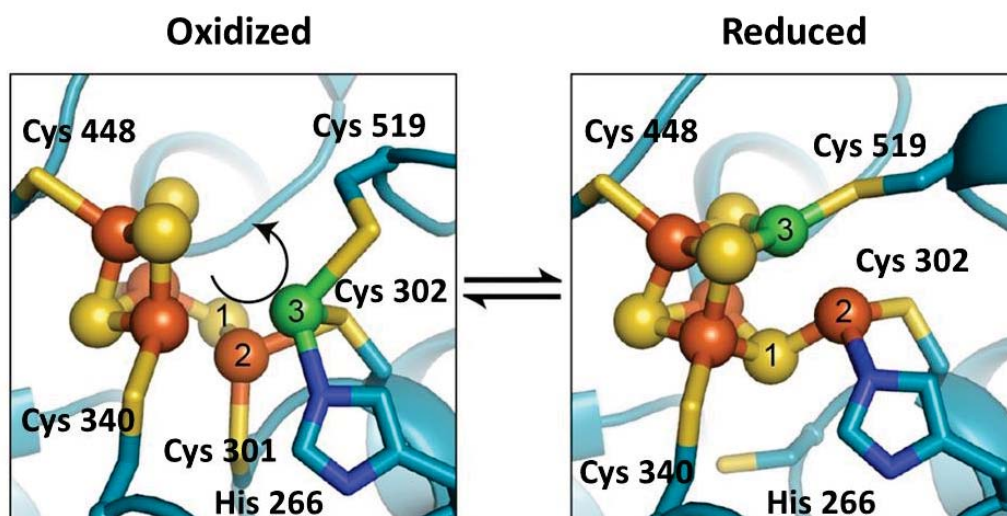


Figure 30. Proposed model for the conversion between oxidized and reduced state for *DvCODH*. The numbers indicates: 1-sulfur, 2-Fe<sub>u</sub> and 3-nickel. As it is possible to see, in the reduced conformation the cluster adopts the canonical conformation. In the oxidized one, the Fe<sub>u</sub>, Ni and one of the sulfurs are in a shifted position and the Cys 301 appears as a new coordinating residue. The arrow and numbers indicate the assumed direction of metal ion movement between the two states. Adapted from <sup>116</sup>.

Regardless, this study is an additional proof of the inherent plasticity of the C-cluster.<sup>116</sup> Additional studies are needed in order to shed light on the C-cluster formation and insertion, but it is nonetheless clear that the key step during the CODH activation is nickel insertion. In the next paragraphs, the characterization of the three nickel chaperones CooC, CooT and CooJ will be addressed.



## CooC

Similar to UreG and HypB, required respectively for the maturation of urease and hydrogenase, CooC is an NTPase involved in CODH maturation (Table 1).<sup>112</sup> As previously shown, these NTPase play an active role in their respective enzyme maturation pathways, actively directing or modulating the interaction between the different partners involved in the nickel delivering process.

Two major groups of CooC have been identified to date, depending on their genomic context (Figure 31):

- “CooC-like” group, positioned in the monofunctional CODH operons;
- “AcsF-like” group, found in the ACS/CODH operons (or gene clusters).

Interestingly, the ACS/CODH gene clusters can have either one CooC or two CooC homologues, one belonging to the “CooC-like” group and the other to the “*acsF*-like” group.<sup>117</sup>

All CooC homologues present a strictly conserved **CXC** motif (*red highlighted region Figure 32*) for nickel binding and a conserved **GKGGVGKS** P-loop nucleotide-binding motif (*cyan highlighted region Figure 32*) in the N-terminal part of the sequence, which catalyzes the hydrolysis of ATP to ADP (Figure 32).

| Protein     | Organism             | P-loop motif | Molecular mass (kDa) | Nucleotide hydrolysis                                |
|-------------|----------------------|--------------|----------------------|--|
| <b>CooC</b> | <i>R. rubrum</i>     | GKGGVGKS     | 27.8                 | ATP $K_M$ 24.4 $\mu$ M<br>GTP $K_M$ 26.0 $\mu$ M     |
| <b>UreG</b> | <i>M. jannaschii</i> | GTPGAGKT     | 26.2                 | $k_{cat}$ 0.026 min <sup>-1</sup>                    |
| <b>HypB</b> | <i>E. coli</i>       | SSPGSGKT     | 31.6                 | $k_{cat}$ 0.2 min <sup>-1</sup><br>$K_M$ 4-7 $\mu$ M |

Table 1 Comparison between CooC<sup>118</sup>, UreG<sup>119</sup> and HypB<sup>120</sup>.

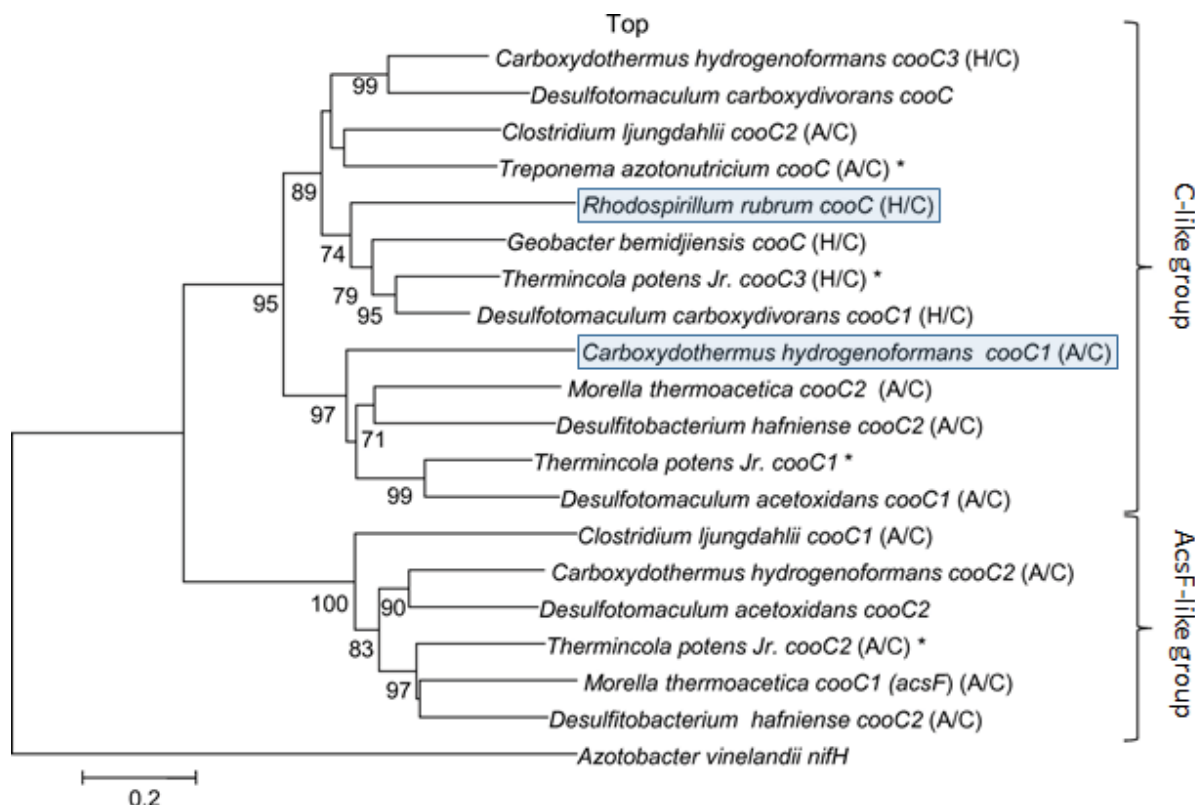


Figure 31. A phylogenetic tree shows the group division between “C-like” and “AcsF-like” groups of CooC homologs. The CooC described in this chapter are highlighted in light blue. Adapted from <sup>117</sup>.

In *C. hydrogenoformans*, CooC1 is located, together with AcsF (formally CooC2), in the CODH-III-ACS gene cluster.<sup>121</sup> Another CooC homologue is present in this organism, *ChCooC3*, found in the CODH-I gene cluster. Initially predicted as involved in CODH maturation, *ChCooC2* has been recently shown to be implicated in ACS maturation, forming a stable 2:1 complex with the enzyme and increasing its activity in the case of heterologous co-expression.<sup>122</sup> Thus, being a member of the “AcsF-like” group<sup>123</sup>, *ChCooC2* has to be correctly named AcsF.

*ChCooC1* is part of the “CooC-like” group involved in the CODH maturation process. *In vitro* studies, demonstrated that the protein is a monomer in the apo state and it dimerizes upon binding of metal (Zn or Ni) and ADP/ATP.<sup>124</sup> The nickel coordination, besides triggering the formation of the dimer, can change upon ATP/ADP binding, providing evidence of a cross-correlation between the nucleotide and metal sites.

### CooC *Rhodospirillum rubrum*

MKIAVTGKGG VGKSTIVGML ARALSDEGWR VMAIDADPDA NLASAIGVPA ERLSALLPIS KMTGLARERT  
 GASETTGTHF ILNPRVDDIP EQFCVDHAGI KLLLMGTVNH AGSGCVCPEH ALVRTLLRHI LTKRKECVLT  
 DMEAGIEHFG RGTIEAVDLL VIVIEPGSRS LQTAAQIEGL ARDLGIKTIC HIANKLASPV DVGFIIDRAD  
 QFDLLGSIPF DSAIQAADQA GLSCYDLSPA CRDKAHALMA ALLERVGPQTQ GVS

### CooC1 *Carboxydotherrmus hydrogenoformans* (PDB:3KJI)

MKLAVAGKGG VGKTTVAAGL IKIMASDYDK IYAVDGGPDS CLGQTLGLSI EEAYAITPLI EMKDEIREKT  
 GDGGLLILNP KVDGDLKYG RYIDDKIFLI RMGEIKKGGG QCYC RENSFL GSVVSALFLD KKEAVVMDMG  
 AGIEHLTRGT AKAVDMMIAV IEPNLNSIKT GLNIEKLAGD LGIKKVRYVI NKVRNIKEEK LIKKHLPEDK  
 ILGIIPYNEL FIELSLKGEE IWQSTNPAFV NLHDIYQKLR LEVG

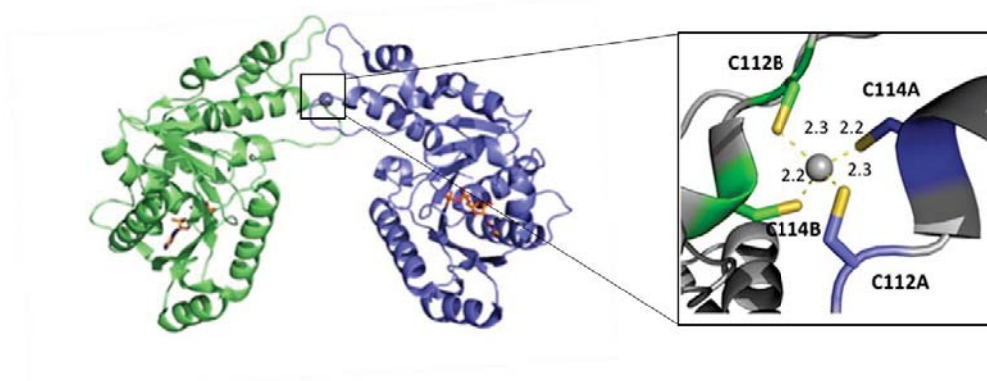


Figure 32 Sequence similarity between CooC from *R. rubrum* (accession number: P31897) and CooC1 from *hydrogenoformans* (accession number: Q3ACS5). The nucleotide-binding site is composed of the P-loop domain in cyan, switch I domain in blue, switch II domain in grey and the metal-binding site is highlighted in red. The overall structure is presented for *ChCooC1* homodimer in complex with Zn and ADP (PDB code: 3KJI). The two monomers are in green and blue, ADP molecules are depicted as orange sticks and Zn is depicted as a grey sphere. The metal-binding site at the dimer interface is in the zoom.

Yet, this alteration in coordination does not happen in the presence of Zn, which also hinders the *ChCooC1* ability to bind Ni. In fact, *ChCooC1* binds Zn more tightly than Ni, as shown by a competitive binding experiment where  $\text{Ni}^{2+}$  was unable to replace the bound  $\text{Zn}^{2+}$  on the protein ( $K_d \text{ Ni} = 0.41 \pm 0.05 \mu\text{M}$ ,  $K_d \text{ Zn}$ , not determined).<sup>124</sup> Since it is not clear whether the role of zinc is physiological or not, these results should not be taken as a proof that *ChCooC1* does not act as a nickel binding or transport protein. Instead, it could be a sign, as it is for HypB, that the protein could use this Zn/Ni binding to regulate the interaction with other partners and modulate the release of Ni.

To better decipher the *solution* results, several crystal structures were collected for *ChCooC1* in the presence or absence of ADP and Zn(II), showing a homodimer that coordinates the metal at the dimeric interface using the conserved cysteine motif **CXC**.<sup>121</sup>

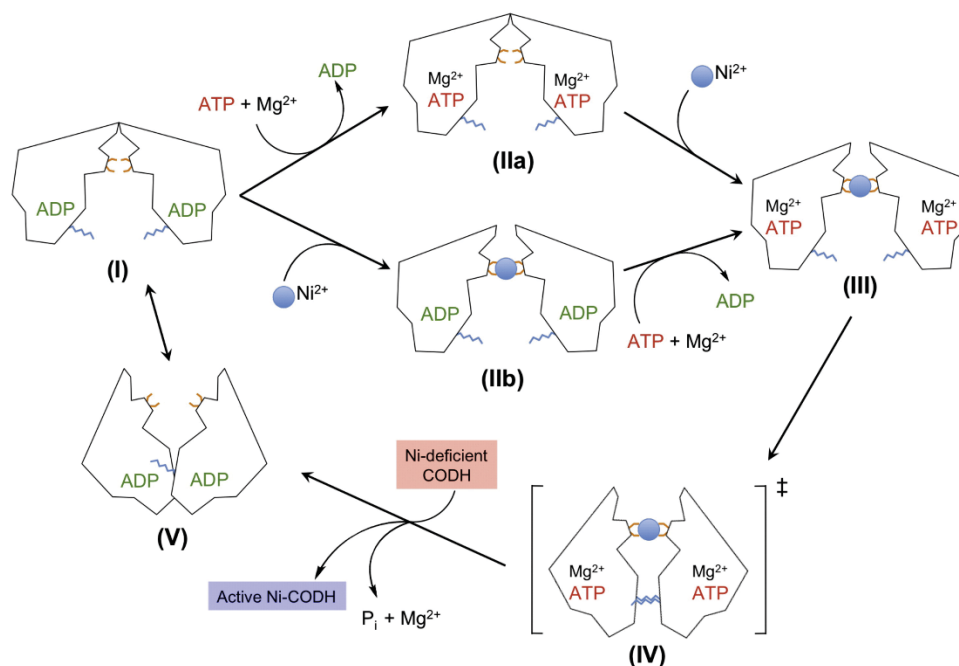


Figure 33. Proposed model for nickel binding and releasing by *ChCooC1*. The ADP-bound dimeric state (I) was described via X-ray crystallography. Considering the impossibility to crystallize the protein with bound Ni, the Zn/ADP-bound *CooC1* (IIa, IIb, and III) is used as a model for the Ni/ADP or ATP-bound dimer. The possible ATP triggered conformational change is shown in the state (IV), followed by state (V) where ATP hydrolysis releases Ni. The cysteines in the C-X-C motif are represented in orange, and the signature lysine is shown in blue. Reported from <sup>121</sup>.

The ATPase motif is spatially separated from the metal-binding motif, on which the Zn is tetrahedrally coordinated by four thiolate groups coming from the two cysteines of each monomer. Considering the inability to obtain a Ni-bound structure, it has been postulated that Ni coordinates with a different geometry (probably square-planar), possibly more “open”, necessary to promote a dynamic cycle between Ni bound state and ATP hydrolysis (Figure 33).

The proposed model for nickel binding and release is the following. Ni(II) binding induces the formation of a stable dimeric *ChCooC1*, in which the P-loop would be in an open conformation. ATP binding promotes a conformational change (Figure 33 step IV), allowing the release of the metal (Figure 33 step V). How this conformational change occurs is not clear, it could be promoted by the ATP hydrolysis that could also play a role in modulating the interaction of *ChCooC1* with other partners. Taken all together, these data suggest that *ChCooC1* cycles dynamically between an open Ni<sup>2+</sup> state and a closed ATP-dependent state that drives the metal release.

In the case of CooC from *R. rubrum*, the available information comes from its purification from the native organism. It is reported to be a homodimer, containing less than 0.1 atom of  $\text{Ni}^{2+}$  or  $\text{Zn}^{2+}$  and being able to catalyze the hydrolysis of both ATP and GTP.<sup>118</sup> A mutation in its P-loop domain demonstrated the importance this region (Figure 32 highlighted in cyan). In fact, the mutant K13Q-*RrCooC* hinders both the ATP hydrolysis *in vitro* and nickel insertion in CODH *in vivo*, suggesting that *RrCooC* promotes the nickel insertion into nickel deficient CODH subsequent to the ATP hydrolysis.<sup>118</sup> This is an additional indication that nickel insertion in the active site is an energy requiring processes.<sup>112</sup> In fact, as it was postulated for *ChCooC1*, the nickel release to the enzyme could depend on an ATP-induced conformational change of *RrCooC*. The ability of *RrCooC* to hydrolyze GTP may lead the reader to suppose a divergence with *ChCooC1* (which shows only ATPase activity). However, when activation tests were performed on nickel depleted whole cell extracts (thus possessing CooCTJ), an enhanced activity was observed upon the addition of ATP+ $\text{Ni}^{2+}$ , whereas GTP+ $\text{Ni}^{2+}$  addition resulted in a loss of activity. This suggests that ATP is the physiological substrate of *RrCooC*, as it is for *ChCooC1*.<sup>118</sup> No structural studies are available for this protein. Obtaining the recombinant *RrCooC* from heterologous expression in *E. coli* could be of great help to fully characterize this protein *in vitro*.<sup>121</sup>

## CooJ

CooJ is a 12.5 kDa metallochaperone that had only few biochemical data prior to our studies.<sup>125</sup> It contains a His-rich domain in the C-terminal region, formed by 16 His and 2 Cys, reported to be able to bind up to four Ni per monomer with a  $K_D$  of 4.3  $\mu$ M.<sup>125</sup> During its purification from *R. rubrum*, the protein co-eluted with the CODH and other proteins, suggesting the formation of a multi-protein complex between CODH, CooJ and additional putative partners.

The presence of a histidine-rich cluster (HRC) in nickel-binding proteins involved in Ni-enzyme maturation, notably in some UreE and HypB homologues, has already been described in the previous chapters and may be related to nickel storage and/or detoxification activities (Table 2). Intriguingly, a deletion of the CooJ region containing the His-tail was still able to activate the CODH *in vivo*, displaying ~41% of wild-type activity but constant and independent from the nickel concentration in the culture medium.<sup>118</sup> This result suggests that the HRC could be dispensable for the activation process and that it may be involved in another function rather than only CODH maturation.

The role of histidine rich regions in proteins still needs to be deciphered. It is of great interest to crosscheck between different organisms if, depending on their Ni-enzyme maturation pathways, they possess HRC proteins. A common feature among the overall class of nickel enzymes seems to be the need of both a nucleotide-binding and a histidine rich domain, in either the same or different proteins, in order to correctly insert the nickel ion in their active centers.

| Protein     | Organism            | Molecular mass (kDa) | His per binding domain | Nickel bound per dimer | $K_d$                     |
|-------------|---------------------|----------------------|------------------------|------------------------|---------------------------|
| <b>CooJ</b> | <i>R. rubrum</i>    | 12.5                 | 16 C-term              | 4                      | 4.3 $\mu$ M               |
| <b>UreE</b> | <i>K. aerogenes</i> | 17                   | 10 N-term<br>2 C-term  | 5-6<br>1-2             | 0.5 $\mu$ M<br>85 $\mu$ M |
| <b>HypB</b> | <i>B. japonicum</i> | 32                   | 24                     | 8-9 per monomer        | 2.3 $\mu$ M               |

Table 2 Comparison between CooJ<sup>125</sup>, UreE<sup>126</sup> and HypB<sup>99</sup>.

## CooT

CooT is a 7.1 kDa protein characterized, in the group prior to my arrival, as a stable homodimer in solution able to bind one nickel per dimer with a dissociation constant of  $9 \pm 4$  nM measured via isothermal titration calorimetry.<sup>127</sup> From its sequence, two amino acids were identified as potential nickel coordinating residue, Cys2 and His41 (Figure 34). Site-directed mutagenesis proved that the Cys2 is essential for nickel binding, considering that a mutation of this amino acid to serine abrogates the nickel binding. In addition, as observed in the structure, dimerization brings the two cysteines in close proximity, while the histidines 41 are too far apart to interact with the cysteine to co-coordinate the metal.<sup>127</sup>

The crystal structure of the apo-protein showed that each monomer is composed of seven strands forming two bent anti-parallel  $\beta$ -sheets ( $\beta$ -sheet I:  $\beta_6, \beta_7, \beta$  and  $\beta_2$ ,  $\beta$ -sheet II:  $\beta_3, \beta_4$  and  $\beta_5$ ), tightly packed on top of each other. According to the CATH classification, CooT belongs to the “mainly  $\beta$ ” class in the “not yet named” superfamily (CATH code: 2.30.30.100). The two monomeric chains form an anti-parallel homodimer, held together by hydrogen bonding and hydrophobic interactions of the side chains. All CooT possess an RNA-binding domain, predicting a role of this protein in RNA metabolism. Moreover, it was shown by electrophoretic mobility shift assay experiments that CooT does not bind RNA.

### CooT *Rhodospirillum rubrum* (PDB: 5N76)

MCM<sup>1</sup>AKVVLTK ADGGRVEIGD VLEVRAEGGA VRVTTLFDEE HAF<sup>2</sup>PGLAIGR VDLRSGVISL IEEQNR

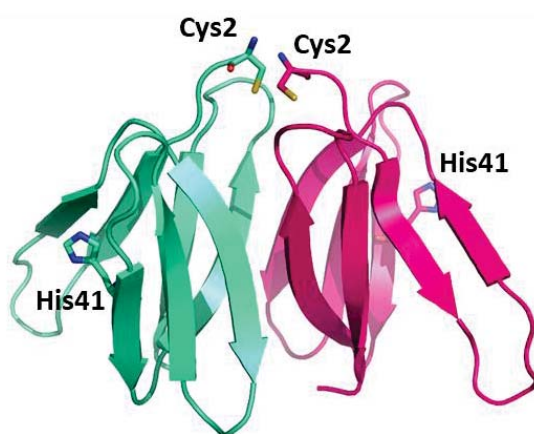


Figure 34. *RrCooT* sequence, the methionine 1 is highlighted in blue, the cysteine 2 in red, the histidine 41 in green. The crystal structure of the apo dimer shows the relative spatial position of His41 and Cys2 in the dimer conformation, while Met1 is not present in the crystal structure.

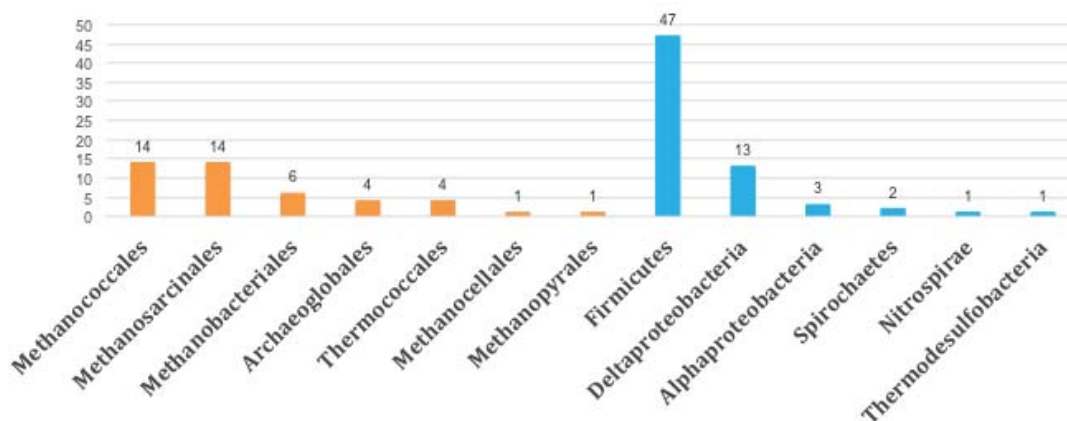


Figure 35 Taxonomic distribution of the 111 CooT homologues. Adapted from<sup>127</sup>.

A phylogenetic analysis performed in 2017 showed the existence of 111 CooT homologues in anaerobic bacteria and archaea. The sequences present a large variability, except for the strictly conserved cysteine in the second position. In 83 proteomes CooT co-occurs with CooS and in the remaining 28 proteomes, CooT co-occurs with CooC and ACDS or with CooC and ACS/CODH in *Thermodesulfovibrio yellowstonii* and *Butyrivibrio proteoclasticus* (Figure 35).



## Proposed maturation process

In *R. rubrum*, two different models were proposed for the role of CooC during nickel insertion in the CODH active site. They are both based on the results obtained from the whole cell extract induced CODH activation upon nickel  $\pm$  ATP exposure and on nickel CODH re-activation with purified *RrCooC* and *RrCooJ*, *in vitro*.<sup>118</sup> In the first model, *RrCooC* works like a nickel insertase to transfer the nickel from *RrCooJ* to CODH using the energy released by the ATP hydrolysis. In this scenario, *RrCooJ* binds the free nickel in the cell but it cannot donate it directly to the CODH and its interaction with *RrCooC* would promote the nickel release to the enzyme.<sup>118</sup> In the second model, *RrCooC* would use the ATP hydrolysis to control the conformation around nickel center and *RrCooJ* would be in charge of nickel delivery to the vacant *RrCODH* site.<sup>118</sup> The role of *RrCooT* in the mechanism was not considered, due to the inability to purify this protein from the organism, and it still needs to be clarified. Regardless, all three chaperones are needed to fully activate the enzyme in *R. rubrum*. The ability of CODH to be activated by an excess of nickel salt, but with a lower activity compared to the wild type protein, is a possible indication that the site is not in the correct conformation in the absence of the Coo-accessory proteins.<sup>118</sup>

In most of CODH-possessing micro-organisms, CooJ and CooT are absent (or have not yet been identified). In some micro-organisms, like in the case of *C. hydrogenoformans*, the different CODH operons do not have any accessory proteins at all (*see chapter 1.3*). This could indicate that either their maturation occurs without the support of any accessory protein or that the accessory proteins associated with different CODH can interplay between each other.

In addition, as we saw previously in the case of *H. pylori*, accessory proteins supposed to be related to a specific maturation system can actually inter-cross and play a role in the maturation process of different Ni-enzymes.

## General remarks

Overall, 2% of the total prokaryotic genomes encodes a least one Ni-CODH. Recently a new structural and phylogenetic analysis on Ni-CODHs was performed, which divides them into seven clades (from A to E Figure 36) that were further classified into 24 structural groups depending on their sequences and structure predictions (C-cluster motifs and additional common features).<sup>54</sup> Several newly identified CODHs harbor a His-rich extension (clade B), possibly playing the role of CooJ in nickel metabolism. In order to identify the structural and functional relationships between the different clades, the authors searched for similarities among the Ni-CODHs, including the associated chaperone proteins. Of interest from our perspective, it was possible to highlight that the presence of metallochaperone proteins (including CooC and CooT) is one of the common features for both methanogens and acetogens, which use the CODH in order to fix the carbon via Wood-Ljungdahl pathway (WLP), and hydrogenogenic carboxydrotrophs that couple the CODH to an energy converting hydrogenase (ECH). Another interesting point is the presence, for some CODH classes, of an additional N-terminal Cys motif (Figure 36 novel structural motifs column). This motif could play a role in the interaction with the accessory proteins or have a metallochaperone role by itself. Among those, we found the sequence of *R. rubrum* CooS. The role of these additional cysteines could be an interesting feature to investigate.

To end this session with an interesting point for the next one, the authors investigated how the structural features of different Ni-CODHs are related to their biological functions in order to create a tool to predict the role of each Ni-CODH. Remarkably, only the Ni-CODHs present in the structural groups of *C. hydrogenoformans* CooSI (clade E-1, F-1) and *R. rubrum* CooS (clade F-2) are close to the [NiFe]-hydrogenase catalytic subunit. This discovery implies that a way to identify Ni-CODH/ECH coupled enzymes that are probably involved in the WGS reaction has been found.

| Structural group | D-cluster      | C-cluster |       |      |      |      |      | Acid-base |     | E- & F-clusters | Novel structural motifs   |
|------------------|----------------|-----------|-------|------|------|------|------|-----------|-----|-----------------|---|
|                  |                | His       | Cys 1 | Cys2 | Cys3 | Cys4 | Cys5 | His       | Lys |                 |   |
| A-1              | type I         | H         | C     | C    | C    | C    | C    | H         | K   | +               | → <i>M. Barkeri</i> ACS α subunit                                 |
| A-2              | –              | H         | C     | C    | C    | C    | C    | H         | K   | +               |   |
| A-3              | –              | H         | C     | C    | C    | C    | C    | H         | K   | –               |   |
| A-3              | –              | H         | C     | C    | C    | C    | C    | H         | K   | –               |   |
| A-4              | –              | D         | P     | C    | C    | C    | C    | Q         | R   | +               |   |
| A-5              | –              | H         | C     | C    | C    | C    | C    | H         | K   | +               | Two Cys × 4 motifs between E- and F-clusters                      |
| A-6              | –              | N         | S     | C    | C    | C    | C    | H         | L   | + <sup>b</sup>  |   |
| B-1              | type II        | H         | C     | D/E  | C    | C    | C    | H         | K   | –               | His-rich region at the N-terminus                                 |
| B-2              | type II        | H         | C     | D    | C    | C    | C    | H         | K   | –               |   |
| B-3              | type II        | H         | C     | D    | C    | C    | C    | H         | K   | –               | His-rich region at the C-terminus                                 |
| B-4              | type II        | H         | C     | C    | C    | C    | C    | H         | K   | –               |   |
| C-1              | type II        | H         | C     | C    | C    | C    | C    | Y         | Q   | –               | A Cys × 2 to 4 motif at the N-terminus                            |
| C-2              | type II        | H         | C     | C    | C    | C    | C    | Y         | Q   | –               |   |
| C-3              | – <sup>a</sup> | H         | C     | C    | C    | C    | C    | Y         | Q   | –               |   |
| C-4              | type II        | H         | C     | C    | C    | C    | C    | Y         | Q   | –               | A Cys × 4 motif between the Tyr and His residues                  |
| D-1              | type III       | H         | E     | C    | C    | C    | C    | Y         | K   | –               |   |
| D-2              | type III       | H         | C     | C    | C    | C    | C    | N         | K   | –               |   |
| D-3              | type II        | H         | C     | C    | C    | C    | C    | Y         | K   | –               | A Cys × 3 motif at the N-terminus                                 |
| D-4              | type II        | H         | C     | C    | C    | C    | C    | Y         | K   | –               |   |
| E-1              | type II        | H         | C     | C    | C    | C    | C    | H         | K   | –               | → <i>T. Onnuriensis</i> CooS                                      |
| E-2              | type III       | H         | C     | C    | C    | C    | C    | H         | K   | –               | → <i>D. Vulgaris</i> CooS   |
| E-3              | type II        | H         | C     | C    | C    | C    | C    | H         | K   | –               | A Cys × 2 to 3 motif at the N-terminus                            |
| F-1              | type II        | H         | C     | C    | C    | C    | C    | H         | K   | –               | → <i>C. Hydrogenoformans</i> CooS                                 |
| F-2              | type II        | H         | C     | C    | C    | C    | C    | H         | K   | –               | A Cys × 2 to 3 motif at the N-terminus<br>→ <i>R. Rubrum</i> CooS |
| G-1              | type III       | H         | T     | C    | C    | C    | C    | Y         | K   | –               |   |

Figure 36. Structural classification of Ni-CODHs. The seven different clades (from A to G) (*clade is defined as a group of organisms with a common ancestor and all its lineal descendants*) are divided into 24 structural groups considering the predicted structural signatures from their sequences. The new motifs identified are reported in the column “novel structural motifs”. Adapted from <sup>54</sup>.



### 1.3 Heterologous production of CODH in *E. coli*

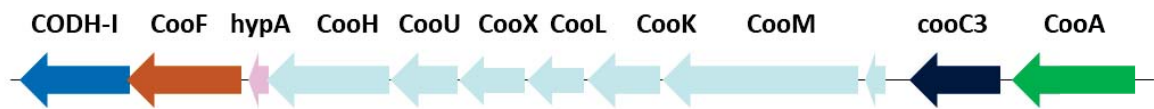
The possibility of heterologous production in the host bacterium *E. coli* has opened a new way to study proteins.<sup>128</sup> Especially for the proteins coming from difficult to grow microorganisms with a final low yield, the capability to overproduce them in large quantities introduces an alternative possibility for their detailed characterization. The knowledge acquired so far about *E. coli* has enabled researchers to master the molecular tools needed to perform, for example, site-directed mutagenesis in a fast and reproducible way.<sup>129</sup> In addition, the opportunity to select different expression plasmids, engineered strains, and cultivation strategies provided the benefits to choosing this “protein factory”.<sup>130</sup>

For all these reasons, the possibility to produce and purify a recombinant protein is one of the most coveted scenario for a biochemist, especially if she/he is working with strictly anaerobic bacteria. Moreover, when the focus is switched to the possible industrial applications, it is essential to have a stable and reproducible expression system for large amounts of protein. Nevertheless, it is important to consider that, in particular for metalloproteins, the establishment of a productive over-expression system has proven to be particularly challenging due to the complexity of the maturation processes required to build and insert their active centers.<sup>131</sup>

As you should have guessed at this point, working with CODH from the native organisms is not an easy nor straightforward task. For this reason, most of the groups handling this enzyme are trying, and some already succeeded, in correctly expressing, producing and purify it from *E. coli*. Hence, I will present the cases where the heterologous expression resulted in active CODH from different source organisms.

Starting from *C. hydrogenoformans*, which possesses 5 different CODH as described previously, a notable success has been achieved in over-producing and purifying RecChCODH-I, II, IV and V (Figure 39).

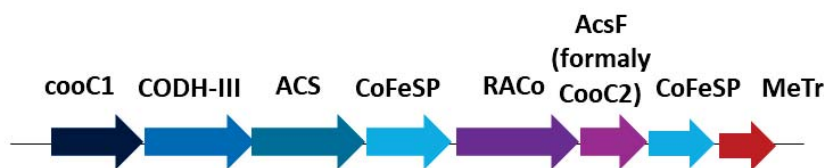
***Carboxydothemus hydrogenoformans* CODH-I**



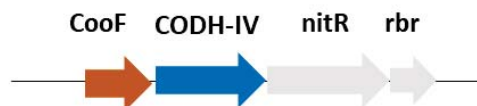
***Carboxydothemus hydrogenoformans* CODH-II operon**



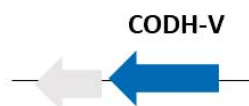
***Carboxydothemus hydrogenoformans* CODH-III operon**



***Carboxydothemus hydrogenoformans* CODH-IV operon**



***Carboxydothemus hydrogenoformans* CODH-V operon**



***Desulfovibrio vulgaris* CODH operon**



***Rhodospirillum rubrum* CODH operon**



Figure 37. Operons encoding CODH from *Carboxydothemus hydrogenoformans*(Ch),*Desulfovibrio vulgaris*(Dv) and *Rhodospirillum rubrum* (Rr). In light blue are depicted the genes encoding for the hydrogenase subunits, in light grey are the genes encoding for unknown proteins.

Listing from publication year, an efficient protocol for the recombinant production of His-tagged RecChCODH-II was published in 2007 by Jeoung and Dobbek. They purified it from anaerobic and nickel supplemented *E. coli* cultures, obtaining a pure and fully active enzyme in one purification step (nearly equal to the native one) (Table 3).<sup>132</sup> In 2012, Inoue et al. published the protocol for the production of RecChCODH-II starting from aerobic conditions prior to switching to anaerobic conditions before IPTG induction, with nickel always present in the culture medium.

Using the same protocol, this group published in 2014 the purification of RecChCODH-I, which displayed only 14% of activity compared to the native one. Co-production with its accessory protein CooC3 resulted in a partial activity recovery, corresponding to 48% of the native one (Table 3).<sup>133</sup> Interestingly both RecChCODH-I and RecChCODH-I-CooC3 samples are not fully nickel-reconstituted, presenting 0.1 and 0.33 mol. per monomer, respectively. These values are low compared to the native one, which has 0.71 mol. per monomer. This could explain the decreased activity of these recombinant CODHs. According to the authors, “As CODH activity levels coincided with the concentration of  $Ni^{2+}$  in the recombinant protein, lower activity of Co-Rec-CODH-I is attributed to the lower concentration of  $Ni^{2+}$  compared to native-CODH-I”.<sup>133</sup> To recover the activity of RecChCODH-I, nickel reconstitution *in vitro* was tried without any success suggesting that co-expression of CooC3 is necessary to integrate the nickel in the enzyme. However, the metal *in vitro* insertion experiment was done without first reducing the cluster. Another possibility is that additional players, others than CooC3, are necessary for the nickel activation and they need to be co-expressed in order to match the native CODH’s activity (Table 3).

| Name             | CooC in the operon | CO oxidation activity (U/mg)                 |          |                |                |                          |                         |                             |                                   |
|------------------|--------------------|--|----------|----------------|----------------|--------------------------|-------------------------|-----------------------------|-----------------------------------|
|                  |                    | Nickel amount after purification (per dimer) |          |                | Native (70 °C) | Recombinant              |                         |                             |                                   |
|                  |                    | Native CODH                                  | Rec CODH | Rec CODH+ CooC |                | CODH As-prepared (70 °C) | CODH + CooC As-prepared | CODH 2 mM nickel incubation | CODH+ CooC 2 mM nickel incubation |
| <b>ChCODH-I</b>  | CooC3              | 0.71   | 0.1      | 0.3            | 16.800         | 2.270                    | 8.060                   | 1.540                       | ×                                 |
| <b>ChCODH-II</b> | ×                  | 1  | NR       | ×              | 13.828         | 12.229                   | ×                       | ×                           | ×                                 |

Table 3. Catalytic properties of CO oxidation for NatChCODH-I<sup>133</sup> and NatChCODH-II<sup>56</sup>. The activities are presented for the native purified enzyme vs the recombinant ones activated with nickel or nickel+CooC.

| Name                | Presence of<br>CooC in the<br>operon | Catalytic properties      |                      |                                     |
|---------------------|--------------------------------------|---------------------------|----------------------|-------------------------------------|
|                     |                                      | $k_{cat}$<br>( $s^{-1}$ ) | $K_M$<br>( $\mu M$ ) | $k_{cat}/K_M$<br>( $M^{-1}s^{-1}$ ) |
| <b>ChCODH-II</b>    | <b>x</b>                             | 31000 (70 °C)             | 18 (70 °C)           | $1.7 \times 10^9$ (70 °C)           |
| <b>RecChCODH-II</b> | <b>x</b>                             | 1021 (25 °C)              | 8 (25 °C)            | $2.09 \times 10^8$ (25-60 °C)       |
| <b>RecChCODH-IV</b> | <b>x</b>                             | 14 (25 °C)                | 0.047 (25 °C)        | $3 \times 10^8$ - $10^9$ (25-60 °C) |

Table 4. Catalytic properties of CO oxidation for ChCODH-II<sup>56</sup>, RecChCODH-II<sup>57</sup> and RecChCODH-IV<sup>57</sup>. The  $K_M$  value of ChCODH-II corresponds to the  $K_M$  value of the RecChCODH-II. ( $k_{cat}$  turnover number,  $K_M$  apparent binding constants of CO,  $k_{cat}/K_M$  catalytic efficiency).

From this study, the authors suggested that when CooC is found in the same operon as the CODH, its co-expression is needed for the full nickel insertion into the recombinant CODH.<sup>133</sup> In support to this theory, RecChCODH-II does not have any dedicated CooC and it is indeed expressed fully active (Figure 37).

The purification of RecChCODH-IV was published in 2017<sup>57</sup>, resulting in an active enzyme able to maintain a partial activity even after air exposure. It is worth mentioning that this enzyme has never been purified native.

In table 4, the comparison of the catalytic properties of ChCODH-II (Nat/Rec) and RecChCODH-IV is reported. Considering that the expected best catalytic efficiency ( $k_{cat}/K_M$ ) for the enzyme is around  $\sim 10^8$ - $10^9 M^{-1}s^{-1}$ , ChCODH-II displays the highest possible catalytic efficiency with a  $k_{cat}/K_M = 1.7 \times 10^9 M^{-1}s^{-1}$ .<sup>134</sup> The  $k_{cat}/K_M$  of RecChCODH-II is constant in the temperature range 25-60 °C, whereas RecChCODH-IV matches the temperature dependence of CO's diffusion coefficient in water.<sup>135</sup> This means that any change in the  $k_{cat}$  is compensated by a change in the  $K_M$ , making the RecChCODH-IV working constantly at the diffusion limit of CO. In addition, RecChCODH-IV is more O<sub>2</sub>-tolerant than RecChCODH-II.<sup>57</sup>

Even if not published in a journal, RecChCODH-V was purified and structurally investigated and the results are presented in the PhD thesis of J. M. Fessler from Dobbek group in 2015.



(Biochemische und strukturelle Untersuchungen der Kohlenmonoxid-Dehydrogenasen CODH-II und CODH-V aus *Carboxydotherrmus hydrogenoformans*. 17.06.2015)

From this study, the crystal structure of *Ch*CODH-V is closely related to the one from *Ch*CODH-II, presenting the same fold, but with clear deviation in the metal cluster composition, which is a [4Fe3S2O] cluster (hybrid cluster) instead of a C-cluster in the as-isolated state. In addition, the enzyme does not show neither CO oxidation nor CO<sub>2</sub> reduction activity; although it has hydroxylamine reduction activity. This means that CODH-V is more related to the hybrid cluster proteins (HCP)<sup>58</sup>, for both catalytic capabilities and active site conformation, than to the CODHs family. The author of the manuscript proposes that *Ch*CODH-V could be the connecting link between the HCPs and the CODHs family.

In the case of *Ch*CODH-III, the bi-functional enzyme in complex with ACS, the complex has been purified from the native *C. hydrogenoformans*<sup>136</sup>, while only the ACS subunit has been heterogeneously produced in *E. coli*.

Another example concerning mono-functional CODHs comes from the heterologous production of CODH from *Desulfovibrio vulgaris* (*Dv*CHODH) in *Desulfovibrio fructosovorans*.<sup>137</sup> For this CODH, its operon harbors a *cooA* and a *cooC* genes. (Figure 38) When the Rec*Dv*CODH is produced in the absence of its nickel chaperone protein *CooC*, it results in an almost inactive form. The activity is recovered by co-expression with *CooC*, indicating that this chaperone is needed for inserting the nickel. Considering that this enzyme has a good homology with *Ch*CODH-I, it could make sense that co-expression with *CooC* is required for a better enzyme activation, as shown from Sako's group for the production of Rec*Ch*CODH-I.<sup>133</sup>

| Name           | Nickel amount after Rec purification (per dimer) |                    | Native | CO oxidation activity (U/mg) |                      |  |  |
|----------------|--|--------------------|--------|------------------------------|----------------------|--|--|
|                |  |                    |        | Recombinant                  |                      |  |  |
|                | CODH   | CODH + <i>CooC</i> |        | CODH*                        | CODH + <i>CooC</i> * | CODH <i>in vitro</i> 2.4 mM nickel (after 1.2 mM NaDT treat) | CODH+ <i>CooC in vitro</i> 2.4 mM nickel (after 1.2 mM NaDT treat) |
| <i>Dv</i> CODH | 0-0.5  | 0.8-1.8            | ×      | < 5                          | 160 ± 40             | 4–60   | 1660   |

\* As-prepared

Table 5. Reported values obtained from the recombinant production of *Dvc*CODH in *Desulfovibrio fructosovorans*.

The RecDvCODH co-produced with CooC can be further activated via incubation with  $\text{NiCl}_2$  and the reducing agent sodium dithionite (NaDTH) resulting in a total activity of  $1660 \pm 340$  units $\cdot\text{mg}^{-1}$ , against the  $160 \pm 40$  units $\cdot\text{mg}^{-1}$  of the expressed CODH without CooC (Table 5). These results are, to me, in agreement with what has been found for RecChCODH-I, where the *in vitro* nickel insertion on the protein did not improve the enzyme activity.<sup>133</sup> However, the experimental conditions used during the *in vitro* nickel insertion were not the same, considering that in the case of RecChCODH-I the samples were not NaDTH reduced to promote the nickel insertion.<sup>133</sup>

The hypothesis that DvCooC could also play an important role in the correct folding of the  $[\text{3Fe-4S-Fe}_u]$  active site has been recently excluded from the latest crystallographic studies on RecDvCODH.<sup>116,138</sup> From the latest study, CooC is most likely involved in nickel insertion and it is not required for the C-cluster formation.<sup>138</sup> Moreover, the nickel insertion is impeded in absence of the D-cluster, suggesting for the latest a redox role rather than a structural role.<sup>138</sup>

## 1.4 Our study model: *Rhodospirillum rubrum*

Commonly known as a toxic molecule, carbon monoxide (CO) plays a crucial role in many microorganism's metabolism<sup>139</sup> and it can be used as a source of energy or carbon by some CO-utilizing microbes (carboxydrotrophs). CO is converted to CO<sub>2</sub>, with the production of a pair of reducing equivalents. As already shown in the previous section, the enzyme responsible for this conversion is CODH and, depending on the aerobic or anaerobic nature of the organism, it presents different structures and metal cofactors.

Focusing on the anaerobic carboxydrotrophs, they couple the electrons derived by the CO/CO<sub>2</sub> oxidation, which has a low redox potential of -558 mV, to reduce various molecules. Some example reactions that exploit these high-energy electrons are: sulfate to sulfide (sulfate-reducer microorganisms), CO<sub>2</sub> to acetate (acetogens) or methane (methanogens) and water to hydrogen (hydrogenogens).<sup>48</sup> In the latter case, the produced electrons from the CO oxidation, catalyzed by the [NiFe]-CODH, are transferred to a [NiFe]-hydrogenase, which uses them to reduce protons and produce hydrogen. The electron transfer between the two enzymes is mediated by a ferredoxin, CooF. The global reaction is known as the water gas shift (WGS) reaction that leads to the production of H<sub>2</sub> and CO<sub>2</sub> from CO and water, as follows:



For the reported reaction, monofunctional CODHs are involved and their expression is regulated by the presence of environmental CO, detected by the transcriptional regulator CooA.<sup>140</sup>

It is among the anaerobic hydrogenogenic carboxydrotrophs that we found our study model, *Rhodospirillum rubrum*.

*R. rubrum* is a purple non-sulfur photosynthetic bacterium. It is a facultative anaerobe, able to obtain energy through a variety of mechanisms with an extreme metabolic versatility.<sup>141</sup> In fact, *R. rubrum* is capable of aerobic or anaerobic respiration, anoxygenic photosynthesis and acid mixed fermentation, depending on the growth conditions. In anaerobic growth, if light is available, it is capable of photoautotrophic growth, using the produced H<sub>2</sub> as an electron donor while carbon fixation occurs through CO<sub>2</sub> reduction using the Calvin-Benson-Bassham cycle (not via acetyl-CoA synthesis). Alternatively, in photoheterotrophic growth it can use carbon compounds as both carbon source and electron donor.<sup>142</sup> During anaerobic growth in the dark or light, in the presence of CO, it can grow using carbon monoxide as sole source of energy. Its ability to switch so quickly depending on the external conditions, constantly maintaining its redox balance, has been of great scientific interest and is one of the reasons why this organism has been deeply investigated.

In the presence of CO,<sup>143</sup> the CooA regulatory protein induces the expression of the CO oxidation/proton reducing systems. Eleven genes, named *coo* due to their role in CO oxidation, are organized in two transcript-regulated groups. Six different genes encode for the hydrogenase subunit and related partner proteins (*cooMKLXUH*), while the remaining five genes encode for the CODH subunit (CooS) and its accessory proteins for the nickel storage-insertion (CooCTJ) and a ferredoxin (CooF).

Other hydrogenases are also present and they are involved in the CO<sub>2</sub>/H<sub>2</sub> fermentative growth on substrates like pyruvate or coupled to nitrogenase to recycle the hydrogen during the nitrogen-fixing conditions.

Due to this high metabolic versatility, *R. rubrum* has different potential uses in biotechnology, such as the in cell accumulation of poly-hydroxy-butric-acid (PHB)<sup>144,145</sup>, a precursor for bioplastic production, or as a model system to study the conversion from light energy to chemical energy. In particular, in our laboratory we are interested in another possibility that directly involves the use of CODH, the biologically-mediated WGS reaction (Figure 38).

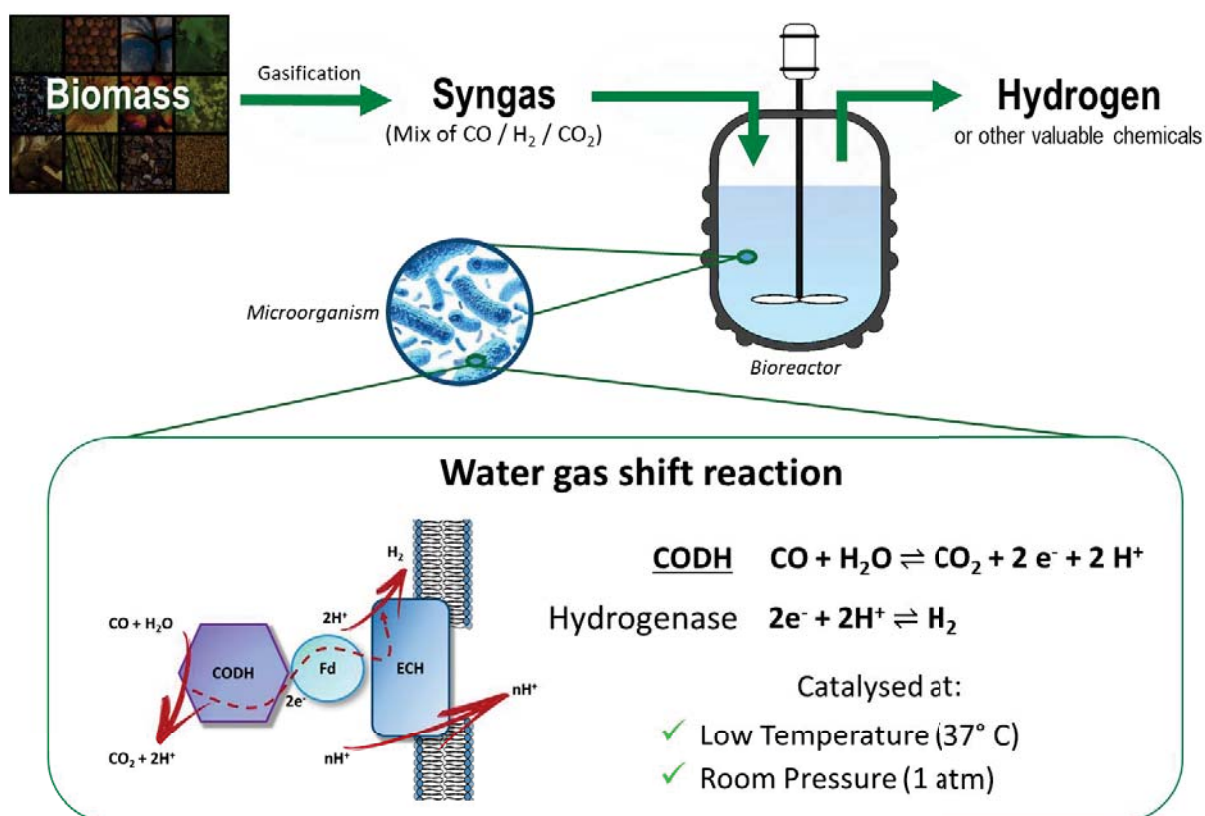


Figure 38. Biological WGS reaction. Schematic representation of what we would like to achieve in our laboratory.

The WGS reaction, discovered in 1780 by F. Fontana as reversible and exothermic, can be described as:



The starting feedstock of the WGS reaction is the synthesis gas, called syngas, which is a mixture of gasses mainly rich in CO and H<sub>2</sub>. The WGS reaction is used to balance the CO/H<sub>2</sub> ratio via a H<sub>2</sub>–enrichment process. Nowadays, industry commonly uses this process for either direct hydrogen production or during the Fischer–Tropsch process to balance the CO/H<sub>2</sub> ratio to further produce liquid hydrocarbons and/or commodity products.

Among the possible starting materials to produce syngas, an increasingly popular source is biomass (Figure 38).<sup>146</sup> This renewable source can be easily gasified, leading to an eco-friendly and renewable syngas, which can be further upgraded via the WGS reaction.

Currently, industries adopt a two stage catalytic process to promote this reaction: a high temperature step (400-500 °C), using iron oxide stabilized in a chromium oxide-based catalyst, followed by a low temperature step (200-220 °C), catalyzed by copper-zinc oxide supported on alumina.<sup>147</sup> As you can imagine, this process is not eco-friendly, hindering the concept of starting from a renewable feedstock as biomass. A possible alternative to these inorganic catalysts and high temperature industrial processes is the use of bacteria able to perform the WGS reaction at room temperature and pressure.

It is in this context that *R. rubrum* is involved. As reported above, this organism is capable of anaerobic growth using the WGS reaction. Due to its highly variable metabolism, it can also tolerate small amounts of oxygen (possibly present in the starting syngas), making it a particularly attractive strain for industrial applications. The various studies performed on *R. rubrum* using syngas as a substrate, showed that a mild-light exposure can increase the growth rate and the hydrogen production at the same time.<sup>124,126</sup> In addition to this, it was reported that an additional carbon source, other than CO, helps the efficient conversion of the latter molecule. The maximum CO conversion was obtained using acetate.<sup>151</sup>

To establish that this organism is capable of directly convert syngas using the WGS reaction, as a proof of concept, the CO conversion of *R. rubrum* grown with acetate and syngas as substrates was evaluated in a stirred bioreactor.<sup>152,153</sup> From this study, the global conversion yield obtained was  $87 \pm 2.4 \%$ , confirming the feasibility of the biological WGS reaction on a small scale. Considering this encouraging result, the industrial and economical workability of these outcomes was evaluated in a recent paper, where the eco-exergy concept, centered on using the maximum useful work possible during a process in a closed environment, seems to be the most suitable tool for large-scale biohydrogen production using living organisms.<sup>154</sup> Even if at the moment the biological WGS reaction cannot be economically competitive, the environmental feedback could be a sufficient incentive for pushing forward this technology.<sup>154</sup> (referee to the review in annex for more information)

With all these considerations, the biological WGS reaction offers a concrete alternative to conventional chemical metal-based industrial processes. First, the use of room temperature and pressure catalysis could reduce the industrial costs, re-orienting it towards a green economy philosophy. In addition, the inorganic materials can be irreversibly poisoned or fouled, being dependent on the purity of the starting syngas. On the other hand, living

organisms can easily re-generate themselves, resulting in a longer life span. Despite these advantages, many improvements need to be achieved in order to make the biological WGS reaction industrially competitive. Different areas can be addressed to improve this issue, such as the bioreactor layout and utilization or the design/selection of new CO-resistant strains better suited for fast growth and maximal syngas conversion. An interesting alternative could be represented by the recent discovery of the possible heterologous production of CODH in the easy-to-grow bacterium *E. coli*. *E. coli* is a robust host microorganism, able to grow quickly and adapt to different environments (aerobic/anaerobic conditions using respiration or fermentation) and it is widely used in modern biotechnology.<sup>130</sup> In addition, more information are available for this strain, making the *E. coli* metabolic and protein engineering a more realistic prospective compared to the one carboxydothrophic bacteria.<sup>155</sup>

The biologically promoted WGS reaction is a reality but it is still limited to the laboratory scale. Nonetheless, the preliminary results and the new possibility opened by recent discoveries makes it a hopeful option worthy of consideration. Taken together, the combination of inorganic chemistry, biochemistry, microbiology and synthetic biology could be of great impact in devising new possibilities to make the biological WGS reaction an industrial reality.





## Statement of purpose

In this manuscript, the results are presented in three main chapters according to the nature of the studied topic.

Chapter 2 is devoted to the chaperone protein CooT. The results are presented in two papers. The first one, describing the nickel coordination site of *R. rubrum* CooT, is entitled "*A solvent-exposed cysteine forms a peculiar Ni(II)-binding site in the metallochaperone CooT from Rhodospirillum rubrum*". For the results presented in this part I would like to acknowledge Giulia Veronesi (LCBM, Grenoble) for her implication in the XAS collections and analysis; Francesco Musiani (University of Bologna) for the computational calculations and models; Stefano Ciurli (University of Bologna) for the NMR data collection and analysis; Barbara Zambelli (University of Bologna) for the ITC data collection and analysis; Luca Signor (IBS, Grenoble) for the mass spectroscopy. I would also like to thank all of them in general for the always useful discussions about all the possible CooT coordination sites and the time they all spent to explain to me the principles of the different techniques and to answer all my questions about data analysis and interpretation.

The second one concerns the characterization of the putative *C. hydrogenoformans* CooT and it is entitled "*Biophysical and structural characterization of the putative nickel chaperone CooT from Carboxydotherrmus hydrogenoformans*". For this part, I would like to acknowledge Patrice Catty (LCBM, Grenoble) and Roger Miras (LCBM, Grenoble) for their help in the estimation of the ChCooT nickel dissociation constant via equilibrium dialysis and CD spectroscopy and Julien Perard for the site-direct mutagenesis.

Chapter 3 is dedicated to the chaperoen protein RrCooJ. The first paper presented the characterization of RrCooJ and its mutants, entitled "*The CO-dehydrogenase accessory protein CooJ is a histidine-rich multidomain dimer with an unexpected additional Ni(II)-binding site*". For making this paper possible, I would like to acknowledge Philippe Carpentier (LCBM, Grenoble) for his help during X-ray data collection and structure determination; Julien Perard (LCBM, Grenoble) for the SAXS data collections and interpretations and site-direct mutagenesis; Christian Basset (LCBM, Grenoble) for his help with in the proteins purifications;

Serge Crouzy (LCBM, Grenoble) for the molecular dynamics simulations; Barbara Zambelli (University of Bologna) for the ITC data collection and analysis.

The second paper, regarding the CooJ oligomerization induced by nickel addition, is entitled “*Nickel-induced oligomerization of the histidine-rich metallochaperone CooJ from Rhodospirillum rubrum*”. For this paper, I would like to acknowledge Julien Perard (LCBM, Grenoble) for the TEM samples preparation and images interpretation.

Chapter 4 concerns the results obtained during the last months of my PhD. The protocols for the purification of recombinant *RrCODH* in *E. coli* are presented, as well as the preliminary *in vitro* nickel activation experiments. All the data presented in this chapter have to be considered as preliminary. Umberto Contaldo, started his PhD in 2018 pursuing the purification of Rec*RrCODH*, working on the project “In vitro biological water-gas shift reaction: Bio-functionalization of carbon nanotubes with carbon monoxide dehydrogenase (CODH) and [NiFe]-hydrogenase”.



# Chapter 2



## CHAPTER 2

### CooT

#### INTRODUCTION

As a quick re-cap from the introduction chapter, *RrCooT* is a small 7.1 kDa protein thought to play a role, with *RrCooJ* and *RrCooC*, in nickel delivery to the active site of *RrCODH*. CooT is present in solution as a homodimer, able to bind one nickel per dimer, with a high affinity ( $K_{D1} = 9 \pm 4$  nM).<sup>127</sup> From its sequence, two amino acids stand out as potential nickel ligands: the cysteine in second position and the histidine in position 41 (Figure 1).

Interestingly, even though not experimentally described in the literature, phylogenetic analysis shown the existence of 91 non-redundant homologues in the proteome databases, all sharing a strictly conserved cysteine 2.<sup>127</sup> The occurrence of the histidine is, however, more enigmatic and is the crucial element to identify the nature of the metal coordination.

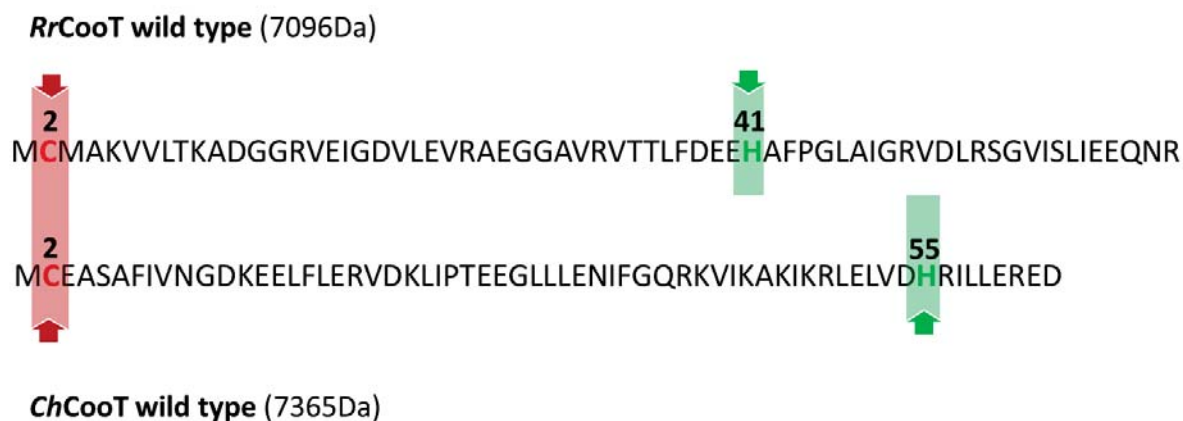


Figure 1. Comparison between the CooT sequences of *R. rubrum* and *C. hydrogenoformans*. These are the representatives of the two possible nickel-binding motifs of the CooT family. The “Cys2” motif, represented by *RrCooT*, and the “Cys2-His55” motif, represented by the *C. hydrogenoformans*. The strictly conserved cysteine 2 is highlighted in red, the histidine, variable or not present in the case of the “Cys2” motif CooT sequences, is highlighted in green.

Two distinct nickel-binding motifs can in fact occur in the CooT family. The **“Cys2-His55” motif** was identified in the putative *C. hydrogenoformans* (CHY\_RS00800) *ChCooT* by X-ray crystallography (see next *ChCooT* paper), where a conserved histidine in position 55 coordinates the Ni in addition to the cysteine 2 (Figure 1). This case counts for ~70% of the CooT sequences, mainly bacteria and 16 archaea.

In the **“Cys2” motif** the histidine is either not present or is in a random position, leaving the cysteine 2 as principal binding residue. This kind of binding motif matches the remaining ~30 % of the identified sequences, to be exact 7 bacterial and 19 archaeal ones.

In the next pages, these two binding motifs are presented, and these studies led to the publication of the papers *“A solvent-exposed cysteine forms a peculiar Ni(II)-binding site in the metallochaperone CooT from Rhodospirillum rubrum”* and *“Biophysical and structural characterization of the putative nickel chaperone CooT from Carboxydothemus hydrogenoformans”*. More emphasis is devoted to the description of our study model, the *R. rubrum* CooT protein.

## 2.1 CooT from *Rhodospirillum rubrum*

Prior to my arrival, the crystal structure of apo-CooT was solved in the laboratory.

Contrarily to what was expected, histidine41 is not close to cysteine2, considered to be essential for the metal coordination as shown by the hindered nickel-binding ability of the Cys2Ser mutant (Figure 2A).<sup>127</sup> As we would assume from the previously described binding motifs, *R. rubrum* CooT falls in the “Cys2 only” coordination.

From these considerations, the hypothesis we formulated is that the nickel coordination of *RrCooT* should be held by its N-terminal motif, certainly involving the thiolate groups of the two Cysteines 2. However, the coordination cannot be maintained by only two ligands and at least two additional ones are required to complete the coordination of Ni(II) in a square-planar or tetrahedral geometry. From the crystal structure, the best candidates in the proximity of Cys2 to fulfil this role are either the terminal amines of the first amino acid (Met1) or the amide groups from Cys2 (Figure 2B).

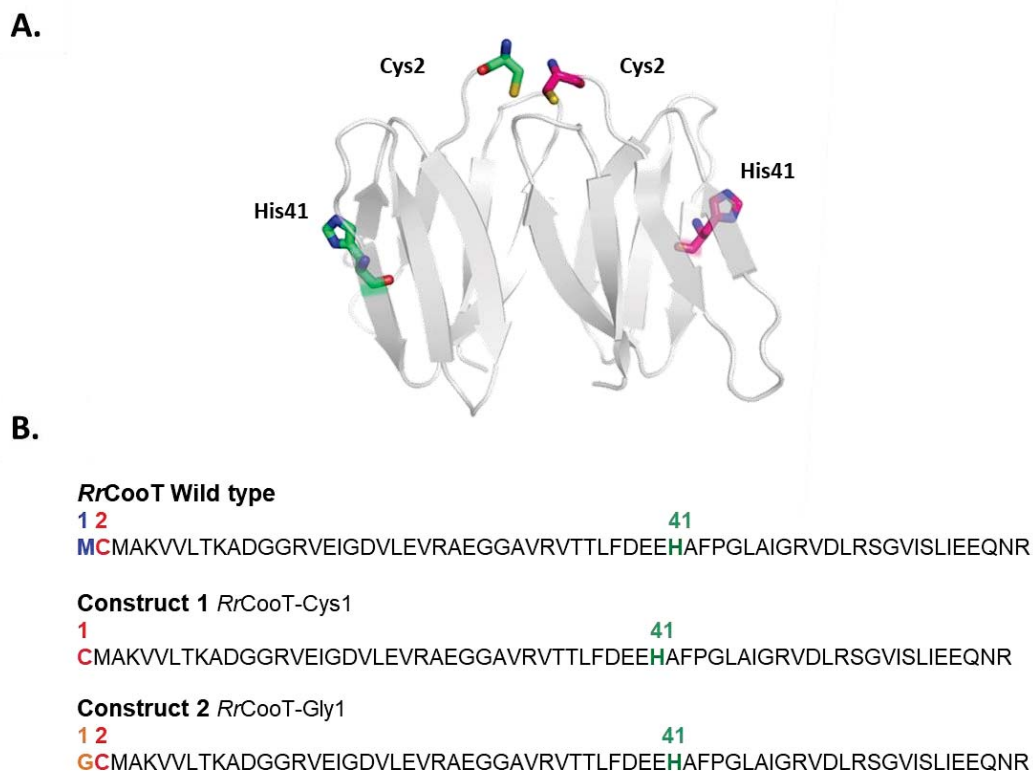


Figure 2. A) Crystal structure of the apo dimer in gray. The relative spatial position of Cys2 and His41 from each monomer is highlighted by colors (green monomer and magenta monomer) and these amino acids are shown in sticks. B) *RrCooT* wild type and its constructs sequences. The methionine 1 is highlighted in blue, the cysteine 2 in red, the histidine 41 in green and the mutated glycine in position 1 in orange.

To verify this, two different approaches have been adopted. The first one was based on site directed mutagenesis and processing of the Met1, where two different constructs were produced to verify the impact of the first amino acid on the nickel binding coordination:

- Construct 1 (*RrCooT-Cys1*): removal of Met1, so cysteine is now in the first position.
- Construct 2 (*RrCooT-Gly1*): substitution of Met1 with a Gly, in this case the side chain of the first amino acid is not present anymore.

Once produced and purified, the behavior of the two constructs was compared *in vitro* to that of the wild type protein, using different spectroscopical techniques.

The second approach was based on EXAFS and modeling data analysis, in which the use of DTF simulation was applied to the EXAFS experimental results to determine the best possible coordinating model.


This multi-methods approach led to the proposition of a square planar nickel coordination geometry, held by the thiolates and amides of the two facing cysteines 2 brought in spatial proximity by the protein dimerization. (*as shown in the graphical abstract of page 88*)

How we got to this suggested model and all the results obtained from the protein characterization are explained in detail in the paper “*A solvent-exposed cysteine forms a novel Ni(II)-binding site in the metallochaperone CooT from Rhodospirillum rubrum*” reported in the next pages.

The session that follows the paper is devoted to present the unpublished data, mainly concerning the unexpected behavior of the metal-binding coordination towards TCEP, the pH dependence, the assays performed on the holo-CooT protein produced in Ni supplemented media and the failed attempts to obtain N-terminal amine transamination reaction.



### 2.1.1 RrCooT paper

 ChemPubSoc  
Europe

DOI: 10.1002/chem.201903492

CHEMISTRY  
A European Journal  
Full Paper

■ Chaperone Proteins | Hot Paper |

**A Solvent-Exposed Cysteine Forms a Peculiar Ni<sup>II</sup>-Binding Site in the Metallochaperone CooT from *Rhodospirillum rubrum***

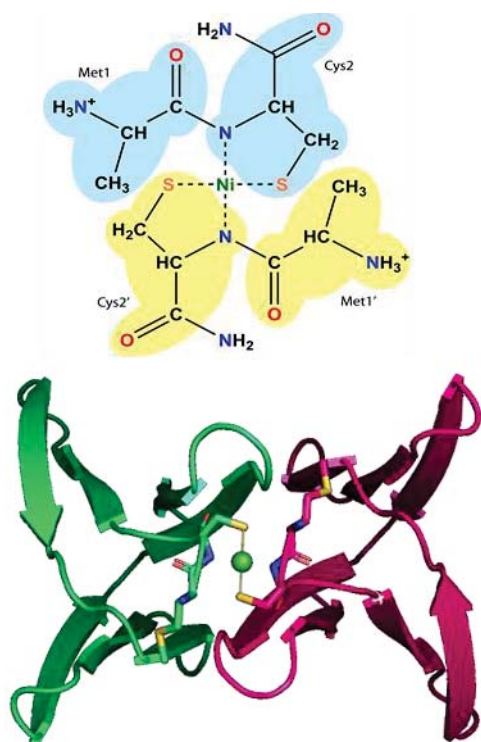
Marila Alfano,<sup>[a]</sup> Giulia Veronesi,<sup>[a]</sup> Francesco Musiani,<sup>[b]</sup> Barbara Zambelli,<sup>[b]</sup> L. Signor,<sup>[c]</sup> O. Proux,<sup>[d]</sup> M. Rovezzi,<sup>[d]</sup> Stefano Ciurli,<sup>\*,[b]</sup> and Christine Cavazza<sup>\*,[a]</sup>

- RrCooT binds one nickel per dimer using the conserved N-terminal residue sequence Met<sub>1</sub>-Cys<sub>2</sub>.
- The characterization of the two constructs RrCooT-Cys<sub>1</sub> and RrCooT-Gly<sub>1</sub> shown that :
  - RrCooT-Cys<sub>1</sub> loses the ability to coordinate Ni(II)
  - RrCooT-Gly<sub>1</sub> displays characteristics similar to the wild type protein

These results provide the evidence that an amino acid in position one is needed in order to bind the metal and that the nature of its side chain does not influence the coordination.

- From the CD in the far UV-Vis region and the absorption UV-Vis spectra a square planar geometry for the Ni(II) coordination is postulated. The possible ligands are the thiolates of the cysteine 2 and the terminal amine or the amide group of cysteine 2.
- The NMR studies confirm the involvement of the N-terminal region in nickel binding, which also increases the protein stability.
- The XANES spectrum confirms a square planar coordination geometry.
- From a combination of density functional theory (DFT) calculation and experimental EXAFS data, a bis-amidate/bis-thiolate coordination environment is postulated.
- The formation of the dimer promotes the spatial proximity of the two cysteine 2, which are the main players in the Ni coordination.
- Nickel binding is triggered by the dimer formation and held by the amide main-chain coordination completed by the side chain of the same amino acid (Cys<sub>2</sub>), making it a unique example for this kind of Ni(II) coordination among the chaperones.

## Graphical Abstract



1

Nickel (II) is bound in a square based geometry coordinated by :

- Thiolate of Cys2;
- Amidate of Cys2.

2

Solvent exposed cysteines are not commonly found in nickel binding site and, as far as we know, a single cysteine motif has never been shown in the literature.

## RESEARCH ARTICLE

# A solvent-exposed cysteine forms a peculiar Ni(II)-binding site in the metallochaperone CooT from *Rhodospirillum rubrum*

Marila Alfano,<sup>[a]</sup> Giulia Veronesi,<sup>[a]</sup> Francesco Musiani,<sup>[b]</sup> Barbara Zambelli,<sup>[b]</sup> L. Signor,<sup>[c]</sup> O. Proux,<sup>[d]</sup> M. Rovezzi,<sup>[d]</sup> Stefano Ciurli,<sup>\*,[b]</sup> Christine Cavazza<sup>\*,[a]</sup>

**Abstract:** In *Rhodospirillum rubrum*, the maturation of carbon monoxide dehydrogenase (CODH) requires three nickel chaperones, namely *RrCooC*, *RrCooT* and *RrCooJ*. Recently, the biophysical characterization of *RrCooT* homodimer and the X-ray structure of its apo-form revealed the existence of a solvent exposed Ni(II)-binding site at the dimer interface, involving the strictly conserved Cys2. Here, a multifaceted approach that used NMR and X-ray absorption spectroscopies, complemented with structural bio-modelling methodologies, was used to characterise the binding mode of Ni(II) in *RrCooT*. This study suggests that Ni(II) adopts a square-planar geometry via a N<sub>2</sub>S<sub>2</sub> coordinating environment that comprises the two thiolate and amidate groups of both Cys2 residues at the dimer interface. The existence of a diamagnetic mononuclear Ni(II) centre with bis-amidate/bis-thiolate ligands, coordinated by a single cysteine motif, is unprecedented in biology and raises the question of its role in the activation of CODH, at the molecular level.

## Introduction

Despite their low occurrence in nature, the importance of Ni-dependent enzymes in Archaea, bacteria, fungi, algae and higher plants has been widely recognized in the last few decades. These enzymes use a Ni-containing active site to catalyse essential reactions such as the reduction of anionic superoxide radical,<sup>[1]</sup> urea hydrolysis,<sup>[2]</sup> methane production,<sup>[3]</sup> or H<sub>2</sub>/CO reversible oxidation.<sup>[4]</sup> In the hydrogenogenic carboxydotroph *Rhodospirillum rubrum*, the nickel-dependent carbon monoxide dehydrogenase (*RrCODH*) is a key enzyme in the energy metabolism.<sup>[5]</sup> Indeed, a characteristic feature of this versatile photosynthetic bacterium is its capacity to use CO as the sole energy source under anaerobic conditions, via the water-gas shift reaction (H<sub>2</sub>O + CO → CO<sub>2</sub> + H<sub>2</sub>).<sup>[6]</sup> In this reaction, *RrCODH* plays a central role by catalysing the oxidation of CO to CO<sub>2</sub>, coupled to the H<sub>2</sub> production catalysed by an energy-converting [NiFe]-hydrogenase.<sup>[7]</sup> *RrCODH* is a homodimer bridged by an intermolecular [4Fe-4S]-cluster (D-cluster), with each monomer containing a [4Fe-4S]-cluster (B-cluster) and a peculiar [Ni-4Fe-4S] active site (C-cluster).<sup>[8]</sup> The CO oxidation occurs at the C-cluster, whereas the B-clusters are involved in the electron transfer between the active site and its physiological partner, the ferredoxin CooF.<sup>[9]</sup> Nickel insertion into the active site is the essential step for the enzyme activation: Ni-deficient CODH, produced in nickel-depleted growth media, presents a fully assembled [Fe<sub>4</sub>S<sub>4</sub>] C-cluster that can be partially reactivated *in vitro* by Ni(II) addition under reductive conditions.<sup>[10]</sup>

A common feature of metalloenzymes is the requirement of auxiliary proteins, typically involved in the metallo-centre biosynthesis and assembly.<sup>[11]</sup> In the case of *RrCODH*, the *cooFSCTJ* operon contains the *cooCTJ* genes encoding three different nickel-binding proteins, CooC, CooT and CooJ, present downstream the structural *cooS* gene that encodes CODH.<sup>[12]</sup>

*RrCooC* is Ni-dependent ATPase, proposed to provide the energy required for the maturation process and able to dimerize by binding Ni(II) via a CXC motif that is conserved among CooC homologs.<sup>[13]</sup> *RrCooJ* is a 25-kDa homodimeric protein containing two distinct metal-binding sites: a C-terminal histidine tail composed by sixteen histidines and two cysteines able to bind four Ni(II) per dimer, and a strictly conserved HX<sub>3</sub>HX<sub>3</sub>H binding site present in the N-terminal region and binding one Ni(II) per dimer.<sup>[14]</sup> Finally, *RrCooT* is a small homodimeric protein of 66 residues able to bind one Ni(II) per dimer with a dissociation constant in the nanomolar range, using Cys2.<sup>[15]</sup> Looking at the amino acid recurrence in proteins, cysteine is one of the least abundant residues.<sup>[16]</sup> However, when present, cysteines are required in essential functional sites, such as structural disulphide bridges, iron-sulphur cluster coordination or sulphur-rich metal coordination sites in metalloproteins.<sup>[16]</sup> Based on the analysis of protein structure databases, this residue shows the characteristic feature of being either highly conserved in buried cysteine clusters, or poorly conserved when isolated and located at the protein surface.<sup>[16]</sup> In *RrCooT*, Cys2 is the only cysteine in the protein sequence and the crystal structure of the apo-protein revealed its solvent exposure.<sup>[15]</sup> Unexpectedly, while the amino acid sequences of the 111 CooT homologues identified in anaerobic archaea and bacteria show a remarkable large variability, Cys2 is the only strictly conserved residue,<sup>[15]</sup> in this way violating the general conservation vs. structure pattern. Thus, the evolutionary conservation of the solvent-exposed cysteine in CooT proteins suggests its importance to meet a specific and essential functional role.

By analogy, several well-described nickel proteins possess Ni-binding sites containing multiple cysteine residues in the proximity of the N-terminus. They form the well-known Ni-hook motif "HCX<sub>2</sub>PCGXY" in the nickel-dependent superoxide dismutase (NiSOD)<sup>[17]</sup> or the high-affinity "MCX<sub>2</sub>CXC" site in HypB,<sup>[18]</sup> a GTPase involved in [Ni-Fe]-hydrogenase for which the Ni(II)-binding mode is not definitively established. The particularity of *RrCooT* is the presence of a Ni-binding site in the N-terminal region, with a single cysteine motif, raising the question of its metal-coordination mode. The X-ray structure of apo-*RrCooT* revealed that its dimerization brings the two Cys2<sub>A</sub> and Cys2<sub>B</sub> residues (belonging to the two monomers A and B) in close proximity, likely triggering the formation of a Ni-binding site. Indeed, mutation of Cys2 to Ser obliterated Ni(II)-binding to CooT, revealing the essential role of the Cys-thiolate groups.<sup>[15]</sup> However, no structural information is yet available for the Ni-binding site in *RrCooT*.

In the present study we investigated the structural details of the Ni(II)-binding site of *RrCooT* using a multifaceted approach that involved NMR and X-ray absorption spectroscopies complemented with structural bio-modelling methodologies. The results support the presence of an unprecedented metal-binding motif that involves a pair of isolated solvent exposed cysteine bound to a square-planar Ni(II) ion in a N<sub>2</sub>S<sub>2</sub> coordination environment, with a bis-thiolate/bis-amidate ligation in a *trans* configuration.

<sup>[a]</sup> University of Grenoble Alpes, CEA, CNRS, IRIG, CBM, F-38000 Grenoble, France

<sup>[b]</sup> Laboratory of Bioinorganic Chemistry, Department of Pharmacy and Biotechnology, University of Bologna, Via Giuseppe Fanin 40, I-40127 Bologna (Italy)

<sup>[c]</sup> University of Grenoble Alpes, CEA, CNRS, IRIG, IBS, F-38000 Grenoble, France.

<sup>[d]</sup> Univ. Grenoble Alpes, CNRS, IRD, Irstea, Météo France, OSUG, FAME, 38000 Grenoble, France

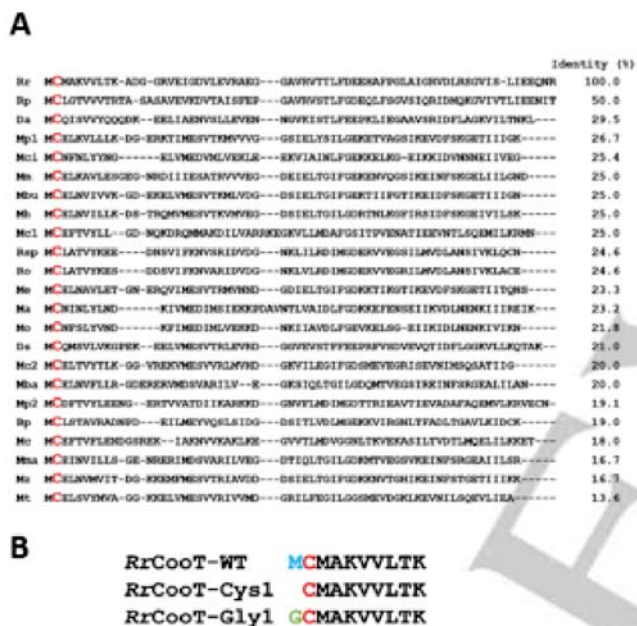


## RESEARCH ARTICLE

## Results

## Metal-binding motifs in the CooT family

Alignment of 91 non-redundant CooT sequences revealed the existence of two distinct nickel-binding motifs in the CooT family: the “Cys2-His55” motif, previously characterized in the putative CooT from *Carboxydotherrmus hydrogenoformans* (ChCooT),<sup>[19]</sup> is conserved in 64 CooT homologues (corresponding to 70.3% of all sequences) (Figure S1). It is mainly found in bacteria (48 sequences) and in some Archaea (16 sequences). On the other hand, the “Cys2” motif is found in 16 archaeal and 7 bacterial CooT sequences (25.3 % of all sequences) (Figure 1A). In addition, in three methanogens, a second cysteine is also found in position 4, forming a putative not characterized “Cys2-Cys4” motif. *RrCooT* belongs to the “Cys2” motif group, as shown in Figure 1A.



**Figure 1.** (A) Sequence alignment of CooT homologues possessing the “Cys2” motif. Cys2 is in red. Rr: *Rhodospirillum rubrum* ATCC11170, Da: *Desulfurivibrio alkaliphilus* AHT2, Ro: *Ruminococcus obeum* A2-162, Mp1: *Methanobolus psychrophilus* R15, Mh: *Methanomethylovorans hollandica* DSM 15978, Mci: *Methanocaldococcus infernus* ME, Me: *Methanohalobium evestigatum* Z-7303, Mm: *Methanohalophilus mahii* DSM 5219, Ma: *Methanococcus aeolicus* Nankai-3, Mc2: *Methanosarcina concilii* 2 GP6, Rsp: *Ruminococcus* sp, Mbu: *Methanococcoides burtonii* DSM 6242, Mo: *Methanothermococcus okinawensis* IH1, Mc1: *Methanosarcina concilii* 1 GP6, Ds: *Desulfocapsa sulfexigens* DSM 10523, Mc: *Methanocella conradii* HZ254, Mp2: *Methanobolus psychrophilus* 2 R15, Bp: *Butyrivibrio proteoclasticus* B316, Mz: *Methanosalsum zhilinae* DSM 4017, Mba: *Methanosarcina barkeri* MS, Mma: *Methanosarcina mazel* S-6, Mt: *Methanosarcina thermophila* PT, Rp: *Rhodopseudomonas palustris* BisB18. Identity percent corresponds to the conserved amino acids between each sequence and *RrCooT* sequence. (B) N-terminal sequences of *RrCooT*-WT, *RrCooT*-Cys1 and *RrCooT*-Gly1.

Modification of the N-terminus of *RrCooT* and Ni(II) binding

The thiolate groups of Cys2<sub>A</sub> and Cys2<sub>B</sub> provide two potential coordinating ligands for Ni(II). However, at least two additional ligands are required to complete the coordination environment of Ni(II) in a square-based, tetrahedral or octahedral geometry. Considering the position of the metal-binding site at the beginning of the protein sequence, the N-terminal amino group would be a logical choice as a Ni donor ligand. However, in all identified sequences shown in Figure 1A the first residue is a methionine, which is often processed by methionine amino peptidases (MetAP), thus raising the question as to whether this residue is

strictly necessary to complement the thiolate groups of Cys2 for Ni binding. No information is available about the processing of Met1 in *RrCooT* by MetAP in *R. rubrum*. Therefore, in order to understand the role of Met1 and the impact of the position of the conserved Cys residue in the sequence at the N-terminus, different variants were constructed. Overproduction of *RrCooT* in *E. coli* grown for 16 h in a minimal medium at 25 °C led to a major Met1-containing form (about 80% of the total protein, as previously described in<sup>[15]</sup>), corresponding to *RrCooT*-WT (Figure 1B). On the other hand, replacement of the minimal medium with a rich-growth medium, raising the temperature to 37 °C and using an induction time of 4 h resulted in a form of *RrCooT* lacking Met1 (*RrCooT*-Cys1), (Figure 1B). The latter protocol was also used to express a synthetic gene featuring a Gly residue inserted between Met1 and Cys2, producing the *RrCooT*-Gly1 variant also lacking the Met residue at the N-terminus (Figure 1B). The identity of these two protein variants was confirmed by ESI-MS (Figure S2). The ESI-MS spectrum of *RrCooT*-Gly1 showed the expected single peak at 7,022.11 Da, confirming the identity and homogeneity of this mutant.

Considering that N-terminal cysteines are known to be highly reactive,<sup>[20]</sup> the integrity of the thiol groups in *RrCooT*-Cys1 was tested after removal of the reducing agent TCEP and overnight exposure to air. In the case of *RrCooT*-Gly1, this treatment led to the appearance of a major species, with a mass of 14,042.24 Da, corresponding to a homodimer induced by the formation of a disulphide bridge between Cys2<sub>A</sub> and Cys2<sub>B</sub> with the loss of 2 Da, due to the oxidation of the two thiol groups (Figure S2). On the contrary, ESI-MS of several samples of *RrCooT*-Cys1 revealed a mixture of two species: the first, at 6,964.93 Da, corresponds to the Met1-processed monomer, while the second showed an unknown adduct of 26 Da. This adduct is never observed when the Met1 is present at the N-terminus of *RrCooT*. Moreover, the amount of the second species varied from 50 to 90% depending on the *RrCooT*-Cys1 samples. The 26 Da-mass is consistent with the formation of a S-CN moiety on Cys2<sup>[21]</sup> and ESI-MS experiments showed that its presence hinders the formation of the disulphide bridge (Figure S2), suggesting that Met1 processing drastically affects the thiol group reactivity.

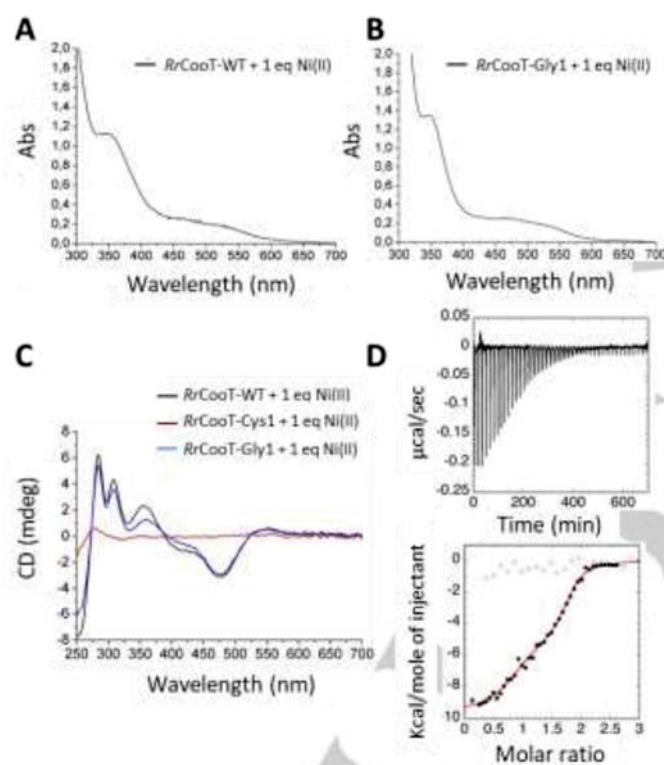
Biophysical characterization of *RrCooT*-Cys1 and *RrCooT*-Gly1

The Ni(II) binding ability of the different constructs was estimated using UV/vis and circular dichroism (CD) spectroscopy in the near-UV/vis range (250–700 nm) and by isothermal titration calorimetry (ITC). UV/vis and CD spectra were recorded under anaerobic conditions to prevent the formation of disulphide bridges, while ITC titrations were conducted in the presence of the reducing agent TCEP. In the UV/vis spectra of *RrCooT*-WT and *RrCooT*-Gly1, transition bands at 345 nm, 427 nm, 475 nm and 530 nm were observed (Figure 2A, B). Similar UV/vis signatures have been previously observed for square planar Ni(II) coordinated to di-amido/di-thiolato ligands in Ni-containing superoxide dismutase (NiSOD) model peptides<sup>[22]</sup> and a NiSOD mutant, in which the amidate/amine coordination was switched to a bis-amidate mode by adding an alanine residue to the N-terminus.<sup>[23]</sup> The CD spectrum of *RrCooT*-Gly1 (Figure 2C) is similar to the previously published spectrum of *RrCooT*-WT,<sup>[15]</sup> with ligand-to-metal charge transfer and d-d transitions at 260 nm (-), 287 nm (+), 310 nm (+), 355 nm (+), ~410 nm (-) and 475 nm (-). The presence of CD bands between 400 and 600 nm suggests the presence of a Ni(II)-bound amide, rather than amine, nitrogen binding, as previously observed for Ni(II)-containing peptides mimicking NiSOD.<sup>[24]</sup>

A calorimetric Ni(II) titration of *RrCooT*-Gly1 showed negative peaks following Ni(II) injections, indicative of exothermic events (Figure 2D). The shape of the binding isotherm obtained from data

## RESEARCH ARTICLE

integration indicated the occurrence of two independent binding events. The first event involves the binding of one Ni(II) ion to the protein dimer with a high affinity ( $K_{A1} = 3 \pm 1 \times 10^7$ ;  $K_{D1} = 50 \pm 20$  nM,  $\Delta H_1 = -9.6 \pm 0.3$  kcal mol $^{-1}$ ;  $\Delta S_1 = +2.10$  kcal mol $^{-1}$  K $^{-1}$ ), while the second binding event shows a lower affinity ( $K_{A2} = 3 \pm 2 \times 10^5$ ;  $K_{D2} = 3 \pm 0.6$   $\mu$ M,  $\Delta H_2 = -4.8 \pm 0.7$  kcal mol $^{-1}$ ,  $\Delta S_2 = +13.2$  kcal mol $^{-1}$  K $^{-1}$ ). These data are similar to those previously obtained for RrCooT-WT,<sup>[15]</sup> indicating that RrCooT-Gly1 contains a preserved high affinity Ni(II) binding site, as well as an additional weaker metal site, most likely non-physiological. The similar metal binding behaviour of RrCooT-Gly1 and RrCooT-WT indicates that the nature of the side chain on the first residue of the sequence does not influence the Ni(II) binding properties of the high affinity site of RrCooT. On the other hand, the RrCooT-Cys1 mutant spectrum does not show any CD signal upon Ni(II) addition, indicating the absence of Ni(II) binding (Figure 2C). This was confirmed by ITC, which showed that the binding event is fully abrogated (Figure 2D). These observations indicate that the position of the conserved Cys as the second residue in the sequence of RrCooT is essential for binding, and consequently that Met1 processing by MetAP is most likely not occurring in *R. rubrum*.



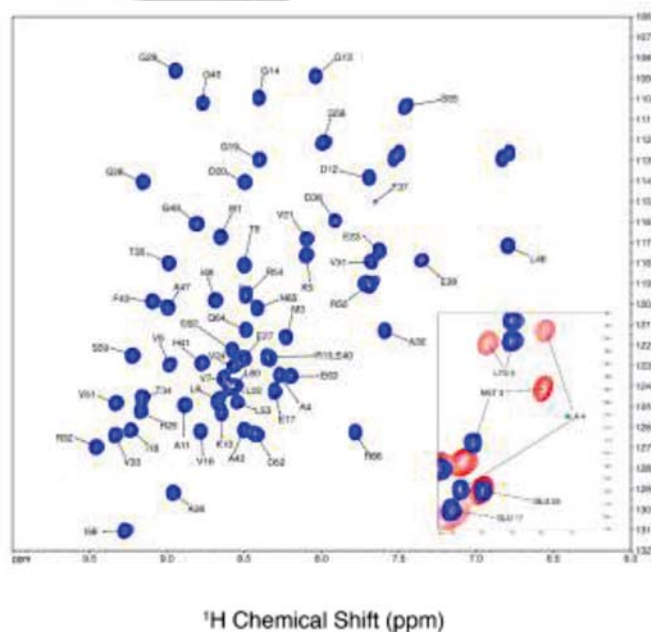
**Figure 2:** (A) UV-vis spectrum of RrCooT-WT (500  $\mu$ M dimer) in the presence of one equivalent of Ni(II), in 50 mM HEPES buffer at pH 7.5, under anaerobic conditions;  $\epsilon_{345} = 2254$  M $^{-1}$  cm $^{-1}$ ,  $\epsilon_{475} = 502$  M $^{-1}$  cm $^{-1}$  and  $\epsilon_{530} = 345$  M $^{-1}$  cm $^{-1}$ . (B) UV-vis spectrum of RrCooT-Gly1 (630  $\mu$ M dimer) in the presence of one equivalent of Ni(II), in 50 mM HEPES buffer at pH 7.5;  $\epsilon_{345} = 2134$  M $^{-1}$  cm $^{-1}$ ,  $\epsilon_{427} = 417$  M $^{-1}$  cm $^{-1}$ ,  $\epsilon_{475} = 396$  M $^{-1}$  cm $^{-1}$ , and  $\epsilon_{530} = 279$  M $^{-1}$  cm $^{-1}$ . (C) CD spectra for RrCooT-WT (black line), RrCooT-Cys1 (red line) and RrCooT-Gly1 (blue line) registered for a 25  $\mu$ M dimer concentration in the presence of one molar equivalent of Ni(II). (D) Top panel: representative raw plot of the ITC data of Ni(II) (240  $\mu$ M) injected onto the RrCooT-Gly1 solution (12  $\mu$ M); bottom panel: integrated heat data as a function of the metal-to-protein molar ratio for RrCooT-Gly1 (filled circles) and RrCooT-Cys1 (hollow circles). The continuous line represents the best fit obtained for RrCooT-Gly1 titration, using a model involving two independent sets of binding sites for Ni(II), as described in the text.

The thiol modification appears to be the most likely reason for RrCooT-Cys1 being unable to bind Ni(II). However, the removal

of the Met1 also changed the amide group of Cys2 to a protonated N-terminal amine, which could also impact Ni(II) binding. To go further into the characterization of the Ni-binding site of RrCooT-WT and the identification of the N donor ligands, NMR spectroscopy experiments were carried out on the apo- and holo-forms of the protein.

### Assignment of the NMR spectra of apo-RrCooT-WT

Heteronuclear 2D and 3D NMR spectra of dimeric RrCooT-WT were recorded and analysed, and backbone assignments were obtained using the scalar connectivities provided by triple resonance experiments (Table S1). The assigned  $^1\text{H}$ ,  $^{15}\text{N}$  HSQC spectrum recorded at 950 MHz  $^1\text{H}$  Larmor frequency is shown in Figure 3. The large spectral dispersion of the signals in the  $^1\text{H}$  dimension unambiguously indicates that the protein is well-folded in solution, while the number of cross-peaks is consistent with the number of residues in a single subunit, supporting a symmetrical average orientation of the two monomers. The identification of amide  $^1\text{H}$  and  $^{15}\text{N}$  peaks was obtained at 950 MHz for 63 out of the 66 residues following a standard sequential assignment procedure (Table S2).



**Figure 3.**  $^1\text{H}$ ,  $^{15}\text{N}$  HSQC spectrum of RrCooT-WT at 950 MHz with the corresponding assignment. The inset shows the overlay between the spectrum of apo-RrCooT-WT (blue) at 950 MHz and holo-RrCooT-WT (red) at 500 MHz in the region covering the N-terminus residues.

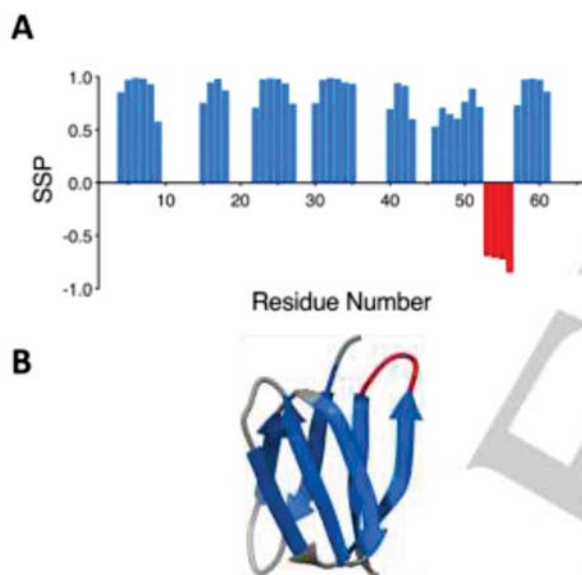
The Pro44 residue is not observable in the  $^1\text{H}$ ,  $^{15}\text{N}$  HSQC spectrum because it does not have the peptide NH group, while the N-terminal Met1 is not observable because of fast proton exchange with water of its  $\text{NH}_3^+$  terminal group with water. The peptide NH signal for the remaining Cys2 residue was also not observable in the spectrum at 950 MHz probably because its NMR signal is broadened beyond detection due to conformational exchange phenomena occurring with rates comparable to the frequency differences among the different conformers. Attempts to observe its signal in  $^1\text{H}$ ,  $^{15}\text{N}$  HSQC spectra taken at 500 MHz were also unsuccessful, indicating that Cys2, and presumably even more so Met1, undergoes conformational changes with correlation times in the sub-microseconds time scale. On the other hand, the peaks corresponding to Met3 and Ala4, while visible at 950 MHz, also disappeared in the spectra taken at 500 MHz, consistently with a progressively slower conformation



## RESEARCH ARTICLE

exchange rate for residues gradually farther away from the N-terminus. Additional attempts to observe the peaks corresponding to Met1 and Cys2 in the  $^1\text{H}$ ,  $^{15}\text{N}$  HSQC spectra, carried out by lowering or raising the temperature at all tested fields, were unsuccessful. However, the  $\text{C}\alpha/\text{H}\alpha$  and  $\text{C}\beta/\text{H}\beta$  pairs of Cys2 were observed in the spectra of apo-*RrCooT*-WT using HNC0/HN(CA)CO and HNCACB/CBCA(CO)NH triple resonance spectra (Table S1).

The program TALOS+<sup>[25]</sup> was used to correct the referencing for  $\text{H}\alpha$ ,  $\text{C}\alpha$ ,  $\text{C}\beta$ , and CO chemical shifts, through an empirical correlation.<sup>[26]</sup> Then, the HN, N,  $\text{H}\alpha$ ,  $\text{C}\alpha$ ,  $\text{C}\beta$ , and CO, chemical shifts assigned to the residues of apo-*RrCooT*-WT were used by TALOS+ to estimate the  $\phi$  and  $\psi$  backbone torsion angles and the consequent secondary structure (Figure 4) using an artificial neural network.<sup>[25]</sup> A negative TALOS+ score indicates a propensity for  $\alpha$ -structure or loop, while a positive TALOS+ score indicates a propensity for  $\beta$ -structure. Figure 4 shows the good agreement between the secondary structure determined by NMR in solution and that obtained by solid-state crystallography of the apo-protein,<sup>[15]</sup> demonstrating the conservation of the protein structure independently of its solid or solution state.

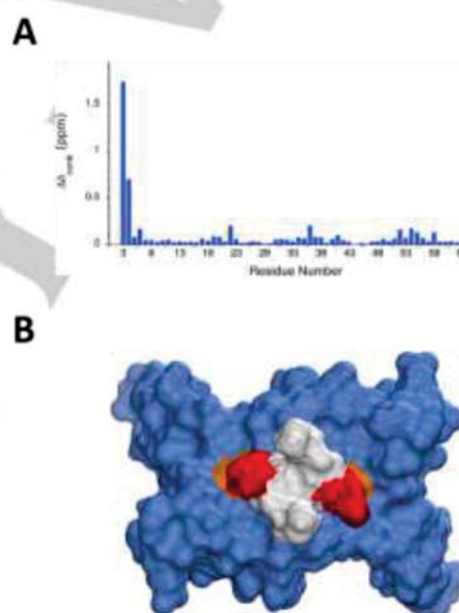


**Figure 4.** (A) Artificial neural network predicted secondary structure (SSP) for apo-*RrCooT*-WT, using  $\text{H}\alpha$ , N,  $\text{C}\alpha$  and  $\text{C}\beta$  backbone secondary NMR chemical shifts obtained at 950 MHz. The blue and red bars indicate beta and helix/loop secondary structure, respectively. (B) The ribbon scheme of the crystal structure of the apo-*RrCooT* monomer (PDB id: 5N76) is shown colored according to the TALOS+ predicted secondary structure to highlight the consistency between the solution NMR and the solid-state crystallographic data.

#### Assignment of the NMR spectra of holo-*RrCooT*-WT

The analysis of the modifications occurring in the  $^1\text{H}$ ,  $^{15}\text{N}$  HSQC spectrum of apo-*RrCooT*-WT at 500 MHz and 950 MHz by subsequent addition of 0.0, 0.5, 1.0 and 2.0 equivalents of Ni(II) per protein dimer revealed that the chemical shift and intensity of most  $^1\text{H}/^{15}\text{N}$  backbone amide cross-peaks were largely unperturbed, and that only few residues were affected by significant chemical shift changes in the slow exchange limit. No modification was observed between the 1.0 and 2.0 titration steps, confirming the stoichiometry of 1 Ni(II) per *RrCooT*-WT dimer determined by ITC for the high-affinity site. The lack of signal disappearance upon Ni(II) addition indicates that the metal ion is diamagnetic in holo-*RrCooT*-WT. The marginally affected cross-peaks were assigned in the Ni(II)-bound *RrCooT*-WT on the basis of a simple closest-neighbour matching approach. The analysis of

the HNCACB/CBCA(CO)NH pair of triple resonance experiments at 500 MHz confirmed this initial assignment and further extended it to comprise all observable signals in the  $^1\text{H}$ ,  $^{15}\text{N}$  HSQC spectrum of holo-*RrCooT*-WT, with the exception of Met1 and Cys2. Significantly, the signals corresponding to Met3 and Ala4, not observable in the 500 MHz spectrum of apo-*RrCooT*-WT, became visible again at this magnetic field upon Ni(II) addition. This observation implicates the hypothesis that Ni(II) binding occurring at the N-terminal region of the protein, causes a significant decrease of its mobility. Finally, the very large chemical shifts induced by Ni(II) binding on Cys2  $\text{C}\alpha$  (from 55.55 ppm to 61.99 ppm) and  $\text{C}\beta$  (from 25.56 to 35.59 ppm), and the very small chemical shifts induced on Met3  $\text{C}\alpha$  (from 52.23 ppm to 52.36 ppm) and  $\text{C}\beta$  (from 31.31 to 31.12 ppm) reveal that Cys2 is part of the metal coordination environment, and that the protein structure is negligibly affected by nickel binding beyond this residue. The Ni(II)-induced chemical shift perturbation (CSP), mapped on the protein surface (Figure 5) confirmed that the region of *RrCooT* at or near the N-terminal residues is involved in Ni(II) binding, with the thiol group of Cys2 being the most likely candidate for direct metal ion coordination.



**Figure 5.** (A) Plot of combined chemical shift perturbation  $\Delta\delta_{\text{comb}}$  upon Ni(II) binding to apo-*RrCooT*-WT, along the protein sequence. (B) The protein surface, calculated from the X-ray structure of apo-*RrCooT* and color-coded according to the size of  $\Delta\delta_{\text{comb}}$  ( $\Delta\delta_{\text{comb}} > 1.0$  red;  $0.5 < \Delta\delta_{\text{comb}} < 1.0$  orange;  $\Delta\delta_{\text{comb}} < 0.5$  blue; non-assigned residues M1 and C2 are shown in white).

This hypothesis was confirmed by the observation that no perturbation in the  $^1\text{H}$ ,  $^{15}\text{N}$  HSQC spectrum was induced by addition of Ni(II) to a sample of apo-*RrCooT*-WT stored for several days in air, a treatment shown by ESI-MS to cause oxidation of the two Cys2 residues in the dimer (see above), while the typical spectrum of holo-*RrCooT*-WT was re-established upon addition of TCEP, known to reduce disulphide bonds. These NMR spectroscopic results indicate that Ni(II) binding in *RrCooT* involves the thiolate groups of the two Cys2 residues, yielding a square-planar or square-pyramidal geometry for a diamagnetic Ni(II) ion in holo-*RrCooT*-WT. In order to further examine the coordination environment of the Ni(II) site, X-ray absorption spectroscopy was applied.

#### Preliminary X-ray absorption spectroscopy analysis

## RESEARCH ARTICLE

Ni K-edge X-ray absorption spectra of *RrCooT*-WT in complex with 0.9 Ni(II) molar equivalent per dimer (*Ni-RrCooT*) were acquired under cryogenic conditions. The X-ray Absorption Near Edge Structure (XANES) spectrum of the holo-protein is shown in Figure 6A, in which a weak pre-edge peak at  $\sim 8332$  eV associated to  $1s \rightarrow 3d$  electronic transitions, and a resolved maximum at  $\sim 8338$  eV, are observed. The latter is associated to  $1s \rightarrow 4p_z$  transitions and has been shown to be indicative of square planar Ni coordination geometry.<sup>[27]</sup>

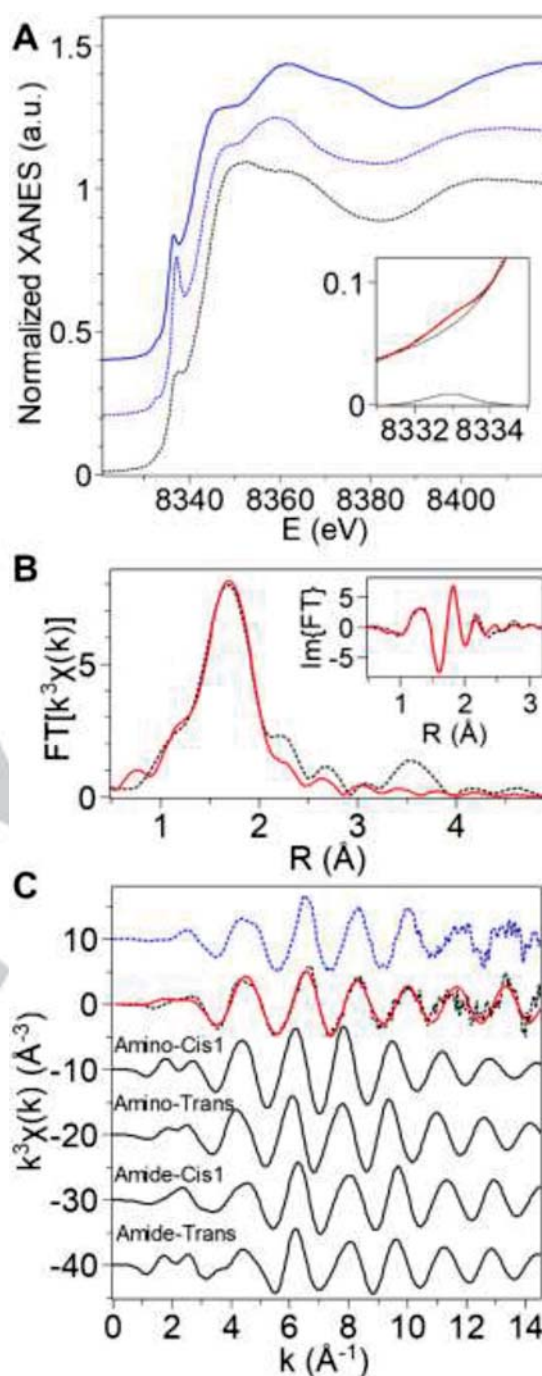
For comparison with *Ni-RrCooT*, we measured the spectra of two Ni(II) tetra-coordinated square planar compounds exhibiting different ligand environments (Figure S3)<sup>[28]</sup> (Figure 6): both  $N_2S_2$  and  $S_4$  coordination models give rise to a resolved peak at  $\sim 8338$  eV as reported in literature, and a  $1s \rightarrow 3d$  feature of comparable intensity with respect to that observed for *Ni-RrCooT*. The area of Gaussian peaks reproducing the  $1s \rightarrow 3d$  transition, which provides information on the coordination geometry at the Ni site, was calculated for protein and reference samples.

The best-fitting curve for *Ni-RrCooT* is reported in the inset of Figure 6A, and the estimated curve parameters in Table S4. The peak areas fall in the range previously described for a collection of four-coordinate, square planar Ni compounds (area  $< 2.9 \cdot 10^{-2}$  eV), and differ dramatically from the areas calculated for tetrahedral and five-coordinated Ni complexes (area  $> 4.2 \cdot 10^{-2}$  eV).<sup>[41]</sup> This analysis, together with the observed resolved  $1s \rightarrow 4p_z$  transition, strongly supports a square planar Ni binding geometry in *Ni-RrCooT*, consistently with NMR spectroscopy.

In order to disclose the nature of the Ni ligands in the *RrCooT* dimer, the first shell peak of the Fourier-transformed EXAFS signal was analysed in the range  $R = 1.0 - 2.2$  Å. Considering the results obtained from the biophysical and spectroscopic characterization of the holo-protein described in previous sections, we expect the Ni-binding atoms to belong to the amino acids in position 1 and 2. Therefore, coordination spheres composed of S and N atoms were considered, given that these atoms can be provided either by the thiolate side chain of Cys2 or by the protein backbone, respectively. Four ligands in planar geometry were considered according to the XANES analysis. The first-shell peak is best fitted by a  $N_2S_2$  coordination (Figure 6B), whereas a N-only coordination fails to reproduce the experimental spectra. The  $N_2S_2$  coordination is also suggested by the similarity in the EXAFS oscillations between the experimental spectrum of *Ni-RrCooT* and of the  $N_2S_2$  reference compound (Figure 6A, dotted lines). Remarkably, two distinct Ni-N distances had to be considered, whereas if a common distance was assigned to both Ni-N scattering paths, the first peak of the FT EXAFS spectrum could not be reproduced. Removing the degeneracy in the Ni-N distances resulted in a net improvement of the agreement between fit and experimental data (R-factor drops from 4.9% to 0.6%), and to the determination of first-shell Debye-Waller factors that were fitted to unphysical values in case of degenerated Ni-N distances (see Table 1, fit number 1 vs 2). In order to obtain reliable structural models for ab initio EXAFS refinement, a well-established computational approach was adopted.<sup>[29]</sup>

### Modelling of the *RrCooT* high affinity Ni(II)-binding site

Two sets of models were generated by considering all the possible configurations for a square planar Ni(II) geometry (Scheme 1). In the first set of models, the Ni(II) Met1 terminal amine N atom and the Cys2 thiolate S atom from each monomer were considered (models Amino-Cis1, Amino-Cis2 and Amino-Trans hereafter). In the second set, the backbone N atom of Cys2 from each monomer was instead examined (models Amide-Cis1, Amide-Cis2 and Amide-Trans hereafter). Initial distances were taken from the fit of the first-shell peak of the Fourier-Transformed EXAFS signal (Table 1). The constraints used in the modelling stages are reported in Table S5 and Table S6).

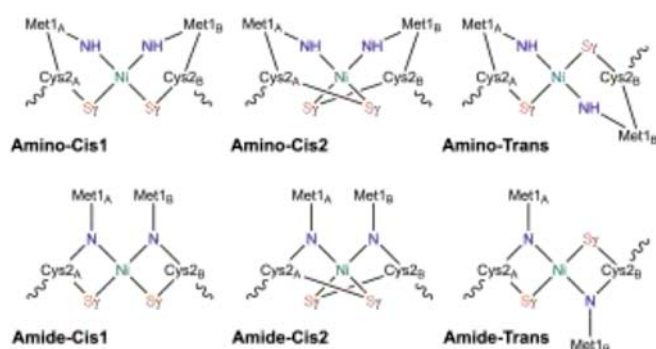


**Figure 6** (A) Ni K-edge XANES spectra of square-planar Ni(II) reference compounds (blue) with  $N_2S_2$  (dotted line) and  $S_4$  (solid line) coordination, and of *Ni-RrCooT* (black dots). Inset: magnification of the  $1s \rightarrow 3d$  transition energy region and fit (red) of the experimental data (black dots). The two components of the fitting curve are represented as black solid lines. (B) Fourier Transformed experimental EXAFS spectrum of *Ni-RrCooT* (black dotted line) and theoretical curve providing the best fit to the first shell peak (red solid curve). The imaginary part of both curves is reported in the inset. (C)  $k^3$ -weighted experimental EXAFS spectra of *Ni-RrCooT* (black dashed line) and of a  $Ni(II)N_2S_2$  reference compound (blue dashes). Theoretical spectra (black solid lines) generated from optimized structures of *Ni-CooT* varying the N donor (amine N of Met1 or amide N of Cys2) and the relative arrangements of first-shell ligands (S atoms in trans or cis position). Best fitting curve (red solid line) for the experimental data starting from the optimized model "amide-trans".



## RESEARCH ARTICLE

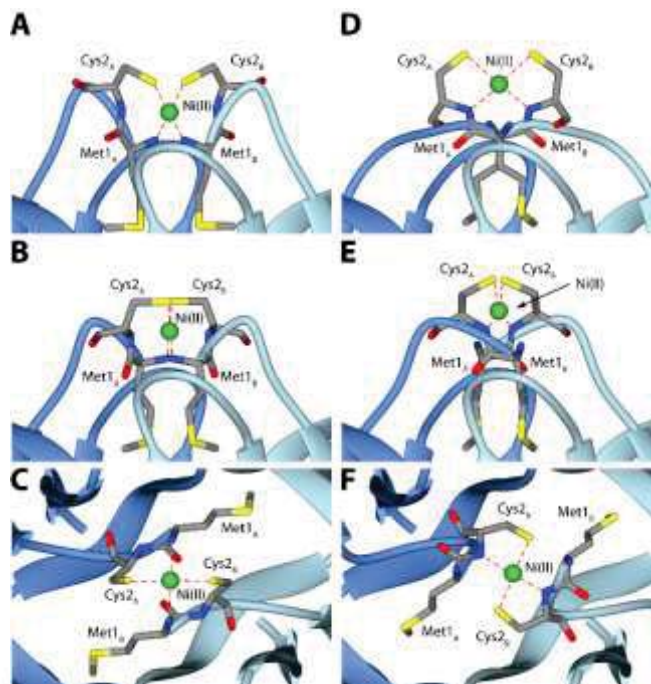
Scheme 1



**Table 1.** Structural and dynamic parameters of the Ni-RrCooT site, extracted from *ab initio* fitting of Ni K-edge EXAFS data based on different models (see text). The  $1\sigma$  standard deviation error relative to the last digit is reported in parentheses

| Fit number | Ni-neighbours   | R (Å)     | $\sigma^2$ ( $10^{-3}$ Å <sup>2</sup> ) | $\Delta E_0$ (eV) | R-factor (%) |
|------------|-----------------|-----------|---|-------------------|--------------|
| 1          | 2N              | 2.07 (4)  | 50 (10)                                 | -10 (2)           | 4.9          |
|            | 2S              | 2.13 (3)  | 3 (2)                                   |                   |              |
| 2          | N               | 1.87 (4)  | 2 (2)                                   | -8.2              | 0.6          |
|            | N               | 2.09 (4)  | 2 (2)                                   | (5)               |              |
|            | 2S              | 2.14 (1)  | 3.0 (6)                                 |                   |              |
| 3          | N               | 1.86 (2)  | 2.0 (10)                                | -9.4              | 5.9          |
|            | N               | 2.10 (2)  | 2.0 (10)                                | (8)               |              |
|            | 2S              | 2.14 (1)  | 2.7 (6)                                 |                   |              |
|            | 6C <sup>+</sup> | 2.90-3.37 | 7 (5)                                   |                   |              |

In all calculated models, the Ni(II) ion binds at the dimer interface that connects the N-terminal portion of each monomer (Figures S4 and S5). The details of the metal binding regions resulting from the modelling are reported in Figure 7. The Amino-Cis1 model (Figure 7A) resulted in a Ni(II) ion in a nearly ideal square planar coordination geometry [root mean square deviation from the ideal geometry ( $\text{rmsd}_{\text{ig}}$ ) = 0.029 Å]. The plane formed by the nickel ligands is perpendicular to the RrCooT dimerization interface, while the metal binding site is fully exposed to the solvent. In the Amino-Cis2 model (Figure 7B), the orientation of the metal binding region appears to be similar to that in the Amino-Cis1 model, but the Ni(II) coordination geometry is strongly distorted ( $\text{rmsd}_{\text{ig}}$  = 0.768 Å). Finally, the Amino-Trans model (Figure 7C) resulted again in nearly ideal square planar metal coordination geometry ( $\text{rmsd}_{\text{ig}}$  = 0.040 Å), and a different disposition of the plane formed by the nickel ion and the four ligands. Indeed, in the Amino-Trans model the plane formed by the Ni ligands is parallel to the dimer's major axis and the Ni(II) ion is exposed to the solvent only on one side. In both Amino-Cis1 and Amino-Cis2 models, the side chains of both Met1<sub>A</sub> and Met1<sub>B</sub> are buried inside the protein, while in the case of model Amino-Trans model, the Met1 side chain from both monomers is exposed to the solvent. The Amide-Cis1 model (Figure 7D) resulted in a slightly distorted square planar metal coordination geometry ( $\text{rmsd}_{\text{ig}}$  = 0.192 Å). As in all previous Cis models, the plane formed by the nickel ion and its ligands is perpendicular to the RrCooT dimerization interface and the metal binding site is fully exposed to the solvent. Also, in the case of the Amide-Cis2 model (Figure 7E), the metal-



**Figure 7.** Detail of the modelled Ni(II) binding site of RrCooT in the case of model Amino-Cis1 (A), Amino-Cis2 (B), Amino-Trans (C), Amide-Cis1 (D), Amide-Cis2 (E), and Amide-Trans (F) (see Scheme 1). The ribbons of RrCooT monomers are colored in blue and light blue, while the Ni(II) ion is reported as a green sphere and Met1 and Cys2 from each monomer are depicted as sticks colored accordingly to the atom type. The metal binding site in panels C and F have been rotated by 90° degrees around the vertical axis with respect to the representations in the other panels. See also Figures S4 and S5 for further details.

binding region is in a similar orientation with respect to the Amide-Cis1 model, but the Ni(II) coordination geometry is strongly distorted ( $\text{rmsd}_{\text{ig}}$  = 0.509 Å) and the metal ion is less exposed to the solvent. Finally, the Amide-Trans model (Figure 7F) resulted in a metal binding region oriented in a similar arrangement with respect to the Amino-Trans model. The plane defined by the metal ligands is parallel to the dimer's major axis and the Ni(II) ion is exposed to the solvent only on one side. The coordination geometry is only slightly distorted with respect to the ideal geometry ( $\text{rmsd}_{\text{ig}}$  = 0.121 Å). As in the previous set of amino models, in both Amide-Cis1 and Amide-Cis2 models the side chains of Met1<sub>A</sub> and Met1<sub>B</sub> are buried inside the protein, while in the case of Amide-Trans model the side chain of Met1 from both monomers lies on the dimer surface. Considering the distortion of the metal binding geometry achieved in both model Amino-Cis2 (Figure 7B) and Amide-Cis2 (Figure 7E), these models were not considered in the following refinement of the structural parameters extracted from the EXAFS data.

## EXAFS structural refinement

Theoretical EXAFS signals were thus simulated for structure-optimized square planar Ni<sub>2</sub>S<sub>2</sub> clusters considering the Amino-Cis1 (Figure 7A) Amino-Trans (Figure 7C), Amide-Cis1 (Figure 7D) Amide-Trans (Figure 7F) models. For simulations of the EXAFS spectra, a shell of 4.5 Å around the Ni absorber was considered. This approach resulted in 20-atoms clusters that included the backbone atoms of Met1 and Cys2, together with the side chain of Cys2. None of the optimized models showed asymmetric Ni-N distances for the two N donors, as instead suggested by the first-shell EXAFS analysis. Consistently, the corresponding simulated EXAFS spectra (Figure 6C), solid black lines) did not reproduce the experimental oscillations, and further



## RESEARCH ARTICLE

refinement was needed. Thus, each of the theoretical EXAFS spectra was fitted to the experimental data in the reciprocal space applying the rigid-body-refinement scheme, i.e. allowing first-shell distances and DW factors to vary, while constraining outer shell atoms to move rigidly following the variation in first-shell distances. Any attempt to assign a single Ni-N distance to the two first-shell Ni-N paths resulted in a significant overestimation of the Ni-N Debye-Waller factors ( $\sigma^2 > 0.02 \text{ \AA}^2$ ); therefore, those fits were rejected, and two distinct Ni-N distances were introduced. Upon refinement, the theoretical signal generated from the Amide-Trans model (Figure 7F) provided the best agreement with the experimental data (fit number 3 in Table 1, red curve in Figure 6C), whereas the other three models failed to reproduce the data (either the fit did not converge, or it provided unphysical parameters).

## DISCUSSION

In proteins, the most common nickel-coordinating amino acid is histidine while the distribution of cysteine is much less important.<sup>[30]</sup> However, cysteine is known to be a special case among amino acids because of its unique properties due to its nucleophilicity, its high affinity for metals and its ability to form disulphide bridges. With its tuneable chemical reactivity, this amino acid plays crucial roles as Ni ligands in active sites of redox Ni enzymes, such as CODH, ACS or [NiFe]-hydrogenases to lower the potential of the reducing couple thanks to the high reactivity of its thiol group. Instead, histidines are preferred to coordinate Ni(II) ions in non-redox proteins.<sup>[16, 20]</sup> However, cysteine-rich coordinating motifs, known to be excellent nickel binders, can be also found in some nickel chaperones, such as HypB<sup>[18]</sup> or HspA from *H. pylori*.<sup>[34]</sup> In the latter case, histidines are also present, suggesting a mixed His/Cys nickel-binding pattern. Interestingly, cysteine exhibits extreme patterns of conservation, being either highly conserved or completely degenerated. When cysteines are conserved, the degree of conservation is typically above 90%, highlighting its essential role. Another feature of cysteine is its tendency to clustering with other cysteines, which has been highly favoured during evolution.

The case of *RrCooT* and its homologues is somehow intriguing: not only this nickel chaperone possesses neither cysteine-rich region nor nearby histidines in the sequence, but also the protein dimerization forms a solvent-exposed nickel-binding site with two cysteine residues via a single cysteine motif. In this way, the homo-dimerization allows the clustering of cysteines from two different peptide chains. Moreover, the exposure of isolated cysteines is associated with a considerable increase in its reactivity, which is generally a drawback for experimental analyses.<sup>[20]</sup> In *RrCooT*, the proximity of the two Cys2 leads, upon oxygen exposure, to the rapid formation of a disulphide bridge with the formation of a homogenous covalent dimer, as shown by ESI-MS. While this phenomenon prevents Ni(II) binding, it protects the protein from intermolecular disulphide bridges formation, multimerization and even precipitation. Moreover, as shown by NMR, the phenomenon is fully reversible by addition of TCEP, recovering the ability of *RrCooT* to bind Ni(II). Since *RrCooT* is involved in anaerobic metabolism, it is likely to easily bind Ni(II) ions present in its natural environment.

Focusing on the nature of the *RrCooT* ligands for Ni(II), the present study strongly indicates that this metal ion binds two S and two N from the two proximal Cys2 side chains and backbones, respectively. The arrangement of the dimer, as supported by a combination of experimental and modelling studies, is such that the two cysteine side chains are in *trans* position with respect to the metal ion.

Although amidate groups are unusual biological ligands,<sup>[32]</sup> Ni(II) is able to induce deprotonation and metal ion coordination of the peptide amide N atoms.<sup>[33]</sup> Among the Ni(II)-N<sub>2</sub>S<sub>2</sub> moieties found in proteins, the redox enzyme NiSOD harbours a Ni(II) ions coordinated by the two thiolates of Cys2 and Cys6, the terminal amine and the amide group of Cys2 in its reduced state<sup>[17]</sup>. Several NiSOD model peptides and complexes have been synthesized in which the N-donor ligands were varied from the natural mixed amine/amidate environment to a bis-amidate one, leading exclusively to catalytically inactive species. One reason would be that bis-amidate complexes are considered to be more reactive towards O<sub>2</sub>.<sup>[34]</sup> However, the effect of bis-amidate vs mixed amine/amidate environment on the oxygen stability of NiSOD model peptides and NiN<sub>2</sub>S<sub>2</sub> complexes is controversial.<sup>[35]</sup> A bis-amidate coordination has also been observed in the Ala0-NiSOD mutant<sup>[23]</sup>, which possesses an Ala N-terminal extension, resulting in a S → Ni LMCT red-shift transition from 380 nm for the wild-type enzyme to 457 nm, this reddish color being similar to the one observed with Ni-*RrCooT*. In the mutant form, the terminal amine coordination observed in WT-NiSOD is replaced by the amide group of His1, resulting in the abolition of the enzyme activity and an increase in air sensitivity with the appearance of disulphide and S-oxygenated species over two weeks.

In biology, the only example of a Ni(II) centre with a bis-amidate/bis-thiolate environment has been identified in the active site of acetyl-coenzyme A synthase.<sup>[36]</sup> This enzyme possesses a dinuclear Ni active site, with one distal nickel (Ni<sub>d</sub>), coordinated by an unusual Cys-Gly-Cys motif in a square planar geometry involving the thiolates of both cysteines and the amide N atoms of a cysteine and a glycine. The same cysteines also coordinate the proximal nickel (Ni<sub>p</sub>), which is the catalytic centre of the enzyme. Ni<sub>d</sub> is postulated not to be involved in redox chemistry and to remain in the Ni(II) state.<sup>[37]</sup> Its role has been proposed either as a supporting ligand to stabilize the proximal nickel in a low-valent redox level<sup>[22, 38]</sup> or for the generation of an open site on the catalytic metal thanks to its hemilabile ring-opening property.<sup>[38-39]</sup> These hypotheses arise from the investigations on Ni-N<sub>2</sub>S<sub>2</sub> complexes that have been shown to be i) good electron donors and ii) able to bind as mono- or bidentate ligands, a property fundamental to hemilability.<sup>[39-40]</sup> Among the library of monomeric square planar Ni(II)-N<sub>2</sub>S<sub>2</sub> complexes with either amino- or amidate N donors, tetra-anionic amido-thiols favor the reduction of Ni(II) to Ni(I) at highly negative potentials while biologically-relevant potentials would stabilize the Ni(II)/Ni(III) couple.<sup>[38]</sup> All these examples show that the presence of N donors as well as their nature tune the chemical properties of the metal centre.

It is worth mentioning that *CooT* likely works in concert with its partners, with the ultimate goal of transferring Ni(II) to CODH. Moreover, the existence of the two distinct metal-binding motifs, namely "Cys2" and "Cys2-His55" raises the question of their physiological role. While the "Cys2" motif in *RrCooT* allows Ni(II) binding with an affinity in the nanomolar range, *ChCooT* possessing the predicted "Cys2His55" motif binds Ni(II) in the micromolar range,<sup>[19]</sup> possibly hinting to distinct functions *in vivo*. In *R. rubrum*, *CooT* has been experimentally shown to be involved in CODH maturation. Therefore, despite a high affinity for Ni(II), the metal release is required and could occur via the interaction with *CooJ*, *CooC* or/and CODH. The solvent exposure of the nickel-binding site in *RrCooT* makes it possible to consider that the cooperation with its partner(s) would induce a change in coordination and affinity, facilitating Ni(II) release. However, the precise role of *RrCooT* has not been fully demonstrated and to date, no information is available regarding the role of other *CooT* homologues.

## RESEARCH ARTICLE

To conclude, a single cysteine motif for Ni(II) makes *RrCooT* a unique example among Ni chaperones. The binding of a mononuclear Ni(II) ion with a  $N_2S_2$  environment via a bis-thiolate/bis-amidate coordination is unprecedented and its role remains to be clarified. It is not fully understood how the  $N_2S_2$  ligands contribute to the physiological role of *RrCooT*.

## Experimental Section

The details of protein production and purification, CD, UV-vis and NMR spectroscopies as well as ITC experiments are provided as Supplementary Information.

**Molecular Modelling:** The Ni(II) bound form of *RrCooT* was generated using Modeller v9.18,<sup>[41]</sup> and the structure of apo-*RrCooT* (PDB id: 5N76) as template.<sup>[15]</sup> A single Ni(II) ion was included in the model, in agreement with metal content analyses and ITC experiments previously reported.<sup>[15]</sup> The van der Waals parameters for Ni(II) were derived from the Zn(II) parameters included in the CHARMM22 force field,<sup>[32]</sup> implemented in the Modeller v9.18 package, by applying a scale factor (1.12) calculated on the basis of the ionic radius of the two divalent ions. The best structure among the 100 generated models was selected on the basis of the lowest value of the DOPE score included in Modeller v9.18,<sup>[42]</sup> and the analysis performed with PROCHECK.<sup>[43]</sup> A loop optimization routine was used to refine the metal binding region. In all modelling calculations, constraints were imposed using a Gaussian-shaped energy potential for distances, angles, and dihedrals in order to correctly position the Ni(II) ions with respect to the tested list of ligated atoms (Table S4). The best model was selected on the basis of the lowest value of the DOPE score. The molecular graphics, together with analysis of the Ni(II) sites, were performed using the UCSF Chimera package.<sup>[44]</sup>

**X-ray Absorption Spectroscopy:** Ni K-edge X-ray absorption spectroscopy (XAS) was performed at the beamline CRG-FAME-BM30B of the European Synchrotron Radiation Facility (ESRF, Grenoble, France).<sup>[45]</sup> Apo-*RrCooT*-WT was concentrated to 1.5 mM and incubated with 0.9 molar eq. of  $NiSO_4$  for 10 minutes before addition of 10 % glycerol as cryoprotectant. Drops of the solution (~50  $\mu$ L) were deposited on the  $\varnothing$ 5mm holes of the sample holder equipped with kapton windows, immediately frozen in liquid nitrogen, then transferred into the liquid He-cryostat and measured at 15 K in order to limit radiation damages. Depending on the Ni concentration, 5 to 10 spectra were acquired for the same sample, then merged. The position of the beam (size:  $205 \times 120 \mu m^2$ , FWHM) was changed for each acquisition in order to limit the effects of X-ray irradiation. The Ni K-edge energy region was scanned in the 8100-9300 eV range with a liquid nitrogen-cooled Si(220) monochromator. The photon energy was calibrated using a metallic Ni foil, by defining the first inflection point of its absorption spectrum at 8333.0 eV. X-ray absorption spectra were recorded in fluorescence mode with a 30-elements Ge solid-state detector (Canberra).

XAS data reduction and normalization was performed with standard methods using the Athena software.<sup>[46]</sup> Peak fitting in the XANES region was performed with the Fityk program, using a Levenberg-Marquardt minimization algorithm.<sup>[47]</sup> A detailed description of the data analysis protocol is reported in the SI.

## Acknowledgements

This work was supported by "the ITERLIS PhD program, CEA Life sciences" for MA's PhD funding, the "FUNBIOCO" project (IDEX-UGA, Initiatives de Recherche stratégiques) and the "COSYNBIO" project (Projets exploratoires, Cellule energie-CNRS). This work has been partially supported by Labex ARCANÉ and CBH-EUR-GS (ANR-17-EURE-0003). The research leading to these results has received funding from the networking support from the COST Action FeSBioNet (Contract CA15133). This work used the platforms of the Grenoble Instruct-ERIC Center (ISBG: UMS3518CNRS-CEA-UGA-EMBL) with support from FRISBI (ANR-10-INBS-05-02) and GRAL (ANR-10-LABX-49-01) within the Grenoble Partnership for Structural Biology (PSB). The authors gratefully acknowledge the European Synchrotron Radiation Facility program committee (Grenoble, France) for provision of beamtime (LS-2473) and the CRG-FAME for the In-house-Research beamtime (IH-LS-2796) and support of their staff. The authors acknowledge the support and the use of resources of Instruct-ERIC. This work was supported by the Department of Pharmacy and Bio- technology of the University of Bologna (SC, BZ, FM). The NMR experiments were partially obtained in the frames of access to NMR infrastructure by EuroBioNMR EEIG (<http://www.eurobionmr.eu/>). The Center for Magnetic Resonance of the University of Florence (CERM) provided access to the high-field NMR spectrometers. Fabio Calogiuri and Massimo Lucci are acknowledged for spectral data collection.

**Keywords:** Carbon monoxide dehydrogenase maturation • Ni chaperone protein •  $Ni(II)N_2S_2$  coordination • CooT • Nickel-binding cysteine

## References

- [1] Y. Sheng, I. A. Abreu, D. E. Cabelli, M. J. Maroney, A.-F. Miller, M. Teixeira, J. S. Valentine, *Chem Rev* **2014**, *114*, 3854-3918.
- [2] M. J. Maroney, S. Ciurli, *Chem. Rev.* **2014**, *114*, 4206-4228.
- [3] P. N. Evans, J. A. Boyd, A. O. Leu, B. J. Woodcroft, D. H. Parks, P. Hugenoltz, G. W. Tyson, *Nature Reviews Microbiology* **2019**, *17*, 219-232.
- [4] M. Can, F. A. Armstrong, S. W. Ragsdale, *Chem Rev* **2014**, *114*, 4149-4174.
- [5] M. Alfano, C. Cavazza, *Sustainable Energy & Fuels* **2018**, *2*, 1653-1670.
- [6] R. L. Kerby, P. W. Ludden, G. P. Roberts, *Journal of bacteriology* **1995**, *177*, 2241-2244.
- [7] J. D. Fox, R. L. Kerby, G. P. Roberts, P. W. Ludden, *J. Bacteriol.* **1996**, *178*, 1515-1524.
- [8] C. L. Drennan, J. Heo, M. D. Sintchak, E. Schreiter, P. W. Ludden, *Proceedings of the National Academy of Sciences* **2001**, *98*, 11973-11978.
- [9] S. W. Singer, M. B. Hirst, P. W. Ludden, *Biochim Biophys Acta* **2006**, *1757*, 1582-1591.
- [10] D. Bonam, M. C. McKenna, P. J. Stephens, P. W. Ludden, *Proceedings of the National Academy of Sciences* **1988**, *85*, 31-35.
- [11] Y. Li, D. B. Zamble, *Chem Rev* **2009**, *109*, 4617-4643.
- [12] R. L. Kerby, P. W. Ludden, G. P. Roberts, *J. Bacteriol.* **1997**, *179*, 2259-2266.
- [13] W. B. Jeon, J. Cheng, P. W. Ludden, *J Biol Chem* **2001**, *276*, 38602-38609.
- [14] M. Alfano, J. Pérard, P. Carpentier, C. Basset, B. Zambelli, J. Timm, S. Crouzy, S. Ciurli, C. Cavazza, *J. Biol. Chem.* **2019**.
- [15] J. Timm, C. Brochier-Armanet, J. Perard, B. Zambelli, S. Ollagnier-de-Choudens, S. Ciurli, C. Cavazza, *Metallomics* **2017**, *9*, 575-583.
- [16] S. M. Marino, V. N. Gladyshev, *Journal of molecular biology* **2010**, *404*, 902-916.
- [17] D. P. Barondeau, C. J. Kassmann, C. K. Bruns, J. A. Tainer, E. D. Getzoff, *Biochemistry* **2004**, *43*, 8038-8047.
- [18] M. R. Leach, S. Sandal, H. Sun, D. B. Zamble, *Biochemistry* **2005**, *44*, 12229-12238.

## RESEARCH ARTICLE

- [19] M. Alfano, J. Perard, R. Miras, P. Catty, C. Cavazza, *J Biol Inorg Chem* **2018**, 23, 809-817.
- [20] L. B. Poole, *Free Radical Biology and Medicine* **2015**, 80, 148-157.
- [21] J. Jeong, Y. Jung, S. Na, J. Jeong, E. Lee, M.-S. Kim, S. Choi, D.-H. Shin, E. Paek, H.-Y. Lee, K.-J. Lee, *Molecular & Cellular Proteomics* **2011**, 10.
- [22] R. Krishnan, C. G. Riordan, *Journal of the American Chemical Society* **2004**, 126, 4484-4485.
- [23] H.-T. Huang, S. Dillon, K. C. Ryan, J. O. Campecino, O. E. Watkins, D. E. Cabelli, T. C. Brunold, M. J. Maroney, *Inorg. Chem.* **2018**, 57, 12521-12535.
- [24] N. Lihi, G. Csire, B. Szakács, N. V. May, K. Várnagy, I. Sóvágó, I. Fábrián, *Inorganic Chemistry* **2019**, 58, 1414-1424.
- [25] Y. Shen, F. Delaglio, G. Cornilescu, A. Bax, *J Biomol NMR* **2009**, 44, 213-223.
- [26] L. Wang, H. R. Eghbalnia, A. Bahrami, J. L. Markley, *J Biomol NMR* **2005**, 32, 13-22.
- [27] G. J. Colpas, M. J. Maroney, C. Bagyinka, M. Kumar, W. S. Willis, S. L. Suib, N. Baidya, P. K. Mascharak, *Inorg. Chem.* **1991**, 30, 920-928.
- [28] a Y. Oudart, V. Artero, J. Pécaut, C. Lebrun, M. Fontecave, *European Journal of Inorganic Chemistry* **2007**, 2007, 2613-2626; b Johanna A. W. Verhagen, M. Lutz, Anthony L. Spek, E. Bouwman, *European Journal of Inorganic Chemistry* **2003**, 2003, 3968-3974; c H. J. Kruger, G. Peng, R. H. Holm, *Inorganic Chemistry* **1991**, 30, 734-742.
- [29] V. Martin-Diaconescu, M. Bellucci, F. Musiani, S. Ciurli, M. J. Maroney, *J Biol Inorg Chem* **2012**, 17, 353-361.
- [30] R. J. Sudan, J. L. Kumari, C. Sudandiradoss, *PLoS One* **2015**, 10, e0126787.
- [31] K. Schauer, C. Muller, M. Carrière, A. Labigne, C. Cavazza, H. De Reuse, *Journal of bacteriology* **2010**, 192, 1231.
- [32] A. MacKerell, D. Bashford, M. Bellot, R. Dunbrack, J. Evanseck, M. Field, *J. Phys. Chem. B* **1998**, 102, 3586-3616.
- [33] D. Zambale, M. Rowińska-Zyrek, H. Kozłowski, in *Metallobiology*, Vol. 10 (Ed.: C. D. Garner), Royal Society of Chemistry, Cambridge (UK), **2017**.
- [34] a V. Mathrubootham, J. Thomas, R. Staples, J. McCracken, J. Shearer, E. L. Hegg, *Inorganic Chemistry* **2010**, 49, 5393-5406; b K. P. Neupane, J. Shearer, *Inorganic Chemistry* **2006**, 45, 10552-10566.
- [35] D. Tietze, J. Sartorius, B. Koley Seth, K. Herr, P. Heimer, D. Imhof, D. Mollenhauer, G. Buntkowsky, *Scientific reports* **2017**, 7, 17194-17194.
- [36] C. Darnault, A. Volbeda, E. J. Kim, P. Legrand, X. Vernède, P. A. Lindahl, J. C. Fontecilla-Camps, *Nature Structural & Molecular Biology* **2003**, 10, 271-279.
- [37] E. L. Maynard, C. Sewell, P. A. Lindahl, *Journal of the American Chemical Society* **2001**, 123, 4697-4703.
- [38] M. L. Golden, M. V. Rampersad, J. H. Reibenspies, M. Y. Darensbourg, *Chemical Communications* **2003**, 1824-1825.
- [39] K. N. Green, S. P. Jeffery, J. H. Reibenspies, M. Y. Darensbourg, *Journal of the American Chemical Society* **2006**, 128, 6493-6498.
- [40] M. L. Golden, C. M. Whaley, M. V. Rampersad, J. H. Reibenspies, R. D. Hancock, M. Y. Darensbourg, *Inorganic Chemistry* **2005**, 44, 875-883.
- [41] M. A. Marti-Renom, A. C. Stuart, A. Fiser, R. Sanchez, F. Melo, A. Sali, *Annu. Rev. Biophys. Biomol. Struct.* **2000**, 29, 291-325.
- [42] M.-y. Shen, A. Sali, *Protein Science* **2006**, 15, 2507-2524.
- [43] R. A. Laskowski, M. W. MacArthur, D. S. Moss, J. M. Thornton, *J. Appl. Crystallograph.* **1993**, 26, 283-291.
- [44] E. F. Pettersen, T. D. Goddard, C. C. Huang, G. S. Couch, D. M. Greenblatt, E. C. Meng, T. E. Ferrin, *J. Comput. Chem.* **2004**, 25, 1605-1612.
- [45] O. Proux, X. Biquard, E. Lahera, J. J. Menthonnex, A. Prat, O. Ulrich, Y. Soldo, P. Trivison, G. Kapoujyan, G. Perroux, P. Taunier, D. Grand, P. Jeantet, M. Deleglise, J. P. Roux, J. L. Hazemann, *Physica Scripta* **2005**, 970.
- [46] B. Ravel, M. Newville, *J. Synchrotron Radiat.* **2005**, 12, 537-541.
- [47] M. Wojdyr, *J. Appl. Cryst.* **2010**, 43, 1126-1128.

SUPPLEMENTARY INFORMATION FOR

**A solvent-exposed cysteine forms a peculiar Ni(II)-binding site in the  
metallochaperone CooT from *Rhodospirillum rubrum***

Marila Alfano,<sup>[a]</sup> Giulia Veronesi,<sup>[a]</sup> Francesco Musiani,<sup>[b]</sup> Barbara Zambelli,<sup>[b]</sup> L. Signor,<sup>[c]</sup> O.

Proux,<sup>[d]</sup> M. Rovezzi,<sup>[d]</sup> Stefano Ciurli\*<sup>[b]</sup>, Christine Cavazza\*<sup>[a]</sup>

<sup>[a]</sup> University of Grenoble Alpes, CEA, CNRS, IRIG, CBM, F-38000 Grenoble, France

<sup>[b]</sup> Laboratory of Bioinorganic Chemistry, Department of Pharmacy and Biotechnology,  
University of Bologna, Via Giuseppe Fanin 40, I-40127 Bologna (Italy)

<sup>[c]</sup> University of Grenoble Alpes, CEA, CNRS, IRIG, IBS, F-38000 Grenoble, France.

<sup>[d]</sup> Univ. Grenoble Alpes, CNRS, IRD, Irstea, Météo France, OSUG, FAME, 38000  
Grenoble, France

\* Corresponding authors:

[christine.cavazza@cea.fr](mailto:christine.cavazza@cea.fr), ORCID: 0000-0002-3657-1302

[stefano.ciurli@unibo.it](mailto:stefano.ciurli@unibo.it), ORCID: 0000-0001-9557-926X



### *Protein production and purification*

*RrCooT* was produced and purified as previously described.<sup>[1]</sup> Two additional constructs were produced by exploiting the natural processing of the N-terminal methionine in *E. coli*. The two constructs, namely *RrCooT*-Gly1 and *RrCooT*-Cys1, were produced in Terrific Broth (TB) medium, supplemented with the appropriate antibiotics, under stirring at 37 °C. The expression was induced using 0.5 mM of IPTG once the culture reached an OD<sub>600</sub> of 0.5-0.6. The cultures were then left at 37 °C and 180 rpm for 4 hours before harvesting by centrifugation. The pellet was re-suspended in 50 mM Tris-HCl at pH 8.5, containing 1 mM EDTA and a protease inhibitor tablet “complete” by Roche, and frozen at -80 °C. The production of the single (<sup>15</sup>N) and double (<sup>13</sup>C, <sup>15</sup>N) labeled proteins used for NMR spectroscopy was achieved using a medium containing 3 g L<sup>-1</sup> Na<sub>2</sub>HPO<sub>4</sub>, g L<sup>-1</sup> KH<sub>2</sub>PO<sub>4</sub>, 0.5 g L<sup>-1</sup> NaCl, 1 g L<sup>-1</sup> <sup>15</sup>N ammonium sulfate, 2 g L<sup>-1</sup> glucose (or <sup>13</sup>C-glucose), 2 mg mL<sup>-1</sup> thiamine, 0.1 mg mL<sup>-1</sup> CaCl<sub>2</sub> and 1 mg mL<sup>-1</sup> MgSO<sub>4</sub>. The cultures were supplemented with the required antibiotic and grown at 37 °C and 180 rpm until an OD<sub>600</sub> of 0.5-0.6 was obtained. *RrcooT*-WT (wild-type) expression was induced by addition of 0.5 mM of IPTG and the cultures were left overnight at 25 °C and 200 rpm before harvesting by centrifugation. All *RrCooT* samples were purified as previously described.<sup>[1]</sup> The protein size and identity, as well as the presence/absence of the N-terminal methionine, were determined by electrospray ionization mass spectrometry (ESI-MS).

### *LC/ESI Mass spectrometry*

Liquid Chromatography Electrospray Ionization Mass Spectrometry (LC/ESI-MS) was performed on a 6210 LC/ESI-TOF mass spectrometer interfaced with an HPLC binary pump system (Agilent Technologies). The mass spectrometer was calibrated with standard calibrants (ESI-L, Low concentration tuning mix, Agilent Technologies) before measurements and mass spectra were recorded in the 300-3200 *m/z* range in the positive ion mode. All solvents used were HPLC grade (Chromasolv, Sigma-Aldrich), trifluoroacetic acid (TFA) was from Acros Organics (puriss., p.a.). Solvent A was 0.03% TFA in water, solvent B was 95% acetonitrile-5% water-0.03% TFA. Just before analysis the protein samples were diluted in acidic denaturing conditions to a final concentration of 5 μM with solution A (0.03% TFA in water). Samples were thermostated at 10°C in the autosampler and the analysis was run by injecting 4 μL of each sample. They were first trapped and desalted on a reverse phase-C8 cartridge (Zorbax 300SB-C8, 5μm, 0.3mm IDx5mm, Agilent Technologies) for 3 minutes at a flow rate of 50 ul/min with 100% solvent A and then eluted with 70% solvent B at flow rate of 50 ul/min for MS detection. The RP-C8 cartridge was then re-equilibrated for 4 min with 100% solvent A at a flow rate of 50 ul/min. MS spectra were acquired

and the data processed with MassHunter workstation software (v. B.02.00, Agilent Technologies) and with GPMAW software (v. 7.00b2, Lighthouse Data, Denmark).

#### *CD and UV-visible spectroscopy.*

The CD spectra were recorded with a JASCO J-1500 circular dichroism spectrometer (JASCO Analytical Instruments). The Ni(II) binding ability of the different constructs was investigated under anaerobic conditions. The proteins were thawed and diluted to reach the final concentration of 25  $\mu$ M dimer in 50 mM HEPES at pH 7.5. NiSO<sub>4</sub>, previously degassed in glovebox, was added. The spectra of the proteins with the addition of 1 molar equivalent of NiSO<sub>4</sub> were recorded from 250 to 700 nm using 10 accumulations and a 10-mm cuvette. The UV-visible spectra were recorded on the same samples, at 500  $\mu$ M dimer for *RrCooT* and 630  $\mu$ M dimer for *RrCooT-Gly1*.

#### *Isothermal Titration Calorimetry*

Ni(II) binding titrations of *RrCooT-Gly1* and *RrCooT-Cys1* were performed at 25 °C using a high-sensitivity VP-ITC microcalorimeter (MicroCal, Northampton, MA, USA) according to the protocol previously described for the wild type protein.<sup>[1]</sup> The raw data were processed using the NITPIC software,<sup>[2]</sup> and the integrated data were fit with the Origin software package (MicroCal) using a nonlinear least-squares minimization algorithm to theoretical titration curves with different binding schemes. The values for the enthalpy change of the reaction ( $\Delta H$ ), the binding affinity constant ( $K_A$ ), and the number of sites ( $n$ ) were parameters of the fit. The reaction entropy was calculated using  $\Delta G = -RT \ln K_A$  ( $R = 1.9872 \text{ cal mol}^{-1} \text{ K}^{-1}$ ,  $T = 298 \text{ K}$ ) and  $\Delta G = \Delta H - T\Delta S$ . The values given for  $\Delta H$  and  $\Delta S$  are apparent values and include contributions not only from metal binding but also from associated events such as protonation/deprotonation of the amino acid residues involved in binding and the consequent change in the buffer ionization state.

#### *NMR spectroscopy data collection and analysis for backbone assignment of apo-*RrCooT**

All NMR spectra were acquired at 25 °C using a Bruker AVANCE III spectrometer operating at the <sup>1</sup>H nominal frequency of 950.20 MHz (22.3 T) and a Bruker DRX spectrometer, operating at the <sup>1</sup>H nominal frequency of 500.13 MHz (11.7 T). The spectrometers were equipped with 5 mm TCI <sup>1</sup>H-<sup>13</sup>C/<sup>15</sup>N/D Z-GRD triple resonance cryoprobes with shielded pulsed field gradients along the z-axis. A typical sample contained 0.7-1.1 mM of <sup>13</sup>C and <sup>15</sup>N labeled apo-*RrCooT* dimer in 25 mM HEPES buffer at pH 7.0, containing 5 mM NaCl and 1 mM TCEP, in 90% H<sub>2</sub>O and 10% D<sub>2</sub>O. kDa), as shown

by light. The experiments used to obtain the sequential backbone resonance assignment of  $^1\text{H}$ ,  $^{13}\text{C}$ , and  $^{15}\text{N}$  nuclei for apo-*RrCooT* are reported in Table S1. In these pulse schemes, water suppression is achieved using watergate with pulsed field gradients associated with a flip-back pulse. The spectra were processed using NMRpipe.<sup>[3]</sup> Spectra recorded in non-uniform sampling (NUS) mode were reconstructed using an implementation of a real channel approach to iterative soft thresholding (IST) routines,<sup>[4]</sup> using NMRPipe.<sup>[3]</sup> Spectral analysis for resonance assignment of apo-*RrCooT* was performed using CARA 1.9.1.1 (<http://cara.nmr.ch>).<sup>[5]</sup> The assignment for apo- *RrCooT* is reported in Table S2.

Apo-*RrCooT* interaction with Ni(II) followed by NMR: The binding of apo-*RrCooT* to Ni(II) was investigated at 25 °C by recording  $^1\text{H}$ , $^{15}\text{N}$  HSQC spectra on samples obtained by adding a  $\text{NiSO}_4$  solution in buffer A to a 10 mg/mL sample of  $^{15}\text{N}$ -labeled apo-*RrCooT* in the same buffer at 0.00, 0.25, 0.50, 1.00, and 2.00 equivalents of metal ion per *RrCooT* dimer, at 950 MHz. The assignment of the backbone nuclei  $^1\text{H}$ ,  $^{15}\text{N}$ ,  $\text{C}\alpha$  and  $\text{C}\beta$  chemical shifts in holo-*RrCooT* (Table S3) was carried using a combination of closest neighbor approach with the 950 MHz  $^1\text{H}$ , $^{15}\text{N}$  HSQC spectrum of apo-*RrCooT* as reference, together with CBCANH/CBCA(CO)NH triple resonance experiments at 500 MHz. The information about the residue-specific changes within *RrCooT*, induced by Ni(II) binding, was obtained by calculating a combined chemical shift perturbation (CSP) index,  $\Delta\delta_{\text{comb}}$ , which comprises the information of the HN, N,  $\text{C}\alpha$  and  $\text{C}\beta$  nuclei each amino acid at position  $j$  in the protein sequence using the formula given below:

$$\Delta\delta_{\text{comb},j} = \sqrt{\sum_{i=1}^4 (w_i \Delta\delta_{ji})^2}$$

This method uses the normalized length of a vector  $E_j$  (Euclidean distance), with its components  $E_{ji}$  defined by the chemical shift differences  $\Delta\delta_{ji}$  for the atoms  $i$  at a specific position  $j$ , with a weighting factor  $w_i$  that accounts for differences in sensitivity of different resonances in an amino acid:

$$w_i = \frac{|\gamma_i|}{|\gamma_H|}$$

where  $\gamma_i$  and  $\gamma_H$  are the magnetogyric ratio of nucleus  $i$  and of the proton, respectively. For  $^1\text{H}$ ,  $^{15}\text{N}$  and  $^{13}\text{C}$  the weighting factors are then 1.000, 0.102, 0.251.<sup>[6]</sup>

### *X-ray Absorption Spectroscopy*

A two-components model, entailing an arctangent curve for background subtraction and a Gaussian curve, was used to reproduce the  $1s \rightarrow 3d$  transition in the range 8320-8336 eV of normalized XANES

spectra. Six free parameters (three for the arctangent function and three for the Gaussian curve) were considered. The coefficient of determination R-squared, used to estimate the goodness-of-fit, was higher than 99.90% for all fits.

Scattering amplitudes and phase shifts for *ab initio* EXAFS refinement were calculated using the code FEFF9, [7] by making use of Self Consistent Field (SCF) calculations. Input clusters for the *ab initio* calculations were built starting from optimized model structures that included atoms from the protein backbone and amino-acid side chains found within 4.5 Å from the Ni absorber. This resulted in input clusters of 20 atoms around the Ni absorber, corresponding to the backbone of Met1 and Cys2 as well as the side chain of Cys2, for both monomers. The amplitude reduction factor  $S_0^2$  was calculated by the FEFF9 program from atomic overlap integrals, and the value of 0.94 was kept constant during the fit. The fits to the experimental data were performed using the Artemis program, minimizing the R-factor with Levenberg-Marquardt non-linear least-squares minimization algorithm.[8] Firstly, only the first shell peak of the Fourier transformed signal was fitted in the real space, in the range 1.0-2.2 Å, providing an estimation of first-shell distances and Debye-Waller factors. Secondly, the  $k^n$ -weighted data were fitted directly in k-space in the range 2.8 - 11 Å<sup>-1</sup>, with multiple k-weights (n = 1, 2, 3). The rigid-body-refinement scheme was applied,[9] in which only first-shell distances were allowed to vary while outer shell atoms were constrained to move rigidly following the variations in first-shell distances. The floating parameters were the first-shell distances and the Debye-Waller (DW) factors, a common DW factor for all second-shell atoms (Cα and Cβ of Cys2 and carbonyl-C of Met1, for both monomers), and a common shift in the energy origin ( $\Delta E_0$ ).

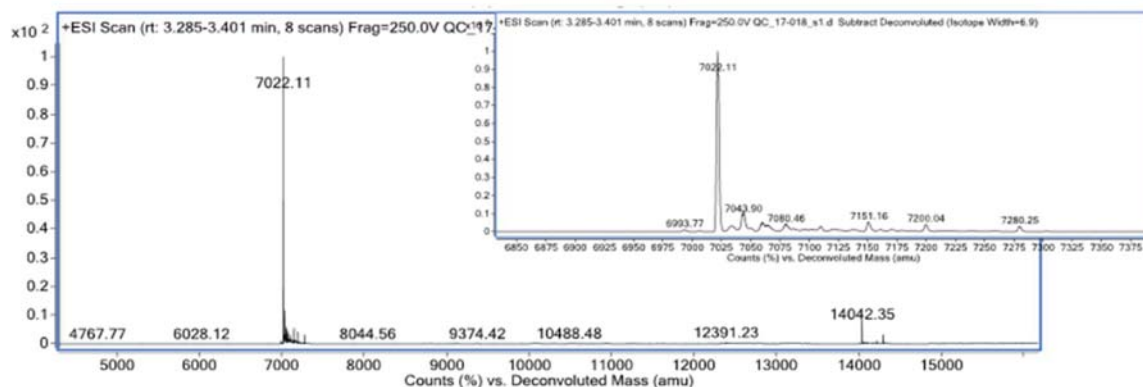


**Figure S1. Clustal 2.1 Multiple sequence alignment of the 64 cooT homologues possessing the “Cys2/His55” motif**

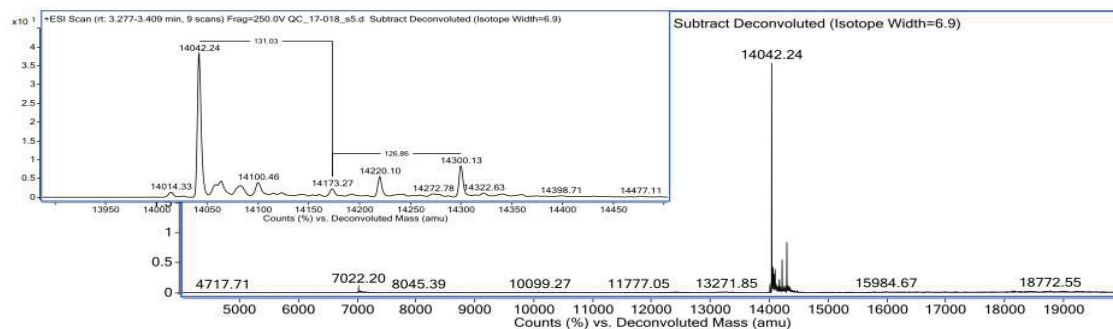
|                                   |  |                                   |                         |
|-----------------------------------|--|-----------------------------------|-------------------------|
| Clostridium-ljungdahlii           | MCESTAYLVTPEGE--R-KIDYVVDIVPKENGKLSLTDILGEEIEE-G | Clostridium-ljungdahlii           | MLKEVRLLEHKIVIEKAM----- |
| Clostridium-autoethanogenum       | MCESTAYLVTPEGE--R-KIDYVVDIVPKENGKLSLTDILGEEIEE-G | Clostridium-autoethanogenum       | MLKEVRLLEHKIVIEKAM----- |
| Clostridium-pasteurianum          | MCESTAYLITPEGE--Y-KIDYVVDIVPKENGKLSLTDILGEEIEE-G | Clostridium-pasteurianum          | MLKEVRLLEHKIVIEKAM----- |
| Eubacterium-limosum               | MCESSAYLVNKGGE--EEKIDYVVDIVPKENGKLSLTDILGEEIEE-G | Eubacterium-limosum               | RLKEVRLLEHKIVIEKAM----- |
| Clostridium-sticklandii           | MCESSAYLVNKGGE--EEKIDYVVDIVPKENGKLSLTDILGEEIEE-G | Clostridium-sticklandii           | RLKEVRLLEHKIVIEKAM----- |
| Alkaliphilus-metallireducens      | MCESSAYLVNKGGE--EEKIDYVVDIVPKENGKLSLTDILGEEIEE-G | Alkaliphilus-metallireducens      | RLKEVRLLEHKIVIEKAM----- |
| Clostridioides-difficile          | MCESSAYLVNKGGE--EEKIDYVVDIVPKENGKLSLTDILGEEIEE-G | Clostridioides-difficile          | RLKEVRLLEHKIVIEKAM----- |
| Tepidanaerobacter-acetatoxydan    | MCESSAYLVNKGGE--EEKIDYVVDIVPKENGKLSLTDILGEEIEE-G | Tepidanaerobacter-acetatoxydan    | RLKEVRLLEHKIVIEKAM----- |
| Thermosediminibacter-oceanii      | MCESSAYLVNKGGE--EEKIDYVVDIVPKENGKLSLTDILGEEIEE-G | Thermosediminibacter-oceanii      | RLKEVRLLEHKIVIEKAM----- |
| Clostridium-sp                    | MCESSAYLVNKGGE--EEKIDYVVDIVPKENGKLSLTDILGEEIEE-G | Clostridium-sp                    | RLKEVRLLEHKIVIEKAM----- |
| Ruminiclostridium-cellulolytic    | MCESSAYLVNKGGE--EEKIDYVVDIVPKENGKLSLTDILGEEIEE-G | Ruminiclostridium-cellulolytic    | RLKEVRLLEHKIVIEKAM----- |
| Clostridium-clariflavum           | MCESSAYLVNKGGE--EEKIDYVVDIVPKENGKLSLTDILGEEIEE-G | Clostridium-clariflavum           | RLKEVRLLEHKIVIEKAM----- |
| Desulfotomaculum-carboxydifera    | MCESSAYLVNKGGE--EEKIDYVVDIVPKENGKLSLTDILGEEIEE-G | Desulfotomaculum-carboxydifera    | RLKEVRLLEHKIVIEKAM----- |
| Thermosediminibacter-oceanii2     | MCESSAYLVNKGGE--EEKIDYVVDIVPKENGKLSLTDILGEEIEE-G | Thermosediminibacter-oceanii2     | RLKEVRLLEHKIVIEKAM----- |
| Candidatus-Desulfurudis           | MCESSAYLVNKGGE--EEKIDYVVDIVPKENGKLSLTDILGEEIEE-G | Candidatus-Desulfurudis           | RLKEVRLLEHKIVIEKAM----- |
| Thermacetogenium-phaeum1          | MCESSAYLVNKGGE--EEKIDYVVDIVPKENGKLSLTDILGEEIEE-G | Thermacetogenium-phaeum1          | RLKEVRLLEHKIVIEKAM----- |
| Desulfotomaculum-kuznetsovii      | MCESSAYLVNKGGE--EEKIDYVVDIVPKENGKLSLTDILGEEIEE-G | Desulfotomaculum-kuznetsovii      | RLKEVRLLEHKIVIEKAM----- |
| Pelotomaculum-thermopropionis     | MCESSAYLVNKGGE--EEKIDYVVDIVPKENGKLSLTDILGEEIEE-G | Pelotomaculum-thermopropionis     | RLKEVRLLEHKIVIEKAM----- |
| Desulfotomaculum-acetoxidans      | MCESSAYLVNKGGE--EEKIDYVVDIVPKENGKLSLTDILGEEIEE-G | Desulfotomaculum-acetoxidans      | RLKEVRLLEHKIVIEKAM----- |
| Desulfotomaculum-gibsoniae        | MCESSAYLVNKGGE--EEKIDYVVDIVPKENGKLSLTDILGEEIEE-G | Desulfotomaculum-gibsoniae        | RLKEVRLLEHKIVIEKAM----- |
| Ammonifex-degensis                | MCESSAYLVNKGGE--EEKIDYVVDIVPKENGKLSLTDILGEEIEE-G | Ammonifex-degensis                | RLKEVRLLEHKIVIEKAM----- |
| Carboxydotherrnus-hydrogenofom    | MCESSAYLVNKGGE--EEKIDYVVDIVPKENGKLSLTDILGEEIEE-G | Carboxydotherrnus-hydrogenofom    | RLKEVRLLEHKIVIEKAM----- |
| Thermicola-potens                 | MCESSAYLVNKGGE--EEKIDYVVDIVPKENGKLSLTDILGEEIEE-G | Thermicola-potens                 | RLKEVRLLEHKIVIEKAM----- |
| Heliohabacterium-modesticaldum    | MCESSAYLVNKGGE--EEKIDYVVDIVPKENGKLSLTDILGEEIEE-G | Heliohabacterium-modesticaldum    | RLKEVRLLEHKIVIEKAM----- |
| Thermacetogenium-phaeum2          | MCESSAYLVNKGGE--EEKIDYVVDIVPKENGKLSLTDILGEEIEE-G | Thermacetogenium-phaeum2          | RLKEVRLLEHKIVIEKAM----- |
| Desulfotomaculum-carboxydifera    | MCESSAYLVNKGGE--EEKIDYVVDIVPKENGKLSLTDILGEEIEE-G | Desulfotomaculum-carboxydifera    | RLKEVRLLEHKIVIEKAM----- |
| Desulfotomaculum-ruminis          | MCESSAYLVNKGGE--EEKIDYVVDIVPKENGKLSLTDILGEEIEE-G | Desulfotomaculum-ruminis          | RLKEVRLLEHKIVIEKAM----- |
| Mahella-australiensis             | MCESSAYLVNKGGE--EEKIDYVVDIVPKENGKLSLTDILGEEIEE-G | Mahella-australiensis             | RLKEVRLLEHKIVIEKAM----- |
| Desulfococcus-oleovorans          | MCESSAYLVNKGGE--EEKIDYVVDIVPKENGKLSLTDILGEEIEE-G | Desulfococcus-oleovorans          | RLKEVRLLEHKIVIEKAM----- |
| Desulfobaculum-toluoelica         | MCESSAYLVNKGGE--EEKIDYVVDIVPKENGKLSLTDILGEEIEE-G | Desulfobaculum-toluoelica         | RLKEVRLLEHKIVIEKAM----- |
| Desulfobacterium-autotrophicum    | MCESSAYLVNKGGE--EEKIDYVVDIVPKENGKLSLTDILGEEIEE-G | Desulfobacterium-autotrophicum    | RLKEVRLLEHKIVIEKAM----- |
| Desulfatibacillum-alkenivorans    | MCESSAYLVNKGGE--EEKIDYVVDIVPKENGKLSLTDILGEEIEE-G | Desulfatibacillum-alkenivorans    | RLKEVRLLEHKIVIEKAM----- |
| Desulfosporosinus-meridiei1       | MCESSAYLVNKGGE--EEKIDYVVDIVPKENGKLSLTDILGEEIEE-G | Desulfosporosinus-meridiei1       | RLKEVRLLEHKIVIEKAM----- |
| Desulfosporosinus-meridiei2       | MCESSAYLVNKGGE--EEKIDYVVDIVPKENGKLSLTDILGEEIEE-G | Desulfosporosinus-meridiei2       | RLKEVRLLEHKIVIEKAM----- |
| Desulfosporosinus-orientis1       | MCESSAYLVNKGGE--EEKIDYVVDIVPKENGKLSLTDILGEEIEE-G | Desulfosporosinus-orientis1       | RLKEVRLLEHKIVIEKAM----- |
| Desulfatobacterium-dehalogenan    | MCESSAYLVNKGGE--EEKIDYVVDIVPKENGKLSLTDILGEEIEE-G | Desulfatobacterium-dehalogenan    | RLKEVRLLEHKIVIEKAM----- |
| Desulfatobacterium-hafnienae      | MCESSAYLVNKGGE--EEKIDYVVDIVPKENGKLSLTDILGEEIEE-G | Desulfatobacterium-hafnienae      | RLKEVRLLEHKIVIEKAM----- |
| Desulfatobacterium-dichloroeli    | MCESSAYLVNKGGE--EEKIDYVVDIVPKENGKLSLTDILGEEIEE-G | Desulfatobacterium-dichloroeli    | RLKEVRLLEHKIVIEKAM----- |
| Desulfosporosinus-acidiphilus     | MCESSAYLVNKGGE--EEKIDYVVDIVPKENGKLSLTDILGEEIEE-G | Desulfosporosinus-acidiphilus     | RLKEVRLLEHKIVIEKAM----- |
| Desulfosporosinus-orientis2       | MCESSAYLVNKGGE--EEKIDYVVDIVPKENGKLSLTDILGEEIEE-G | Desulfosporosinus-orientis2       | RLKEVRLLEHKIVIEKAM----- |
| Desulfarculus-baarsii             | MCESSAYLVNKGGE--EEKIDYVVDIVPKENGKLSLTDILGEEIEE-G | Desulfarculus-baarsii             | RLKEVRLLEHKIVIEKAM----- |
| Desulfobacca-acetoxidans          | MCESSAYLVNKGGE--EEKIDYVVDIVPKENGKLSLTDILGEEIEE-G | Desulfobacca-acetoxidans          | RLKEVRLLEHKIVIEKAM----- |
| Syntrophobacter-fumaroxidans      | MCESSAYLVNKGGE--EEKIDYVVDIVPKENGKLSLTDILGEEIEE-G | Syntrophobacter-fumaroxidans      | RLKEVRLLEHKIVIEKAM----- |
| Thermodesulfobacterium-yellowston | MCESSAYLVNKGGE--EEKIDYVVDIVPKENGKLSLTDILGEEIEE-G | Thermodesulfobacterium-yellowston | RLKEVRLLEHKIVIEKAM----- |
| Desulfomonile-tiedjei             | MCESSAYLVNKGGE--EEKIDYVVDIVPKENGKLSLTDILGEEIEE-G | Desulfomonile-tiedjei             | RLKEVRLLEHKIVIEKAM----- |
| Thermodesulfobacterium-narugense  | MCESSAYLVNKGGE--EEKIDYVVDIVPKENGKLSLTDILGEEIEE-G | Thermodesulfobacterium-narugense  | RLKEVRLLEHKIVIEKAM----- |
| Methanothermobacter-thermautot    | MCESSAYLVNKGGE--EEKIDYVVDIVPKENGKLSLTDILGEEIEE-G | Methanothermobacter-thermautot    | RLKEVRLLEHKIVIEKAM----- |
| Methanothermobacter-marburgensis  | MCESSAYLVNKGGE--EEKIDYVVDIVPKENGKLSLTDILGEEIEE-G | Methanothermobacter-marburgensis  | RLKEVRLLEHKIVIEKAM----- |
| Methanobacterium-lacus            | MCESSAYLVNKGGE--EEKIDYVVDIVPKENGKLSLTDILGEEIEE-G | Methanobacterium-lacus            | RLKEVRLLEHKIVIEKAM----- |
| Methanobacterium-paludis          | MCESSAYLVNKGGE--EEKIDYVVDIVPKENGKLSLTDILGEEIEE-G | Methanobacterium-paludis          | RLKEVRLLEHKIVIEKAM----- |
| Methanothermobacter-fervidus      | MCESSAYLVNKGGE--EEKIDYVVDIVPKENGKLSLTDILGEEIEE-G | Methanothermobacter-fervidus      | RLKEVRLLEHKIVIEKAM----- |
| Methanococcus-vannielii           | MCESSAYLVNKGGE--EEKIDYVVDIVPKENGKLSLTDILGEEIEE-G | Methanococcus-vannielii           | RLKEVRLLEHKIVIEKAM----- |
| Methanococcus-maripaludis         | MCESSAYLVNKGGE--EEKIDYVVDIVPKENGKLSLTDILGEEIEE-G | Methanococcus-maripaludis         | RLKEVRLLEHKIVIEKAM----- |
| Archaeoglobus-sulfatocalidus      | MCESSAYLVNKGGE--EEKIDYVVDIVPKENGKLSLTDILGEEIEE-G | Archaeoglobus-sulfatocalidus      | RLKEVRLLEHKIVIEKAM----- |
| Ferroglobus-placidus              | MCESSAYLVNKGGE--EEKIDYVVDIVPKENGKLSLTDILGEEIEE-G | Ferroglobus-placidus              | RLKEVRLLEHKIVIEKAM----- |
| Archaeoglobus-veneficus           | MCESSAYLVNKGGE--EEKIDYVVDIVPKENGKLSLTDILGEEIEE-G | Archaeoglobus-veneficus           | RLKEVRLLEHKIVIEKAM----- |
| Thermodesulfatator-indicus        | MCESSAYLVNKGGE--EEKIDYVVDIVPKENGKLSLTDILGEEIEE-G | Thermodesulfatator-indicus        | RLKEVRLLEHKIVIEKAM----- |
| Thermococcus-onnurineus           | MCESSAYLVNKGGE--EEKIDYVVDIVPKENGKLSLTDILGEEIEE-G | Thermococcus-onnurineus           | RLKEVRLLEHKIVIEKAM----- |
| Thermococcus2                     | MCESSAYLVNKGGE--EEKIDYVVDIVPKENGKLSLTDILGEEIEE-G | Thermococcus2                     | RLKEVRLLEHKIVIEKAM----- |
| Thermococcus-banophilus           | MCESSAYLVNKGGE--EEKIDYVVDIVPKENGKLSLTDILGEEIEE-G | Thermococcus-banophilus           | RLKEVRLLEHKIVIEKAM----- |
| Thermococcus1                     | MCESSAYLVNKGGE--EEKIDYVVDIVPKENGKLSLTDILGEEIEE-G | Thermococcus1                     | RLKEVRLLEHKIVIEKAM----- |
| Methanopyrus-kandleri             | MCESSAYLVNKGGE--EEKIDYVVDIVPKENGKLSLTDILGEEIEE-G | Methanopyrus-kandleri             | RLKEVRLLEHKIVIEKAM----- |

Figure S2. ESI-MS of RrCooT-Gly1 and RrCooT-Cys1

RrCooT-Gly1 + 1 mM TCEP



RrCooT-Gly1 (no TCEP)



RrCooT-Cys1 (no TCEP)

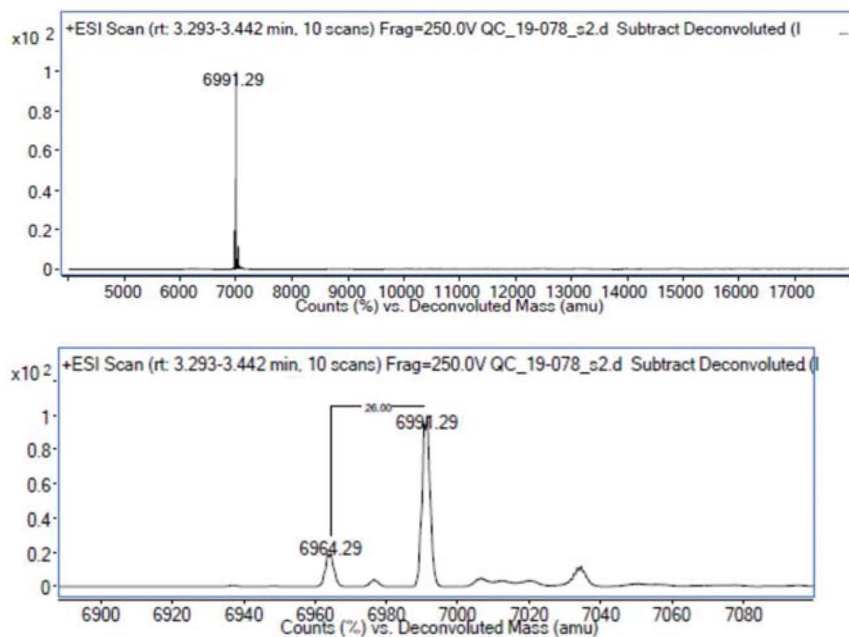
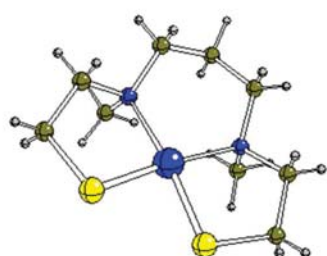


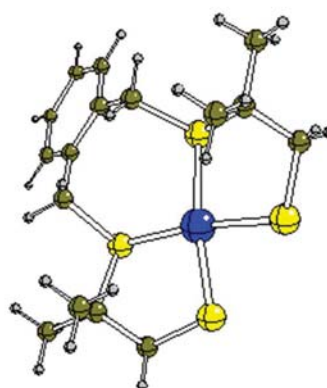
Figure S3

## Structures of Ni-CooT complexes (Reference compounds)

Ni(II)-N2S2

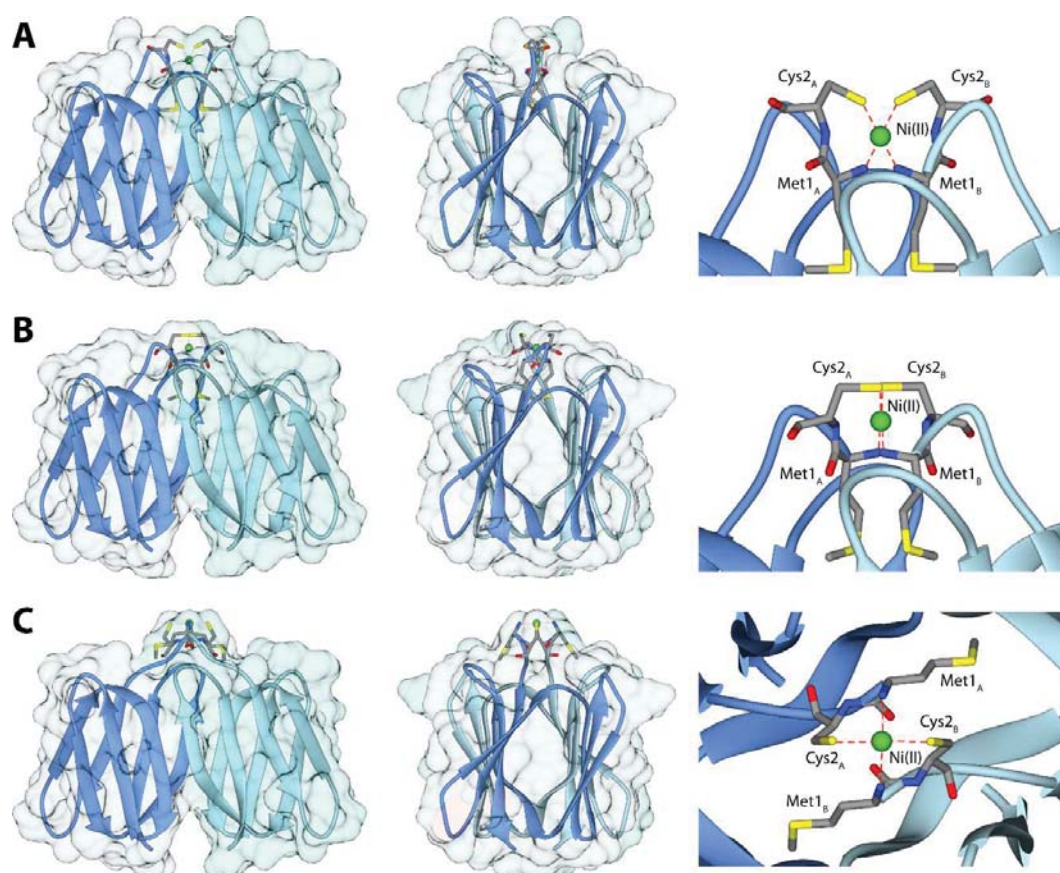


Ni(II)-S4

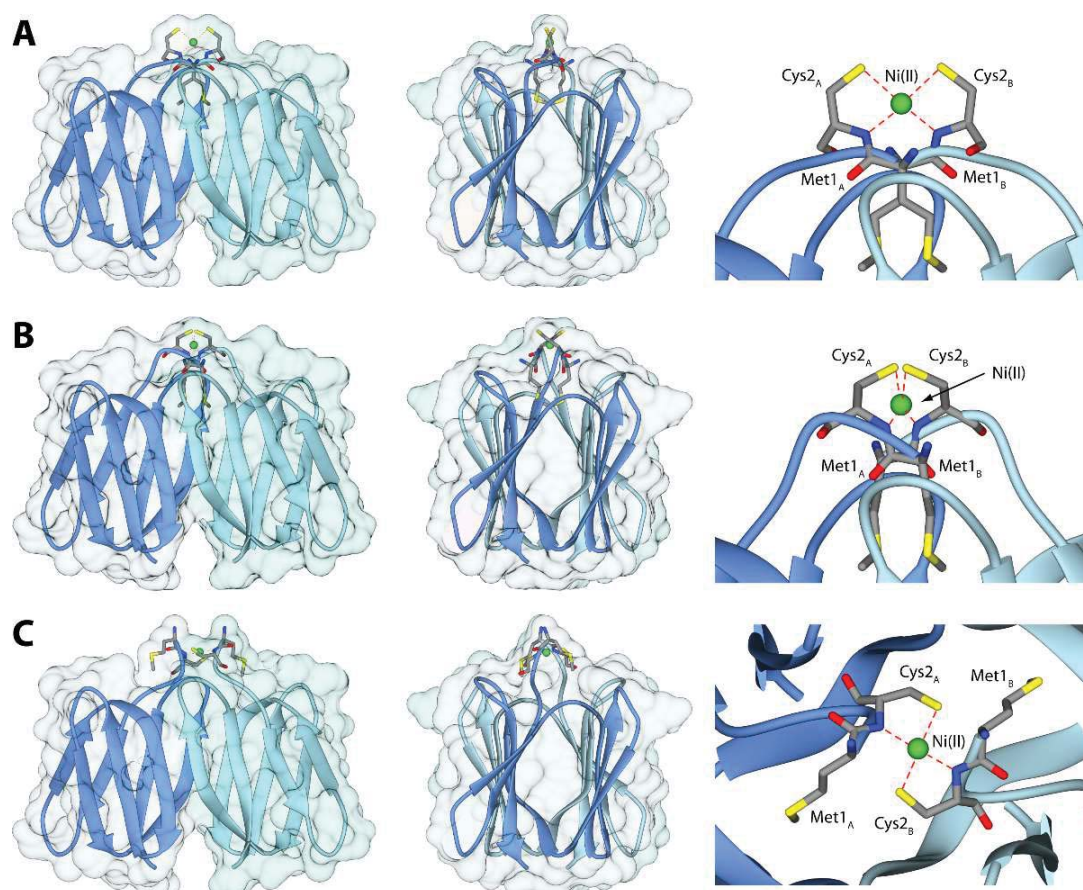




**Figure S4.** Ribbon diagrams and solved exposed surface (left and central panels) together with a detail of the Ni(II) binding site (right panels) of the Ni(II)-bound *Rr*CooT Amino-Cis1, Amino-Cis2 and Amino-Trans model structures (panels **A**, **B**, and **C**, respectively; see Scheme 1 in the main text). The *Rr*CooT monomers are colored in blue and light blue, while the Ni(II) ion is reported as a green sphere and Met1 and Cys2 from each monomer are depicted as sticks colored accordingly to the atom type. The ribbons in the central panels have been rotated by 90° degrees around the vertical axis with respect to the representations in the left panels. The orientation reported in the right panel of the Amino-Trans model has been rotated by 90° degrees around the horizontal axis with respect to the representations in the left panel.



**Figure S5.** Ribbon diagrams and solved exposed surface (left and central panels) together with a detail of the Ni(II) binding site (right panels) of the Ni(II)-bound *Rr*CooT Amide-Cis1, Amide-Cis2 and Amide-Trans model structures (panels **A**, **B**, and **C**, respectively; see Scheme 1 in the main text). The *Rr*CooT monomers are colored in blue and light blue, while the Ni(II) ion is reported as a green sphere and Met1 and Cys2 from each monomer are depicted as sticks colored accordingly to the atom type. The ribbons in the central panels have been rotated by 90° degrees around the vertical axis with respect to the representations in the left panels. The orientation reported in the right panel of the Amide-Trans model has been rotated by 90° degrees around the horizontal axis with respect to the representations in the left panel.



**Table S1.** NMR experiments carried out on apo-*R*-CooT

| <b>Protein Isotope Labeling</b>   | <b>NMR experiment</b>               | <b>Field (MHz)</b> | <b>T (K)</b> | <b>NUS</b> | <b>Bruker Pulse sequence</b> | <b>Acquired points</b> |
|-----------------------------------|-------------------------------------|--------------------|--------------|------------|------------------------------|------------------------|
| $^{15}\text{N}$ , $^{13}\text{C}$ | $^1\text{H}$ , $^{15}\text{N}$ HSQC | 950                | 298          | -          | hsqcfpf3gp phwg              | 1280x128               |
| $^{15}\text{N}$ , $^{13}\text{C}$ | $^1\text{H}$ , $^{13}\text{C}$ HSQC | 950                | 298          | -          | hsqcedetgpsisp               | 1280x256               |
| $^{15}\text{N}$ , $^{13}\text{C}$ | HNCACB                              | 950                | 298          | -          | nhcacbgpwg3d                 | 1792x48x128            |
| $^{15}\text{N}$ , $^{13}\text{C}$ | CBCA(CO)NH                          | 950                | 298          | -          | cbcaconhgpwg3d               | 1792x48x128            |
| $^{15}\text{N}$ , $^{13}\text{C}$ | HBHANH                              | 950                | 298          | -          | hbhanhgpwg3d                 | 1792x48x128            |
| $^{15}\text{N}$ , $^{13}\text{C}$ | HBHA(CO)NH                          | 950                | 298          | -          | hbhaconhgpwg3d               | 1792x48x128            |
| $^{15}\text{N}$ , $^{13}\text{C}$ | HNCO                                | 950                | 298          | 40%        | hncogpwg3d                   | 1792x64x128            |
| $^{15}\text{N}$ , $^{13}\text{C}$ | HN(CA)CO                            | 950                | 298          | 40%        | hncacogpwg3d                 | 1792x64x128            |
| $^{15}\text{N}$ , $^{13}\text{C}$ | HNCA                                | 950                | 298          | 40%        | hncagpwg3d                   | 1792x64x128            |
| $^{15}\text{N}$ , $^{13}\text{C}$ | HN(CO)CA                            | 950                | 298          | 40%        | hncocagpwg3d                 | 1792x64x128            |
| $^{15}\text{N}$ , $^{13}\text{C}$ | $^1\text{H}$ , $^{15}\text{N}$ HSQC | 500                | 298          | -          | hsqcfpf3gp phwg              | 1280x128               |
| $^{15}\text{N}$ , $^{13}\text{C}$ | HNCACB                              | 500                | 298          | -          | nhcacbgpwg3d                 | 1024x72x108            |
| $^{15}\text{N}$ , $^{13}\text{C}$ | CBCA(CO)NH                          | 500                | 298          | -          | cbcaconhgp3d                 | 1024x72x108            |

**Table S2.** Chemical shifts (ppm) of backbone nuclei of apo-*RrCooT*

| Residue | H    | N      | C      | C $\alpha$ | C $\beta$ | H $\alpha$ 2 | H $\alpha$ 3 | H $\beta$ | H $\beta$ 2 | H $\beta$ 3 |      |
|---------|------|--------|--------|------------|-----------|--------------|--------------|-----------|-------------|-------------|------|
| Met1    |      |        |        |            |           |              |              |           |             |             |      |
| Cys2    |      |        | 170.11 | 55.55      | 25.56     | 4.45         |              |           | 2.81        |             |      |
| Met3    | 8.26 | 121.53 | 172.15 | 52.23      | 31.31     | 4.48         |              |           | 1.91        |             |      |
| Ala4    | 8.31 | 123.35 | 172.45 | 47.79      | 18.93     | 4.77         |              |           | 1.00        |             |      |
| Lys5    | 8.14 | 117.54 | 172.53 | 51.30      | 32.87     | 5.11         |              |           | 1.55        |             |      |
| Val6    | 9.02 | 122.88 | 171.66 | 56.99      | 30.90     | 5.16         |              |           | 1.65        |             |      |
| Val7    | 8.67 | 123.51 | 171.44 | 57.04      | 32.03     | 4.99         |              |           | 1.73        |             |      |
| Leu8    | 8.70 | 124.59 | 171.79 | 51.45      | 41.26     | 4.58         |              |           |             | 1.36        | 1.19 |
| Thr9    | 8.54 | 118.02 | 171.71 | 58.59      | 66.92     | 4.68         |              |           | 3.85        |             |      |
| Lys10   | 8.68 | 125.24 | 175.54 | 52.74      | 30.98     | 4.24         |              |           |             | 1.85        | 1.75 |
| Ala11   | 8.92 | 124.84 | 175.63 | 51.98      | 15.29     | 3.99         |              |           | 1.34        |             |      |
| Asp12   | 7.73 | 113.75 | 174.23 | 50.37      | 36.95     | 4.38         |              |           |             | 2.90        | 2.48 |
| Gly13   | 8.08 | 108.80 | 172.23 | 41.90      |           |              | 4.25         | 3.45      |             |             |      |
| Gly14   | 8.44 | 109.86 | 169.36 | 42.47      |           |              | 3.90         | 3.62      |             |             |      |
| Arg15   | 8.39 | 122.48 | 173.31 | 52.36      | 29.67     | 4.96         |              |           | 1.62        |             |      |
| Val16   | 8.82 | 126.12 | 170.85 | 58.69      | 31.76     | 4.11         |              |           | 1.66        |             |      |
| Glu17   | 8.34 | 124.17 | 172.87 | 51.27      | 29.78     | 5.27         |              |           | 1.73        |             |      |
| Ile18   | 9.27 | 126.04 | 172.11 | 57.50      | 37.67     | 4.10         |              |           | 1.37        |             |      |
| Gly19   | 8.44 | 112.87 | 170.22 | 40.92      |           |              | 4.48         | 3.67      |             |             |      |
| Asp20   | 8.53 | 113.96 | 172.04 | 52.43      | 36.65     | 4.20         |              |           |             | 2.87        | 2.64 |
| Val21   | 8.14 | 116.73 | 172.20 | 61.68      | 29.50     | 3.54         |              |           | 1.58        |             |      |
| Leu22   | 8.63 | 124.00 | 173.00 | 52.34      | 42.18     | 4.50         |              |           | 1.30        |             |      |
| Glu23   | 7.66 | 117.31 | 171.14 | 52.13      | 30.69     | 5.16         |              |           |             | 1.99        | 1.84 |
| Val24   | 8.61 | 122.91 | 171.38 | 57.75      | 32.00     | 4.93         |              |           | 1.64        |             |      |
| Arg25   | 9.21 | 125.13 | 170.47 | 51.98      | 31.41     | 5.08         |              |           | 1.70        |             |      |
| Ala26   | 9.00 | 129.10 | 172.97 | 48.08      | 16.53     | 5.16         |              |           | 1.36        |             |      |
| Glu27   | 8.54 | 122.54 | 174.15 | 52.81      | 29.62     | 4.49         |              |           |             | 2.07        | 1.85 |
| Gly28   | 9.19 | 113.98 | 172.65 | 44.21      | 44.36     | 3.77         |              |           |             |             |      |
| Gly29   | 8.98 | 108.55 | 170.77 | 42.52      |           |              | 4.00         | 3.73      |             |             |      |
| Ala30   | 7.63 | 121.21 | 173.14 | 47.87      | 19.52     | 4.94         |              |           | 1.13        |             |      |
| Val31   | 7.71 | 117.86 | 172.21 | 58.27      | 30.75     | 4.84         |              |           | 1.71        |             |      |
| Arg32   | 9.49 | 126.88 | 173.04 | 51.64      | 29.50     | 5.25         |              |           |             | 1.68        | 1.54 |
| Val33   | 9.37 | 126.31 | 171.64 | 58.38      | 32.31     | 4.71         |              |           | 2.14        |             |      |
| Thr34   | 9.20 | 124.46 | 172.68 | 58.84      | 67.51     | 5.26         |              |           | 3.90        |             |      |
| Thr35   | 9.02 | 117.93 | 172.62 | 57.10      | 69.56     | 5.17         |              |           | 4.64        |             |      |
| Leu36   | 8.56 | 119.02 | 173.42 | 53.82      | 40.60     | 3.82         |              |           | 1.18        |             |      |
| Phe37   | 7.69 | 114.92 | 171.91 | 54.04      | 35.01     | 4.65         |              |           |             |             |      |
| Asp38   | 7.95 | 115.84 | 172.24 | 52.41      | 36.57     | 4.27         |              |           |             | 2.83        | 2.66 |
| Glu39   | 7.39 | 117.78 | 172.39 | 52.89      | 28.22     | 4.33         |              |           |             | 1.89        | 1.81 |
| Glu40   | 8.38 | 122.60 | 172.82 | 52.08      | 29.42     | 5.10         |              |           | 1.71        |             |      |
| His41   | 8.81 | 122.78 | 170.83 | 51.68      | 30.91     | 4.52         |              |           | 2.70        |             |      |
| Ala42   | 8.54 | 126.04 | 173.50 | 47.88      | 18.34     | 5.05         |              |           | 1.06        |             |      |
| Phe43   | 9.13 | 119.78 | 170.95 | 51.86      | 38.06     | 4.96         |              |           |             | 2.92        | 2.68 |
| Pro44   |      |        | 175.51 | 60.25      | 29.47     | 4.56         |              |           |             | 2.31        | 1.94 |
| Gly45   | 8.81 | 110.12 | 170.76 | 43.50      |           |              | 4.04         | 3.78      |             |             |      |
| Leu46   | 6.83 | 117.07 | 171.21 | 50.14      | 40.90     | 5.05         |              |           |             | 1.50        | 1.33 |
| Ala47   | 9.03 | 120.09 | 173.23 | 46.99      | 19.40     | 4.25         |              |           | 1.08        |             |      |
| Ile48   | 8.72 | 119.73 | 171.93 | 59.63      | 34.46     | 3.80         |              |           | 1.63        |             |      |
| Gly49   | 8.84 | 116.00 | 170.56 | 44.04      |           |              | 4.21         | 3.52      |             |             |      |
| Arg50   | 7.72 | 119.07 | 171.02 | 52.86      | 30.94     | 4.98         |              |           |             | 1.80        | 1.66 |
| Val51   | 9.37 | 124.76 | 170.39 | 58.58      | 30.82     | 4.20         |              |           | 1.91        |             |      |
| Asp52   | 8.47 | 126.24 | 173.78 | 49.83      | 39.00     | 4.99         |              |           |             | 3.07        | 2.30 |
| Leu53   | 8.59 | 124.68 | 175.15 | 52.87      | 35.76     | 4.05         |              |           | 1.87        |             |      |
| Arg54   | 8.53 | 119.52 | 175.29 | 56.05      | 26.82     | 4.15         |              |           |             | 2.00        | 1.86 |
| Ser55   | 7.50 | 110.28 | 172.97 | 55.03      | 61.49     | 4.34         |              |           | 3.67        |             |      |
| Gly56   | 8.03 | 112.06 | 170.74 | 43.95      |           |              | 4.07         | 3.58      |             |             |      |
| Val57   | 7.74 | 118.90 | 172.86 | 59.22      | 30.95     | 4.60         |              |           | 1.74        |             |      |
| Ile58   | 9.31 | 130.95 | 171.59 | 56.98      | 36.87     | 4.85         |              |           | 1.80        |             |      |
| Ser59   | 9.26 | 122.43 | 171.85 | 54.17      | 61.24     | 5.17         |              |           | 3.66        |             |      |
| Leu60   | 8.59 | 123.89 | 173.02 | 50.14      | 39.45     | 5.04         |              |           |             | 1.64        | 1.28 |
| Ile61   | 8.69 | 116.63 | 171.35 | 56.14      | 39.60     | 4.63         |              |           | 1.79        |             |      |
| Glu62   | 8.61 | 122.14 | 174.03 | 54.16      | 27.34     | 4.22         |              |           | 1.96        |             |      |
| Glu63   | 8.24 | 123.41 | 173.60 | 54.38      | 27.24     | 3.97         |              |           | 1.70        |             |      |
| Gln64   | 8.53 | 121.17 | 172.84 | 52.92      | 26.49     | 4.19         |              |           |             | 1.97        | 1.83 |
| Asn65   | 8.45 | 120.11 | 171.30 | 50.40      | 35.80     | 4.60         |              |           |             | 2.74        | 2.59 |
| Arg66   | 7.82 | 126.16 | 178.07 | 54.58      | 28.48     | 4.04         |              |           |             | 1.72        | 1.60 |

**Table S3.** Chemical shifts (ppm) of backbone nuclei of holo-*RrCooT*

| Residue | H    | N      | C $\alpha$ | C $\beta$ |
|---------|------|--------|------------|-----------|
| Met1    |      |        |            |           |
| Cys2    |      |        | 61.99      | 35.59     |
| Met3    | 8.00 | 119.46 | 52.36      | 31.12     |
| Ala4    | 7.99 | 117.16 | 47.98      | 19.24     |
| Lys5    | 8.18 | 117.67 | 51.52      | 33.24     |
| Val6    | 8.82 | 122.36 | 56.99      | 30.94     |
| Val7    | 8.61 | 123.45 | 57.03      | 32.12     |
| Leu8    | 8.64 | 124.18 | 51.60      | 41.37     |
| Thr9    | 8.50 | 117.78 | 58.65      | 67.03     |
| Lys10   | 8.64 | 125.06 | 52.80      | 31.04     |
| Ala11   | 8.86 | 124.79 | 52.25      | 15.17     |
| Asp12   | 7.71 | 113.78 | 50.43      | 37.08     |
| Gly13   | 8.05 | 108.81 | 42.00      |           |
| Gly14   | 8.40 | 109.82 | 42.51      |           |
| Arg15   | 8.35 | 122.53 | 52.48      | 29.65     |
| Val16   | 8.78 | 126.26 | 58.78      | 31.90     |
| Glu17   | 8.31 | 124.43 | 51.37      | 29.94     |
| Ile18   | 9.21 | 126.15 | 57.49      | 37.61     |
| Gly19   | 8.34 | 112.76 | 41.08      |           |
| Asp20   | 8.57 | 114.30 | 52.58      | 36.97     |
| Val21   | 7.85 | 116.70 | 61.80      | 29.58     |
| Leu22   | 8.68 | 124.54 | 51.61      | 41.51     |
| Glu23   | 7.66 | 117.44 | 52.33      | 31.03     |
| Val24   | 8.56 | 123.11 | 57.87      | 31.58     |
| Arg25   | 9.22 | 125.17 | 51.92      | 31.60     |
| Ala26   | 8.95 | 128.80 | 48.24      | 16.54     |
| Glu27   | 8.50 | 122.47 | 52.84      | 29.71     |
| Gly28   | 9.16 | 113.90 | 44.35      | 44.36     |
| Gly29   | 8.98 | 108.55 | 42.65      |           |
| Ala30   | 7.59 | 121.22 | 47.91      | 19.53     |
| Val31   | 7.63 | 118.06 | 58.37      | 30.68     |
| Arg32   | 9.49 | 126.89 | 51.71      | 29.59     |
| Val33   | 9.37 | 126.38 | 58.36      | 32.35     |
| Thr34   | 9.11 | 123.88 | 58.75      | 67.55     |
| Thr35   | 8.96 | 117.51 | 57.16      | 69.72     |
| Leu36   | 8.59 | 118.91 | 54.13      | 41.00     |
| Phe37   | 7.60 | 115.14 | 53.92      | 35.18     |
| Asp38   | 7.92 | 115.61 | 52.48      | 36.39     |
| Glu39   | 7.24 | 117.25 | 52.89      | 28.44     |
| Glu40   | 8.27 | 122.33 | 52.26      | 29.60     |
| His41   | 8.86 | 124.07 | 51.65      | 31.44     |
| Ala42   | 8.46 | 125.51 | 47.92      | 18.47     |
| Phe43   | 9.10 | 119.71 | 51.89      | 38.17     |
| Pro44   |      |        | 60.25      | 29.47     |
| Gly45   | 8.79 | 110.27 | 43.50      |           |
| Leu46   | 6.82 | 117.23 | 50.28      | 40.93     |
| Ala47   | 8.99 | 120.10 | 47.01      | 19.48     |
| Ile48   | 8.68 | 119.76 | 59.71      | 34.51     |
| Gly49   | 8.88 | 116.07 | 44.19      |           |
| Arg50   | 7.73 | 119.24 | 52.87      | 31.17     |
| Val51   | 9.30 | 124.58 | 58.71      | 31.17     |
| Asp52   | 8.33 | 126.57 | 49.83      | 39.15     |
| Leu53   | 8.55 | 124.34 | 52.71      | 35.89     |
| Arg54   | 8.44 | 119.09 | 55.81      | 26.64     |
| Ser55   | 7.47 | 110.36 | 55.34      | 61.57     |
| Gly56   | 7.95 | 112.17 | 43.99      |           |
| Val57   | 7.81 | 119.17 | 59.35      | 31.92     |
| Ile58   | 9.17 | 131.02 | 56.97      | 36.96     |
| Ser59   | 9.24 | 122.19 | 54.31      | 61.22     |
| Leu60   | 8.62 | 123.86 | 50.16      | 39.45     |
| Ile61   | 8.67 | 116.57 | 56.24      | 39.56     |
| Glu62   | 8.58 | 122.17 | 54.25      | 27.36     |
| Glu63   | 8.20 | 123.40 | 54.38      | 27.44     |
| Gln64   | 8.50 | 121.09 | 53.01      | 26.73     |
| Asn65   | 8.42 | 120.07 | 50.45      | 35.93     |
| Arg66   | 7.77 | 126.18 | 54.65      | 28.59     |



**Table S4.** Best-fit parameters of the Gaussian curved used to model the  $1s \rightarrow 3d$  transition peak of the Ni K-edge X-ray absorption spectra. The errors relative to the last digit are reported in parentheses. HWHM: Half Width at Half Maximum.

|   | <b>Center (eV)</b> | <b>High (a.u.)</b> | <b>HWHM (eV)</b> | <b>Area (<math>10^{-2}</math> eV)</b> |
|---|--------------------|--------------------|------------------|---------------------------------------|
| <b>Ni(II)N<sub>2</sub>S<sub>2</sub></b> | 8332.64 (2)        | 0.018 (1)          | 0.66 (3)         | 2.6                                   |
| <b>Ni(II)S<sub>4</sub></b>              | 8332.78 (5)        | 0.015 (1)          | 0.75 (8)         | 2.3                                   |
| <b>Ni-CooT</b>                          | 8332.92 (7)        | 0.009 (1)          | 0.70 (8)         | 1.3                                   |

**Table S5.** Distances, angle and dihedral constraints used to guide the modeling of *RrCooT* Ni(II) binding site in the case of the amine models (Amine-Cis1, Amine-Cis2 and Amine-Trans). The constraints are listed in the form “mean  $\pm 1\sigma$ ”.

| Model              | Constrained atoms   | Distance (Å)    |
|--------------------|---|-----------------|
| All amine          | Ni(II) - Met1 <sub>A</sub> (N) / Ni(II) - Met1 <sub>B</sub> (N)   | 2.00 $\pm$ 0.10 |
|                    | Ni(II) - Cys2 <sub>A</sub> (S $\gamma$ ) / Ni(II) - Cys2 <sub>B</sub> (S $\gamma$ )   | 2.15 $\pm$ 0.05 |
| Angle (degrees)    |   |                 |
| All amine          | Ni(II) - Met1 <sub>A</sub> (N) - Met1 <sub>A</sub> (C $\alpha$ ) / Ni(II) - Met1 <sub>B</sub> (N) - Met1 <sub>B</sub> (C $\alpha$ )                   | 109 $\pm$ 5     |
|                    | Ni(II) - Cys2 <sub>A</sub> (S $\gamma$ ) - Cys2 <sub>A</sub> (C $\beta$ ) / Ni(II) - Cys2 <sub>B</sub> (S $\gamma$ ) - Cys2 <sub>B</sub> (C $\beta$ ) | 109 $\pm$ 5     |
| Amine-Cis1         | Met1 <sub>A</sub> (N) - Ni(II) - Met1 <sub>B</sub> (N)  | 90 $\pm$ 5      |
|                    | Cys2 <sub>A</sub> (S $\gamma$ ) - Ni(II) - Cys2 <sub>B</sub> (S $\gamma$ )  | 90 $\pm$ 5      |
|                    | Met1 <sub>A</sub> (N) - Ni(II) - Cys2 <sub>A</sub> (S $\gamma$ )  | 90 $\pm$ 5      |
|                    | Met1 <sub>B</sub> (N) - Ni(II) - Cys2 <sub>B</sub> (S $\gamma$ )  | 90 $\pm$ 5      |
|                    | Met1 <sub>A</sub> (N) - Ni(II) - Cys2 <sub>B</sub> (S $\gamma$ )  | 180 $\pm$ 5     |
|                    | Met1 <sub>B</sub> (N) - Ni(II) - Cys2 <sub>A</sub> (S $\gamma$ )  | 180 $\pm$ 5     |
| Amine-Cis2         | Met1 <sub>A</sub> (N) - Ni(II) - Met1 <sub>B</sub> (N)  | 90 $\pm$ 5      |
|                    | Cys2 <sub>A</sub> (S $\gamma$ ) - Ni(II) - Cys2 <sub>B</sub> (S $\gamma$ )  | 90 $\pm$ 5      |
|                    | Met1 <sub>A</sub> (N) - Ni(II) - Cys2 <sub>A</sub> (S $\gamma$ )  | 180 $\pm$ 5     |
|                    | Met1 <sub>B</sub> (N) - Ni(II) - Cys2 <sub>B</sub> (S $\gamma$ )  | 180 $\pm$ 5     |
|                    | Met1 <sub>A</sub> (N) - Ni(II) - Cys2 <sub>B</sub> (S $\gamma$ )  | 90 $\pm$ 5      |
|                    | Met1 <sub>B</sub> (N) - Ni(II) - Cys2 <sub>A</sub> (S $\gamma$ )  | 90 $\pm$ 5      |
| Amine-Trans        | Met1 <sub>A</sub> (N) - Ni(II) - Met1 <sub>B</sub> (N)  | 180 $\pm$ 5     |
|                    | Cys2 <sub>A</sub> (S $\gamma$ ) - Ni(II) - Cys2 <sub>B</sub> (S $\gamma$ )  | 180 $\pm$ 5     |
|                    | Met1 <sub>A</sub> (N) - Ni(II) - Cys2 <sub>A</sub> (S $\gamma$ )  | 90 $\pm$ 5      |
|                    | Met1 <sub>B</sub> (N) - Ni(II) - Cys2 <sub>B</sub> (S $\gamma$ )  | 90 $\pm$ 5      |
|                    | Met1 <sub>A</sub> (N) - Ni(II) - Cys2 <sub>B</sub> (S $\gamma$ )  | 90 $\pm$ 5      |
|                    | Met1 <sub>B</sub> (N) - Ni(II) - Cys2 <sub>A</sub> (S $\gamma$ )  | 90 $\pm$ 5      |
| Dihedral (degrees) |   |                 |
| Amine-Cis1         | Met1 <sub>A</sub> (N) - Ni(II) - Cys2 <sub>A</sub> (S $\gamma$ ) - Cys2 <sub>B</sub> (S $\gamma$ )  | 180 $\pm$ 10    |
|                    | Met1 <sub>B</sub> (N) - Ni(II) - Met1 <sub>A</sub> (N) - Cys2 <sub>A</sub> (S $\gamma$ )  | 180 $\pm$ 10    |
|                    | Cys2 <sub>A</sub> (S $\gamma$ ) - Ni(II) - Cys2 <sub>B</sub> (S $\gamma$ ) - Met1 <sub>B</sub> (N)  | 180 $\pm$ 10    |
|                    | Cys2 <sub>B</sub> (S $\gamma$ ) - Ni(II) - Met1 <sub>B</sub> (N) - Met1 <sub>A</sub> (N)  | 180 $\pm$ 10    |
| Amine-Cis2         | Met1 <sub>A</sub> (N) - Ni(II) - Met1 <sub>B</sub> (N) - Cys2 <sub>A</sub> (S $\gamma$ )  | 180 $\pm$ 10    |
|                    | Met1 <sub>B</sub> (N) - Ni(II) - Cys2 <sub>A</sub> (S $\gamma$ ) - Cys2 <sub>B</sub> (S $\gamma$ )  | 180 $\pm$ 10    |
|                    | Cys2 <sub>A</sub> (S $\gamma$ ) - Ni(II) - Cys2 <sub>B</sub> (S $\gamma$ ) - Met1 <sub>A</sub> (N)  | 180 $\pm$ 10    |
|                    | Cys2 <sub>B</sub> (S $\gamma$ ) - Ni(II) - Met1 <sub>A</sub> (N) - Met1 <sub>B</sub> (N)  | 180 $\pm$ 10    |
| Amine-Trans        | Met1 <sub>A</sub> (N) - Ni(II) - Cys2 <sub>B</sub> (S $\gamma$ ) - Met1 <sub>B</sub> (N)  | 180 $\pm$ 10    |
|                    | Met1 <sub>B</sub> (N) - Ni(II) - Cys2 <sub>A</sub> (S $\gamma$ ) - Met1 <sub>A</sub> (N)  | 180 $\pm$ 10    |
|                    | Cys2 <sub>A</sub> (S $\gamma$ ) - Ni(II) - Met1 <sub>A</sub> (N) - Cys2 <sub>B</sub> (S $\gamma$ )  | 180 $\pm$ 10    |
|                    | Cys2 <sub>B</sub> (S $\gamma$ ) - Ni(II) - Met1 <sub>B</sub> (N) - Cys2 <sub>A</sub> (S $\gamma$ )  | 180 $\pm$ 10    |

**Table S6.** Distances, angle and dihedral constraints used to guide the modeling of *RrCooT* Ni(II) binding site in the case of the amide models (Amide-Cis1, Amide-Cis2 and Amide-Trans). The constraints are listed in the form “mean  $\pm$  1 standard deviation”.

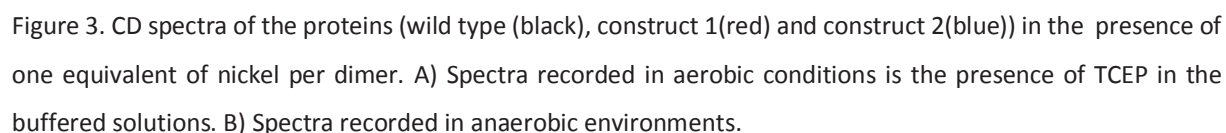
| Model              | Constrained atoms   | Distance (Å)    |
|--------------------|---|-----------------|
| All amide          | Ni(II) - Cys2 <sub>A</sub> (N) / Ni(II) - Cys2 <sub>B</sub> (N)   | 2.00 $\pm$ 0.10 |
|                    | Ni(II) - Cys2 <sub>A</sub> (S $\gamma$ ) / Ni(II) - Cys2 <sub>B</sub> (S $\gamma$ )   | 2.15 $\pm$ 0.05 |
| Angle (degrees)    |   |                 |
| All amide          | Ni(II) - Cys2 <sub>A</sub> (N) - Cys2 <sub>A</sub> (C $\alpha$ ) / Ni(II) - Cys2 <sub>B</sub> (N) - Cys2 <sub>B</sub> (C $\alpha$ )                   | 120 $\pm$ 5     |
|                    | Ni(II) - Cys2 <sub>A</sub> (N) - Met1 <sub>A</sub> (C) / Ni(II) - Cys2 <sub>B</sub> (N) - Met1 <sub>B</sub> (C)                                       | 120 $\pm$ 5     |
|                    | Ni(II) - Cys2 <sub>A</sub> (S $\gamma$ ) - Cys2 <sub>A</sub> (C $\beta$ ) / Ni(II) - Cys2 <sub>B</sub> (S $\gamma$ ) - Cys2 <sub>B</sub> (C $\beta$ ) | 109 $\pm$ 5     |
|                    |   |                 |
| Amide-Cis1         | Cys2 <sub>A</sub> (N) - Ni(II) - Cys2 <sub>B</sub> (N)  | 90 $\pm$ 5      |
|                    | Cys2 <sub>A</sub> (S $\gamma$ ) - Ni(II) - Cys2 <sub>B</sub> (S $\gamma$ )  | 90 $\pm$ 5      |
|                    | Cys2 <sub>A</sub> (N) - Ni(II) - Cys2 <sub>A</sub> (S $\gamma$ )  | 90 $\pm$ 5      |
|                    | Cys2 <sub>B</sub> (N) - Ni(II) - Cys2 <sub>B</sub> (S $\gamma$ )  | 90 $\pm$ 5      |
|                    | Cys2 <sub>A</sub> (N) - Ni(II) - Cys2 <sub>B</sub> (S $\gamma$ )  | 180 $\pm$ 5     |
|                    | Cys2 <sub>B</sub> (N) - Ni(II) - Cys2 <sub>A</sub> (S $\gamma$ )  | 180 $\pm$ 5     |
| Amide-Cis2         | Cys2 <sub>A</sub> (N) - Ni(II) - Cys2 <sub>B</sub> (N)  | 90 $\pm$ 5      |
|                    | Cys2 <sub>A</sub> (S $\gamma$ ) - Ni(II) - Cys2 <sub>B</sub> (S $\gamma$ )  | 90 $\pm$ 5      |
|                    | Cys2 <sub>A</sub> (N) - Ni(II) - Cys2 <sub>A</sub> (S $\gamma$ )  | 180 $\pm$ 5     |
|                    | Cys2 <sub>B</sub> (N) - Ni(II) - Cys2 <sub>B</sub> (S $\gamma$ )  | 180 $\pm$ 5     |
|                    | Cys2 <sub>A</sub> (N) - Ni(II) - Cys2 <sub>B</sub> (S $\gamma$ )  | 90 $\pm$ 5      |
|                    | Cys2 <sub>B</sub> (N) - Ni(II) - Cys2 <sub>A</sub> (S $\gamma$ )  | 90 $\pm$ 5      |
| Amide-Trans        | Cys2 <sub>A</sub> (N) - Ni(II) - Cys2 <sub>B</sub> (N)  | 180 $\pm$ 5     |
|                    | Cys2 <sub>A</sub> (S $\gamma$ ) - Ni(II) - Cys2 <sub>B</sub> (S $\gamma$ )  | 180 $\pm$ 5     |
|                    | Cys2 <sub>A</sub> (N) - Ni(II) - Cys2 <sub>A</sub> (S $\gamma$ )  | 90 $\pm$ 5      |
|                    | Cys2 <sub>B</sub> (N) - Ni(II) - Cys2 <sub>B</sub> (S $\gamma$ )  | 90 $\pm$ 5      |
|                    | Cys2 <sub>A</sub> (N) - Ni(II) - Cys2 <sub>B</sub> (S $\gamma$ )  | 90 $\pm$ 5      |
|                    | Cys2 <sub>B</sub> (N) - Ni(II) - Cys2 <sub>A</sub> (S $\gamma$ )  | 90 $\pm$ 5      |
| Dihedral (degrees) |   |                 |
| All amide          | Ni(II) - Cys2 <sub>A</sub> (N) - Met1 <sub>A</sub> (C) - Met1 <sub>A</sub> (O)  | 180 $\pm$ 10    |
|                    | Ni(II) - Cys2 <sub>B</sub> (N) - Met1 <sub>B</sub> (C) - Met1 <sub>B</sub> (O)  | 180 $\pm$ 10    |
| Amide-Cis1         | Cys2 <sub>A</sub> (N) - Ni(II) - Cys2 <sub>A</sub> (S $\gamma$ ) - Cys2 <sub>B</sub> (S $\gamma$ )  | 180 $\pm$ 10    |
|                    | Cys2 <sub>B</sub> (N) - Ni(II) - Cys2 <sub>A</sub> (N) - Cys2 <sub>A</sub> (S $\gamma$ )  | 180 $\pm$ 10    |
|                    | Cys2 <sub>A</sub> (S $\gamma$ ) - Ni(II) - Cys2 <sub>B</sub> (S $\gamma$ ) - Cys2 <sub>B</sub> (N)  | 180 $\pm$ 10    |
|                    | Cys2 <sub>B</sub> (S $\gamma$ ) - Ni(II) - Cys2 <sub>B</sub> (N) - Cys2 <sub>A</sub> (N)  | 180 $\pm$ 10    |
| Amide-Cis2         | Cys2 <sub>A</sub> (N) - Ni(II) - Cys2 <sub>B</sub> (N) - Cys2 <sub>A</sub> (S $\gamma$ )  | 180 $\pm$ 10    |
|                    | Cys2 <sub>B</sub> (N) - Ni(II) - Cys2 <sub>A</sub> (S $\gamma$ ) - Cys2 <sub>B</sub> (S $\gamma$ )  | 180 $\pm$ 10    |
|                    | Cys2 <sub>A</sub> (S $\gamma$ ) - Ni(II) - Cys2 <sub>B</sub> (S $\gamma$ ) - Cys2 <sub>A</sub> (N)  | 180 $\pm$ 10    |
|                    | Cys2 <sub>B</sub> (S $\gamma$ ) - Ni(II) - Cys2 <sub>A</sub> (N) - Cys2 <sub>B</sub> (N)  | 180 $\pm$ 10    |
| Amide-Trans        | Cys2 <sub>A</sub> (N) - Ni(II) - Cys2 <sub>B</sub> (S $\gamma$ ) - Cys2 <sub>B</sub> (N)  | 180 $\pm$ 10    |
|                    | Cys2 <sub>B</sub> (N) - Ni(II) - Cys2 <sub>A</sub> (S $\gamma$ ) - Cys2 <sub>A</sub> (N)  | 180 $\pm$ 10    |
|                    | Cys2 <sub>A</sub> (S $\gamma$ ) - Ni(II) - Cys2 <sub>A</sub> (N) - Cys2 <sub>B</sub> (S $\gamma$ )  | 180 $\pm$ 10    |
|                    | Cys2 <sub>B</sub> (S $\gamma$ ) - Ni(II) - Cys2 <sub>B</sub> (N) - Cys2 <sub>A</sub> (S $\gamma$ )  | 180 $\pm$ 10    |

- [1] J. Timm, C. Brochier-Armanet, J. Perard, B. Zambelli, S. Ollagnier-de-Choudens, S. Ciurli, C. Cavazza, *Metallomics* **2017**, *9*, 575-583.
- [2] S. Keller, C. Vargas, H. Zhao, G. Piszczek, C. A. Brautigam, P. Schuck, *Anal. Chem.* **2012**, *84*, 5066-5073.
- [3] F. Delaglio, S. Grzesiek, G. W. Vuister, G. Zhu, J. Pfeifer, A. Bax, *J. Biomol. NMR* **1995**, *6*, 277-293.
- [4] A. S. Stern, J. C. Hoch, *Magn Reson Chem* **2015**, *53*, 908-912.
- [5] R. L. J. Keller, Swiss Federal Institute of Technology of Zurich (Zurich), **2004**.
- [6] a M. Geyer, C. Herrmann, S. Wohlgemuth, A. Wittinghofer, H. R. Kalbitzer, *Nat Struct Biol* **1997**, *4*, 694-699; b R. A. Williamson, M. D. Carr, T. A. Frenkiel, J. Feeney, R. B. Freedman, *Biochemistry* **1997**, *36*, 13882-13889; c M. P. Williamson, *Progress in Nuclear Magnetic Resonance Spectroscopy* **2013**, *73*, 1-16.
- [7] J. J. Rehr, J. J. Kas, M. P. Prange, A. P. Sorini, Y. Takimoto, F. Vila, *CR Phys.* **2009**, *10*, 548-559.
- [8] a E. A. Stern, M. Newville, B. Ravel, Y. Yacoby, D. Haskel, *Physica B* **1995**, *208-209*, 117-120; b M. Newville, *J Synchrotron Radiat.* **2001**, *8*, 322-324; c B. Ravel, M. Newville, *J. Synchrotron Radiat.* **2005**, *12*, 537-541.
- [9] L. Giachini, G. Veronesi, F. Francia, G. Venturoli, F. Boscherini, *J. Synchrotron Rad.* **2010**, *17*, 41-45.

### 2.1.2 Circular Dichroism characterization of CooT WT, CooT-Cys1 and CooT-Gly1

Experiments under aerobic conditions in the presence of TCEP were performed in parallel. The comparison between anaerobic and aerobic conditions brought to light an interesting feature: the experimental conditions affect the spectral signatures. In fact, the holo samples (1 eq. of nickel per dimer) present a different CD signal depending on their aerobically (*RrCooT*- O<sub>2</sub>-TCEP) vs anaerobically (*RrCooT*-ana) preparation mode (Figure 3).

The main differences between the two experimental conditions are registered in the regions 300-400 nm and 450-500 nm, usually corresponding to a N-Ni LMCT and d-d transitions <sup>156</sup>, respectively. In addition, the bands around 320–380 nm, characteristic of a S→Ni<sup>2+</sup> charge transfer transition, differ distinctly between the two collected CD spectra (Figure 3).



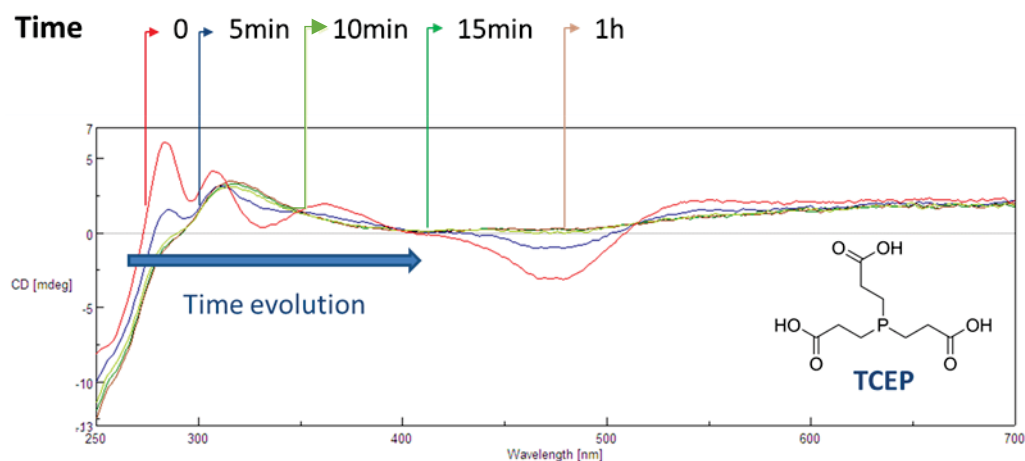


Figure 4. Signal evolution in time upon TCEP addition to the previously anaerobically titrated *RrCooT-gly1*. It is possible to see that the signal modification appears shortly after the TCEP addition. At time 0 the sample presents the previously observed signals (red curve), which then evolves in a time-dependent manner to give the previously observed *RrCooT-O<sub>2</sub>-TCEP* behavior (green curve) (Figure 3).

To explain the reason why this happens, the aerobic experimental conditions was mimicked adding 1 equivalent of TCEP to the samples previously prepared in glovebox (*RrCooT-ana*). The result is a time dependent signal shape modification shortly after the TCEP addition, which results in a CD spectrum identical to the one of *RrCooT-O<sub>2</sub>-TCEP* after 10 minutes of incubation (Figure 4). This is a proof that the TCEP is interacting with the holo-protein, allegedly changing the metal-protein geometry and provoking the difference in signals observed for the two samples.

The difference between the presence/absence of TCEP is also evinced by the UV-Vis absorption spectra, with the presence of bands at 427 nm and 534 nm for *RrCooT-O<sub>2</sub>-TCEP* and 345 nm, 427 nm, 475 nm and 530 nm for *RrCooT-ana* (Figure 5).

This is not the first time that evidence of a nickel-TCEP interaction is observed. In fact it has been, proven that TCEP can interact with thiol-bound metal ions. In specific, Ni(II) presents a particular ability to form such a complex, like in the Ni(II)-glutathione system.<sup>157</sup>

Another interesting feature is the reversibility of the TCEP addition. In fact, TCEP removal by dialysis restores the original signal observed for the anaerobic samples.

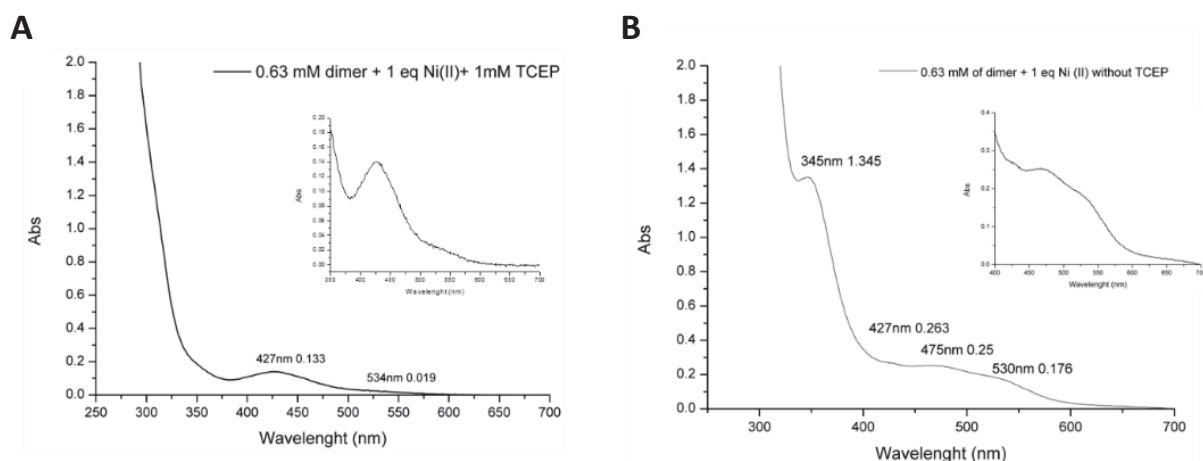


Figure 5. A) RrCooT-O<sub>2</sub>-TCEP sample in a 50 mM HEPES pH 7.5 and 1 mM TCEP buffer. The protein concentration is 0.63 mM dimer, metallated with one equivalent of Ni<sup>2+</sup>. B) RrCooT-ana in a 50 mM HEPES pH 7.5 buffer. The protein concentration is 0.63 mM dimer, metallated with one equivalent of Ni<sup>2+</sup>.

The behavior of the protein with TCEP is intriguing. Considering that the Ni-binding site is solvent exposed, the possibility of CooT to interact with external molecules or partners is not farfetched and the TCEP behavior could uncover the presence of this exchangeable low-affinity position. What if TCEP is playing the role of a possible CooT partner in solution?

TCEP is used as a reducing agent due to free electron pair of the phosphorus, which can break the S-S bonds forming S-P bonds that are then replaced by a water oxygen to make the phosphorus-oxygen bond in an irreversible reaction. Thus, in the case of absence of S-S bridges (remember that the nickel bound to the protein will prevent disulfide bridge formation), can the free electrons of TCEP play a role of an electron donor mimicking the interaction that CooT could have with a physiological partner?

In order to investigate this theory, I tried to reproduce the effect of TCEP using other reducing agents (ammonium acetate, NaH<sub>2</sub>PO<sub>4</sub>, NaBr) and I also checked how the protein would behave in presence of CO. The signal of the anaerobic protein is unaffected for all the tested chemicals.

The interaction of CooT with a possible additional ligand was also calculated via DFT (Figure 6). The metal binding site of the holo-protein model with the Ni<sup>2+</sup> in the trans configuration was used as an input for the quantum mechanical calculations after the addition of the hydrogen atoms and a portion of TCEP (acetate or Me<sub>3</sub>P or Me<sub>3</sub>P=O) or of a water molecule

or a hydroxide ion. The quantum mechanical calculations were done in the gas phase without including any solvent molecule around the system. From this simulation, the direct binding of an additional ligand on the metal axial position is not allowed.

Another other hypothesis is that TCEP could cause a distortion in the Ni geometry, but without direct binding to the metal center. With this in mind, the Ni-N and Ni-S distances in the models were checked considering the presence of either a water molecule or different TCEP portions to see if they would cause some sort of asymmetry in the Ni(II) coordination. Large variations in the Ni(II)-N distances were not obvious inside the same model, however a variation in the Ni(II)-SY distances was observed in the case of the model including a molecule of (Me)<sub>3</sub>P=O (Table 1).

It is important to keep in mind that this analysis cannot provide a complete answer because the models are calculated in the gas phase with only one molecule added to the metal binding site and additional water molecules should be added in the model in order to have a reliable picture of the protein's metal binding site in solution. In addition, also the whole protein environment should be considered.

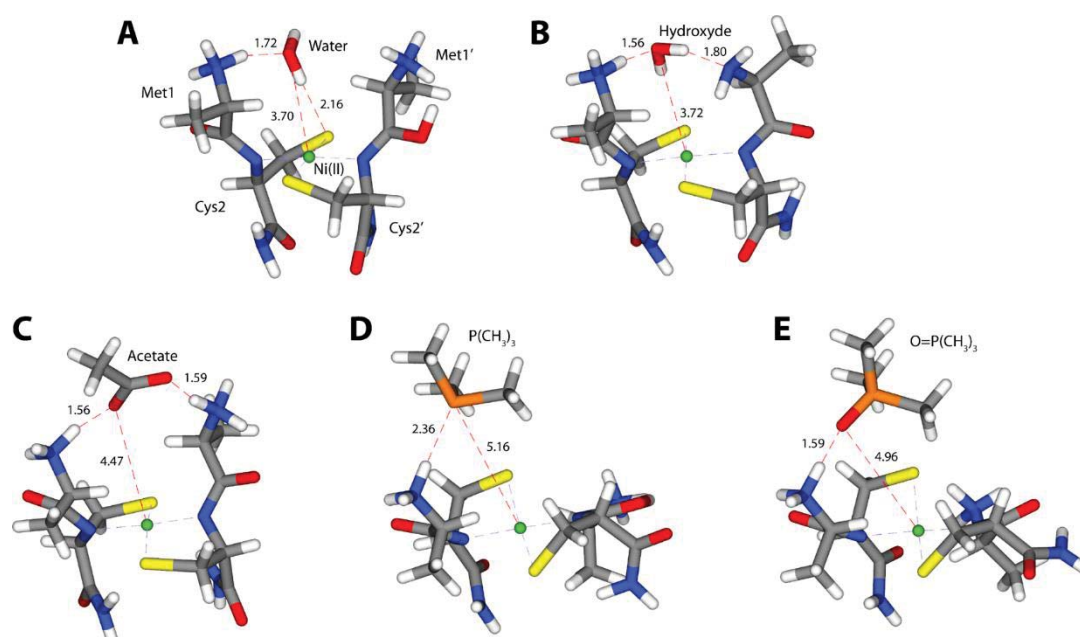


Figure 6 .Optimized geometries of RrCooT Ni-binding site model structures with the inclusion of a water molecule (A), a hydroxide ion (B), an acetate ion (C), a P(CH<sub>3</sub>)<sub>3</sub> molecule (D), and a O=P(CH<sub>3</sub>)<sub>3</sub> molecule (E). Atoms are colored accordingly to the atom type, O red, N blue, S yellow, P orange.



| Model                               | Ni(II)-<br>Cys2(N) | Ni(II)-<br>Cys2'(N) | $ \Delta d $ (N) | Ni(II)-<br>Cys2(S $\gamma$ ) | Ni(II)-<br>Cys2'(S $\gamma$ ) | $ \Delta d $ (S $\gamma$ ) | RMSD  |
|-------------------------------------|--------------------|---------------------|------------------|------------------------------|-------------------------------|----------------------------|-------|
| H <sub>2</sub> O                    | 1.90               | 1.93                | 0.03             | 2.27                         | 2.23                          | 0.04                       | 0.088 |
| OH <sup>-</sup>                     | 1.92               | 1.93                | 0.01             | 2.27                         | 2.24                          | 0.03                       | 0.095 |
| CH <sub>3</sub> -COO <sup>-</sup>   | 1.93               | 1.92                | 0.01             | 2.25                         | 2.26                          | 0.01                       | 0.188 |
| (CH <sub>3</sub> ) <sub>3</sub> P   | 1.93               | 1.94                | 0.01             | 2.24                         | 2.26                          | 0.02                       | 0.172 |
| (CH <sub>3</sub> ) <sub>3</sub> P=O | 1.95               | 1.93                | 0.02             | 2.33                         | 2.22                          | 0.11                       | 0.241 |

Table 1. Calculated distances for RrCooT Ni-binding site model structures with the inclusion of a water molecule, a hydroxide ion, an acetate ion, a P(CH<sub>3</sub>)<sub>3</sub> molecule, and a O=P(CH<sub>3</sub>)<sub>3</sub> molecule .

The hypothesis of a “geometrical distortion” is also suggested from the experimental EXAFS data of the protein in presence of TCEP. The comparison between the preparations with and without TCEP show differences in the XANES’s relative intensity spectral features at ~8338 eV and post-edge at ~8350 eV, suggesting a modification of the coordination sphere of Ni. (Figure 7 A) However, the pre-peak analysis of the 1s→3d transition, which provides information about the coordination geometry of the Ni site, results in areas indicative of square planar coordination for both preparations <sup>158</sup>, as the one reported in the paper. (Figure 7 A, Table 2) The extracted EXAFS spectra do not show any significant difference in terms of oscillation frequency and amplitude (Figure 7 B), suggesting that the nature of Ni donors is the same in the two protein preparations. The comparison between the Ni k-edges XANES spectra and the extracted EXAFS for the two conditions is reported in Figure 7.

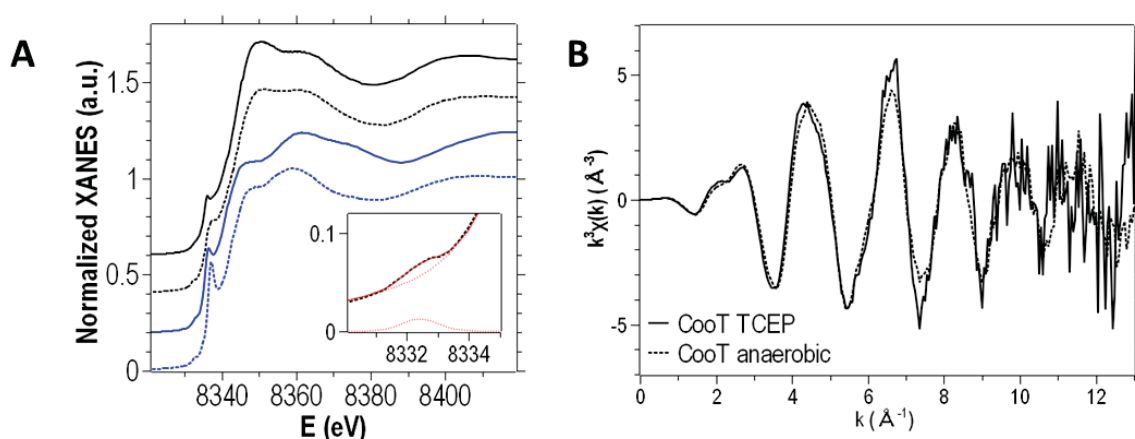


Figure 7. A) Ni K-edge XANES spectra of square-planar Ni(II) reference compounds (blue) with N<sub>2</sub>S<sub>2</sub> (solid line) and S<sub>4</sub> (dashed line) coordination, RrCooT-O<sub>2</sub>-TCEP solid black line and RrCooT-ana dashed black line. Inset: magnification of the 1s → 3d transition energy region and fit (red) of the experimental data (black). The two components of the fitting curve are represented as red dashed lines. B) Ni K-edge EXAFS spectrum of the RrCooT-O<sub>2</sub>-TCEP (solid line) or RrCooT-ana (dashed line).

|   | Center (eV)  | High (a.u.) | HWHM (eV) | Area (10 <sup>-2</sup> eV) |
|---|--------------|-------------|-----------|----------------------------|
| <b>Ni(II)N<sub>2</sub>S<sub>2</sub></b> | 8332.64 (2)  | 0.018 (1)   | 0.66 (3)  | 2.6                        |
| <b>Ni(II)S<sub>4</sub></b>              | 8332.78 (5)  | 0.015 (1)   | 0.75 (8)  | 2.3                        |
| <b>Ni-CooT TCEP</b>                     | 8332.74 (10) | 0.009 (1)   | 0.70 (8)  | 1.4                        |
| <b>Ni-CooT anaer.</b>                   | 8332.42 (5)  | 0.013 (1)   | 0.70 (8)  | 1.9                        |

Table 2. Best-fitting parameters of the Gaussian curves used to model the 1s → 3d transition peak of Ni K-edge X-ray absorption spectra. The errors relative to the last digit are reported in parenthesis. The peak areas fall in the range described by Colpas et al. for a collection of four-coordinate, square planar Ni compounds (area < 2.9·10<sup>2</sup> eV).<sup>158</sup> This analysis, together with the observed resolved 1s→4p<sub>z</sub> transition, allows us to state that the Ni binding geometry is square planar in Ni-CooT complexes, regardless of the preparation.

Whatever is the cause of the differences observed in presence or absence of TCEP in the protein spectra, it still needs to be clarified. From the previous analysis a “solvent induced distortion” in the Ni coordination sphere is supposed, but without direct binding to the metal. Notably, another possibility not considered from the computational models the EXAFS calculations is a possible modification of the nature of the nitrogen ligands from a bisamidate nature to a mix amine/amidate or amide/amidate or even an influence on the protonation state of the thiolate groups.

As you may remember from the introduction, the active site of NiSOD is one of the best-known examples of Ni<sup>2+</sup> coordination in a bis-thiolate and mixed amine/amidate configuration, in its reduced state (*see section 1.1.5*). In order to decipher its peculiar reaction mechanism and evaluate the role of the N-donors, several NiSOD model peptides have been synthesized and studied to mimic the His1Cys2xxxCys6 Ni-binding motif. From these analyses it was concluded that the nucleophilicity nature of the S-donor atoms is strictly related to the coordinating nature of the N-donors.<sup>30</sup> Indeed, the redox-potential of the site increases going from the bisamine to the mix amine/amide ending with the bis amidate couple.<sup>32</sup> In our case, CooT does not have any redox role but we can use these studies for our assumptions. We could imagine that the presence of an electron rich reductant species could perturb the S-donor nucleophilicity inducing them to switch from a bis-amidate coordination towards a more stable amidate/amide or amidate/amine coordination. This hypothesis is consistent with the studies performed by Yuen et al. on the CXC coordination motif, which is supposed to coordinate the nickel in a square planar geometry via the cysteines thiolates and the amide.<sup>86</sup>

In fact, when the peptide CGC-NH<sub>2</sub> (presence of terminal amine) and the protected N-terminus version Ac-CGC-NH<sub>2</sub> were characterized via CD and UV-Vis spectroscopy they registered a similar behavior to our CooT samples (Table 3).<sup>86</sup> From the analysis of the transitions of the NiH<sub>2</sub>L species of Ac-CGC-NH<sub>2</sub>, a di-amidato di-thiolato {2S<sup>-</sup>,2N<sup>-</sup>} coordination mode is proposed (Table 3). When the analysis concerned the NiH<sub>2</sub>L species of CGC-NH<sub>2</sub>, similar transitions with the protected counterpart were registered but with some distinctions that induced the writers to propose either a {NH<sub>2</sub>, 2N<sup>-</sup>, S<sup>-</sup>} or a {2S, 2N<sup>-</sup>} binding mode. It is interesting to find the same behavior in Ni-CooT-ana and Ni-CooT+TCEP cases, allowing the hypothesis of a switch between bisamidate Ni-CooT-ana coordination towards an amine-amidate Ni-CooT-TCEP-O<sub>2</sub>.

Otherwise, could it play a role in the protonation state of the thiolate, influencing the Ni-S distance as supposed from the EXAFS and computational models? The thiolate protonation is suggested to be involved in an “oxygen protective role” of the thiolate S atom in the NiSOD model.<sup>32</sup>

Why the presence of the electron rich TCEP should induce this conformational change is unclear. If this effect can be correlated to a physiological role for the ligand distortion still needs to be investigated. Could the TCEP mimic the presence of the electron rich CO in the bacterium? This could maybe suggest a CO-induced nickel release due to a change in the metal binding coordination that could maybe promote the metal insertion in the CODH active site.

| Ac-CGC-NH <sub>2</sub> <sup>86</sup> |         | CGC-NH <sub>2</sub> <sup>86</sup> |             | CooT-ana |             | CooT-TCEP |         |
|--------------------------------------|---------|-----------------------------------|-------------|----------|-------------|-----------|---------|
| UV-Vis                               | CD      | UV-Vis                            | CD          | UV-Vis   | CD          | UV-Vis    | CD      |
|                                      | 547 (-) |                                   | 536 (-)     |          | 548 (+)     |           |         |
| 550                                  | 444 (+) |                                   | 432 (+)     | 530      | 475 (-)     |           | 521 (-) |
| 431                                  | 322 (+) | 417                               | 333(sh) (+) | 475      | 416(sh) (-) |           | 412 (-) |
| 333                                  | 307 (+) |                                   | 300 (+)     | 427      | 359 (+)     | 427       | 316 (+) |
|                                      | 266 (-) |                                   | 267 (-)     | 345      | 308 (+)     |           | 280 (-) |
|                                      |         |                                   | 238 (+)     |          | 280 (+)     |           |         |

Table 3. Comparison of a published CXC peptide vs RrCooT Ni-binding site with and without TCEP.

### 2.1.3 pH dependence

With the same spirit, the pH dependent metal binding was studied for both *RrCooT*-O<sub>2</sub>-TCEP and *RrCooT*-ana.

#### 2.1.3.1 Experimental procedure

For the study at different pH, 50 mM HEPES buffer was prepared in a pH range between 5.4-10.8 using a 0.2 pH interval (total of 21 points). For the anaerobic experiment, each buffer was degassed in glovebox overnight and the proteins were incubated with 1 equivalent of nickel under anaerobic conditions to a final concentration of 14  $\mu$ M of dimer. The CD spectra were recorded using a J-1500 circular dichroism spectrometer from JASCO Analytical Instruments from 190 to 250 nm using a 1 mm cuvette, with six accumulations to increase the signal-to-noise ratio. The spectra were analyzed using the JASCO spectra manager software (version 2).

#### 2.1.3.2 Results and discussions

Interestingly, for pH lower than 5.9, the protein does not bind nickel and the signal intensity increased according to the pH values. However, the signal shape presents always the same main peaks regardless of the chosen conditions, indicating that the binding geometry is not affected (Figure 8A). As it is possible to see from the data plot (pH values vs CD signal of the specific peak absorption), the protein does not bind nickel up to pH 6 (Figure 8B). The signal then increases constantly up to pH 7, from where it remains constant up to 8.2, at which point it is possible to observe a drop, followed by an additional increase (Figure 8B).

It is interesting to notice that the nickel is stably coordinated throughout the whole 6-10.8 pH range, while only the peak intensity varies. The first jump in absorption could correspond to thiol deprotonation. It is worth mentioning that the pKa value of exposed cysteines is supposed to be around the physiological pH (~7.5), lower compared to the expected value for a buried one (~9).<sup>159,160</sup> This means that exposed cysteines could be easily affected even by marginal increases in pH, changing their protonation state and increasing thiolate reactivity, making it a reactive solvent-accessible group.

Further interpretations about the behavior of the protein at higher pH were not possible due to the protein denaturation and the difficulty to correctly interpret the pH variations.

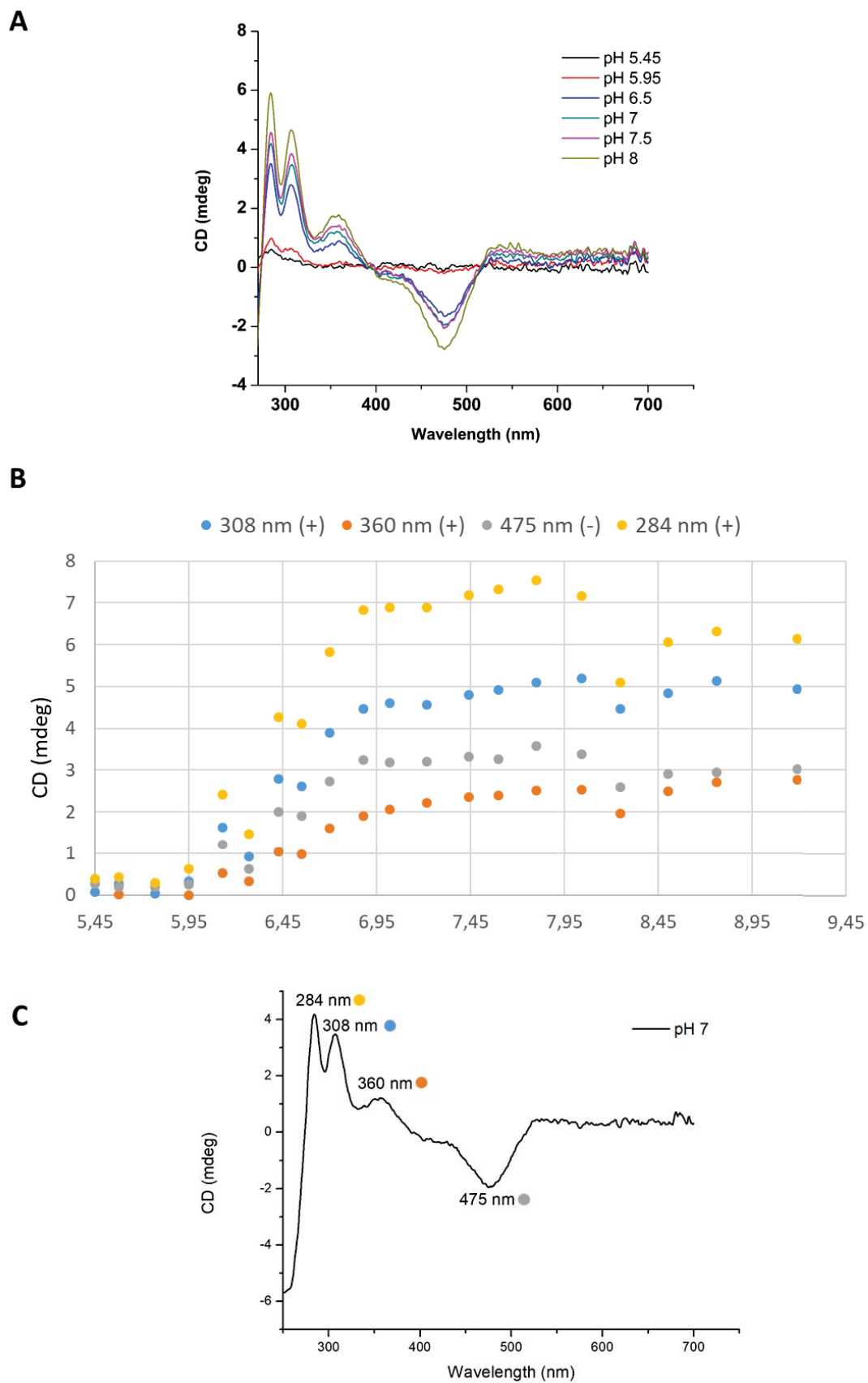


Figure 8. A) CD signals obtained at different pH for *RrCooT-ana*. B) Plotted pH dependency versus the absorption values of the peaks registered via CD spectroscopy. C) Example of a recorded CD spectrum showing the peaks used for the plot of the graph B.

#### 2.1.4 Crystallization assays for holo-*RrCooT*

In order to answer to the question “what are the ligands involved in the nickel coordination” the most straightforward answer would be to crystallize the holo-form of the protein. However, this statement has proven to be really challenging to achieve and, up to date, we are still not able to crystallize the protein with its physiological nickel bound. Different attempts, performed prior to my arrival, were done crystallizing the protein in the glovebox. Yet, the crystals presented a not explicable nickel coordination, involving the non-physiological formation of a tetramer. For this reason, these collections have been considered as artifacts.

Starting from the crystallisation conditions (50 mM sodium acetate at pH 4.6, 100 mM CaCl<sub>2</sub> and 16% (v/v) 2-propanol) of the apo-protein different trials were performed, both in a glovebox and under aerobic conditions, to crystallize the holo-protein without any success. Considering that the wild type protein is always present in a mixture of +/- Met1, a possibility for the lack of success was that the inhomogeneity of the sample could cause a preference for the apo-protein crystallization. This mixture of species does not occur during the production of the *RrCooT*-gly1 and, considering that we demonstrated (*see paper*) that this construct presents the same nickel binding coordination motif as the one of the wild type, we decided to move forward using this protein for the crystallization of the holo-form.

The holo- *RrCooT*-gly1 protein was sent to the HTXLab crystallization platform (EMBL Grenoble) to search for other possible crystallization conditions. The result was a clear trend of the protein to crystallize at low pH (around 4-5). The best promising conditions were reproduced manually in the laboratory but in all cases the protein kept on crystallizing in the apo form, as expected from the behavior of the protein at low pH (not binding at pH<6). Confident to have identified the problem, I focused my efforts on the crystallization conditions with a pH higher than 6.5. At these pH values the protein crystallizes in a needle cluster form, which had to be optimized via micro seeding and additive screens before data collection. However, once again, the crystals were all in the apo form. (The same procedure was also reproduced in glovebox, with the same discouraging result).

In parallel, I tried to insert the nickel into the apo-crystals by the soaking method.

In order to understand if the pH could play a role in the insertion of the metal, the crystals were soaked in different buffer solutions using multiple nickel concentrations and different soaking times (100 mM NiSO<sub>4</sub> for 10 sec, 30 sec, 1 minute; 10 mM NiSO<sub>4</sub> for 30 sec, 1 minute, 3 minutes; 5 mM NiSO<sub>4</sub> for 10 minutes). Once again, this did not give any positive results.

At this point we thought that the nickel insertion could be hindered by the formation of the S-S bridges on the apo crystals, despite the presence of TCEP in the protein stock. For this reason, the apo-crystals were first reduced using TCEP and then soaked on the nickel-containing buffer. This time something happened, the crystals looked like they changed in color (from transparent to a slightly yellow color, consistent with the registered spectra for the TCEP-Ni bound protein) but they completely lost the diffraction ability. At the same time, we started noticing the uncommon behavior of TCEP with our protein (*see previous session*), for this reason we decided to abandon this method.

Another approach that I tried was to purify the *RrCooT-gly1* protein in a holo-form directly from the over production in *E. coli*. For the production of this so called “nickel-*in vivo*” protein, the same expression protocol reported in the paper was used, except that the TB media was supplemented with 0.25 mM of NiSO<sub>4</sub>. In addition, during the purification 1 mM of NiSO<sub>4</sub> was added in all the purification buffers. The resulting protein is a red color protein, in agreement with what was observed from the UV-vis characterization described in the paper (Figure 9).

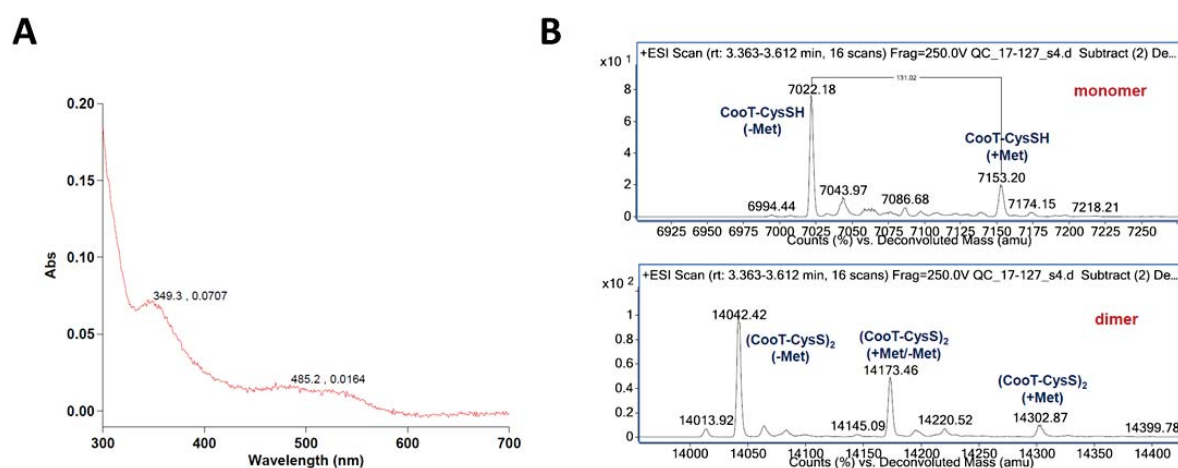


Figure 9. A) UV-Vis absorption spectra for the “nickel in vivo” protein. It does present the same main absorption bands as the apo-titrated protein in anaerobic condition (see Figure 5B). B) The mass spectrum confirms that the protein is in a mixture between the species (Met1-Gly2-Cys3) and (Gly1-Cys2).



This new purified protein was also sent to the HTXLab automatic crystallization platform. Compared to the in vitro inserted metal, almost the same conditions were found and the results obtained in a pH range > than 7 were reproduced in the lab. For this pool, the condition at 0.8 M Tri-sodium citrate, 0.1 M CHES pH 9.5 for a protein concentration of 15 mg/ml gave the most interesting results. However, once again, the crystals obtained were transparent and in an apo-form.

A crystal screening of the “nickel in vivo” protein was also started in glovebox without any interesting conditions obtained.

This could be due to the fact that with in this production method, the protein is present in a mixture of two different forms (Met1-Gly2-Cys3) and (Gly1-Cys2), which is not the case for the apo-protein (Figure 9). This point could be interesting, it looks like that the presence of the metal prevents the met1 cleavage.

The inability to crystallize the holo-form is not so surprising. It is important to remember that the nickel binding site is highly solvent exposed, probably hindering a correct holo-crystal formation. Similar sites, as for example the exposed N-terminal site of serum albumin, are known to be difficult to characterize due to the high mobility of the motif that impedes the crystallization.

### 2.1.5 N-terminal amine transamination reaction

The main question about CooT's coordination site remained whether the amide of the cys2 or the amine of the first amino acid is the one involved, together with cysteine 2 thiolate, in the metal binding.

In the paper, we answer to this question considering our best data interpretation, presenting a model that matches all our different spectroscopic data. However, in order to obtain a further confirmation, I tried to perform a transamination reaction on the N-terminal amine. Theoretically, if the amine is not involved in the nickel coordination, this modification should not affect the CooT's nickel binding abilities.

After literature research, I found that one of the best promising way to perform such a reaction, without attacking the close by cysteine or other amino acids, in a protein is via Pyridoxal 5'-Phosphate (PLP). This reaction has been reported to be specific for the oxidation of the N-terminal amine to a ketone (Figure 10).<sup>161 162</sup>

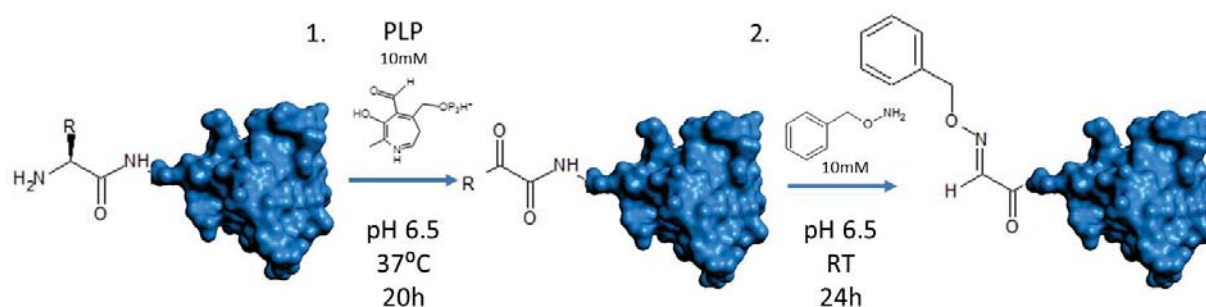


Figure 10. Reaction scheme. (A) In the first step, the protein is incubated with PLP, which transaminates the N-terminus to form a ketone or an aldehyde.<sup>162</sup>

#### 2.1.5.1 Experimental procedure

To a protein concentration of 50  $\mu$ M protein monomer, final PLP concentrations of 10 mM and 100 mM are added to a 50  $\mu$ L volume of 25 mM phosphate buffer at pH 6.5. The reaction is performed at 37 °C for 20 hours without agitation. A control reaction was started in parallel. Due to the absorbance of PLP (bright yellow color), it had to be removed and we used the concentration-dilution method using a vivaspin with a 3 kDa cutoff. At this point, half of the

protein was incubated with a final concentration of 10 mM benzyloxyamine and left at 25° C for 20 hours. The samples were then frozen at -80° C before characterization by ESI-MS.

#### **2.1.5.2 Results and discussions**

In order to assure that the removal of the amine does not play a role in the nickel binding, a high reaction yield is required to be confident about any possible observable coordination differences between the control sample and the transaminated one. In fact, if the protein is present in a mixture between keto-form and amine form it will not be possible to ensure the real effect on the nickel binding coordination. For this reason, the efficiency of the reaction was checked via CD spectroscopy, looking if the protein retains its ability to bind nickel, and via ESI-MS to ensure the presence of homogeneous species.

In the case of the transamination reaction, the formation of a ketone would cause a loss of - 1 Da coming from the formation of the oxygen double bond. This change in mass will be probably too small to ensure a correct interpretation via the ESI-MS analysis. For this reason, it was decided to perform an oxime reaction from the ketone form, which would bring a theoretical gain of 107 Da (Fig 10 step 2).

All the performed reactions did not work and the protein could not be detected via ESI-MS. The control was stable, indicating that the incubation time and temperature are not deleterious for the protein and that the problem is most likely related to the presence of the PLP.

The experiment was then conducted under mild conditions, using 10 mM of PLP at room temperature for 18 hours. Yet, the observed mass between the control and the protein in the reaction mixture are identical, suggesting that the reaction did not occur (Figure 11).

Considering the discouraging preliminary results and the premise that we are looking for a reaction yield as high as possible we decided to abandon this idea and to focus on other parts of the project.

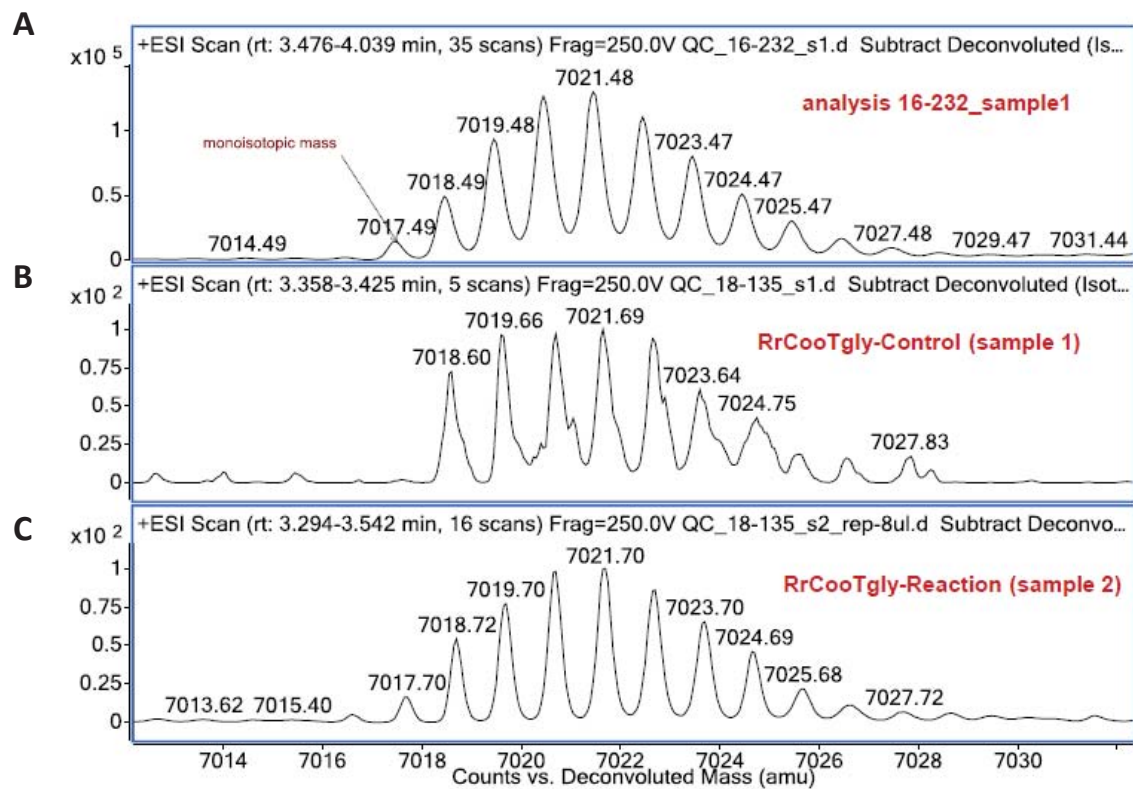


Figure 11. A) Mass *RrCooT-gly1* as purified B) Negative control of *RrCooT-gly1* C) *RrCooT-gly1* after PLP treatment (room temperature for 18 hours). No difference in mass is recorded for the three samples.



## 2.2 C. hydrogenoformans CooT

Switching now to the *C. hydrogenoformans* organism, the possibility to heterogeneously produce CODHs in *E. coli* has been proven (see introduction 1.3 ). However, it is still enigmatic to which accessories proteins are needed in order for the enzyme to be fully active.

The studies on heterologous production of *Ch*CODHs in *E. coli*, showed that the efficiency of production with the related accessory proteins is mainly dependent on the nature of each CODH. The *Ch*CODH-I operon possesses a gene that encodes for a CooC homologue (*cooC*<sub>3</sub>), whose co-expression with *Ch*CODH-I is essential to increase the enzyme activity. Yet, even with the co-expression of *CooC*<sub>3</sub>, *RecCh*CODH-I could not reach the native enzyme activity, a sign that other factors are likely playing an important role during the enzyme activation. In contrast, an excess of nickel in the culture medium is sufficient to activate *RecCh*CODH-II and *RecCh*CODH-IV (which do not have *CooC* in the operon).

Could the absence of another accessory protein be the cause of the partial activity of *RecCh*CODH-I? It is important to remember that *Ch*CODH-I is the one most closely related to *Rr*CODH, both catalyzing the WGS reaction. As we observed for the activity variation of CODH in *R. rubrum* with cultures missing one of its chaperones (*CooCTJ*), could the co-production of a potential accessory protein re-establish the activity of *RecCh*CODH-I?

A possible answer to these questions was provided by the discovery of a putative *CooT* in the *C. hydrogenoformans* proteome. Contrary to *Rr*CODH, this putative *ChCooT* is not in the operon with the *Ch*CODH-I gene. Its position is not close to any other *CooS* and *CooC* genes.

We then decided to characterize this protein to confirm that it belongs to the *CooT* family. Interestingly, the characterization of this protein led us to the formulation of the presence of two different nickel-binding motifs in the *CooT* family. In *ChCooT*, the solved structure revealed the proximity between the Cys2 and His55, allocating this protein in the “**Cys2-His55**” motif subgroup. The production and characterization of this protein resulted in a paper titled “*Biophysical and structural characterization of the putative nickel chaperone CooT from Carboxydotherrmus hydrogenoformans*” reported in the next pages.





### 2.2.1 ChCooT paper

JBIC Journal of Biological Inorganic Chemistry (2018) 23:809–817  
<https://doi.org/10.1007/s00775-018-1576-2>

ORIGINAL PAPER



#### Biophysical and structural characterization of the putative nickel chaperone CooT from *Carboxydothemus hydrogenoformans*

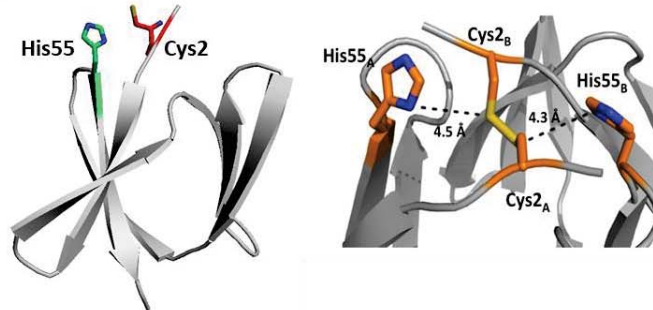
M. Alfano<sup>1</sup> · J. Pérard<sup>1</sup> · R. Miras<sup>1</sup> · P. Catty<sup>1</sup> · C. Cavazza<sup>1</sup>

#### Paper highlights

- *Carboxydothemus hydrogenoformans* possesses a putative CooT.
- ChCooT is in a dimeric conformation in apo-form and it binds one nickel per dimer.
- The structure of ChCooT was solved by X-ray crystallography and even though it has only 19% identity with *R. rubrum* CooT, they share an identical protein conformation. (Seven strands forming two bent antiparallel  $\beta$ -sheets tightly packed on top of each other).
- Contrary to RrCooT, in ChCooT the localization of the unique histidine residue is in close proximity to cysteine 2, supporting a possible cooperation of these two amino acids in the nickel-binding.
- Site-directed mutagenesis of cysteine 2 and histidine 55 revealed the essential role of both residues in nickel-binding. The mutants ChCooT-C2S and ChCooT-C2S-H55A were produced and characterized.
- Only the double mutation hinders the protein nickel binding, while ChCooT-C2S is still able to coordinate the metal, but differently to the WT.
- This putative CooT from *C. hydrogenoformans* resembles RrCooT, despite a potentially different nickel-binding motif that allowed us to identify the presence of a second binding motif in the CooT family.
- Further studies need to be done in order to certify its involvement in the maturation of ChCODH.

**ChCooT wild type (7365Da)**

<sup>2</sup>  
M<sup>C</sup>EASAFIVNGDKEELFLERV<sup>55</sup>DKLIPTEEGLLLENIFGQRKVIKAKIKRLELVD<sup>H</sup>RILLERED





# Biophysical and structural characterization of the putative nickel chaperone CooT from *Carboxydothemus hydrogenoformans*

M. Alfano<sup>1</sup> · J. Pérard<sup>1</sup> · R. Miras<sup>1</sup> · P. Catty<sup>1</sup> · C. Cavazza<sup>1</sup>

Received: 24 February 2018 / Accepted: 1 June 2018 / Published online: 7 June 2018  
© SBIC 2018

## Abstract

*Carboxydothemus hydrogenoformans* is a model microorganism for the study of [NiFe]–CODH, a key enzyme of carbon cycle in anaerobic microorganisms. The enzyme possesses a unique active site (C-cluster), constituted of a distorted [NiFe<sub>3</sub>S<sub>4</sub>] cubane linked to a mononuclear Fe(II) center. Both the biogenesis of the C-cluster and the activation of CODH by nickel insertion remain unclear. Among the three accessory proteins thought to play a role in this latter step (CooC, CooJ, and CooT), CooT is identified as a nickel chaperone involved in CODH maturation in *Rhodospirillum rubrum*. Here, we structurally and biophysically characterized a putative CooT protein present in *C. hydrogenoformans* (pChCooT). Despite the low sequence homologies between CooT from *R. rubrum* (RrCooT) and pChCooT (19% sequence identity), the two proteins share several similarities, such as their overall structure and a solvent-exposed Ni(II)-binding site at the dimer interface. Moreover, the X-ray structure of pChCooT reveals the proximity between the histidine 55, a potential nickel-coordinating residue, and the cysteine 2, a highly conserved key residue in Ni(II)-binding.

**Keywords** Carbon monoxide dehydrogenase · Nickel-binding protein · Enzyme maturation · Multi-metallic active site

## Introduction

Several complex metallocluster-containing enzymes, such as hydrogenase, acetylCoA synthase (ACS), or carbon monoxide dehydrogenase (CODH), are likely to have played a central role in the emergence of life in an environment where hot oceans contained high concentrations of transition metal ions, such as iron or nickel, and the atmosphere was rich in gases like H<sub>2</sub>, CO, or CO<sub>2</sub> [1]. Since H<sub>2</sub>/CO oxidation or CO<sub>2</sub> reduction took place in an anoxic environment, the oxygen-sensitive metalloenzymes catalyzing these reactions are generally confined to anaerobic environments such as the bottoms of lakes or volcanic hot springs. A remarkable feature of these enzymes is the presence of surprising cofactors, composed of heteronuclear FeS clusters with metal atoms or

diatomic ligands, buried in the protein [2]. Given their complexity, the assembly and insertion of these cofactors into the target enzyme require specific and elaborated maturation machineries [3]. Understanding the biogenesis of multi-metallic active sites is quite challenging, since it requires the identification of the dedicated accessory proteins, the characterization of the protein complexes involved in metal and ligand insertion and the sequence of events leading to the final active enzyme.

Hydrogenogenic carboxydotrophs such as *Rhodospirillum rubrum* or *Carboxydothemus hydrogenoformans* are able to use CO as a sole energy source thanks to the water–gas shift reaction (WGS):  $\text{CO} + \text{H}_2\text{O} \rightarrow \text{CO}_2 + \text{H}_2$  [4]. The oxidation of CO to CO<sub>2</sub>, catalyzed by [NiFe]–CODH, is coupled to an energy-conserving [NiFe]–hydrogenase catalyzing the production of H<sub>2</sub>. Five genes encoding CODH have been identified in *C. hydrogenoformans* with different proposed functions depending on their genomic context [5]: CODH-I, involved in WGS; CODH-II, linked to NADH generation; CODH-III, coupled to ACS in the acetyl CoA pathway, and CODH-IV potentially related to the oxidative stress response [6]. No function has been found for CODH-V. In 2001, the crystal structures of CODH from *R. rubrum* (RrCODH) [7] and *C. hydrogenoformans* (ChCODH-II) [8] were solved

**Electronic supplementary material** The online version of this article (<https://doi.org/10.1007/s00775-018-1576-2>) contains supplementary material, which is available to authorized users.

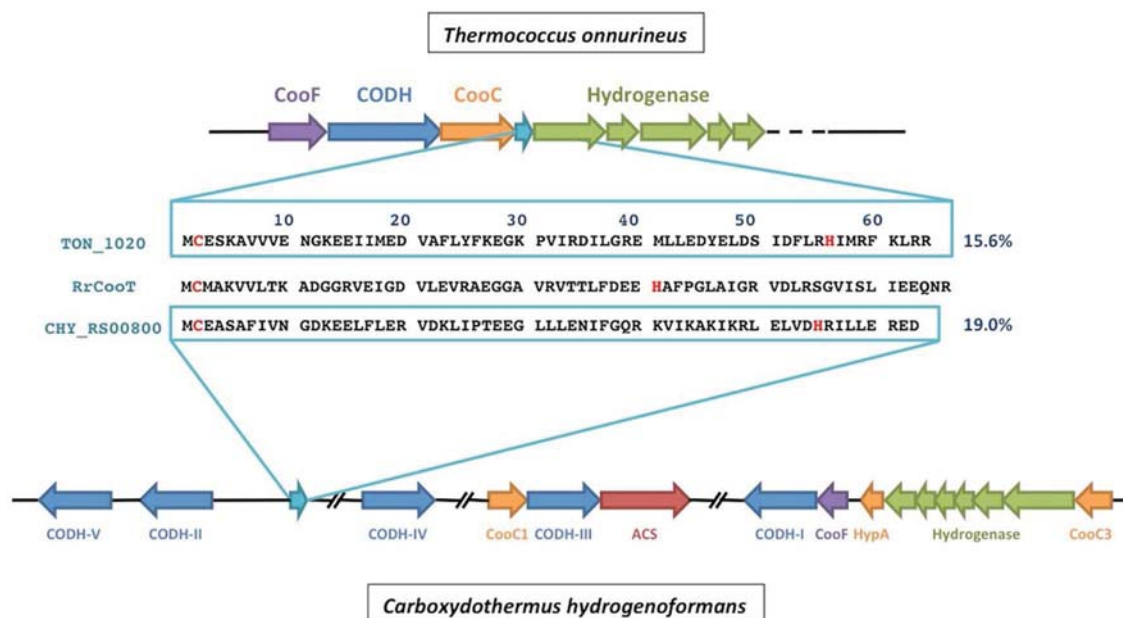
✉ C. Cavazza  
christine.cavazza@cea.fr

<sup>1</sup> University of Grenoble Alpes, CEA, CNRS, BIG, CBM, 38000 Grenoble, France

revealing the unique structure of the active site (named C-cluster), constituted of a  $[\text{NiFe}_3\text{S}_4]$  cluster coordinated to a mononuclear Fe(II) site. While the reaction mechanism of the enzyme has been extensively studied in the last decade [4, 9–11], very little information is available concerning the biosynthesis of the C-cluster and its insertion into apo-CODH. Several studies showed that, when *R. rubrum* is grown in an Ni(II)-depleted growth medium, the produced FeS–CODH form is inactive. The enzyme thus purified can be partly activated in vitro by addition of Ni(II) salts [12, 13], highlighting nickel insertion as a key step in CODH activation.

NTPases are recognized to play a central role in the activation of several Ni enzymes such as the well described GTPases HypB [14, 15] and UreG [16] involved in [NiFe]–hydrogenase and urease maturation pathways, respectively. Their activity depends on their cooperation with other accessory proteins to form a “nickel-delivery complex” responsible for nickel insertion into the target enzyme (HypAHypB  $\pm$  SlyD complex for hydrogenase and Ure(D/H)EFG complex for urease) [17]. In the case of CODH, despite highly similar overall structures, the maturation pathways of RrCODH, ChCODH-II, and ChCODH-IV present different characteristics. *C. hydrogenoformans* contains three ATPases, named CooC1 [18], CooC2, and CooC3 [5]. CooC1 and CooC2 are located in the same gene cluster as CODH-III, while CooC3 is co-expressed with CODH-I (Fig. 1). Recently, CooC2 was purified and characterized as an AcsF-type enzyme, with a function associated to the

maturation of ACS [19]. Interestingly, the activation of recombinant ChCODH-II or ChCODH-IV heterologously produced in *E. coli* does not require any accessory protein. For example, supplying an excess of Ni(II) (500  $\mu\text{M}$ ) in the growth medium is sufficient to correctly activate ChCODH-II, with a specific activity of 12,229 U/mg, comparable to that of the native ChCODH-II (13,823 U/mg). This does not, however, exclude the involvement of accessory proteins under physiological growth conditions, since the high nickel concentration can compensate for their lack. On the contrary, in the case of ChCODH-I heterologously produced in *E. coli*, nickel supply is not sufficient to correctly mature the enzyme. Its co-production with CooC3 led to an approximately fourfold increase in CO oxidation specific activity (8060 U/mg), a value half of that described for native CODH-I (15,756 U/mg) [20]. Among different possibilities, this could indicate that other accessory proteins are required for the full activation of CODH-I in vivo in *C. hydrogenoformans*. On the other hand, in *R. rubrum*, the *coo* operon encodes five different proteins [21]: a ferredoxin CooF that transfers electrons from CODH to hydrogenase, the CODH CooS, CooC [22] and two additional proteins involved in nickel insertion in vivo [23], the metallochaperone CooJ and a small protein of 66 residues, CooT. CooJ has only been described in *R. rubrum* and its exact role in nickel insertion is yet to be determined [24], we recently characterized RrCooT biochemically and structurally [25]. This dimeric protein specifically binds Ni(II) with an affinity in the nanomolar range via an N-terminal binding motif



**Fig. 1** Genomic organization of Mrp–Mbh–Codh operon in *Thermococcus onnurineus* (adapted from 31) and CODH operons in *Carboxythermus hydrogenoformans*. Sequence alignments of RrCooT

with TON\_1020 and CHY\_RS00800 gene products. Putative nickel-coordinated residues are depicted in red

involving the two cysteines 2 from both monomers. While CooT was proposed to be poorly distributed, our phylogenetic analysis highlighted the existence of 111 CooT homologues present mainly in anaerobic bacteria but also in some anaerobic archaea [25]. Cysteine 2, the key residue involved in nickel binding, is strictly conserved. Among the 111 complete proteomes, 83 of them also contain *cooS* and *cooC* genes. In the remaining 28 proteomes, CooT co-occurs either with CooC and acetyl-CoA decarboxylase/synthase (*cdhA* gene), mainly found in methanogens (22 proteomes), or with CooC and ACS/CODH in *Thermodesulfobacterium yellowstonii* and *Butyrivibrio proteoclasticus*. The presence of CooT is not related to the presence of a CODH enzyme and CooC in only four proteomes. These results point strongly towards the role of CooT in CODH maturation [25].

We present here the genomic organization of CODH operons in *C. hydrogenoformans*, as well as the biophysical and the structural characterization of the CHY\_RS00800 gene product to investigate deeper the newly discovered CooT family. Site-directed mutagenesis of cysteine 2 and histidine 55 revealed the essential role of histidine in nickel-binding. Therefore, we have found that this putative CooT from *C. hydrogenoformans* resembles RrCooT, despite a potentially different nickel-binding motif.

## Materials and methods

### Site-directed mutagenesis, protein production and purification

The synthetic CHY\_RS00800 gene, coding for the putative CooT protein from *C. hydrogenoformans* (pChCooT), was codon-optimized for expression in *E. coli* and cloned in the pET15b vector by Genscript. pChCooT–C2S, pChCooT–H55A, and pChCooT–C2S/H55A mutants were obtained by site-directed mutagenesis. PCR was done in the presence of appropriate primers (C2SF: 5'-accatgggctcgaagcgagc-3'/C2SR: 5'-atatctccttattaaagtttaggaccccttagtc-3'; H55AF: 5'-actgtgtgacgctcgtattctgc-3'/H55AR: 5'-tccagacgcttgatttc-3') with Phusion polymerase-GC (NEB) at recommended T<sub>m</sub> (60 °C for single mutants and 65 °C for the double mutant) determined by the AmplifX 1.7.0 software (by Nicolas Jullien; CNRS, Aix-Marseille University). PCR samples were loaded on a 1% agarose gel in TAE buffer to confirm the amplification before incubation with a reaction buffer containing 2 µl of DpnI, 10 µl of T4 DNA Ligase, 1 mM ATP, and 2 µl PNK in PNK buffer from NEB for 15 min at room temperature. Top10 ultracompetent cells were transformed with the PCR products. Each mutant was DNA-sequenced (MWG) before expression and purification as described.

*Escherichia coli* BL21 (DE3) cells harbouring pChCooT–WT, pChCooT–C2S, pChCooT–H55A, and pChCooT–C2S/H55A were cultured in LB medium, supplemented with kanamycin, at 37 °C and 180 rpm. Protein expression was induced with 0.5 mM IPTG at OD<sub>600</sub> ~ 0.6 and left overnight at 25 °C and 180 rpm before harvesting by centrifugation. Bacterial pellets from 2 L cultures were thawed on ice and re-suspended in buffer A (50 mM HEPES pH 7.0, 100 mM NaCl, 1 mM TCEP) containing one complete Protease Inhibitor cocktail tablet (Roche) per 100 mL buffer and disrupted by French press cell. The cell debris were removed by centrifugation at 25,000 rpm for 20 min at 4 °C. Taking into account the thermophilic nature of the protein, a temperature precipitation step was carried out at 65 °C for 5 min, followed by centrifugation at 15,000 rpm for 15 min. The production of the four proteins was checked by SDS-PAGE, revealing that the pChCooT–H55A mutant was not correctly produced. pChCooT–WT, pChCooT–C2S, and pChCooT–C2S/H55A were purified as follows: the lysate was precipitated by addition of ammonium sulphate (100% saturation), re-centrifuged at 20,000 rpm for 15 min and re-suspended in buffer A. A dialysis step in buffer A was performed. The cleared lysate was loaded at 0.8 mL min<sup>-1</sup> onto a Superdex 75 16/600 size-exclusion column (GE Healthcare) equilibrated in buffer A. The fractions were then collected, analyzed by SDS-PAGE, and those containing pChCooT–WT, pChCooT–C2S, or pChCooT–C2S/H55A were pooled and concentrated before flash freezing in liquid nitrogen and stored at –80 °C.

The protein size and identity were determined by electrospray ionization mass spectrometry (ESI-MS). ICP-AES was used to verify that the purified protein corresponded to a metal-free apo-form.

### Size-exclusion chromatography and multi-angle light scattering (SEC–MALLS)

Purified apo-pChCooT, its mutants and pChCooT incubated with 5 molar equivalent of Ni(II), were manually injected into an SEC–MALLS system (Wyatt Dawn HELEOS-II 18-angle light-scattering detector and Wyatt Optilab rEX refractive index monitor linked to a Shimadzu HPLC system comprising an LC-20AD pump and SPD20A UV/Vis detector) using a Superdex 200 10/300 increase size-exclusion column eluted with buffer containing 50 mM HEPES pH 7.0, 150 mM NaCl and 1 mM TCEP ± 5 mM EDTA or 1 mM NiSO<sub>4</sub>. The data were analyzed using the ASTRA software (version 6) and the molecular masses were calculated for each sample. Protein concentration in all samples was determined by integration of the differential refractive index (dRI) peak after injection of 20 µL of sample.



## Ni(II) binding in solution using circular dichroism (CD) spectroscopy and equilibrium dialysis

After thawing, purified apo-pChCooT and its mutants were diluted to 50  $\mu\text{M}$  (dimer concentration) in CD buffer (50 mM HEPES pH 7.5, 1 mM fresh TCEP). Stock solutions of  $\text{NiSO}_4$  were prepared at 10 mM concentration in the same CD buffer. CD spectra were recorded on a JASCO J-1500 circular dichroism spectrometer (JASCO Analytical Instruments). To observe metal–ligand charge transfer bands and d–d transitions of pChCooT, 0, 0.2, 0.4, 0.6, 0.8, 1.0, 1.2, 1.4, 1.6, 2, and 3 molar equivalents of  $\text{NiSO}_4$  were added to the apo-protein. Spectra were recorded from 250 to 700 nm, with ten spectra accumulations, using a 10 mm cuvette.

The binding stoichiometry was determined by equilibrium dialysis conducted at 4 °C under argon atmosphere for 21 h. Dialysis cassettes from Thermo Fisher (Slide-A-Lyzer™ 2 K MWCO) were utilized. In the protein chamber, the apo-pChCooT was diluted to reach a final concentration of 50  $\mu\text{M}$  (monomer) in 500  $\mu\text{L}$  of 50 mM HEPES pH 7.5, 1 mM TCEP and the metal chamber was filled with 250 mL of the same buffer containing 100  $\mu\text{M}$  of  $\text{NiSO}_4$ . The amount of Ni(II) bound to the protein was determined using 4-(2-pyridylazo) resorcinol (PAR), following the absorbance at 495 nm, due to the formation of the metal–PAR<sub>2</sub> complex.

## Crystallization

Initial crystallization screens were carried out using the facilities of the HTXLab at EMBL Grenoble. They were set up on CrystalDirect plates in a sitting-drop vapor diffusion format with 100 nL protein solution  $\pm$  5 molar equivalents of Ni(II) at 18 mg mL<sup>−1</sup> plus an equal volume of precipitant. Crystals were optimized manually with 2  $\mu\text{L}$  protein solution at 18 mg mL<sup>−1</sup> plus an equal volume of precipitant in hanging drop format. The best diffracting crystal of apo-form (crystal 1), corresponding to the monomeric form, was grown with a protein concentration of 18 mg mL<sup>−1</sup> in 47% (v/v) 2-methylpentane-2,4-diol, 2% (v/v) 2-methyl-2-propanol. The crystal was frozen by flash-cooling in liquid nitrogen. The best diffracting crystal of the Ni form (crystal 2), corresponding to the dimeric form, was grown aerobically in the presence of 1 mM TCEP and 5 molar equivalent of Ni(II), with a protein concentration of 18 mg mL<sup>−1</sup> in 15% PEG3350, 250 mM magnesium formate. Microseeding was carried out, according to Hampton protocol (“seed-bead” kit), to increase the size of the crystals. Crystal 2 was cryo-protected in mother liquor containing 25% glycerol before flash-cooling in liquid nitrogen.

## Data collection and X-ray structure solution

Diffraction data of both crystals were collected at the European Synchrotron Radiation Facility (ESRF) on beamline BM30-A. Two X-ray diffraction data sets were collected at the following X-ray wavelengths:  $\lambda_{\text{Se}} = 0.979,790$  Å (maximum  $f''$  for selenium) for crystal 1 and  $\lambda_{\text{Ni}} = 1.488037$  Å (maximum  $f''$  for nickel) for crystal 2. Data reduction was carried out using XDS [26]. Details of data collection, crystallographic data and statistics are summarized in Table S1. The structure was solved with the data set collected from crystal 2 (Ni form) by single anomalous dispersion using AUTOSOL and AUTOBUILD (Phenix) [27]. Three Ni(II) ions were initially located in the asymmetric unit (AU) and the automated model building resulted in a nearly complete model with five monomers in the AU. The model was completed and the protein sequence corrected manually in COOT [28]. This model was used for molecular replacement to obtain the apo-form (monomer) structure at a resolution of 2.0 Å. Refinement was performed using CCP4 and PHENIX [27]. The final models were validated using the graphic software COOT, MolProbity [29] and PDB\_redo [30]. The closest structural homologues to pChCooT were identified using the PDBeFold server (EMBL-EBI). Subsequent Secondary Structure Matching (SSM) superposition and calculation of the overall RMSDs of the respective proteins were carried out in ccp4.

## Results and discussion

### Genomic organization of CODH operons

On the basis of their secondary structures, all the 111 CooT homologues are classified as members of the RNA\_bind\_2 protein family Pfam 10133 [32], even though, as we previously showed, RrCooT does not bind nucleic acids and is likely not involved in RNA metabolism. Moreover, with only 66 amino acids, the ORFs encoding CooT are only rarely recognized. A recent example is the paper from Adams and collaborators, who described the 16-gene putative operon encoding the CO-oxidizing CODH complex in *Thermococcus onnurineus*. Downstream of CooC, they identified the TON\_1020 gene (annotated as a RNA-binding protein in the genome) encoding an “unknown protein” potentially involved in CODH maturation with CooC [31]. Based on its sequence and presence in the operon, we propose that the TON\_1020 gene product corresponds to a CooT protein (Fig. 1). Thanks to our phylogenetic analysis, we identified another hypothetical CooT in *C. hydrogenoformans* (CHY\_RS00800 gene product or pChCooT). In contrast to *R. rubrum* and *T. onnurineus*, here, the gene is neither present in operon nor in the vicinity of any *CooS* or *CooC* (Fig. 1).

Nevertheless, despite differences in their primary structures, the three proteins possess the conserved cysteine 2, which is the only signature of the CooT family. In addition, both putative CooT possess a histidine at position 55/56 not present in RrCooT, where the only histidine is at position 41.

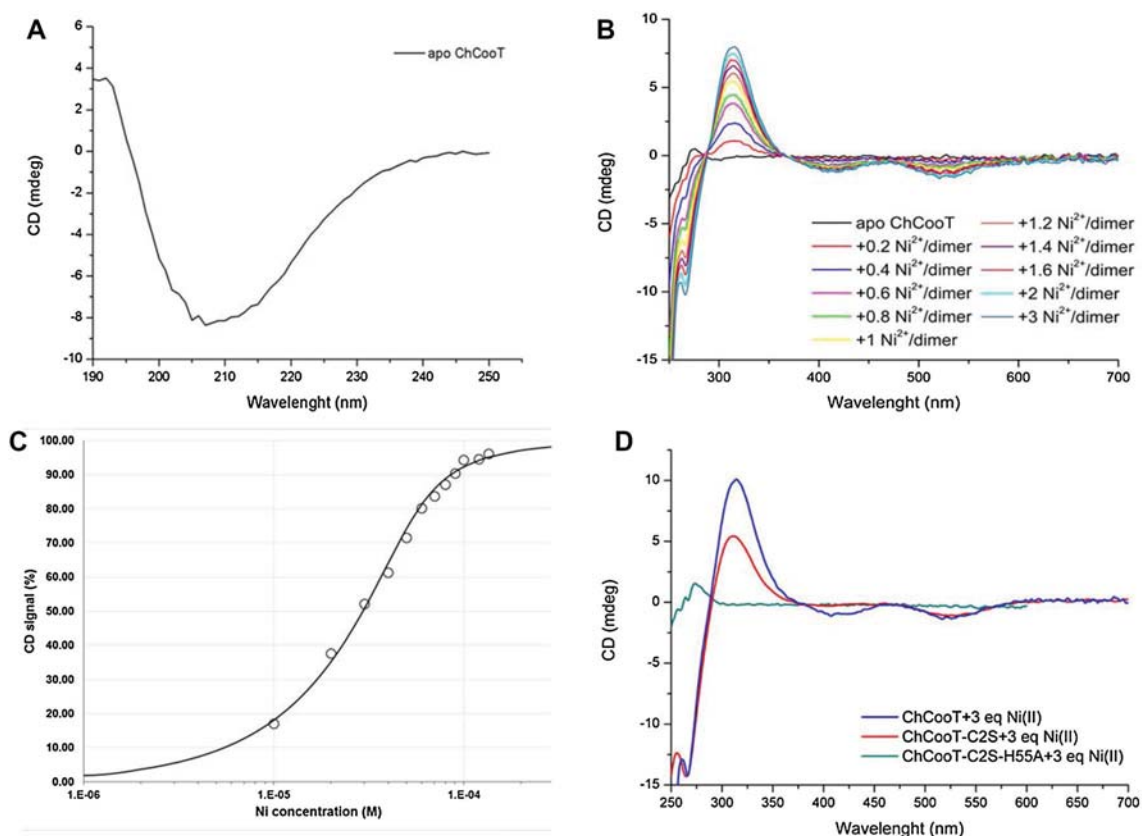
### Biophysical characterization of apo-pChCooT and its mutants

ESI–MS experiments showed that under denaturing conditions, pChCooT, pChCooT-C2S and pChCooT-C2S/H55A are mostly present as monomers (Fig. S1) in a mixture of two forms, with a major peak corresponding to the N-terminal Met-processed protein with molecular masses of 7291.70, 7275.67, and 7209.68 Da, respectively. Circular dichroism (CD) on pChCooT in the far-UV range shows the typical signature of a well-folded  $\beta$ -strand protein (Fig. 2a). The oligomerization state of pChCooT and its mutants was revealed using size-exclusion chromatography coupled to multi-angle

laser light scattering (SEC–MALLS) in the presence of the reducing agent TCEP. As RrCooT, apo- and Ni(II)-pChCooT are both dimers in solution (Fig. 3a, b). Moreover, the single mutation Cys2Ser, as well as the double mutation Cys2Ser/His55Ala do not affect the dimerization of the protein (Fig. 3a). For the wild type, the oligomerization state is only slightly impacted by the addition of an excess of NiSO<sub>4</sub> (1 mM) in the running buffer, with the formation of 13% of tetramer at most (Fig. 3c), revealing that the physiological form of pChCooT likely corresponds to a dimer.

### Nickel-binding properties in solution

With the aim to investigate the ability of pChCooT to bind metals, circular dichroism (CD) experiments in the near-UV/visible range were conducted. Upon the addition of Ni(II) to apo-pChCooT, CD signals appeared at 315 (+), 266 (–), 420 (–), and 535 nm (–) (Fig. 2b). The titration experiment was conducted under both aerobic (in the presence

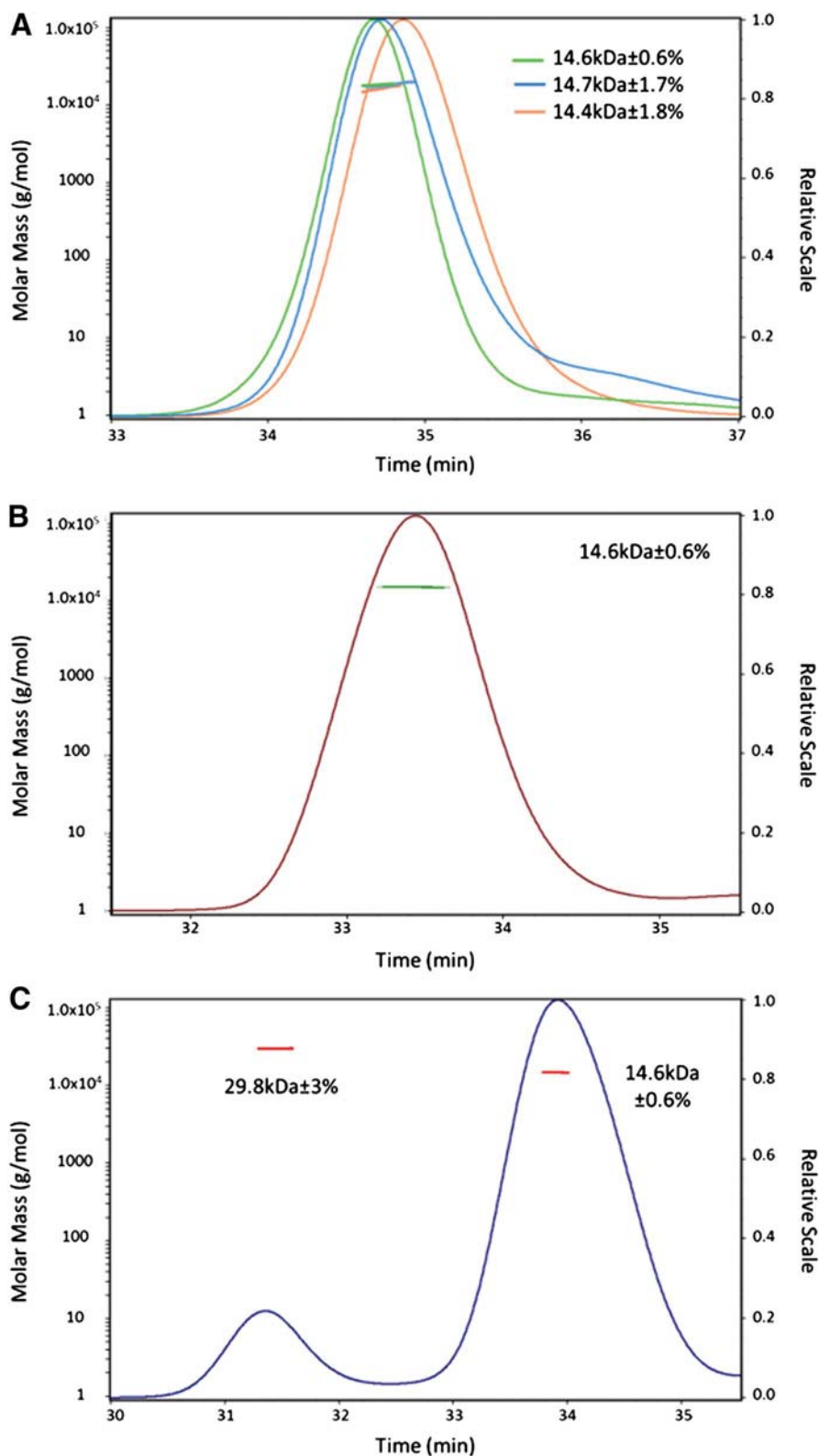


**Fig. 2** **a** CD spectrum of 10  $\mu$ M pChCooT in the far-UV region. Spectra of protein solutions were recorded from 190 to 250 nm with 10 accumulations using a 1 mm cuvette. (Buffer: 50 mM HEPES pH 7.0, 1 mM TCEP). **b** CD measured in near-UV/visible region of Ni(II) titration of pChCooT at 50  $\mu$ M of dimer using a 10 mm cuvette. **c** Data were fitted using a single ligand-binding model described as follows:  $P_1 + L_1 \leftrightarrow PL$  with  $P_1$ : free apoChCooT;  $L_1$ : Ni(II); and  $PL$ :

apo-pChCooT–Ni(II) complex. The concentration of PL ([PL]) is obtained by resolving the following equation where  $K$  (apparent dissociation constant) is the adjustable variable:  $[PL]^2 - [PL](p + l + K) + pl = 0$ .  $p$ : total protein concentration;  $l$ : total ligand concentration. **d** CD measured in near-UV/visible region with five accumulations of pChCooT and mutants at 50  $\mu$ M of dimer after 15 min incubation with 150  $\mu$ M NiSO<sub>4</sub> using a 10 mm cuvette



**Fig. 3** SEC-MALLS profiles for pChCooT and its mutants. **a** apo-pChCooT 550  $\mu$ M of dimer (green), apo-pChCooT-C2S, 500  $\mu$ M of dimer (orange), apo-pChCooT-C2S-H55A, 840  $\mu$ M of dimer (blue) in 50 mM HEPES pH 7.0, 150 mM NaCl, 1 mM TCEP, 5 mM EDTA. **b** apo-pChCooT (550  $\mu$ M of dimer) incubated with 2.5 molar equivalents of Ni(II) per dimer prior to injection, 50 mM HEPES, pH 7.0, 150 mM NaCl, 1 mM TCEP; **c** apo-pChCooT (1.2 mM of dimer) incubated with 2.5 molar equivalents of Ni(II) per dimer prior to injection, 50 mM HEPES pH 7.0, 150 mM NaCl, 1 mM TCEP, 1 mM NiSO<sub>4</sub>



of 1 mM TCEP) and anaerobic conditions in a glove box, resulting in identical spectra and metal saturation. As previously observed with RrCooT, CD signals were also present upon Co(II) addition, but appeared at an approximately

tenfold higher Co(II) concentration than those observed with Ni(II) (Fig. S2). To deeper characterize the nickel-binding site of pChCooT, the Ni(II)-binding stoichiometry was first determined by equilibrium dialysis: with a value of  $0.6 \pm 0.1$

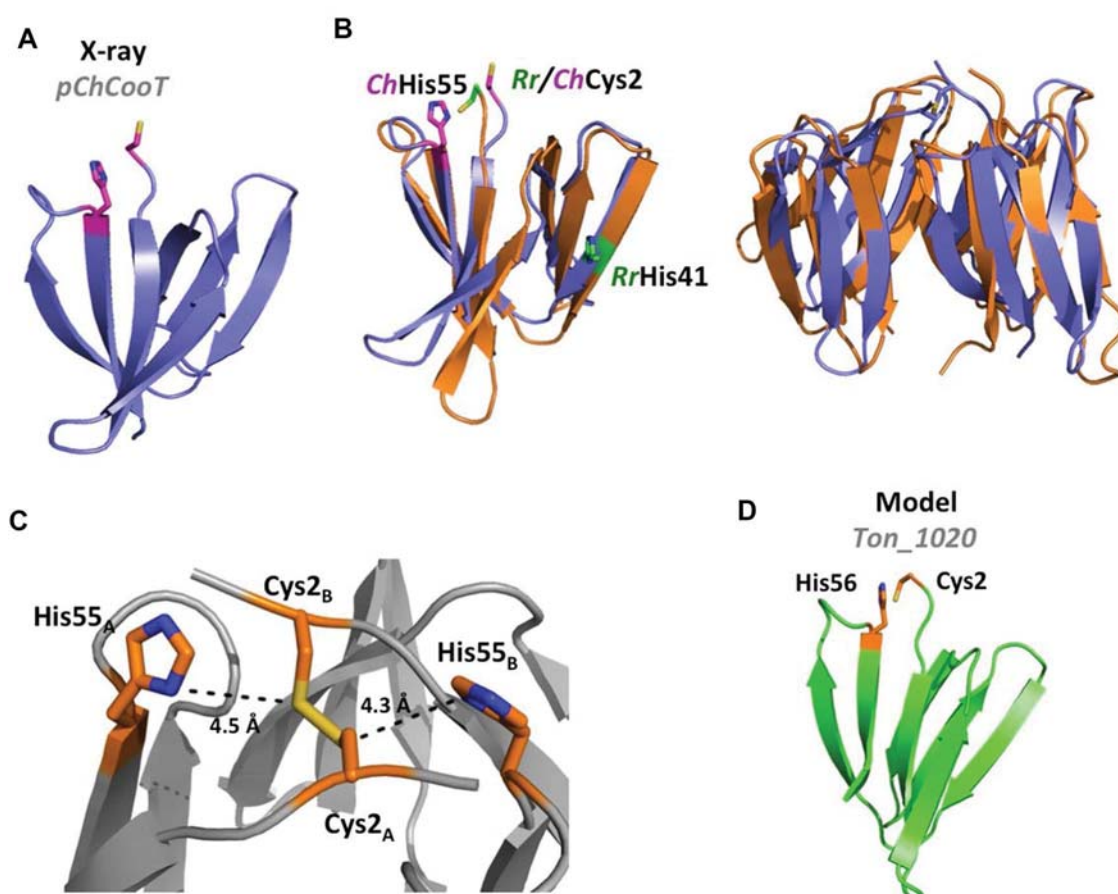
Ni(II) per dimer, pChCooT is assumed to bind 1 Ni(II) per dimer. Second, from the signal of the CD-based curves at 315 nm and taking into account the stoichiometry, an apparent binding constant was roughly estimated to be in the micromolar range (Fig. 2c). Third, the nickel-binding capability of the two mutants, bearing mutations on the amino acids most likely involved in the metal binding, was tested. In comparison with RrCooT, where the single mutation of cysteine2 into serine abolishes totally its capability to bind Ni(II), the pChCooT–C2S spectrum is only slightly different to that of the wild type, with a lower intensity and the disappearance of the 420 nm (–) signal (Fig. 2d). In the case of the pChCooT–C2S/H55A double mutant, the spectrum does not show any CD signal upon Ni(II) addition, highlighting the role of histidine 55 in Ni(II) binding.

### Crystallographic analyses of pChCooT

Pure pChCooT crystallized in the absence (Crystal 1) and in the presence of nickel (crystal 2). The structure was solved

by single-wavelength anomalous diffraction (SAD) using data of a single isomorphous Ni-containing crystal (crystal 2) diffracting to 2.7 Å. Five molecules are present in the asymmetric unit. Assembly analyses using PDBePISA showed that among the 5 molecules, monomers A and B, monomers C and D, and monomers E and E' form stable dimers with an interface surface area of 1640, 1710, and 1730 Å<sup>2</sup>, respectively. The best crystal (crystal 1) diffracted up to 2.0 Å (Table S1, Fig. 4a). Even though the sequence identity with RrCooT is low, the overall structures of monomers and dimers are similar with root-mean-square deviation values of 1.589 and 1.852 Å, respectively, considering all C $\alpha$ -atoms of all structures (Fig. 4b).

Further structure alignment analyses using PDBeFold showed that the most similar structure of pChCooT monomer is the RrCooT monomer. Both proteins adopt an identical topology with seven strands forming two bent anti-parallel  $\beta$ -sheets made of  $\beta$ 6,  $\beta$ 7,  $\beta$ 1, and  $\beta$ 2 ( $\beta$ -sheet I), and  $\beta$ 3,  $\beta$ 4, and  $\beta$ 5 ( $\beta$ -sheet II), respectively, tightly packed on top of each other. The pdb secondary structure analysis, using the



**Fig. 4** **a** X-ray structure of apo-pChCooT monomer (blue) (PDB code: 6FAN). **b** Superposition of monomer and dimer with RrCooT (PDB code: 5N76) in orange. **c** Proposed nickel-binding site in pCh-

CooT, **d** TON\_1020 model (green). Cysteine and histidine residues are in sticks. The structure figures were made with PyMOL (The PyMOL molecular graphic system, Version 2.0 Schrödinger, LCC)

DSSP method available in the 2Struc software, gave 72.6% of  $\beta$ -strand and 27.4% of others. pChCooT displays a Sm-like  $\beta$ -barrel fold present in Sm-like proteins, with distinct characteristics also found in RrCooT, namely, the dimeric architecture and the presence of the strand  $\beta_6$ .

The remarkable difference between RrCooT and pChCooT is the localization of the unique histidine residue. In RrCooT, His41 is present in the  $\beta_5$  strand belonging to  $\beta$ -sheet II, excluding its implication in Ni(II) coordination (Fig. 4b). On the contrary, in pChCooT, His55 is present at the edge of the  $\beta_7$  strand belonging to  $\beta$ -sheet I, at a distance of about 4.3–4.5 Å from Cysteine 2, suggesting a role in Ni(II)-binding, in agreement with the results obtained by CD on the pChCooT mutants (Fig. 4c). Unfortunately, the crystallization of pChCooT in complex with Ni(II) under anaerobic conditions in glove box did not lead to the formation of diffracting crystals, despite the use of crystallization screenings. Under aerobic conditions, the incubation of the protein with the reducing agent TCEP and Ni(II) prior to crystallization (Crystal 2) did not prevent the formation of disulfide bridges, present between Cys2A and Cys2B and Cys2C and Cys2D. This might not happen in vivo, where the reducing conditions of the cytoplasm should prevent thiol oxidation and let them able to coordinate Ni(II). Two Ni(II) ions are bound between His55A and His55C and His55B and His55D, forming a non-physiological tetramer in the structure. It is worth mentioning that the characterization of solvent-exposed metal-binding sites using protein crystallography is quite challenging. First, the crystallization conditions may destabilize metal binding, and second, the crystal packing may easily induce non-physiological binding of metal with amino acids belonging to neighboring molecules. Using Swiss model, modelling of Ton\_1020 was performed with its amino-acid sequence as target. When the automodel mode was running, one model was found (Fig. 4d) and the template chosen was the structure of RrCooT (PDB code: 5N76), suggesting that the product of the TON\_1020 gene could belong to the CooT family.

## Conclusions

In this study, we experimentally showed that the CHY\_RS00800 gene product is a nickel-binding protein belonging to the structural CooT family. In addition, we firmly predict the TON\_1020 gene product to be a CooT homologue. The presence of a histidine at position 55/56, not conserved in all CooT homologues, suggests the existence of different nickel-binding modes in this protein family. Moreover, functional studies are necessary to determine the role of CooT as an active player together with CooC in the CODH maturation pathway.

**Acknowledgements** This work was supported by “the ITERLIS PhD program, CEA Life sciences” for MA’s PhD funding, the “FUN-BIOCO” project (IDEX-UGA, Initiatives de Recherche stratégiques) and the “COSYNBIO” project (Projets exploratoires, Cellule énergie-CNRS). This work has been partially supported by the Labex ARCANE (ANR-11-LABX-0003-01). The research leading to these results has also received funding from the European Community (iNEXt support, Project ID no. 2310). We thank the High Throughput Crystallization Laboratory (HTX Lab) at the EMBL Grenoble for pChCooT crystallization. We thank the staffs from the BM-30A beamline of the European Synchrotron Facility in Grenoble. We thank Dr. Luca Signor (IBS, Grenoble) for mass spectrometry experiments in denaturing conditions (Integrated Structural Biology Grenoble platform).

## References

- Fontecilla-Camps JC, Amara P, Cavazza C, Nicolet Y, Volbeda A (2009) *Nature* 460:814–822
- Drennan CL, Peters JW (2003) *Curr Opin Struct Biol* 13:220–226
- Peters JW, Schut GJ, Boyd ES, Mulder DW, Shepard EM, Broderick JB, King PW, Adams MWW (2015) *Bba-Mol Cell Res* 1853:1350–1369
- Kung Y, Drennan CL (2011) *Curr Opin Chem Biol* 15:276–283
- Wu M, Ren QH, Durkin AS, Daugherty SC, Brinkac LM, Dodson RJ, Madupu R, Sullivan SA, Kolonay JF, Nelson WC, Tallon LJ, Jones KM, Ulrich LE, Gonzalez JM, Zhulin IB, Robb FT, Eisen JA (2005) *PLoS Genet* 1:563–574
- Domnik L, Merrouch M, Goetzl S, Jeoung JH, Leger C, Dementin S, Fourmond V, Dobbek H (2017) *Angew Chem Int Ed* 56:15466–15469
- Drennan CL, Heo JY, Sintchak MD, Schreiter E, Ludden PW (2001) *Proc Natl Acad Sci USA* 98:11973–11981
- Dobbek H, Svetlitchnyi V, Gremer L, Huber R, Meyer O (2001) *Science* 293:1281–1285
- Fesseler J, Jeoung JH, Dobbek H (2015) *Angew Chem Int Ed* 54:8560–8564
- Ribbe MW (2015) *Angew Chem Int Ed* 54:8337–8339
- Ciaccafava A, Tombolelli D, Domnik L, Jeoung JH, Dobbek H, Mroginski MA, Zebger I, Hildebrandt P (2017) *Angew Chem Int Ed* 56:7398–7401
- Spangler NJ, Lindahl PA, Bandarian V, Ludden PW (1996) *J Biol Chem* 271:7973–7977
- Jeon WB, Singer SW, Ludden PW, Rubio LM (2005) *J Biol Inorg Chem* 10:903–912
- Sydor AM, Lebrette H, Ariyakumaran R, Cavazza C, Zamble DB (2014) *J Biol Chem* 289:3828–3841
- Maier T, Lottspeich F, Bock A (1995) *Eur J Biochem* 230:133–138
- Yuen MH, Fong YH, Nim YS, Lau PH, Wong KB (2017) *Proc Natl Acad Sci USA* 114:10890–10898
- Zeier-Wanklyn CJ, Zamble DB (2017) *Curr Opin Chem Biol* 37:80–88
- Jeoung JH, Giese T, Grunwald M, Dobbek H (2009) *Biochemistry* 48:11505–11513
- Gregg CM, Goetzl S, Jeoung JH, Dobbek H (2016) *J Biol Chem* 291:18129–18138
- Inoue T, Takao K, Fukuyama Y, Yoshida T, Sako Y (2014) *Biosci Biotech Bioch* 78:582–587
- Kerby RL, Hong SS, Ensign SA, Coppoc LJ, Ludden PW, Roberts GP (1992) *J Bacteriol* 174:5284–5294
- Jeon WB, Cheng JJ, Ludden PW (2001) *J Biol Chem* 276:38602–38609
- Kerby RL, Ludden PW, Roberts GP (1997) *J Bacteriol* 179:2259–2266
- Watt RK, Ludden PW (1998) *J Biol Chem* 273:10019–10025

25. Timm J, Brochier-Armanet C, Perard J, Zambelli B, Ollagnier-de-Choudens S, Ciurli S, Cavazza C (2017) *Metallomics* 9:575–583
26. Kabsch W (2010) *Acta Crystallogr Sect D-Biol Crystallogr* 66:125–132
27. Adams PD, Afonine PV, Bunkoczi G, Chen VB, Davis IW, Echols N, Headd JJ, Hung LW, Kapral GJ, Grosse-Kunstleve RW, McCoy AJ, Moriarty NW, Oeffner R, Read RJ, Richardson DC, Richardson JS, Terwilliger TC, Zwart PH (2010) *Acta Crystallogr Sect D-Biol Crystallogr* 66:213–221
28. Emsley P, Lohkamp B, Scott WG, Cowtan K (2010) *Acta Crystallogr Sect D-Biol Crystallogr* 66:486–501
29. Chen VB, Arendall WB, Headd JJ, Keedy DA, Immormino RM, Kapral GJ, Murray LW, Richardson JS, Richardson DC (2010) *Acta Crystallogr Sect D-Biol Crystallogr* 66:12–21
30. Joosten RP, Long F, Murshudov GN, Perrakis A (2014) *IUCrJ* 1:213–220
31. Schut GJ, Lipscomb GL, Nguyen DMN, Kelly RM, Adams MWW (2016) *Front Microbiol* 7
32. Finn RD, Coggill P, Eberhardt RY, Eddy SR, Mistry J, Mitchell AL, Potter SC, Punta M, Qureshi M, Sangrador-Vegas A, Salazar GA, Tate J, Bateman A (2016) *Nucleic Acids Res* 44:D279–D285

Supplementary information for

**Biophysical and structural characterization of the putative nickel chaperone CooT  
from *Carboxydothemus hydrogenoformans***

M. Alfano<sup>1</sup>, J. Pérard<sup>1</sup>, R. Miras<sup>1</sup>, P. Catty<sup>1</sup> and C. Cavazza<sup>1\*</sup>

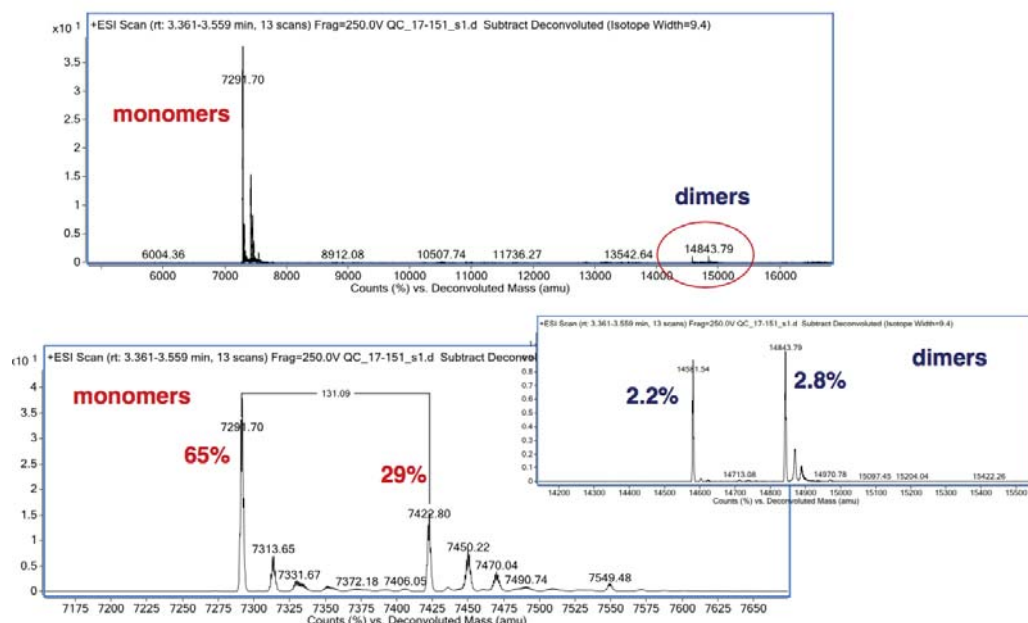
1. Univ. Grenoble Alpes, CEA, CNRS, BIG, CBM, F-38000 Grenoble

\* To whom correspondence should be addressed: [christine.cavazza@cea.fr](mailto:christine.cavazza@cea.fr)

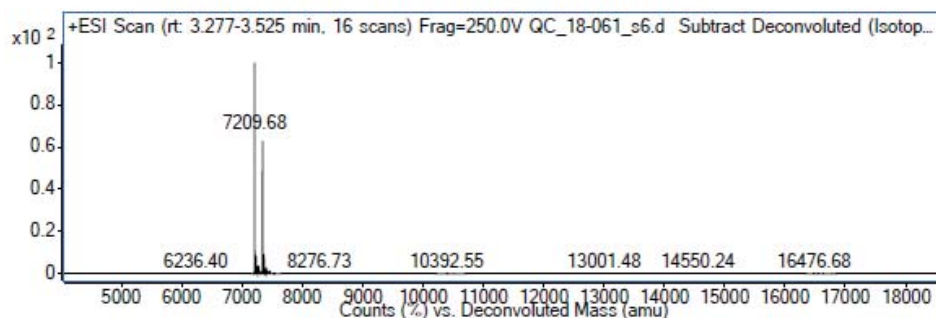


## Supplementary Figures

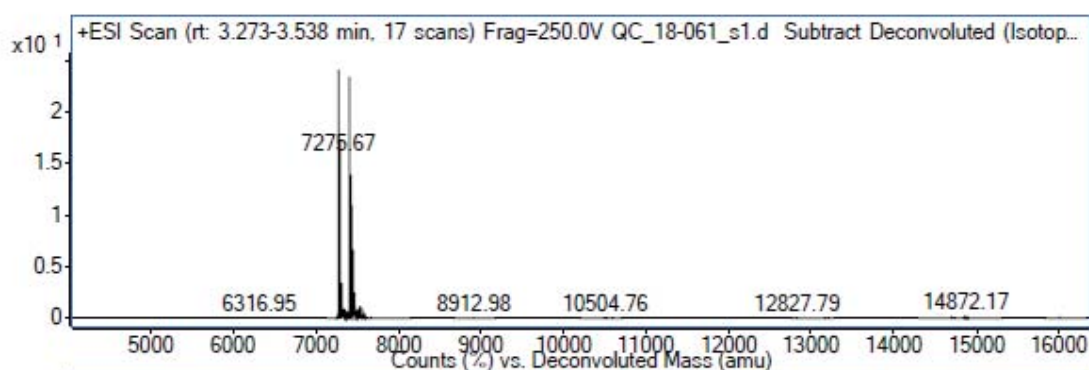
**Sample 1 : ChCooT**  
Diluted to 5uM in 0.03% TFA aq



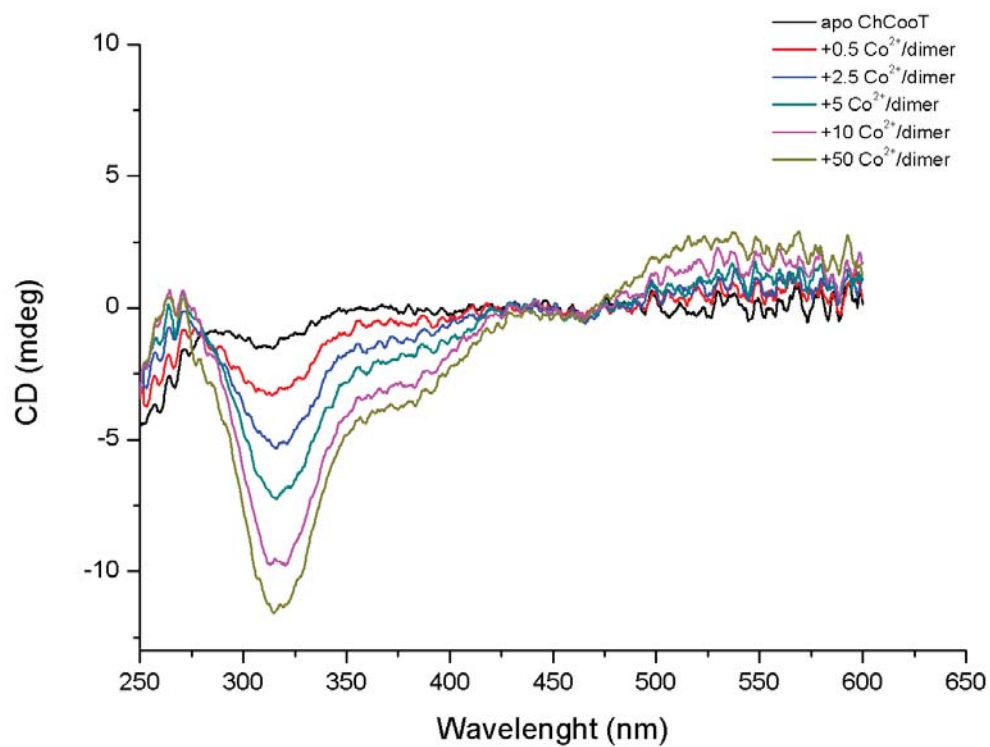
**Sample 6 : ChCooT-H55A-C2S**  
12mg/ml, 50 mM HEPES pH 7.5, 300 mM NaCl, 1 mM TCEP



**Sample 1 : ChCooT-C2S**  
4mg/ml, 50 mM HEPES pH 7.5, 300 mM NaCl, 1 mM TCEP



**Figure S1: ESI-MS spectra of pChCooT, pChCooT-C2S and pChCooT-C2S/H55A**  
(Buffer: 50 mM HEPES pH 7.0, 1 mM TCEP)



**Figure S2: CD spectra of in the near-UV visible region of CHY\_RS00800 titration with Co(II).** Spectra of protein solutions were recorded from 250 to 700 nm with 5 accumulations using a 10 mm cuvette. (Buffer: 50 mM HEPES pH 7.0, 1 mM TCEP).



## Supplementary Table

**Table S1: Crystallographic data and statistics for the pChCooT crystals.** The dimer-crystal was used for phasing and molecular replacement of the monomer-crystal.

| <b>Data collection</b>                               | <b><i>Crystal 1</i></b>   | <b><i>Crystal 2</i></b>   |
|--|---|---|
| Beamline   | BM30A (ESRF)  | BM30A (ESRF)  |
| Wavelength (Å)                                       | 0.979790  | 1.488037  |
| Space group  | <i>P</i> 2 <sub>1</sub>   | <i>C</i> 2  |
| Unit cell parameters                                 | a = 41.261 Å<br>b = 61.35 Å<br>c = 81.118 Å<br>$\alpha=\gamma=90^\circ$<br>$\beta=100.41^\circ$ | a = 170.073 Å<br>b = 54.136 Å<br>c = 37.425 Å<br>$\alpha=\gamma=90^\circ$<br>$\beta=92.071^\circ$ |
| Resolution range (Å)                                 | 33.85-2.0 [2.1-2.0]   | 39.07-2.7 [2.8-2.7]   |
| <i>R</i> <sub>sym</sub> (%)                          | 3.3 [56.6]  | 13.0 [93.4]   |
| Mean I/σ (I)   | 13.41 [1.61]  | 8.6 [1.7]   |
| Completeness (%)                                     | 97.6 [96.1]   | 97.3 [97.6]   |
| Redundancy   | 1.9 [1.9]   | 3.75 [3.8]  |
| <i>N</i> <sub>measured</sub>                         | 97738 [13204]   | 34661 [3596]  |
| <i>N</i> <sub>unique</sub>                           | 51640 [6956]  | 9224 [937]  |
| CC 1/2   | 99.8 [68.5]   | 99.6 [69.9]   |
| <b>Refinement</b>                                    |   |   |
| <i>R</i> factor/ <i>R</i> <sub>free</sub> factor (%) | 23.5 / 27.5   | 23.4/27.8   |
| No. atoms  | 3093  | 2540  |
| No. water molecules                                  | 239   | 15  |
| Average B-factor (Å <sup>2</sup> )                   | 48  | 63.4  |
| Rmsd bonds (Å)                                       | 0.0166  | 0.0108  |
| Rmsd angles (°)                                      | 1.951   | 1.597   |
| Ramachandran plot:                                   |   |   |
| Residues in most favorable region (%)                | 96.35   | 95.3  |
| Residues in disallowed region (%)                    | 0.0   | 0.3   |
| <b>PDB CODE</b>                                      | <b>6FAN</b>   |   |

## 2.3 Comparison of CooT with HypC

Interestingly, the N-terminal sequence (Met-Cys-Met-Ala) of *RrCooT* is similar to the motif found in the chaperone-like *E. coli* HypC. In this protein, the cysteine is not involved in Ni-binding and is needed to form the HypC-HycE complex and maintain the hydrogenase subunit HycE in a nickel-accessible conformation. Cys2 mutants of HypC could not interact anymore with HycE and accumulates the nickel-deficient precursor form of hydrogenase.<sup>163</sup> Notably, the Met1 of HypC is post-translationally cleaved in order to obtain a reactive thiolate, able to interact with precursor form of the large subunit HycE.<sup>163</sup> In the case of CooT we show that the Met1 is strictly needed in order to fully coordinate the metal. This point is intriguing: what if CooT acts both as a Ni chaperone and forms a complex with CODH, as observed for HypC? Could CooT be the possible linker needed to form the CooCTJ-CODH complex in order to fully activate the enzyme? As a possible proof of this supposition, it was postulated that the lack of ATP-stimulated activation of apoCODH during incubation with only CooC and CooJ could be related to the absence of CooT. In fact, the wild-type cell extract shows an ATP-enhanced activation, suggesting that CooT is one of the missing piece in the puzzle.<sup>112</sup> (*see introduction on CooC*)

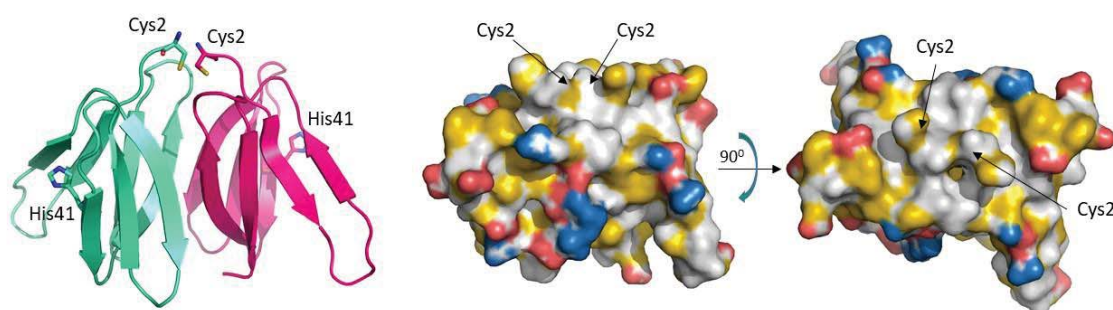
However, the crystal structures of two proteins clearly shown that CooT and HypC have a different conformation and do not belong to the same structural family (Figure 12).<sup>164</sup> Moreover they are not phylogenetically related. HypC possesses two separate domains, a N-terminal  $\beta$ -barrel, similar to CooT, with two additional C-terminal helices, absent in CooT. This region is important to guarantee the conformational flexibility, which could facilitate the association with their target proteins and facilitate the open enzyme conformation needed for the nickel insertion. Remarkably, in the case of *E. coli* HypC there is a His51 in close position to the Cys2, as we reported for the case of *C. hydrogenoformans* CooT (*see ChCooT section*). Cys2 is located in a hydrophobic region, common among the HypC homologues, postulated to facilitate the protein-protein interaction. The hydrophobic region are highlighted in Figure 12. HypC also interacts with HypD in order to coordinate the  $\text{Fe}(\text{CN})_2\text{CO}$  cofactor via the HypD-Cys41 and HypC-Cys2.<sup>165</sup> The HypDCE complex from *Thermococcus kodakarensis* is held through hydrophobic interactions, where the C-terminal region of HypC is not involved.<sup>166</sup> It has been shown recently that HypC C-terminal region is most likely involved in the interaction

with the large hydrogenase subunit but a mutant with a truncation in this region was still able to mature hydrogenase 3.<sup>167</sup> A truncation in this region brings the CooT and HypC structures closer together, both presenting a  $\beta$ -barrel conformation.

The similarity between these two proteins are intriguing and they could provide a possible clue to understand the CooT role during the CODH maturation process.

#### CooT *Rhodospirillum rubrum* (PDB: 5N76)

MCMAKVVLTK ADGGRVEIGD VLEVRAEGGA VRVTTLFDEE HAFPGLAIGR VDLRSGVISL IEEQNR



#### HypC *Escherichia coli* (PDB: 2OT2)

MCIGVPGQIR TIDGNQAKVD VCGIQRDVL TLVGSCDENG QPRVGQWVLV HVGFMASVIN EAEARDTLDA  
LQNMFDVEPD VGALLYGEEK

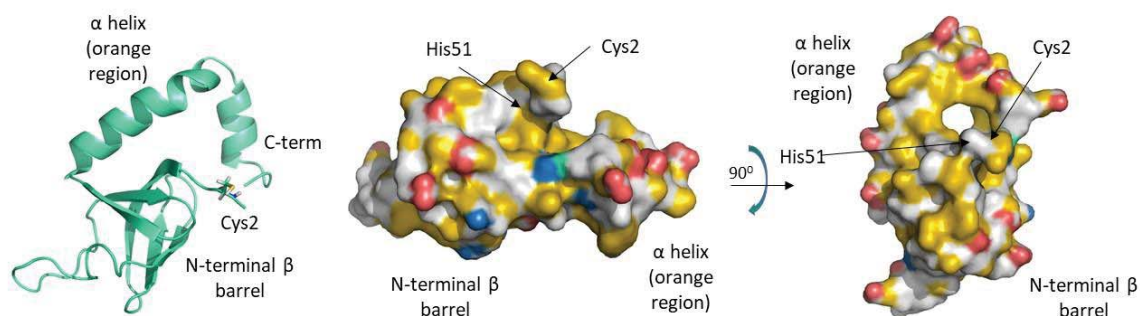
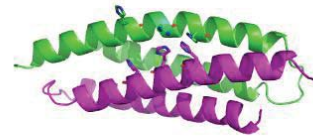


Figure 12. Sequence and crystal structures of CooT from *R. rubrum* and HypC from *E. coli*. The conserved sequence MC is highlighted in red, in orange for HypC is highlighted the part eliminated to produce the mutant. Notably this truncated HypC version was still able to mature correctly the hydrogenase. To visualize hydrophobic and charge contributions to protein interfaces we use the following highlighting scheme at atomic level using the colors yellow, red, and blue (YRB). In this scheme, all carbon atoms not bound to nitrogen and oxygen atoms are highlighted in yellow, nitrogen atoms in the side chains of lysine and arginine are blue, oxygen atoms in the side chains of glutamate and aspartate are red and all remaining atoms white<sup>168</sup>.



# Chapter 3



## CHAPTER 3

### *RrCooJ*

#### INTRODUCTION

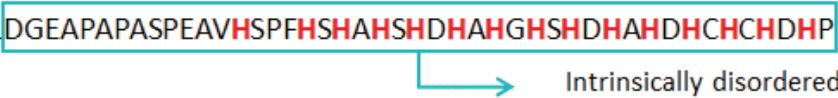
*RrCooJ* is a 12.6kDa protein with a nickel-binding domain containing 16 histidine and 2 cysteines at the C-terminus. As previously described in the introduction chapter, this protein is supposed to be involved in the *R. rubrum* CODH's nickel insertion.<sup>109</sup>

When we started working on *RrCooJ*, very few biochemical data (one publication in 1998<sup>125</sup>) and no structural information were available for this chaperone. One of the reason is because the protein was purified directly from *R. rubrum* cultures, a demanding strain to work with, with low yields (1.4 mg from 20L culture) that makes its characterization more difficult.<sup>125</sup> In addition, *RrCooJ* is classified as an intrinsically disordered protein, due to the presence of the long C-terminal histidine tail, a drawback for structural studies.

To overcome these problems, we first decided to produce *RrCooJ* in *E. coli*. After optimization of the purification protocol, high yields of pure *RrCooJ* (about 25 mg/L of culture) were easily obtained. With the heterologous expression in *E. coli*, site-directed mutagenesis was also more feasible, enabling us to remove the so defined “intrinsically disordered” region and produce a truncated version of the wild type protein called *RrCooJ*-Δ.

#### *RrCooJ* wild type (12638 Da)

MGTESPERGRKRLGIYLAHFLDHVEGHGEIGVQRDALAEDARLGALIDRALADMAVARASLNAV  
LRDL DGEAPAPASPEAVHSPFHSHAHSHDHAHGSHDHAHDHCHCHDHP



Intrinsically disordered region  
(psipred prediction)

#### *RrCooJ* Δ (7510 Da)

MGTESPERGRKRLGIYLAHFLDHVEGHGEIGVQRDALAEDARLGALIDRALADMAVARASLNAV  
LRDL

Figure 1. *CooJ* wild type and its truncated mutant. The predicted disordered region is boxed in cyan. The histidines are highlighted in red and it is possible to notice the presence of the preserved His18, 22 and 26 are in both sequences.

At this point, considering that 3 histidines are still present in the N-terminal part of CooJ-Δ (Figure 1), we had two main questions that needed to be addressed:

- 1) In the absence of the histidines tail, will the truncated version maintain its ability to bind nickel?
- 2) Considering that the disordered region was truncated, will now this version of the protein be able to crystallize?

To answer these questions we started the characterization of the *RrCooJ*-Δ mutant, comparing its behavior with the wild type one using *in cellulo* and *in vitro* tests. These results led us to produce several site-directed mutants that are described in the paper related to this part of the project entitled “*The CO-dehydrogenase accessory protein CooJ is a histidine-rich multidomain dimer with an unexpected additional Ni(II)-binding site*”.

In the following sections I will detail only the data not presented in this paper, mainly concerning the problems encountered during the crystallization of apo-*RrCooJ*-Δ.

The analysis of the *RrCooJ* stabilization and oligomerisation induced by nickel enabled us to publish an additional paper entitled “*Nickel-induced oligomerization of the histidine-rich metallochaperone CooJ from Rhodospirillum rubrum*”, reported at the end of this chapter.



### 3.1 Crystal structure of *RrCooJ-Δ*

When the apo *RrCooJ* was correctly produced and purified, we attempted to crystallize the protein in both the presence and absence of nickel. As expected, we did not obtain any promising condition from the crystals screening, suggesting that the predicted disordered his-tail hinders the protein crystallization.

Once the purified apo*RrCooJ-Δ* protein was obtained, we immediately started to investigate if this construct would crystallize. During my PhD, the CODH project was funded by the iNEXT consortium (Infrastructure for NMR, EM and X-rays for Translational Research), which enabled us to benefit from the HTXLab crystallization platform, present at the EMBL of Grenoble. In this way, multiple crystallization conditions could be automatically screened and imaged leading to the identification of the best possible working conditions. Once we obtained promising and satisfying results from the screening, I could reproduce the best crystallizations conditions in our laboratory.

The most promising condition was found in the classic\_suite screen as:

PEG 8K 10%, Ethylene glycol 8% and HEPES 0.1 M pH 7.5.

for a protein concentration of 9.9 mg/ml.

As it is possible to see in Figure 4, the apo*RrCooJ-Δ* protein crystallized in a needle-cluster fashion, difficult to exploit. Therefore, it was essential to manually optimize this condition to obtain bigger and improved crystals.

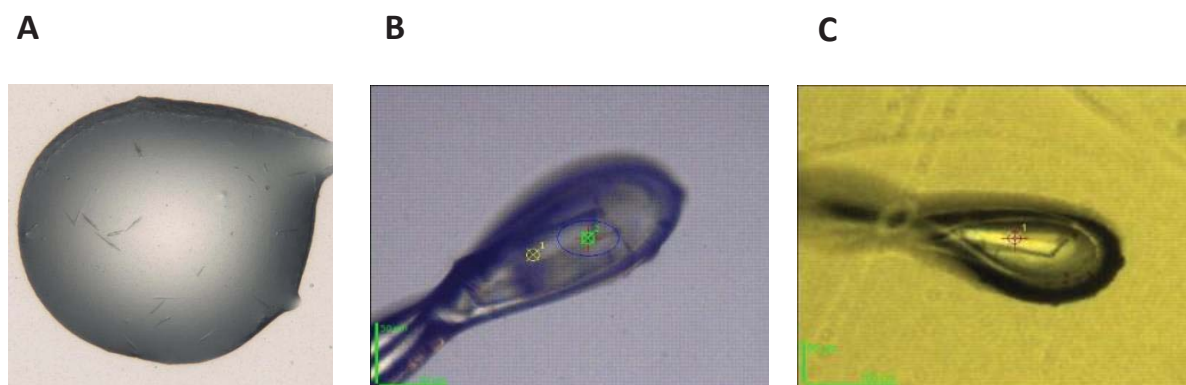


Figure 4. A) Htxlab best crystallization condition of apo-CooJΔ. B) Best apo-diffracting crystal after handmade optimization. A data set was collected at 2.9 Å resolution. C) Best diffracting crystal of holo-SeIMet CooJΔ. A data set was collected at 1.9 Å resolution.

To achieve this, finding the most accurate crystallization condition is the crucial step. The positive conditions obtained from the platform have to be considered as a preliminary idea of the promising percentage of the various components (precipitant and buffer) needed for the protein to crystallize. The first step in the laboratory is to play around with these concentrations and check how the protein crystallization is impacted. In our case, I checked a PEG range from 8 to 16%, Ethylene glycol from 6-10 % and HEPES 0.1 M pH 7.5, finding that the best condition was for PEG 8K 12%, Ethylene glycol 6% and HEPES 0.1 M pH 7.5. The starting protein concentration could also affect the crystallization process. For this reason, I also tested different protein concentrations (8.5-7.6-7-6-5 mg/ml) which led to either the same crystals typology or no crystal at all.

At this point, the crystals were still growing in a 'needle' shaped form, thus different methods were adopted to increase the crystal size. I first started testing the classical procedures for crystal's size increase, such as:

- Temperature control growth. The crystal plate is left at 4° C. Theoretically, a lower temperature should promote a slower crystallization, leading to bigger crystals.
- Micro seeding. An already crystallized protein (in our case needle shaped) is collected and placed in 'seed bead' tube containing a crystal stabilizer solution, composed of a mixture of protein sample and the crystallization reagent. In this way the crystal would neither dissolve nor growth. The tube is vortexed in order to crush the crystals and produce a seed stock that will be used as a nucleation center. Serial dilutions of this seed stock are prepared and used to make a new crystallization plate using the optimized crystallization conditions, one part of seed stock and one part of protein. In this way the crystallization is displaced from the metastable region, thus the protein cannot spontaneously nucleate but the seeds will grow larger crystals.

The micro seeding technique gave the best results, leading to a suitable crystal that could be collected at 2.9 Å resolution (*space group*  $P 6_2 2 2$ ) (Figure 4B).

Unfortunately, it was not possible to solve the structure using this single data set. In fact, considering that this protein does not have any structural homologues in the pdb, it was not possible to determine its crystal structure by molecular replacement. The phasing method has to be employed, for which the collection of the apo crystal is clearly not enough, as explained

in the methods session (*Chapter 6*). Soakings with different heavy compounds were tried to enable the phasing of different collections. However the crystals did not support any kind of soaking, even with its physiological metal nickel, losing their diffraction ability. For this reason, the protein was purified following Se-Met labelling. The crystallization conditions were re-optimized and new crystals were grown and tested. At this point, we realized that we had a problem: all the data collections performed so far were affected by twinning.

Twinning is an anomaly in the crystal growth, resulting in multiple domains partially sharing the crystal lattice in a symmetrical way that does not belong to the symmetry operation of the crystal point group. In other words, this leads to the intergrowth of two separate crystals sharing a surface in which the lattice points are partially or completely in common. This results in an overlap of the diffracted intensity during the X-ray diffraction collection that can be problematic for the space group determination. In order to solve this problem it is essential to identify the twinning typology and try to decode the related twinning law, related to the spatial orientation of the different species relative to each other. This twin law classify four different kinds of twins:

- 1) Merohedral twin. The lattices of two individual crystals are parallel to each other and they overlap entirely, hindering the twin identification. An additional symmetry is present in the crystal system, which is not present in the structure.
- 2) Pseudo-Merohedral twin. In this case, the crystal lattice of each individual has an extra element of pseudo-symmetry, which increases the system of the structure. Depending on the degree of twinning, it will overlap completely or partially with the reciprocal lattice.
- 3) Reticular merohedral twin. The lattice of each individual crystal are not parallel but they coincide in different geometries. The lattice possesses one or more element of symmetry than the original one.
- 4) Non-Merohedral twin. In this systems, the twin law does not belong to neither the crystal class of the structure or to the geometrical symmetry of the unit cell. This leads to a not complete overlap among the reciprocal lattices, making the data collection problematic.

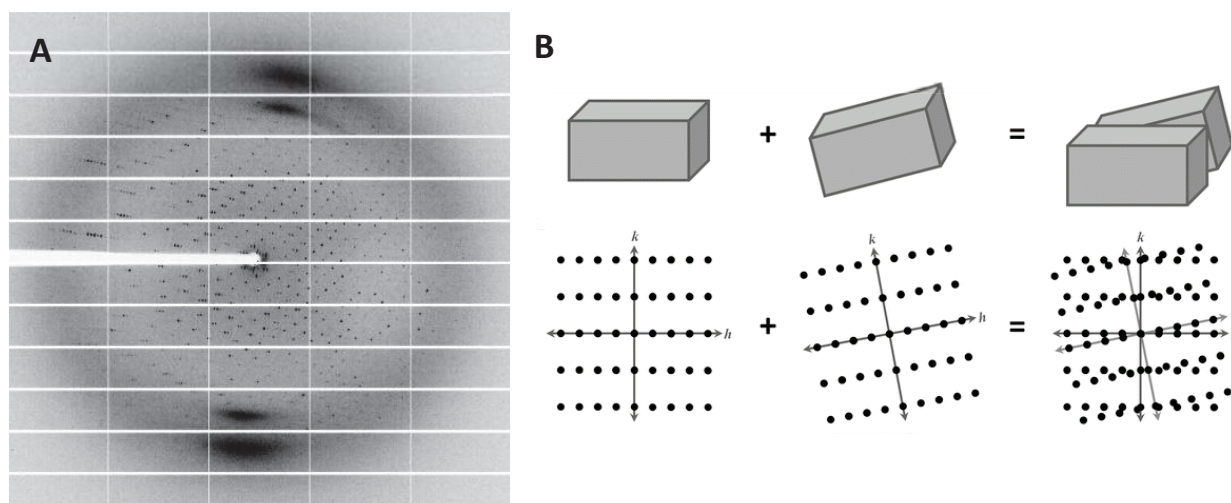


Figure 5. A) Diffraction of apo- *RrCooJ-Δ* crystal indexed as P 6 2 2 (89.1, 89.1, 222.8, 90, 90, 120). B) Nonspecifically overlapping crystals leading to a non-merohedral twin.

In our case, it resulted in a non-merohedral twin, with warning signs of an unusual large unit cell (89.17, 89.17, 223.04) and close and split reflections (Figure 5A). We did not manage to unveil the describing twinning law, needed to further correct and phase the data collection. It is important to remember that the structure needs to be solved via SAD phasing, making the twinning an unsolvable problem in this specific case.

At this point, our main option was to try to re-crystallize the protein by forcing its crystallization in a more “ordered” way. With the notion that a slower crystallization procedure might guarantee a different crystal packing, resulting in a more ordered crystal, I tested different techniques expected to promote this phenomena. Notable conditions were:

- Crystallization using mineral oil. The classical hanging-drop method is combined with the use of a thin layer of mineral oil above the crystallization mother liquid reservoir (Figure 6A). The oil plays a role in the diffusion rate between the crystallization drop and the reservoir, slowing down the equilibration and thus reaching the supersaturation more slowly. In this way, the speed of crystallization is reduced, allowing the growth of bigger and more reproducible crystals.<sup>169</sup>
- Crystallophore induced crystal formation. In this concept, a crystallophore (Tb-xo4 in Figure 6B) is used as a nucleating agent to improve the crystal nucleation and at the same time guarantee the possible phasing of the structure by using its anomalous signal.<sup>170</sup>

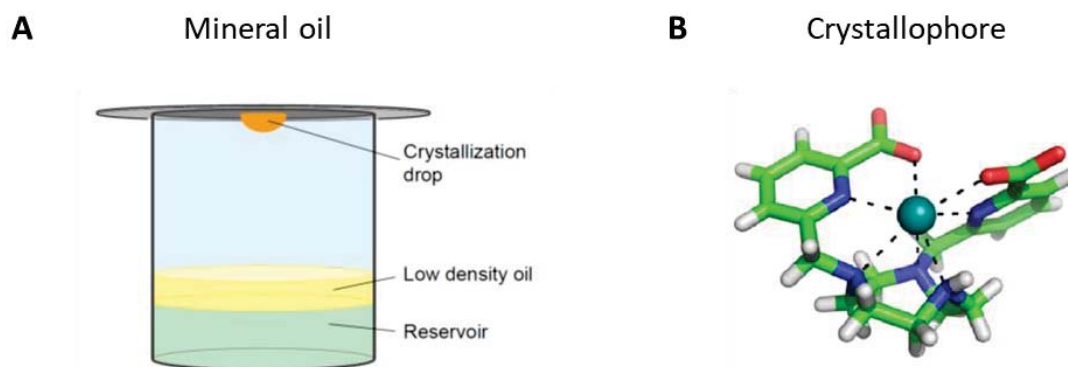


Figure 6. A) Hanging drop set-up containing an oil barrier lying on the reservoir crystallization solution.<sup>169</sup> B) Crystallophore Tb-XO4 is used to induce both protein nucleation and structure determination (via phasing using the anomalous scattering of terbium).<sup>170</sup>

Even though both methods sound promising, they both failed to change the crystal packing of apo-*RrCooJ-Δ* crystals. It is important to point out that, while these crystallization trials were performed, the *in vitro* characterization of the protein was going on in parallel. It was thanks to this that we realized an important characteristic of the protein. In fact, as you may have notice in the previous lines, I did not present the possibility of crystallizing the protein starting from the holo-form. This is due to the fact that the addition of nickel to a concentrated sample, the protein has to be at 9.9 mg/ml in order to crystallize, would cause immediate protein precipitation. However, if the metal was added to a diluted sample that is then concentrated, this process would let the protein remain soluble in solution. The discovery that the truncated protein binds nickel specifically and with a high affinity, made most of the crystallographic efforts focus on obtaining *RrCooJ-Δ* crystals with its natural bound metal. The holo- *RrCooJ-Δ* structure is described in the paper, therefore I will not discuss it, but it was the essential missing piece to be able to solve the twinned apo structure.

Once the holo- *RrCooJ-Δ* crystal structure was solved, we had a homology model that could be used to solve the apo-protein data set via molecular replacement, hoping that the two structures would not be dramatically different. What I did was to identify the real space group playing on lowering the possible symmetry of the twinned space group, re-scale the data set and run each of them using SHELX to find a plausible match with the phases of the holo-structure. Once this space group (we moved from a theoretical  $P 6_2 2 2$  to  $P 3_1 2 1$ ) was found, the phase could be confidently determined and the structure was solved via molecular

replacement. The result is a final structure of the apo protein at 3 Å resolution with a high B-factor. Due to the high displacement values, the position of the amino acids in the electron density cannot be confidently assigned. It is, however, possible to see that the positions of the nickel binding amino acids (His 18-22-26) are oriented accordingly to the holo structure (Figure 7A). As it is possible to see from the superimposition of the apo and holo structures, no major conformational changes are observed for this protein upon nickel addition (Figure 7C-D). This confirms that *RrCooJ-Δ* presents the same conformation in the presence or absence of nickel *in crystallo*.

(An acknowledgment for helping me solving the twinned structure goes to the 2017-CCP4 workshop, in particular to Doc. **Andrey Lebedev**.)

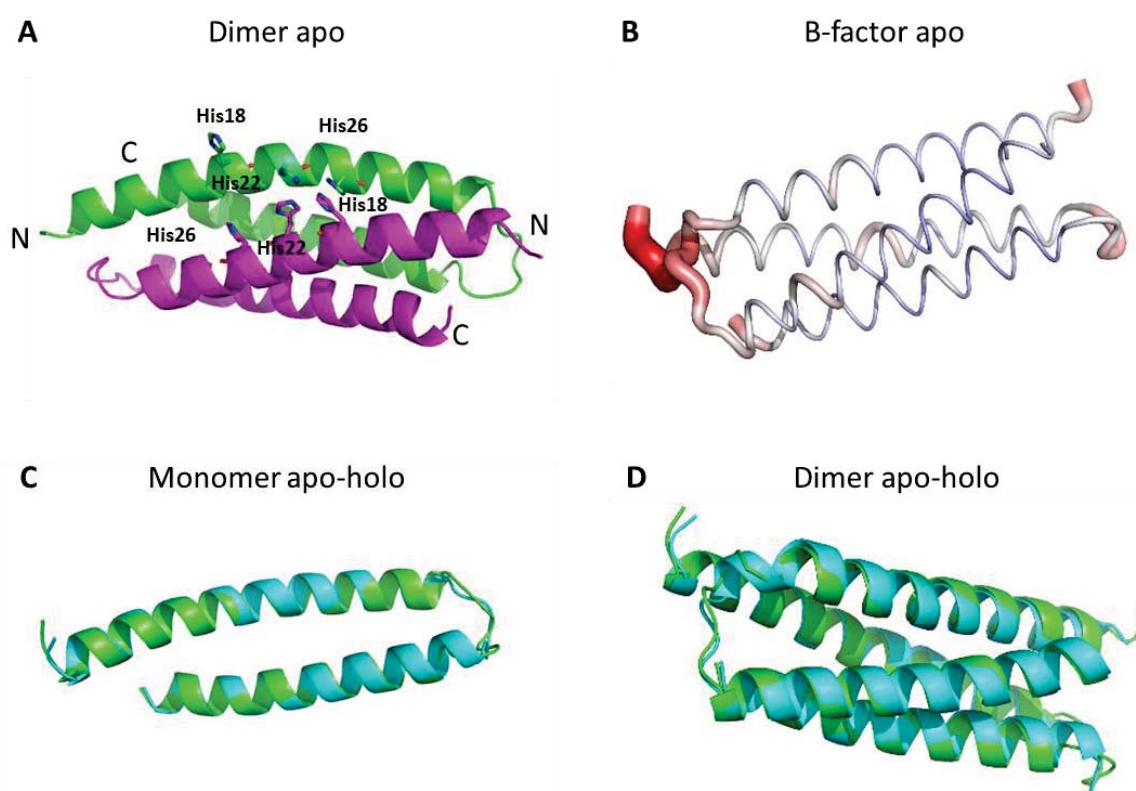


Figure 7. A) Apo- *RrCooJ-Δ* dimer, the amino acids involved in the nickel binding are depicted as sticks. B) B-factor representation of the structure (red-to-blue  $B_{min}=20$ ,  $B_{max}=200$ ), it is possible to see the most disordered regions are the loops and the N and C termini C) Overlay of the apo (blue) and holo (green) monomers. D) Overlay of the apo (blue) and holo (green) dimers.

### 3.2 The mystery of the 315 nm absorption peak

Another peculiar observation found during the UV-visible characterization of apo *RrCooJ*, is the presence of a 315nm-band, absent in *RrCooJ*-Δ. (Figure 8 A, B) Notably, this peak was also observed by Kerby et al.<sup>125</sup> when CooJ was purified from *R. rubrum* cultures, excluding it as a possible artifact due to the heterologous production in *E. coli* (Figure 8C). In order to remove and identify the nature of this peak, the protein was treated with the following methods:

- EDTA treatment (100 mM for 72 hours) in a glove box
- Guanidine treatment 6 M pH 7.2 ( O/N dialysis at 4 °C)
- Heat treatment of the guanidine treated samples (35 °C, 40 °C, 45 °C, 50 °C, 55 °C, 60 °C, 65 °C, 80 °C, 90 °C x 15 min)
- In glovebox dialysis in order to remove the TCEP from the apo-protein solution.

In addition, *RrCooJ* was purified under anaerobic conditions, to check a possible effect of the reducing agents used during the purification (DTT-TCEP) or if the purification in anaerobic conditions could have an impact. Surprisingly, after all these attempts the 315 nm peak is still present. Moreover, we did not observe any mass modification during the ESI-MS analysis and the two cysteines present in the sequence are reduced, excluding the hypothesis of an intra or inter molecular S-S bridge. The contribution of the cysteines to this signal is, however, strongly conceivable. In fact, the mutant lacking these cysteines (*RrCooJ*-2CS) does not present this peak in the Uv-vis spectrum.

Interestingly, Zamble et al. observed a similar absorption band at 320 nm in the histidine-rich protein SlyD from *E. coli*.<sup>171</sup> Also in their case the band did not cause a mass difference and it was not removed via EDTA treatment.

Despite several attempts from different research groups to identify it, the nature of this peak is still unknown and is currently under investigation in our group.



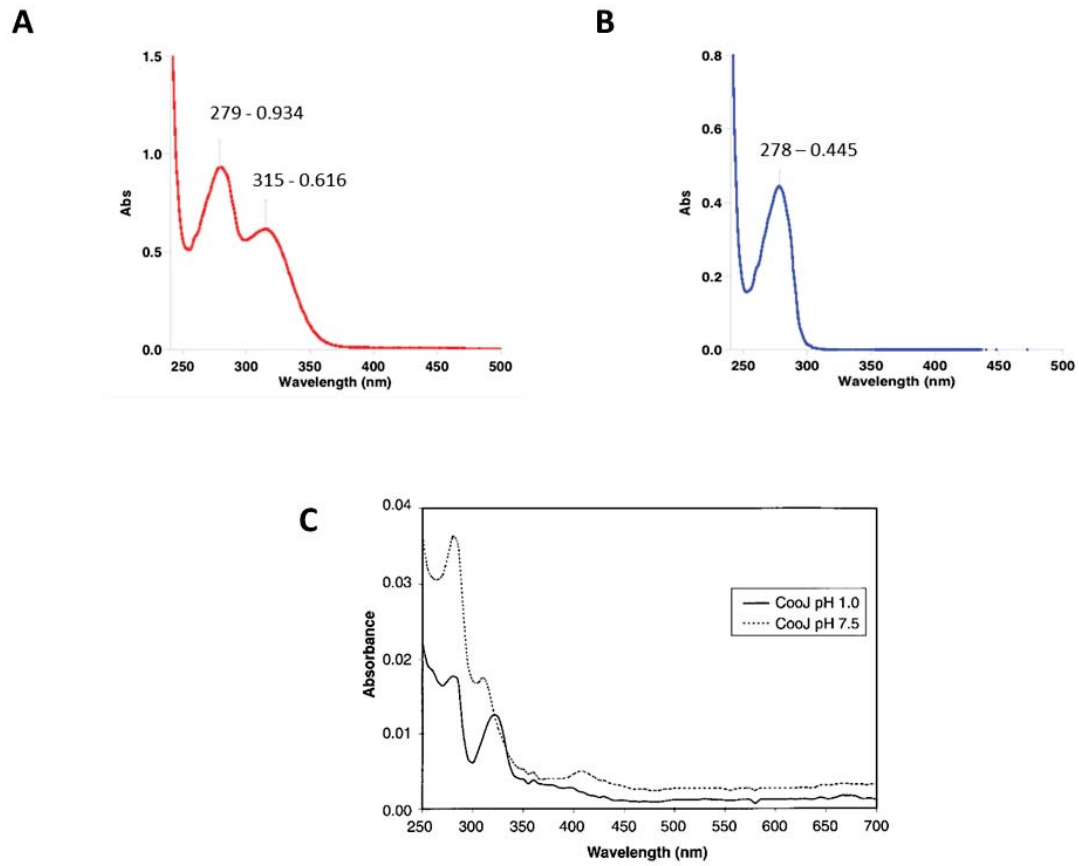


Figure 8. UV-Vis absorption spectra. A) Apo CooJ-WT shows the presence of an unexpected peak at 315 nm. B) Apo CooJ-Δ shows the expected peak at 278 nm due to the single tyrosine present in the sequence. C) Spectra for native CooJ purified from *R. rubrum*.<sup>125</sup>

### 3.3 RrCooJ papers

#### 3.3.1 Paper I

##### JBC ARTICLE



## The carbon monoxide dehydrogenase accessory protein CooJ is a histidine-rich multidomain dimer containing an unexpected Ni(II)-binding site

Received for publication, February 14, 2019, and in revised form, March 9, 2019. Published, Papers in Press, March 11, 2019, DOI 10.1074/jbc.RA119.008011

Marila Alfano<sup>‡</sup>, Julien Pérard<sup>‡</sup>, Philippe Carpentier<sup>‡</sup>, Christian Basset<sup>‡</sup>, Barbara Zambelli<sup>§</sup>, Jennifer Timm<sup>‡1</sup>, Serge Crouzy<sup>‡</sup>, Stefano Ciurli<sup>§</sup>, and Christine Cavazza<sup>‡2</sup>

From the <sup>‡</sup>Laboratory of Chemistry and Biology of Metals, Université Grenoble Alpes, CEA, CNRS, F-38000 Grenoble, France and the <sup>§</sup>Laboratory of Bioinorganic Chemistry, Department of Pharmacy and Biotechnology, University of Bologna, I-40127 Bologna, Italy

#### Paper highlights

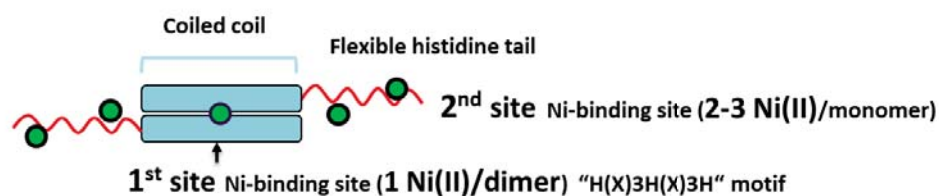
- *RrCooJ* possesses 16 histidines and 2 cysteines in the C-terminus and it forms a homodimer in solution possessing a coiled coil with two independent and highly flexible His-tails.
- *RrCooJ* binds several nickel (four) per dimer on the histidine rich motif, with a dissociation constant of 1.6  $\mu$ M.
- Surprisingly, a truncated version of the protein, *RrCooJ*- $\Delta$ , retains the ability to bind one nickel per dimer with a dissociation constant of 13 nM.
- The structure of *RrCooJ*- $\Delta$  was solved by X-ray crystallography, showing the additional nickel binding site at the protein interface and unveiling the H(W/F)X<sub>2</sub>HX<sub>3</sub>H motif as the finger print to recognize the CooJ family.
- 46 *RrCooJ* bacterial homologues have been identified thanks to this motif, which is strictly conserved.
- *RrCooJ* possesses two separate nickel-binding sites: one able to bind 4 nickel per dimer involving the flexible and variable histidine tails and a second one able to bind 1 nickel per dimer using the H(W/F)X<sub>2</sub>HX<sub>3</sub>H motif.
- The presence of a double set of sites could indicate a dual function for CooJ, as nickel chaperone and nickel storage protein.

## Graphical Abstract

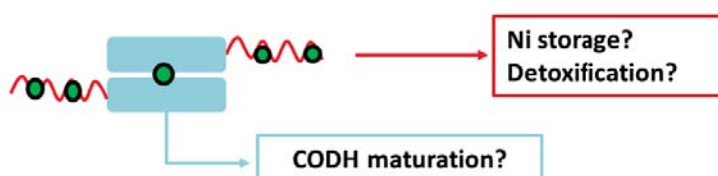
RrCooJ-Δ



RrCooJ-WT



Dual function?



# The carbon monoxide dehydrogenase accessory protein CooJ is a histidine-rich multidomain dimer containing an unexpected Ni(II)-binding site

Received for publication, February 14, 2019, and in revised form, March 9, 2019. Published, Papers in Press, March 11, 2019. DOI 10.1074/jbc.RA119.008011

Marila Alfano<sup>‡</sup>, Julien Pérard<sup>‡</sup>, Philippe Carpentier<sup>‡</sup>, Christian Basset<sup>‡</sup>, Barbara Zambelli<sup>§</sup>, Jennifer Timm<sup>‡1</sup>, Serge Crouzy<sup>‡</sup>, Stefano Ciarli<sup>§</sup>, and Christine Cavazza<sup>‡2</sup>

From the <sup>‡</sup>Laboratory of Chemistry and Biology of Metals, Université Grenoble Alpes, CEA, CNRS, F-38000 Grenoble, France and the <sup>§</sup>Laboratory of Bioinorganic Chemistry, Department of Pharmacy and Biotechnology, University of Bologna, I-40127 Bologna, Italy

Edited by Ruma Banerjee

Activation of nickel enzymes requires specific accessory proteins organized in multiprotein complexes controlling metal transfer to the active site. Histidine-rich clusters are generally present in at least one of the metallochaperones involved in nickel delivery. The maturation of carbon monoxide dehydrogenase in the proteobacterium *Rhodospirillum rubrum* requires three accessory proteins, CooC, CooT, and CooJ, dedicated to nickel insertion into the active site, a distorted [NiFe<sub>3</sub>S<sub>4</sub>] cluster coordinated to an iron site. Previously, CooJ from *R. rubrum* (RrCooJ) has been described as a nickel chaperone with 16 histidines and 2 cysteines at its C terminus. Here, the X-ray structure of a truncated version of RrCooJ, combined with small-angle X-ray scattering data and a modeling study of the full-length protein, revealed a homodimer comprising a coiled coil with two independent and highly flexible His tails. Using isothermal calorimetry, we characterized several metal-binding sites (four per dimer) involving the His-rich motifs and having similar metal affinity ( $K_D = 1.6 \mu\text{M}$ ). Remarkably, biophysical approaches, site-directed mutagenesis, and X-ray crystallography uncovered an additional nickel-binding site at the dimer interface, which binds Ni(II) with an affinity of 380 nM. Although RrCooJ was initially thought to be a unique protein, a proteome database search identified at least 46 bacterial CooJ homologs. These homologs all possess two spatially separated nickel-binding motifs: a variable C-terminal histidine tail and a strictly conserved H(W/F)X<sub>2</sub>HX<sub>3</sub>H motif, identified in this study, suggesting a dual function for CooJ both as a nickel chaperone and as a nickel storage protein.

*Rhodospirillum rubrum* is a purple nonsulfur photosynthetic bacterium, able to obtain energy through a variety of mechanisms depending on the growth conditions. This versatile bacterium is a facultative anaerobe, capable of using CO as a sole energy source during anaerobic growth in darkness through the water–gas shift reaction ( $\text{H}_2\text{O} + \text{CO} \rightarrow \text{H}_2 + \text{CO}_2$ ) (1). The biologically mediated water–gas shift reaction is catalyzed by two nickel-dependent enzymes, namely a monofunctional [NiFe]-carbon monoxide dehydrogenase (CODH)<sup>3</sup> coupled to a CO-tolerant energy-conserving [NiFe]-hydrogenase (2). CO is oxidized to CO<sub>2</sub> by CODH, releasing electrons used for the reduction of two protons to yield H<sub>2</sub>, the latter reaction being catalyzed by the hydrogenase. This CO-oxidizing complex contains also CooF, a ferredoxin that transfers electrons from CODH to hydrogenase. As a result, the implication of these two essential nickel enzymes in CO metabolism requires a highly controlled and effective nickel supply to ensure their correct activation and catalytic efficiency.

The crystal structures of CODH from *R. rubrum* (RrCODH) (3) and *Carboxydotherrmus hydrogenoformans* (4) were first reported in 2001, revealing the unique architecture of its active site, called C-cluster, constituted of a [NiFe<sub>3</sub>S<sub>4</sub>] cluster and a mononuclear Fe(II) site. Previous studies have shown that the iron/sulfur unit is inserted prior to nickel. Indeed, when *R. rubrum* is grown in nickel-depleted medium, the purified RrCODH corresponds to a stable, but inactive, nickel-deficient form (5, 6). The addition of nickel salts to nickel-free CODH in the presence of CO under reductive conditions is sufficient to convert the nickel-free C-cluster into a functional C-cluster, proving that nickel insertion is the key step of CODH activation (7). However, the mechanism of nickel import, trafficking and accumulation prior to its insertion into CODH, remains a subject that has been poorly investigated to date.

In *R. rubrum*, a CODH operon, named *cooFSCTJ* (8), is located downstream the hydrogenase operon, named *coo-MKLXUH* (9). Both operons are CO-induced via the CO-sens-

This work was supported by the ITERLIS Ph.D. program, CEA Life Sciences Ph.D. funding (to M. A.), the FUNBIOCO Project (IDEX-UGA, Initiatives de Recherche Stratégiques), the COSYNBIO project (Projets Exploratoires, Cellule Energie–CNRS), and the Department of Pharmacy and Biotechnology of the University of Bologna. This work has also been supported in part by Labex ARCAN and CBH-EUR-GS Grant ANR-17-EURE-0003. The research leading to these results has received funding from the European Community's Seventh Framework Programme H2020 under iNEXT H2020 Grant 653706 and networking support from COST Action FeSBioNet Contract CA15133. The authors declare that they have no conflicts of interest with the contents of this article.

The atomic coordinates and structure factors (code 6HK5) have been deposited in the Protein Data Bank (<http://www.pdb.org/>).

This article contains supporting text, Tables S1–S3, and Figs. S1–S4.

<sup>1</sup> Present address: Dept. of Biochemistry and Molecular Pharmacology, University of Massachusetts Medical School, Worcester, MA 01605.

<sup>2</sup> To whom correspondence should be addressed. Tel.: 33-4-38-78-91-16; E-mail: [christine.cavazza@cea.fr](mailto:christine.cavazza@cea.fr).

<sup>3</sup> The abbreviations used are: CODH, carbon monoxide dehydrogenase; SAXS, small-angle X-ray scattering; SEC, size-exclusion chromatography; MALLS, multiple-angle laser light scattering; RI, refractive index; ITC, isothermal titration calorimetry; NSD, normalized spatial discrepancy; TCEP, tris(2-carboxyethyl)phosphine; CATH, class (C) architecture (A) topology (T), and homologous superfamily (H).

## A novel Ni(II)-binding site in the nickel chaperone *CooJ*

ing transcriptional activator, *CooA*, found downstream of the *cooFSCTJ* operon (10). The five open-reading frames present in the CODH operon encode the ferredoxin *CooF*, CODH (*CooS*), and three additional nickel-dependent proteins (*CooC*, *CooT*, and *CooJ*) shown to function specifically for nickel insertion into CODH (8). However, the precise role of the latter three proteins in the maturation pathway, from nickel acquisition to its transfer into the enzyme, is not well-understood.

*CooC* (11) is widely distributed in microorganisms possessing a [NiFe]-CODH. The protein is an ATPase, analogous to the GTPases *HypB* (12) and *UreG* (13) required for nickel-processing for hydrogenase and urease, respectively. Today, NTPases are generally recognized as common components of the maturation processes of nickel enzymes, potentially playing a regulatory role by affecting protein–interactions in multimetric chaperonin complexes and/or modulating nickel ion affinity for intermediate metal-binding sites.

*CooT* from *R. rubrum* (*RrCooT*) was recently shown to be a homodimer of 14 kDa that specifically binds one Ni(II) per dimer (14). Initially identified in *R. rubrum* and *Rhodospseudomonas palustris* genomes, phylogenetic analyses have identified 111 *CooT* homologs in anaerobic bacteria and Archaea (14). In all cases, the presence of *CooT* is related to an anaerobic and metal-related metabolism.

*RrCooJ* is a metallochaperone with only few biochemical data and no structural information available to date (15). This 12.5-kDa protein contains a C-terminal nickel-binding domain with 16 histidines and 2 cysteines in the final 34 amino acids (see Fig. 1A). The protein purified directly from *R. rubrum* cultures was reported to bind four Ni(II) per monomer with a  $K_D$  value of 4.3  $\mu\text{M}$ , as shown by equilibrium dialysis (15). During the purification, *RrCooJ* co-eluted with *RrCODH* and other additional proteins, suggesting the formation of a multiprotein complex bound to CODH (15). Interestingly, the histidine-rich region of *RrCooJ* is partially dispensable for its physiological function: a truncated version of *RrCooJ* with only six histidines and no cysteine in the C-terminal part is still able to activate CODH *in vivo*, suggesting that the His tail could play another role in the bacterium, in addition to CODH maturation (8). Ni(II) transport occurs in *R. rubrum* grown either in the presence or in the absence of CO, excluding a role for *CooC*, *CooT*, and *CooJ* in Ni(II) import. However,  $^{63}\text{Ni(II)}$  accumulates more rapidly in CO-induced cells, essentially because of nickel accumulation in CODH, which is highly dependent on the presence of *CooC*, *CooT*, and *CooJ*, proving that all three proteins must be present for *in vivo* nickel insertion into CODH (16).

Here, we report a detailed characterization of *RrCooJ*, which reveals the existence of an unexpected nickel-binding site in the N-terminal region. Biophysical approaches using the WT protein and site-directed mutants allowed the determination of nickel-binding sites in the N-terminal region and in the C-terminal histidine-rich tail. Structural analyses using X-ray crystallography, small-angle X-ray scattering (SAXS), and modeling provided detailed information on the structure of *RrCooJ* and allowed the characterization of the novel Ni(II)-binding

site. Based on the presence of two spatially separated Ni(II)-binding sites, this study suggests a double function for *RrCooJ* as a Ni(II) chaperone and a Ni(II) storage protein.

## Results

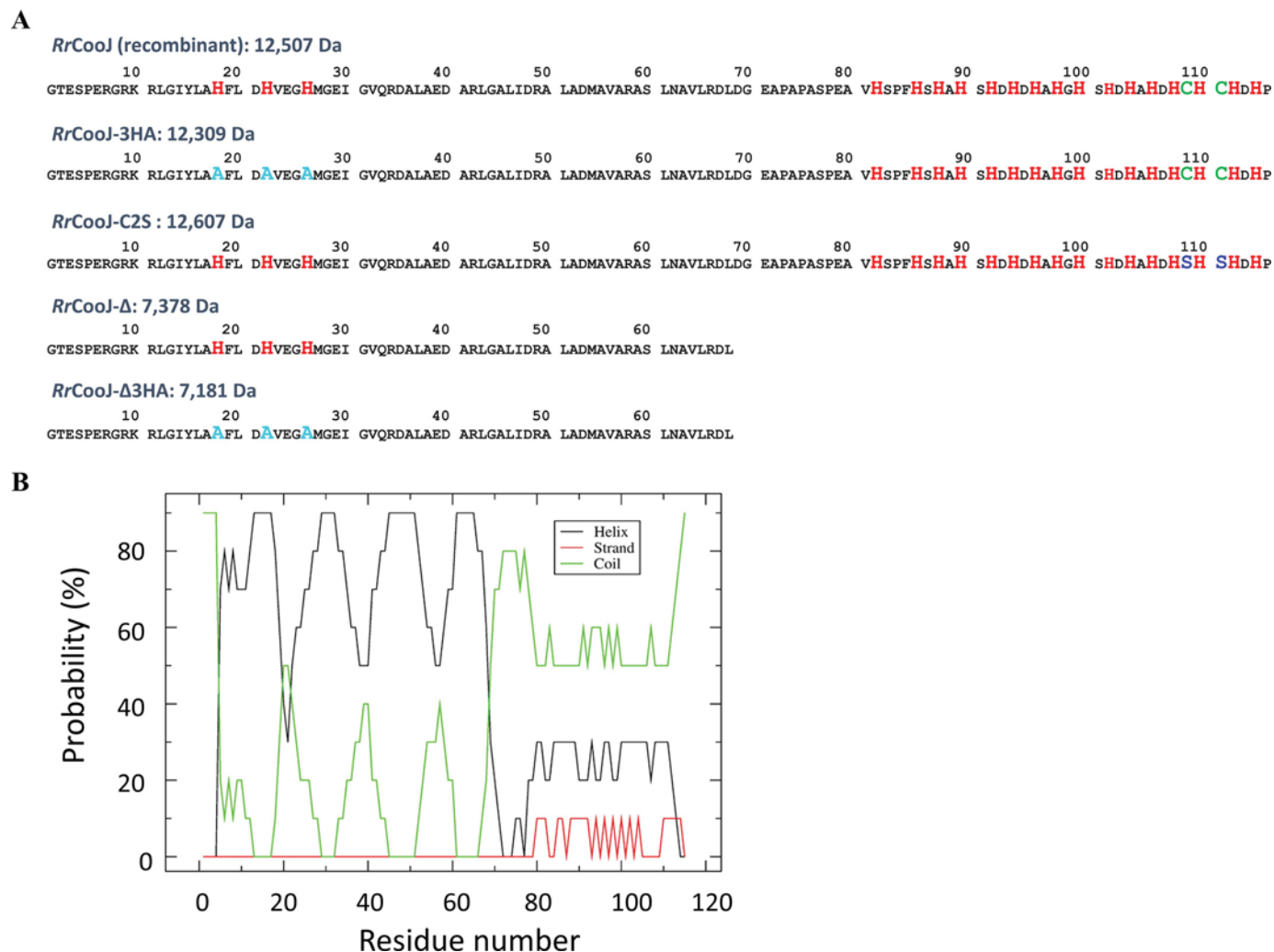
### *RrCooJ* is an $\alpha$ -helical dimeric protein

WT *RrCooJ* is a 12,581-Da protein (comprising Met1) possessing a histidine-rich tail that features 16 histidines and 2 cysteines. The recombinant protein (12,507 Da) (Fig. 1A) was overproduced in *Escherichia coli* and purified to homogeneity, as confirmed by electrospray ionization–MS (Fig. S1). Size-exclusion chromatography coupled with multiple-angle laser light scattering and refractive index (SEC-MALLS-RI) reveals a homodimeric state of the apoprotein in solution, with the peak eluting at  $\sim 25$  kDa (Fig. 2A). The results of secondary structure prediction using PredictProtein (Fig. 1B), show four  $\alpha$ -helical motifs in the N-terminal region from residues 4–68, followed by a mostly disordered region from residues 69 to 115, as expected from its high content in histidines. However, a weak  $\alpha$ -helical and  $\beta$ -strand secondary structure propensity is also expected in the C-terminal portion. The CD spectrum of apo-*RrCooJ* in the far-UV region is consistent with this prediction (Fig. 2B). A truncated form of the protein, lacking the C-terminal region from Asp-69 to Pro-115 (*RrCooJ*- $\Delta$ ; Fig. 1A), was also constructed (Table S1), as confirmed by electrospray ionization–MS (Fig. S1B). Similar to *RrCooJ*, *RrCooJ*- $\Delta$  behaved as a homodimer in solution, as shown by SEC-MALLS-RI (Fig. 2A) with an eluting peak at  $\sim 15$  kDa, and it mainly contains  $\alpha$ -helices as revealed by the far-UV CD spectrum (Fig. 2B).

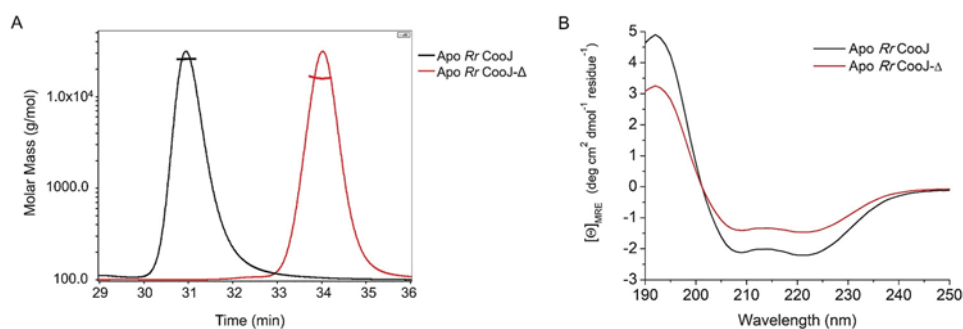
### The nickel-binding properties of *RrCooJ* and its mutants *in vitro*

Inductively coupled plasma atomic emission spectroscopy measurements showed that purified *RrCooJ* and its mutants were devoid of any bound metal ion (Table S2). CD and isothermal titration calorimetry (ITC), coupled with site-directed mutagenesis, were used to investigate the Ni(II)-binding properties of *RrCooJ*. CD spectra in the nearly UV-visible range revealed the appearance of a ligand-to-metal charge transfer transition at 270 (–) nm, with saturation at  $\sim 7$  Ni(II) molar equivalent per dimer upon addition of Ni(II) to *RrCooT*. A second signal at 330 (+) nm appeared when the Ni(II) concentration reached 4 molar equivalents per dimer and saturated at  $\sim 7$  molar equivalents per dimer (Fig. 3A). The signal at 330 (+) nm most likely corresponds to a nickel-thiolate ligand-to-metal charge transfer, because it is not observed in the CD spectrum of the *RrCooJ*-C2S mutant (Fig. 3B), in which Cys-109 and Cys-111 are mutated to serine (Fig. 1A). However, the 270 (–) nm signal is still present in the CD spectrum of *RrCooJ*-C2S, even though at lower intensity (Fig. 3B). This suggests that  $\sim 4$  Ni(II) ions initially bind to the dimeric *RrCooJ* via histidine residues, followed by the binding of additional Ni(II) ions with the involvement of the two cysteine residues. The truncation of the histidine-rich tail in *RrCooJ*- $\Delta$ , does not completely abolish Ni(II)-binding, as indicated by the appearance of the CD signal at 270 (–) nm, saturating at 1 Ni(II) molar equivalent per dimer upon Ni(II) addition (Fig. 3C). To investigate the role of the





**Figure 1. Amino acid sequences and secondary structure prediction of *RrCooJ*.** A, amino acid sequences of recombinant *RrCooJ* and mutants. Histidines are in red. Cys-109 and Cys-111 are in green. In *RrCooJ*-3HA and *RrCooJ*-Δ3HA mutants, His-18, His-22, and His-26 are mutated in Ala (in cyan). In *RrCooJ*-C2S mutant, Cys-109 and Cys-111 are mutated in Ser (in blue). B, secondary structure elements of *RrCooJ*-WT predicted using the Predict-Protein server (<http://www.predictprotein.org>).<sup>4</sup>



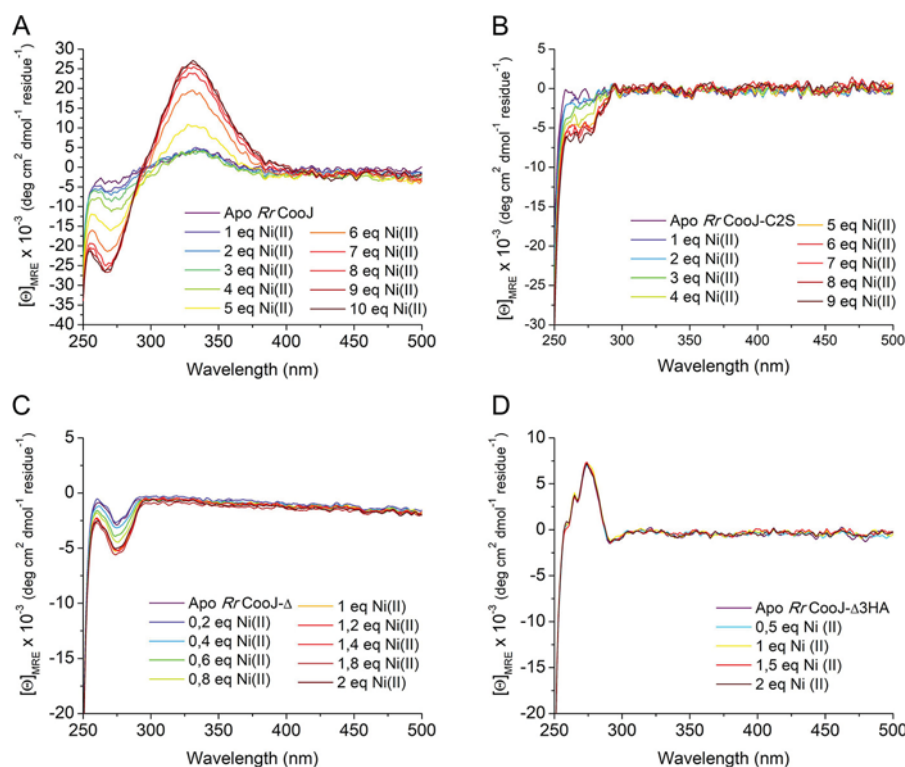
**Figure 2. SEC-MALLS-RI and CD in far UV of apo*RrCooJ* and apo*RrCooJ*-Δ.** A, SEC-MALLS-RI profile for apo*RrCooJ* (black) and apo*RrCooJ*-Δ (red). Both proteins are in a dimeric conformation in solution with a mass of  $25.7 \pm 0.2$  kDa for apo*RrCooJ* and  $15.7 \pm 0.2$  kDa for apo*RrCooJ*-Δ. The SEC-MALLS-RI spectra were recorded for a protein concentration of 750  $\mu$ M dimer (apo*RrCooJ*) and 285  $\mu$ M dimer (apo*RrCooJ*-Δ) in 50 mM HEPES, pH 7.5, 300 mM NaCl, 1 mM TCEP. B, the CD spectra in the 190–250-nm range shows the secondary structure composition of the protein in solution. The apo*RrCooJ* spectrum (black) was recorded at 12  $\mu$ M dimer concentration in 8 mM HEPES, pH 7.2, 10 mM NaCl. The apo*RrCooJ*-Δ spectrum (red) was recorded at 14  $\mu$ M dimer concentration in 9 mM HEPES, pH 7.2, 10 mM NaCl.

three putative nickel-coordinating residues present in the N-terminal part of *RrCooJ*, namely His-18, His-22, and His-26, the triple mutant *RrCooJ*-Δ3HA (Fig. 1A) was constructed and purified. This mutant does not show any CD signal (Fig. 3D), demonstrating that the observed transition band in the dimeric

*RrCooJ*-Δ mutant is due to the coordination of a single Ni(II) to His-18, His-22, and/or His-26.

ITC was applied to measure the Ni(II) binding properties of *RrCooJ* and its variants. In all cases but *RrCooJ*-Δ3HA, Ni(II) titrations produce negative peaks, indicating the occurrence of

## A novel Ni(II)-binding site in the nickel chaperone CooJ



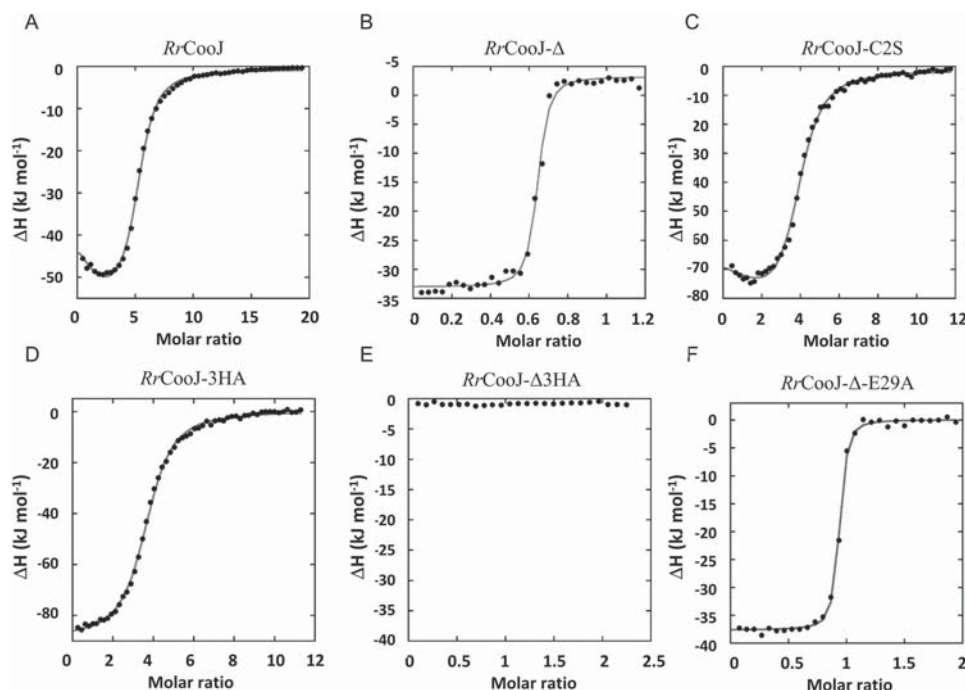
**Figure 3.** Binding properties of Ni(II) to *RrCooJ* and its variants using CD in the near-UV visible region (250–700 nm). A–D, Ni(II) titration of *RrCooJ* at 20  $\mu\text{M}$  dimer (A), *RrCooJ*-C2S at 20  $\mu\text{M}$  dimer (B), *RrCooJ*- $\Delta$  at 100  $\mu\text{M}$  dimer (C), and *RrCooJ*- $\Delta$ 3HA at 50  $\mu\text{M}$  dimer (D). The buffer contained 50 mM HEPES, pH 7.5, 300 mM NaCl (with the addition of 1 mM TCEP for spectra; A).

an exothermic binding reaction. A fit of the integrated heat data for *RrCooJ* (Fig. 4A and Fig. S2A) using the “one-set-of-sites” model was not satisfactory ( $\chi^2_v = 2.01 \text{ kJ}^2 \text{ mol}^2$ ) as the curve showed two inflection points, suggesting the presence of two Ni(II)-binding events. Therefore, the “two-sets-of-sites” model was applied with significantly improved statistics ( $\chi^2_{v \text{ WT}} = 0.453 \text{ kJ}^2 \text{ mol}^2$ ). The results of the fit indicate that one Ni(II) ion binds to the protein dimer with higher affinity ( $K_{D1 \text{ WT}} = 380 \pm 30 \text{ nM}$ ), followed by the interaction of  $\sim 4$  Ni(II) ions with lower affinity ( $K_{D2 \text{ WT}} = 1.60 \pm 0.02 \mu\text{M}$ ). Both events are characterized by favorable enthalpic values ( $\Delta H_{1 \text{ WT}} = -33 \pm 3 \text{ kJ mol}^{-1}$  and  $\Delta H_{2 \text{ WT}} = -54.4 \pm 0.7 \text{ kJ mol}^{-1}$ ) and different entropic contributions ( $\Delta S_{1 \text{ WT}} = +12.9 \text{ J mol}^{-1} \text{ K}^{-1}$  and  $\Delta S_{2 \text{ WT}} = -71.5 \text{ J mol}^{-1} \text{ K}^{-1}$ ). As observed for CD titrations, ITC experiments on *RrCooJ*- $\Delta$  confirmed that the truncated protein retains Ni(II)-binding properties. However, differently from the full-length protein, the integrated heat data show a single inflection point, indicating the occurrence of a single binding event (Fig. 4B and Fig. S2B). A fit of the integrated data with the one-set-of-sites model ( $\chi^2_{v \Delta} = 1.20 \text{ kJ}^2 \text{ mol}^2$ ) shows that a single Ni(II) ion binds to the protein dimer with  $K_{D\Delta} = 13 \pm 2 \text{ nM}$ ,  $\Delta H_{\Delta} = -36.3 \pm 0.4 \text{ kJ mol}^{-1}$ ,  $\Delta S_{\Delta} = +29.3 \text{ J mol}^{-1} \text{ K}^{-1}$ . The calculated thermodynamic parameters are similar to those provided for the first binding event in *RrCooJ*, suggesting that the latter does not involve the C-terminal tail, whereas the subsequent binding of four Ni(II) ions, absent in the *RrCooJ*- $\Delta$  protein, occurs at the C-terminal His-rich tail. These results are consistent with the CD-based titrations and provide improved stoichiometry for the different events, as well as a quantitative determination of the equilibrium binding constants. The affin-

ity of the internal site in the truncated protein is one order of magnitude higher as compared with the same site in the full-length protein, suggesting that the C-terminal region somehow decreases the Ni(II) affinity for the internal binding site, possibly competing for Ni(II) binding. Mutation of the two Cys residues (*RrCooJ*-C2S mutant) (Fig. 4C and Fig. S2C) led to a decreased stoichiometry from  $\sim 5$  to  $\sim 4$  Ni(II) ions per protein dimer. In particular, the thermodynamic parameters obtained with a fit using a two-sets-of-sites model ( $\chi^2_{v \text{ C2S}} = 1.72 \text{ kJ}^2 \text{ mol}^2$ ) are  $K_{D1 \text{ C2S}} = 380 \pm 10 \text{ nM}$ ,  $\Delta H_{1 \text{ C2S}} = -62 \pm 2 \text{ kJ mol}^{-1}$ ,  $\Delta S_{1 \text{ C2S}} = -84.2 \text{ J mol}^{-1} \text{ K}^{-1}$  and  $K_{D2 \text{ C2S}} = 1.07 \pm 0.01 \mu\text{M}$ ,  $\Delta H_{2 \text{ C2S}} = -82 \pm 2 \text{ kJ mol}^{-1}$ ,  $\Delta S_{2 \text{ C2S}} = -159 \text{ J mol}^{-1} \text{ K}^{-1}$ . On the other hand, mutation of the three internal histidines (His-18, His-22, and His-26) produces the *RrCooJ*-3HA variant (Fig. 1A) that lacks the N-terminal binding site and only retains the ability to bind four Ni(II) per protein dimer (Fig. 4D and Fig. S2D). Coherently, the binding isotherm shows a single inflection point and can be fitted with a one-set-of-sites model ( $\chi^2_v = 1.01 \text{ kJ}^2 \text{ mol}^2$ ) and with  $K_{D3 \text{ HA}} = 940 \pm 40 \text{ nM}$ ,  $\Delta H_{3 \text{ HA}} = -88.7 \pm 0.6 \text{ kJ mol}^{-1}$ ,  $\Delta S_{3 \text{ HA}} = -182 \text{ J mol}^{-1} \text{ K}^{-1}$ . Finally, as observed with CD, the triple mutant *RrCooJ*- $\Delta$ 3HA does not feature any significant heat of binding (Fig. 4E and Fig. S2E), further confirming that the first binding event involves Ni(II) coordination to the histidine residues of the folded N-terminal domain.

To study the impact of Ni(II) on *RrCooJ* conformation, SEC-MALLS-RI experiments were conducted. Increasing addition of Ni(II) from 0 to 4 Ni(II) molar equivalents per protein dimer led to elution volume variations, whereas the molecular masses agree with a dimeric protein, regardless of Ni(II) concentration.





**Figure 4. Binding properties of Ni(II) to *RrCooJ* and its variants determined using isothermal titration calorimetry.** A, integrated heat data of Ni(II) titration over *RrCooJ* as a function of metal/protein molar ratio. The continuous line represents the best fit obtained using a two-sets-of-sites model. B, integrated heat data of Ni(II) titration over *RrCooJ*-Δ as a function of metal/protein molar ratio. The continuous line represents the best fit obtained using a one set of site model. C, integrated heat data of Ni(II) titration over *RrCooJ*-C2S. The continuous line represents the best fit obtained using a two-sets-of-sites model. D, integrated heat data of Ni(II) titration over *RrCooJ*-3HA. The continuous line represents the best fit obtained using a one-set-of-sites model. E, integrated heat data of Ni(II) titration over *RrCooJ*-Δ3HA. F, integrated heat data of Ni(II) titration over *RrCooJ*-ΔE29A. The continuous line represents the best fit obtained using a one-set-of-sites model.

This suggests an effect of metal binding on the protein conformation (Fig. 5A). Above  $\sim 100 \mu\text{M}$   $\text{NiSO}_4$ , a slight modification of the total injected protein was observed, suggesting protein aggregation at high Ni(II) concentrations. Notably, the addition of 5 mM EDTA to a *RrCooJ* solution preincubated with 3 molar equivalents of Ni(II) led to the recovery of the apo*RrCooJ* dimeric elution peak, demonstrating that Ni(II) binding and nickel-induced conformational change of *RrCooJ* are reversible processes (Fig. 5B). In the case of *RrCooJ*-Δ, the elution volume is independent from Ni(II) concentration (Fig. 5C), revealing that the conformational change of *RrCooJ*-WT is induced by Ni(II)-binding to the His-rich tail.

#### X-ray structure of *RrCooJ*-Δ in complex with Ni(II)

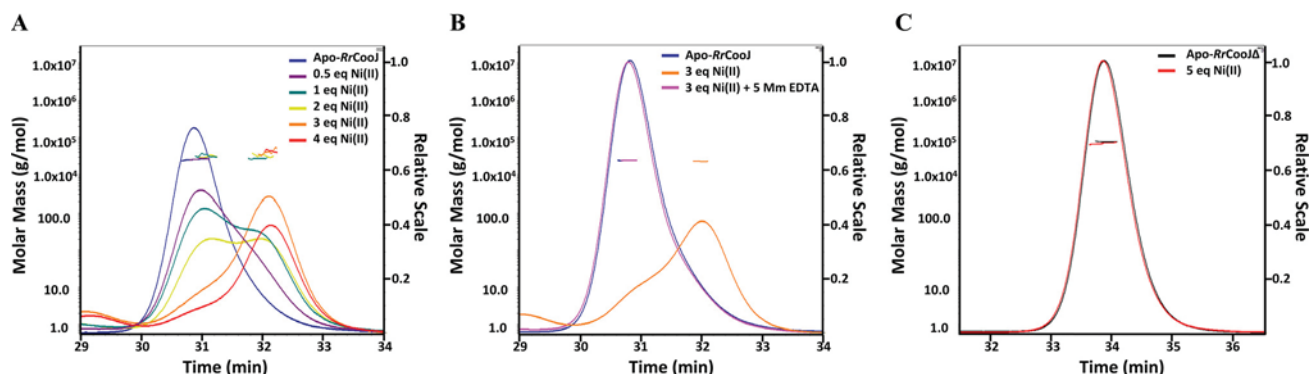
The crystal structure of Ni(II)-loaded *RrCooJ*-Δ (Ni-*RrCooJ*-Δ) was solved at  $2.04 \text{ \AA}$  resolution in the space group  $P2_1$ . Crystallographic data statistics of the Ni-*RrCooJ*-Δ structure are summarized in Table S3. In the structure, eight molecules are present in the asymmetric unit. Ni-*RrCooJ*-Δ forms a homodimer (Fig. 6A) in agreement with the SEC-MALLS-RI analyses. Each monomer is composed of two  $\alpha$ -helices (from Glu-3 to Ala-36 and from Leu-43 to Leu-68) connected by a short turn (from Leu-37 to Arg-42). The homodimer adopts a coiled-coil architecture comprising four helices, with the two monomers arranged antiparallel to each other, indicating that the two histidine tails of the full-length protein are expected to be top-to-tail. The degree of mobility (B-factor) varies along the helix bundle with a central more rigid region and two flexible edges (Fig. 6B). Remarkably, four methionine residues (Met-27<sub>A/B</sub> and Met-54<sub>A/B</sub>) form a methionine-rich cluster (Fig. 6A).

This kind of cluster was previously observed in helix-bundle structures (17, 18) and plays a role in helix-helix packing patterns. According to the CATH protein structure classification database (structural classification UCL), the *RrCooJ*-Δ monomer belongs to the “mainly  $\alpha$ ” class, with a helix hairpin topology (CATH code 1.10.287). *RrCooJ*-Δ shares structural similarities with the “helix hairpin bin” superfamily, although no function can be deduced from it. Moreover, the investigation of the function of *RrCooJ* from its sequence revealed that the protein does not belong to any CATH functional family and that no homolog domain was found in the PFAM database. Assembly analyses using PDBePISA (19) showed that the 8-amino acid chains present in the asymmetric unit form four stable dimers, with interface surface areas of  $\sim 7000 \text{ \AA}^2$ . One Ni(II) ion is bound to each dimer. In addition, two BCEG-Ni<sub>2</sub> and ADFH-Ni<sub>2</sub> tetramers, predicted to be unstable in solution, were suggested.

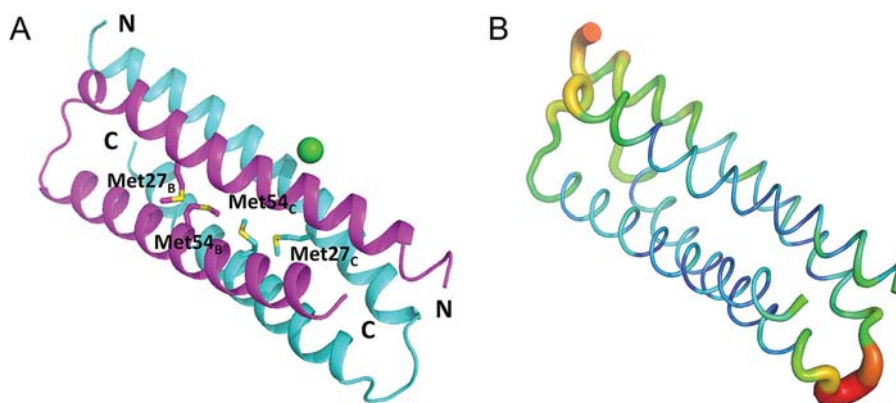
#### Structural characterization of the novel nickel-binding site

The X-ray structure of Ni-*RrCooJ*-Δ allowed the identification and the characterization of the Ni(II)-binding site, found in a largely solvent-exposed site. One peak attributed to nickel is present in the X-ray fluorescence spectrum (Fig. S3), corresponding to four peaks with Ni(II) at 100% occupancy in the anomalous difference electron density map, at the surface of the four dimers. For clarity, only the BC dimer will be discussed here. The Ni(II) ion is hexacoordinated and adopts an octahedral coordination. The metal is bound to the dimer via the four Nε of imidazole groups of His-22<sub>B</sub>, His-26<sub>B</sub>, His-18<sub>C</sub>, and His-22<sub>C</sub>. The two additional solvent-exposed coordination sites are

## A novel Ni(II)-binding site in the nickel chaperone CooJ



**Figure 5. Ni(II) impact on *RrCooJ* and *RrCooJ*- $\Delta$  conformation using SEC-MALLS-RI.** A, chromatogram of SEC-MALLS-RI analysis of *RrCooJ* in the presence of Ni(II) (from 0 to 4 molar eq). 20  $\mu$ l of protein samples were injected at 45  $\mu$ M of dimer in a Superdex 200 increase equilibrated in 50 mM HEPES, pH 7.5, 300 mM NaCl, 1 mM TCEP. The lines show Apo *RrCooJ* alone (blue) or in the presence of 22.5  $\mu$ M of Ni(II) (0.5 eq) (violet), 45  $\mu$ M of Ni(II) (1 eq) (green), 90  $\mu$ M of Ni(II) (2 eq) (yellow), 135  $\mu$ M of Ni(II) (3 eq) (orange), and 180  $\mu$ M of Ni(II) (4 eq) (red). The final protein concentration in the dimeric form was determined by the ASTRA 6 program and found equal to 44  $\mu$ M dimer (100% of injected protein) for the apo *RrCooJ*, and *RrCooJ* preincubated with 0.5 eq Ni(II) or 1 eq Ni(II), 36  $\mu$ M (82% of injected protein) for *RrCooJ* preincubated with 2 eq Ni(II) and *RrCooJ* preincubated with 3 eq Ni(II) and 28  $\mu$ M (64% of injected protein) for preincubated with 4 eq Ni(II). B, chromatogram of SEC-MALLS-RI analysis of apo *RrCooJ* (blue), in the presence of 3 eq Ni(II) (orange) and 3 eq of Ni(II) + 5 mM EDTA (magenta). Samples were injected at 50  $\mu$ M dimer. C, chromatogram of SEC-MALLS-RI analysis of *RrCooJ*- $\Delta$  at 100  $\mu$ M dimer in the absence (black) or preincubated with 5 molar eq of Ni(II) (red).



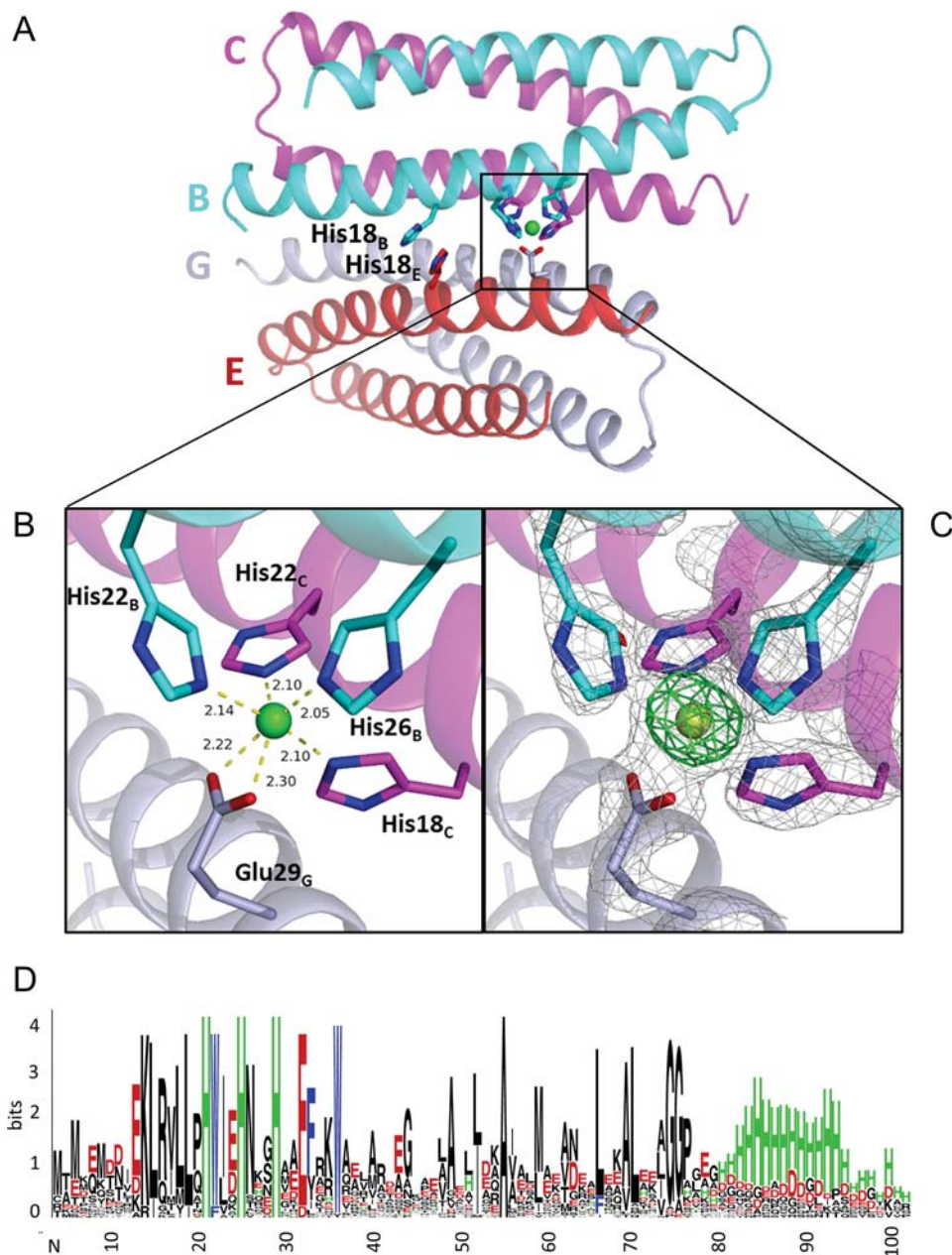
**Figure 6. X-ray structure of Ni-*RrCooJ*- $\Delta$  homodimer.** A, the two monomers are in magenta and cyan. Ni(II) is depicted as a green sphere. B, B-factor diagram of Ni-*RrCooJ*- $\Delta$  dimer by the B-factor putty program in PyMOL. The B-factors are illustrated by color, ranging from low (blue, B-factor = 20) to high (red, B-factor = 60).

filled by the bidentate carboxylate group of Glu-29<sub>G</sub> (Fig. 7, A–C). Thus, in the X-ray structure, a tetramer BCEG is formed via the interaction of Glu-29<sub>G</sub> with the Ni(II) ion bound to the dimer BC. In addition, a  $\pi$ -stacking between His-18<sub>B</sub> and His-18<sub>E</sub> is also observed (Fig. 7A). However, these interactions are insufficient to stabilize the tetrameric conformation in solution, as proven by SEC-MALLS-RI and should be regarded as a solid-state effect. This conclusion is reinforced by ITC experiments on the *RrCooJ*- $\Delta$ E29A mutant, which indicate that mutation of Glu-29 to Ala does not affect Ni(II) binding in solution (Fig. 4F and Fig. S2F). Indeed, the fit of the binding isotherm ( $\chi^2_{\nu \Delta E29A} = 0.325 \text{ kJ}^2 \text{ mol}^2$ ) shows that one Ni(II) ion binds per protein dimer with thermodynamic parameters similar to those of *RrCooJ*- $\Delta$  ( $K_{D \Delta E29A} = 18 \pm 2 \text{ nM}$ ,  $\Delta H_{\Delta E29A} = -37.6 \pm 0.3 \text{ kJ mol}^{-1}$ ,  $\Delta S_{\Delta E29A} = 22.4 \text{ J mol}^{-1} \text{ K}^{-1}$ ).

Unexpectedly, CooJ homologs are present in five different bacterial phyla, with size varying from 70 to 117 residues: they were identified in the phyla *Proteobacteria*, *Planctomycetes*, *Lentisphaerae*, *Chloroflexi*, and *Nitrospirae* and in groundwater and marine sediment metagenomes. In *proteobacteria*, they are found in alphaproteobacteria (*R. rubrum*, *R. palustris*, *Pleo-*

*morphomonas* sp., *Telmatospirillum siberiense*), in unclassified *Deltaproteobacteria* and in sulfate-reducing *deltaproteobacteria* (unclassified *Desulfobulbaceae*, unclassified *Desulfobacteriales*, *Desulfofustis glycolicus*, *Desulforhopalus singaporensis*, *Desulforivibrio alkaliphilus*, *Desulfobulbus* sp., *Desulfotalea* sp., *Desulfopila aestuarii*, *Desulfocapsa sulfexigens*). In *planctomycetes*, they are found in unclassified *Planctomycetes* and in *Candidatus scalindua*. In *Lentisphaerae*, *Chloroflexi*, and *Nitrospirae*, they were identified in unclassified bacteria. The Ni(II)-binding residues His-18, His-22, and His-26 in the *R. rubrum* sequence are strictly conserved among at least 46 putative CooJ sequences (Fig. S4).

Moreover, an acidic residue (Asp or Glu) is also conserved (in position 29 in *R. rubrum*). The new H(W/F)<sub>2</sub>X<sub>2</sub>HX<sub>3</sub>H Ni(II)-binding motif is revealed by the present study as a signature of the CooJ family. This result highlights the importance of the newly characterized Ni(II)-binding site in *RrCooJ*. The alignment of their amino acid sequences by ClustalW further shows that they all possess a histidine-rich motif in the C-terminal part, although their lengths are variable, containing from 3 to 18 histidines, whereas cysteines are rarely present (Fig. 7D).



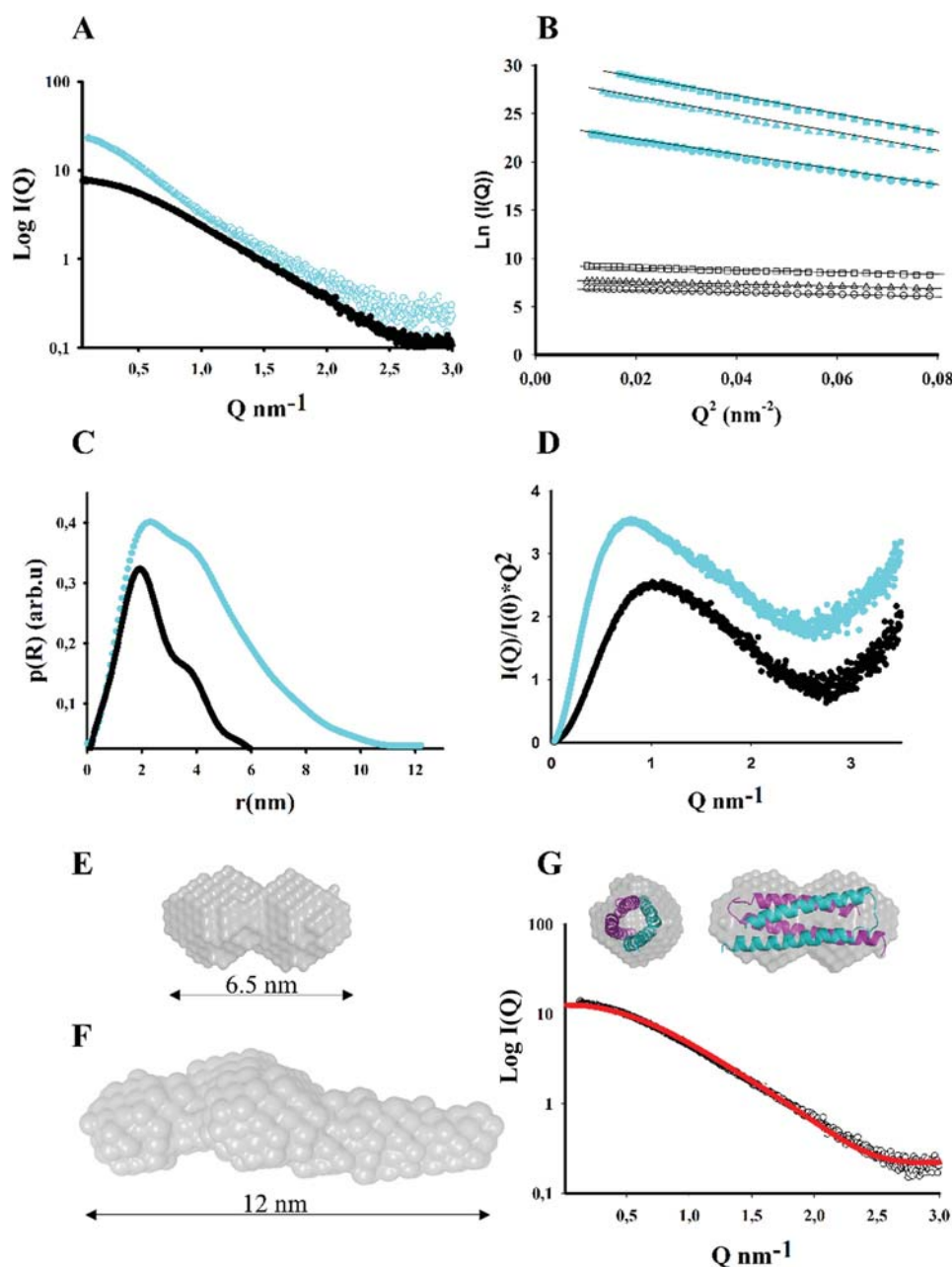
**Figure 7. The novel Ni(II) binding site in RrCooJ.** A, tetrameric conformation of Ni-RrCooJ-Δ, composed of the two homodimers BC (in magenta and cyan) and GE (in gray and red). B, structure of the Ni(II)-binding site. Ni(II) is depicted as a green sphere, and coordinating residues are shown as sticks. C, In gray is the  $2F_o - F_c$  map contoured at  $2.0 \sigma$ , and in green is the difference anomalous map contoured at  $3.0 \sigma$ . D, logo showing the conservation of the 46 CooJ homologs identified by pBlast. For clarity, nonconserved residues forming gaps in ClustalW alignment were removed. The logo of CooJ homologs was generated using the WebLogo tool (50).

#### Analysis of apoRrCooJ-Δ and apoRrCooJ by SAXS

As expected, the presence of a large disordered region in the C-terminal part of RrCooJ hindered its crystallization. To get information on the conformation of the full-length protein in solution, SAXS experiments were performed. The final scattering curves of apoRrCooJ and apoRrCooJ-Δ are shown in Fig. 8A. The linearity of the Guinier plots at several protein concentrations confirms that both proteins are monodisperse in solution (Fig. 8B). The radius of gyration ( $R_g$ ) for apoRrCooJ ( $R_g$ ,  $3.05 \pm 0.12$  nm) and apoRrCooJ-Δ ( $R_g$ ,  $1.99 \pm 0.05$  nm) were measured. Moreover, the calculated molecular mass determined by  $I(0)$  is in good agreement with SEC-MALLS-RI data ( $MM_{SAXS} =$

$26 \pm 0.5$  kDa for apoRrCooJ and  $MM_{SAXS} = 14.9 \pm 0.1$  kDa for apoRrCooJ-Δ). The pair distance distribution patterns of both proteins indicate that apoRrCooJ, with a  $D_{max}$  of  $12 \pm 1$  nm, is more elongated than apoRrCooJ-Δ, which displays a  $D_{max}$  of  $6.5 \pm 0.4$  nm (Fig. 8C). The flexibility of each protein was investigated with the analysis of the Kratky plots ( $s^2I(s)$  versus  $s$ ). The bell-shaped curves of the plots indicate that both proteins are partially folded (Fig. 8D). However, apoRrCooJ shows a decay to zero with a  $Q_{max}$  of 0.7 nm, indicating the presence of more flexible parts compared with apoRrCooJ-Δ with a  $Q_{max}$  of 1 nm. In addition, the lower intensity of the plateau phase of apoRrCooJ-Δ compared with that of apoRrCooJ strongly suggests



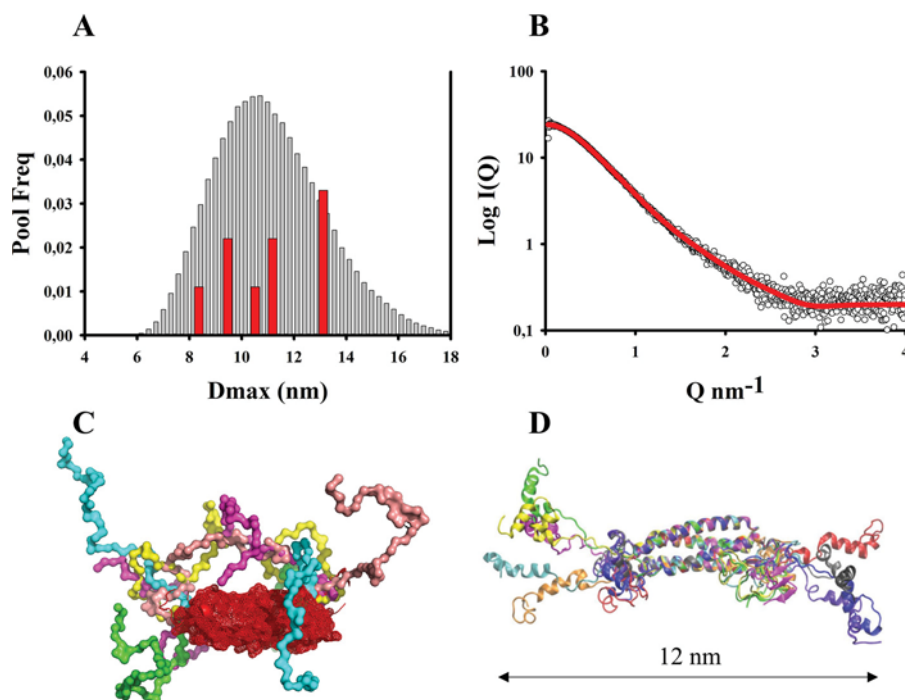


**Figure 8. Study of apoRrCooJ and apoRrCooJ-Δ by SAXS.** A, scattering curves of experimental data of apoRrCooJ in cyan and apoRrCooJ-Δ in black in solution. B, linear dependence of  $\ln(I(Q))$  versus  $Q^2$  determined by Guinier plot at several concentrations:  $\circ$ , 56  $\mu\text{M}$ ;  $\triangle$ , 112  $\mu\text{M}$ ;  $\square$ , 240  $\mu\text{M}$  dimer for apoRrCooJ (cyan); and  $\circ$ , 34  $\mu\text{M}$ ;  $\triangle$ , 190  $\mu\text{M}$ ;  $\square$ , 340  $\mu\text{M}$  dimer for apoRrCooJ-Δ (black), with standard  $R_g$  limits (0.3 to 1.0), indicating the absence of aggregation for both proteins. C, pair distribution function  $p(R)$  in arbitrary units (arb.u) versus  $r$  (nm) determined by GNOM of apoRrCooJ (cyan) with a  $D_{\text{max}}$  of 12 nm and apoRrCooJ-Δ (black) with a  $D_{\text{max}}$  of 6.5 nm. D, Kratky plots of apoRrCooJ (cyan) and apoRrCooJ-Δ (black). E, final *ab initio* model of apoRrCooJ-Δ generated with DAMMIF and merged with DAMAVER (NSD < 0.7). F, final *ab initio* model of apoRrCooJ generated with DAMMIF and merged with DAMAVER (NSD = 0.8–1.5). G, superimposition of the X-ray structure of Ni-RrCooJ-Δ on *ab initio* model of apoRrCooJ-Δ (done with the SUPCOMB software). Fit of the SAXS experimental scattering curve of apoRrCooJ-Δ (black) with the theoretical scattering curve from the crystal structure generated by CRY SOL (red).

the presence of additional flexible regions in the full-length protein. Taken together, these data revealed that apoRrCooJ is less structured than apoRrCooJ-Δ.

To determine the three-dimensional model of apoRrCooJ in solution, *ab initio* models were generated without any constraint in slow mode. First, *ab initio* models of apoRrCooJ-Δ were built. A solution model was selected by DAMAVER (21), with normalized spatial discrepancy (NSD) values < 0.7, corresponding to a homogenous solution in agreement with a compact and folded protein (Fig. 8E). To go further, *ab initio* models

of the full-length protein were generated using the same procedure. In this case, a heterogeneous solution was obtained, with the presence of several clusters with NSD values varying from 0.8 and 1.5, in favor of the existence of flexible regions in apoRrCooJ (Fig. 8F). Taking advantage of the X-ray structure of Ni-RrCooJ-Δ, the SAXS experimental scattering curve of apoRrCooJ-Δ was fitted with the theoretical scattering curve from the crystal structure generated by CRY SOL, giving a  $\chi^2$  of 1.8 (Fig. 8F). Moreover, the Ni-RrCooJ-Δ X-ray structure was superimposed with the best *ab initio* model of apoRrCooJ-Δ



**Figure 9.** *In silico* modeling of apoRrCooJ using EOM2.0 and CHARMM. **A**, distribution ( $D_{\max}$ ) of end-to-end distances computed from pools containing 10,000 RrCooJ structural models generated by EOM2.0 in gray. The five selected models are depicted in red. **B**, fit of experimental data in black compared with final combination model of RrCooJ generated by EOM with a  $\chi^2$  of 0.8. **C**, superposition of the five final models of RrCooJ: coiled-coil domains are in red. For the flexible regions, each color corresponds to an individual model. **D**, modeling of apoRrCooJ. Superposition of the minimum energy models of apoRrCooJ in each of the nine clusters obtained after replica-exchanging self-guided Langevin dynamics simulation with CHARMM. Each model is represented in a different color. The figure was drawn with VMD (51).

using SUPCOMB (Fig. 8G). Our data show that the truncated protein shares the same conformation in solution and in the crystalline state.

#### Molecular modeling of apoRrCooJ

To better characterize the apoRrCooJ dimer in solution, a hybrid approach was adopted combining both the Ni-RrCooJ- $\Delta$  X-ray structure and the SAXS data using the Ensemble Optimization Method (EOM 2.0) (20). From our SAXS analyses, the X-ray structure of the truncated holo-protein can be confidently used to model the N-terminal region of the full-length protein. EOM 2.0 was thus used to generate a pool of 10,000 conformers using both random chains for the C-terminal part and the atomic coordinates of the Ni-RrCooJ- $\Delta$  crystal structure. The Gaussian statistics of the data set resulted in a best final ensemble of five independent models (a  $\chi^2 = 0.82$ ) in a highly variable range of  $D_{\max}$  from 8.3 to 13.1 nm, describing the scattering properties of inherently dynamic macromolecules (Fig. 9, A and B). This result indicates that the apoRrCooJ dimer contains two highly flexible regions corresponding to the two C-terminal histidine rich clusters modeled as polyasparagine by EOM 2.0 (Fig. 9C). The anti-parallel spatial arrangement of the two monomers, observed in the X-ray structure of Ni-RrCooJ- $\Delta$  dimer, is confirmed for the full-length protein in solution.

To obtain more detailed information on the flexible histidine-rich C-terminal tails, a pool of individual models of apoRrCooJ was built using the CHARMM force field and the RrCooJ amino acid sequence, with the atomic coordinates of Ni-RrCooJ- $\Delta$  structure and the biophysical parameters

( $R_g$ ,  $D_{\max}$ ) from SAXS data. To generate individual models, the central ordered domain of RrCooJ dimer was considered similar to the RrCooJ- $\Delta$  dimer. Each model was treated using dynamic replica exchange (RXGLD) between 300 and 400 K covering 8 ns. The last 3 ns of RXGLD dynamics (3000 frames recorded every ps) were analyzed. The calculated average  $R_g$  was  $3.14 \pm 0.04$  nm compatible with the 3.12-nm target determined by SAXS, meaning that the modeled trajectory yields a correct representation of the flexibility of the C-terminal region. Frames were grouped into nine clusters corresponding to coordinate root-mean-square deviation larger than 6.0 Å. The minimum energy models of each cluster are shown superimposed in Fig. 9D, indicating the appearance of transient  $\alpha$ -helix and  $\beta$ -sheet motifs in the C-terminal region. In particular,  $\alpha$ -helices appear between His-92 and Cys-109, whereas transient  $\beta$ -sheets are found in the Leu-68–Asp-69, Pro-84–Phe-85, His-96–Gly-99, His-104–Asp-107, and His-110–Asp-113 regions.

#### Discussion

Nickel is an important trace element for many prokaryotic microorganisms, where it participates to a variety of cellular processes, being part of the active site of essential enzymes. Because the free metal ion is toxic in the intracellular environment, the cells control its availability with transporters, storage proteins, and metalloregulators, working in concert with accessory proteins to deliver the metal ion to its specific target sites. In *R. rubrum*, two key enzymes involved in CO metabolism, namely hydrogenase and CODH, require nickel for their activ-

## A novel Ni(II)-binding site in the nickel chaperone CooJ

ity. Among the specific accessory proteins expressed by this bacterium, *RrCooJ* is an active player in nickel homeostasis and, more specifically, in CODH maturation. In this work, the use of a multidisciplinary approach provided a detailed characterization of the nickel chaperone *RrCooJ*.

*RrCooJ* forms a multidomain dimeric protein, with a coiled coil and two flexible regions. A combination of X-ray crystallographic and SAXS data revealed that the two histidine-rich tails at the C termini are independent to each other. The role of histidine-rich regions in proteins is generally proposed to be related to nickel storage and/or detoxification, because of their ability to quickly bind and release nickel ions. When present, histidine-rich clusters are located either at the C-terminal position as in CooJ for CODH maturation, in UreE for urease maturation (21), and in SlyD for hydrogenase maturation (22) or at the N-terminal region as in HypB for hydrogenase maturation (23), in UreG for urease maturation (24), and in LarC for lactate racemase maturation (25). Differently, *Helicobacter pylori* expresses two unique proteins, Hpn and Hpn-2, composed by histidines and cysteines for half of the total amino acid sequence. The two proteins play a central role in controlling the intracellular nickel trafficking and participate in a common pathway to the maturation of urease, a virulence factor (26). Altogether, these studies suggest that histidine-rich motifs might satisfy a nickel storage role for the bacteria rather than being part of the mechanism of metal insertion into enzymes.

SAXS and modeling experiments showed that the C-terminal region of *RrCooJ*, initially predicted from the sequence as intrinsically disordered, is partly and transiently structured even in the absence of nickel. This region appears to be highly flexible and linked to a more rigid N-terminal domain corresponding to the coiled coil. The two His-rich regions are able to bind four Ni(II) ions per dimer with dissociation constants in the micromolar range. The effect of Ni(II) binding to the C terminus tail onto the protein structure is difficult to evaluate because of the presence of several metal-binding sites with similar affinity and the tendency of the protein to aggregate at high nickel concentrations.

In addition to the characterization of Ni(II) binding to the His-rich tail of *RrCooJ*, an unexpected Ni(II)-binding site was detected with a dissociation constant in the nanomolar range. This site involves conserved histidines located at the dimer interface of the protein. In most nickel-dependent accessory proteins, an exposed nickel-binding site is found, compatibly with fast metal transfer to the physiological partner(s). The Ni(II) ion is coordinated by four histidines, two from each monomer, with two additional ligands from a bidentate Glu-29 of the neighboring dimer completing an octahedral coordination geometry. With a solvent exposed Ni(II)-binding site, water molecules can easily replace the glutamate residue and complete the coordination sphere in solution, with a nitrogen/oxygen ligand environment.

With two independent types of Ni(II)-binding sites, *RrCooJ* thus resembles the known behavior of HypB and UreE nickel-chaperones. Indeed, in addition to the nonconserved histidine-rich cluster, two metal-binding sites have been identified in HypB: a conserved low-affinity Ni(II)/Zn(II) binding site, pro-

posed to be responsible for the direct Ni(II) transfer to its physiological partner HypA (27), and a nonconserved high affinity Ni(II)-binding site at the N terminus (28). Similarly, the dimeric UreE presents a conserved metal-binding site at the dimer interface, involving a conserved His residue from each monomer located in the C-terminal domains, as well as a second His residue in the disordered C-terminal tail, in addition to other metal-binding sites located in the histidine-rich segment, found in some UreE homologs (29).

Using the ever continuously updating of genome databases, several CooJ homologs were identified in recently sequenced genomes. They all present a histidine-rich tail, featuring highly variable size and composition, whereas the strict conservation of the N-terminal binding site reinforces its essential role in nickel metabolism. Although its role was predicted to be minor and restricted to *R. rubrum*, the occurrence of CooJ in various bacteria also raises the question of its implication in the overall maturation pathway of CODH.

Altogether, this study has provided a structural and biophysical characterization of *RrCooJ* having more general implications in the field of nickel trafficking by metallochaperones containing His-rich regions. On the basis of the knowledge acquired on nickel chaperones, a double function can be proposed for CooJ, in both CODH maturation and nickel storage. Further studies, required to clarify the molecular details of protein–protein interactions involving CooJ, CooT, and CooC leading to Ni(II) transfer and CODH activation, are underway in our laboratories.

### Experimental procedures

*RrCooJ* and its mutants (Table S1) were produced in *E. coli* and purified. Cell pellets from 2 L of *E. coli* culture overexpressing *RrCooJ*, *RrCooJ*-3HA, or selenomethionine-labeled *RrCooJ*- $\Delta$  were thawed and resuspended in buffer QA (50 mM Tris-HCl, pH 8.5, containing 1 mM EDTA and 5 mM DTT) complemented with one complete Protease Inhibitor mixture tablet (Roche) per 50 ml of buffer, and lysed by sonication. The cell debris were removed by centrifugation at 40,000 rpm for 30 min at 4 °C. The supernatant was loaded at 0.7 ml min<sup>-1</sup> onto a Q-Sepharose HP 16/10 column (GE Healthcare) equilibrated with buffer QA. After loading, the column was washed with QA buffer until the baseline was reached, and the protein was eluted with a linear gradient (0–50% in six column volumes) of buffer QB (50 mM Tris-HCl, pH 8.5, 1 M NaCl, 1 mM EDTA, 5 mM DTT). The fractions of the main elution peak were analyzed by SDS-PAGE, and those containing *RrCooJ* proteins were pooled and concentrated. The protein was then loaded onto a Superdex75 16/600 size-exclusion column (GE Healthcare) equilibrated in buffer S (50 mM HEPES, pH 7.5, containing 300 mM NaCl, 1 mM TCEP). Fractions were collected and analyzed by SDS-PAGE. The fractions containing *RrCooJ* were pooled and concentrated up to 10.0 mg ml<sup>-1</sup> before flash freezing in liquid nitrogen and stored at -80 °C. *RrCooJ*- $\Delta$ , *RrCooJ*- $\Delta$ 3HA, *RrCooJ*-2CS, and *RrCooJ*- $\Delta$ E29A were purified using the same protocol except that DTT and TCEP were excluded from buffers QA, QB, and S, respectively. Experimental extinction coefficient at 280 nm was calculated from the concentrations determined by SEC-MALLS-RI on three independent sam-



ples for apoRrCooJ ( $4.62 \text{ mM}^{-1}\cdot\text{cm}^{-1}$ ) and apoRrCooJ- $\Delta$  ( $1.52 \text{ mM}^{-1}\cdot\text{cm}^{-1}$ ).

Inductively coupled plasma atomic emission spectroscopy (Shimadzu ICP 9000 instrument with Mini plasma Torch in axial reading mode) was used to measure the metal content. The samples were incubated in 65%  $\text{HNO}_3$  overnight at  $95^\circ\text{C}$  before measurement (30). Secondary structure elements of RrCooJ were predicted using the PredictProtein program (<http://www.predictprotein.org>) (52),<sup>4</sup> an automatic server that searches up-to-date public sequence databases, creates alignments, and predicts aspects of protein structure and function.

Purified and frozen RrCooJ and RrCooJ- $\Delta$  were thawed, diluted to  $\sim 50 \text{ }\mu\text{M}$  (dimer), and incubated in the absence and presence of  $\text{NiSO}_4$  (0.5, 1.0, 2.0, 3.0, or 4.0 molar equivalents per dimer) for 15 min at room temperature prior to injection onto the SEC-MALLS-RI system (Wyatt Dawn HELEOS-II 18-angle light scattering detector and Wyatt Optilab rEX refractive index monitor linked to a Shimadzu HPLC system comprising a LC-20AD pump, a SPD20A UV-visible detector, and a Superdex 200 10/300 column (GE Healthcare)). Injections were carried out using a 20- $\mu\text{l}$  loop. The size-exclusion column was equilibrated using a 50 mM HEPES, pH 7.5, 300 mM NaCl, and 1 mM TCEP. Protein concentration in all samples was determined by integration of the differential refractive index peak. It is important to note that all used samples are coming from the same stock dilution; thus a modification on the final protein concentration can be attributed to protein aggregation. The data were analyzed using ASTRA software (version 6).

CD spectra were recorded using a J-1500 CD spectrometer from JASCO Analytical Instruments. A stock solution of 10 mM  $\text{NiSO}_4$  was used to monitor metal-ligand charge transfer and d-d transitions of RrCooJ and its mutants upon Ni(II) titration (from 0 to 10 molar equivalents of  $\text{NiSO}_4$ ). Spectra were recorded from 250 to 700 nm using a 10-mm cuvette, with five accumulations to increase the signal-to-noise ratio. Proteins were thawed and diluted to 20–100  $\mu\text{M}$  dimer in CD buffer (50 mM HEPES, pH 7.5, 300 mM NaCl, 1 mM TCEP). CD spectra were also recorded using a 1-mm cuvette in the 190–250-nm range to determine the secondary structure content of apoRrCooJ and apoRrCooJ- $\Delta$ , with 10 accumulations on samples containing 7.5  $\mu\text{M}$  protein dimer in CD buffer (8 mM HEPES, pH 7.2, 10 mM NaCl, and 1 mM TCEP). The spectra were analyzed using the JASCO spectra manager software (version 2).

RrCooJ and its variant (7–10  $\mu\text{M}$  dimer) and RrCooJ- $\Delta$  and its variants (12–17  $\mu\text{M}$  dimer) diluted in 50 mM HEPES, pH 7.2, 300 mM NaCl, were titrated with  $\text{NiSO}_4$  (100–500  $\mu\text{M}$ ) diluted in the same buffer. The data were analyzed using the MicroCal PEAQ-ITC analysis software to provide the isotherm of binding and to derive the thermodynamic parameters. Integrated data obtained for each titration were fitted using a nonlinear least-squares minimization algorithm to a theoretical titration curve, with either the one set of sites model or the two-sets-of-sites model.  $N$  (stoichiometry of binding),  $\Delta H$  (reaction enthalpy change,  $\text{cal mol}^{-1}$ ), and  $K_a$  (binding constant,  $\text{M}^{-1}$ ) were the thermodynamic fitting parameters. The reaction entropy was

calculated using the relationships  $\Delta G = -RT\ln K_a$  ( $r = 1.9872 \text{ cal mol}^{-1} \text{ K}^{-1}$ ,  $T = 298 \text{ K}$ ) and  $\Delta G = \Delta H - T\Delta S$ . The reliability of the obtained fits was evaluated using the reduced  $\chi^2$  ( $\chi_v^2$ ) parameter.

Initial crystallization screens were carried out using the HTXLab facilities at EMBL Grenoble set up using a sitting-drop vapor diffusion 1:1 ratio format with 100 nl of protein solution  $\pm 1$  molar eq of Ni(II) at 400  $\mu\text{M}$  dimer. Crystals were optimized manually in the laboratory by the hanging-drop method, using 2  $\mu\text{l}$  of holo-protein solutions at 400  $\mu\text{M}$  dimer added with an equal volume of precipitant. The best diffracting crystal of Ni-RrCooJ- $\Delta$  was obtained with a protein concentration of 400  $\mu\text{M}$  dimer in the presence of one equivalent of Ni(II) per dimer, in 16% (v/v) PEG3350 and 0.2 M calcium chloride. Micro-seeding was carried out, according to the Hampton protocol ("seed-bead" kit), to increase the size of the crystals. A cryo-protection solution was made starting from the mother liquor composition by addition of 25% glycerol before flash-cooling in liquid nitrogen.

Diffraction data were collected on the ID30B Beamline at the European Synchrotron (ESRF, Grenoble, France) (31). Fluorescence spectra were recorded to confirm the presence of selenium and nickel atoms in the crystals. The best crystal diffracted at  $2.04 \text{ \AA}$  resolution, and a single anomalous diffraction data set was recorded above the selenium edge ( $\lambda = 0.9786 \text{ \AA}$ ) to benefit from the anomalous contributions of both selenium and nickel atoms ( $f'' = \sim 3.8$  and  $2e^-$ , respectively). The data were integrated in the P21 space group using XDS and scaled/merged using XSCALE (32). The asymmetric unit was estimated to contain eight RrCooJ- $\Delta$  molecules (68 residues/molecule) based on the Matthews coefficient using the CCP4 suite (33). The structure was solved by the SAD method using the SHELX suite (34). The whole heavy atoms substructure, including 16 selenium and 4 nickel atoms, was solved using SHELXD. The experimental phases were calculated and further improved by alternating density-modification and auto-tracing using SHELXE. Then Phenix-Autobuilt (35) was able to automatically build 78% of the model (422 of 544 residues located, with  $R_{\text{work}} = 28\%$  and  $R_{\text{free}} = 32\%$ ). The model was subsequently rebuilt and corrected manually using COOT (36) and finally refined using Phenix Refine (37), resulting in a final model with  $R_{\text{work}} = 20.0\%$  and  $R_{\text{free}} = 25.9\%$  (refinement statistics in Table S3).

ApoRrCooJ and apoRrCooJ- $\Delta$  were filtered extemporaneously before each experiment, using size-exclusion chromatography with a Superdex 200 increase column (GE Healthcare) equilibrated in 50 mM HEPES, pH 7.5, 300 mM NaCl, 1 mM TCEP. Protein concentration were 33–240  $\mu\text{M}$  of dimer. SAXS data were collected at the European Synchrotron Radiation Facility (Grenoble, France) on the BM29 Beamline at BioSAXS. The data were processed using standard procedures with the ATSAS v2.8.3 suite of programs (38). The *ab initio* determination of the molecular shape of the proteins was done as previously described, using DAMMIF (39). Radius of gyration ( $R_g$ ) and forward intensity at zero angle ( $I(0)$ ) were determined with the programs PRIMUS (40) by using the Guinier approximation at low  $Q$  value, in a  $Q\cdot R_g$  range  $< 1.5$ .

<sup>4</sup> Please note that the JBC is not responsible for the long-term archiving and maintenance of this site or any other third party hosted site.



$$\ln I(Q) = \ln I(0) - \frac{R_g^2 Q^2}{3} \quad (\text{Eq. 1})$$

Porod volumes and Kratky plot were determined using the Guinier approximation and the PRIMUS programs (40). The radius of gyration and the pairwise distance distribution function  $P(r)$  were calculated by indirect Fourier transform with the program GNOM (41). The maximum dimension ( $D_{\text{max}}$ ) value was adjusted so that the  $R_g$  value obtained from GNOM agreed with that obtained from Guinier analysis.

To build *ab initio* models, several independent DAMMIF (39) models were calculated in slow mode with pseudo chain option and merged using the program DAMAVER (42). The program CRY SOL (43) was used to generate the theoretical scattering curves from the Protein Data Bank coordinates of Ni-RrCooJ-Δ.

The program suite EOM 2.0 (20) was used to get information on the predicted intrinsically disordered region: an ensemble of 10,000 conformers was generated using genetic algorithms to study the flexibility of the C-terminal domain of RrCooJ. A starting model, containing the atomic coordinates of Ni-RrCooJ-Δ structure and a polyasparagine mimicking the C-terminal region, was generated as input to the EOM program. The genetic algorithm compared the averaged theoretical scattering intensity from  $N$  independent ensembles of conformations against the scattering data of RrCooJ. The ensemble that best fitted the experimental SAXS data were selected with a maximum number of 10 individual models.

To study the flexibility of the C-terminal part of the protein in relation with SAXS data, molecular dynamics simulations were run using the molecular dynamics program CHARMM (44). A dimer of RrCooJ was modeled with the all-atom force field *all22* for proteins (45, 46). Initial coordinates were taken from the X-ray structure of the Ni-RrCooJ-Δ dimer solved from Glu-3 to Leu-68 for chain A and Pro-5 to Leu-68 for chain B. All missing atom coordinates were built with CHARMM using internal coordinates, and the structure was energy-minimized down to a gradient of 0.1 kcal/mol/Å. With 100 residues unresolved in the dimer, the resulting initial model needed a powerful algorithm able to sample the conformational space of the protein like replica-exchanging self-guided Langevin dynamics (47). Self-guided Langevin dynamics enhances conformational search efficiency through acceleration of low frequency motions in a molecular system (48). Different types of restraints were applied on the protein to maintain the position of backbone atoms resolved in the X-ray structure and prevent fraying of α-helices at their N and C terminus during high temperature dynamics. Finally, the radius of gyration of the solvated proteins were estimated to 19.8 and 34 Å, respectively, for RrCooJ-Δ and RrCooJ dimer by SAXS. The corresponding radius of gyration calculated by CHARMM for the RrCooJ-Δ unsolvated system being 17 Å, the radius of gyration of the RrCooJ unsolvated dimer was estimated to be  $34 \pm 2.8 = 31.2$  Å. To reproduce the experimental observations, a radius of gyration restraint ( $R_{\text{GYR}}$ ) of 2 kcal/mol/Å<sup>2</sup>, calculated on α-carbon atoms, was thus introduced in the simulations. A restraining radius of 32.6 Å was chosen. (This value was shown

in test simulations to best approach the 31.2 Å target). The Screened Coulomb potentials implicit solvent model (49) was used in this work involving 16 replicas. Additional details are provided in the supporting information.

### Data deposition

The atomic coordinates for Ni-CooJ-Δ have been deposited in the Protein Data Bank under accession code 6HK5.

**Author contributions**—M. A., J. P., P. C., C. B., B. Z., J. T., and S. Crouzy data curation; M. A., J. P., P. C., B. Z., J. T., and S. Crouzy formal analysis; M. A., J. T., and C. C. investigation; S. Ciurli validation; S. Ciurli and C. C. writing-review and editing; C. C. conceptualization; C. C. supervision; C. C. writing-original draft.

**Acknowledgments**—We thank the High Throughput Crystallization Laboratory (HTX Lab) at the EMBL Grenoble for RrCooJ-Δ crystallization. We thank the staffs from the BM-30A, BM29-BioSAXS, and ID-30B Beamlines of the European Synchrotron Facility in Grenoble. We thank Dr. Luca Signor (Institute for Structural Biology, Grenoble, France) for MS experiments in denaturing conditions (Integrated Structural Biology Grenoble platform).

### References

- Schultz, J. E., and Weaver, P. F. (1982) Fermentation and anaerobic respiration by *Rhodospirillum rubrum* and *Rhodopseudomonas capsulata*. *J. Bacteriol.* **149**, 181–190 [Medline](#)
- Alfano, M., and Cavazza, C. (2018) The biologically mediated water–gas shift reaction: structure, function and biosynthesis of monofunctional [NiFe]-carbon monoxide dehydrogenases. *Sustain. Energy Fuels* **2**, 1653–1670 [CrossRef](#)
- Drennan, C. L., Heo, J., Sintchak, M. D., Schreiter, E., and Ludden, P. W. (2001) Life on carbon monoxide: X-ray structure of *Rhodospirillum rubrum* Ni–Fe–S carbon monoxide dehydrogenase. *Proc. Natl. Acad. Sci. U.S.A.* **98**, 11973–11978 [CrossRef](#) [Medline](#)
- Dobbek, H., Svetlitchnyi, V., Gremer, L., Huber, R., and Meyer, O. (2001) Crystal structure of a carbon monoxide dehydrogenase reveals a [Ni-4Fe-5S] cluster. *Science* **293**, 1281–1285 [CrossRef](#) [Medline](#)
- Bonam, D., McKennat, M. C., Stephenst, P. J., and Ludden, P. W. (1988) Nickel-deficient carbon monoxide dehydrogenase from *Rhodospirillum rubrum*: *in vivo* and *in vitro* activation by exogenous nickel (hydrogenase/iron-sulfur protein/electron paramagnetic resonance). *Biochemistry* **85**, 31–35
- Spangler, N. J., Lindahl, P. A., Bandarian, V., and Ludden, P. W. (1996) Spectroelectrochemical characterization of the metal centers in carbon monoxide dehydrogenase (CODH) and nickel-deficient CODH from *Rhodospirillum rubrum*. *J. Biol. Chem.* **271**, 7973–7977 [CrossRef](#) [Medline](#)
- Jeon, W. B., Singer, S. W., Ludden, P. W., and Rubio, L. M. (2005) New insights into the mechanism of nickel insertion into carbon monoxide dehydrogenase: Analysis of *Rhodospirillum rubrum* carbon monoxide dehydrogenase variants with substituted ligands to the [Fe3S4] portion of the active-site C-cluster. *J. Biol. Inorg. Chem.* **10**, 903–912 [CrossRef](#) [Medline](#)
- Kerby, R. L., Ludden, P. W., and Roberts, G. P. (1997) *In vivo* nickel insertion into the carbon monoxide dehydrogenase of *Rhodospirillum rubrum*: molecular and physiological characterization of cooCTJ. *J. Bacteriol.* **179**, 2259–2266 [CrossRef](#) [Medline](#)
- Fox, J. D., He, Y., Shelver, D., Roberts, G. P., and Ludden, P. W. (1996) Characterization of the region encoding the CO-induced hydrogenase of *Rhodospirillum rubrum*. *J. Bacteriol.* **178**, 6200–6208 [CrossRef](#) [Medline](#)
- He, Y., Shelver, D., Kerby, R. L., and Roberts, G. P. (1996) Characterization of a CO-responsive transcriptional activator from *Rhodospirillum rubrum*. *J. Biol. Chem.* **271**, 120–123 [CrossRef](#) [Medline](#)
- Jeon, W. B., Cheng, J., and Ludden, P. W. (2001) Purification and characterization of membrane-associated CooC protein and its functional role in

- the insertion of nickel into carbon monoxide dehydrogenase from *Rhodospirillum rubrum*. *J. Biol. Chem.* **276**, 38602–38609 [CrossRef](#) [Medline](#)
12. Sydor, A. M., Lebrette, H., Ariyakumaran, R., Cavazza, C., and Zamble, D. B. (2014) Relationship between Ni(II) and Zn(II) coordination and nucleotide binding by the *helicobacter pylori* [NiFe]-hydrogenase and urease maturation factor HypB. *J. Biol. Chem.* **289**, 3828–3841 [CrossRef](#) [Medline](#)
13. Zambelli, B., Stola, M., Musiani, F., De Vriendt, K., Samyn, B., Devreese, B., Van Beeumen, J., Turano, P., Dikiy, A., Bryant, D. A., and Ciurli, S. (2005) UreG, a chaperone in the urease assembly process, is an intrinsically unstructured GTPase that specifically binds  $Zn^{2+}$ . *J. Biol. Chem.* **280**, 4684–4695 [CrossRef](#) [Medline](#)
14. Timm, J., Brochier-Armanet, C., Perard, J., Zambelli, B., Ollagnier-de-Choudens, S., Ciurli, S., and Cavazza, C. (2017) The CO dehydrogenase accessory protein CooT is a novel nickel-binding protein. *Metallomics* **9**, 575–583 [CrossRef](#) [Medline](#)
15. Watt, R. K., and Ludden, P. W. (1998) The identification, purification, and characterization of CooJ. *J. Biol. Chem.* **273**, 10019–10025 [CrossRef](#) [Medline](#)
16. Watt, R. K., and Ludden, P. W. (1999)  $Ni^{2+}$  transport and accumulation in *Rhodospirillum rubrum*. *J. Bacteriol.* **181**, 4554–4560 [Medline](#)
17. Sutton, R. B., Fasshauer, D., Jahn, R., and Brunger, A. T. (1998) Crystal structure of a SNARE complex involved in synaptic exocytosis at 2.4 Å resolution. *Nature* **395**, 347–353 [CrossRef](#) [Medline](#)
18. Sliz, P., Engelmann, R., Hengstenberg, W., and Pai, E. F. (1997) The structure of enzyme IIA<sub>lactose</sub> from *Lactococcus lactis* reveals a new fold and points to possible interactions of a multicomponent system. *Structure* **5**, 775–788 [CrossRef](#) [Medline](#)
19. Krissinel, E., and Henrick, K. (2007) Inference of macromolecular assemblies from crystalline state. *J. Mol. Biol.* **372**, 774–797 [CrossRef](#) [Medline](#)
20. Tria, G., Mertens, H. D., Kachala, M., and Svergun, D. I. (2015) Advanced ensemble modelling of flexible macromolecules using X-ray solution scattering. *IUCr* **2**, 207–217 [CrossRef](#) [Medline](#)
21. Colpas, G. J., and Hausinger, R. P. (2000) *In vivo* and *in vitro* kinetics of metal transfer by the *Klebsiella aerogenes* urease nickel metallochaperone, UreE. *J. Biol. Chem.* **275**, 10731–10737 [CrossRef](#) [Medline](#)
22. Zhang, J. W., Butland, G., Greenblatt, J. F., Emili, A., and Zamble, D. B. (2005) A role for SlyD in the *Escherichia coli* hydrogenase biosynthetic pathway. *J. Biol. Chem.* **280**, 4360–4366 [CrossRef](#) [Medline](#)
23. Olson, J. W., and Maier, R. J. (2000) Dual roles of Bradyrhizobium japonicum nickel protein in nickel storage and GTP-dependent Ni mobilization. *J. Bacteriol.* **182**, 1702–1705 [CrossRef](#) [Medline](#)
24. Real-Guerra, R., Staniscuasi, F., Zambelli, B., Musiani, F., Ciurli, S., and Carlini, C. R. (2012) Biochemical and structural studies on native and recombinant glycine max UreG: a detailed characterization of a plant urease accessory protein. *Plant Mol. Biol.* **78**, 461–475 [CrossRef](#) [Medline](#)
25. Desguin, B., Soumillion, P., Hols, P., and Hausinger, R. P. (2016) Nickel-pincer cofactor biosynthesis involves LarB-catalyzed pyridinium carboxylation and LarE-dependent sacrificial sulfur insertion. *Proc. Natl. Acad. Sci. U.S.A.* **113**, 5598–5603 [CrossRef](#) [Medline](#)
26. Vinella, D., Fischer, F., Vorontsov, E., Gallaud, J., Malosse, C., Michel, V., Cavazza, C., Robbe-Saule, M., Richaud, P., Chamot-Rooke, J., Brochier-Armanet, C., and De Reuse, H. (2015) Evolution of *Helicobacter*: acquisition by gastric species of two histidine-rich proteins essential for colonization. *PLoS Pathog.* **11**, e1005312 [CrossRef](#) [Medline](#)
27. Lacasse, M. J., Douglas, C. D., and Zamble, D. B. (2016) Mechanism of selective nickel transfer from HypB to HypA, *Escherichia coli* [NiFe]-hydrogenase accessory proteins. *Biochemistry* **55**, 6821–6831 [CrossRef](#) [Medline](#)
28. Khorasani-Motlagh, M., Lacasse, M. J., and Zamble, D. B. (2017) High-affinity metal binding by the *Escherichia coli* [NiFe]-hydrogenase accessory protein HypB is selectively modulated by SlyD. *Metallomics* **9**, 482–493 [CrossRef](#) [Medline](#)
29. Banaszak, K., Martin-Diaconescu, V., Bellucci, M., Zambelli, B., Rypniewski, W., Maroney, M. J., and Ciurli, S. (2012) Crystallographic and X-ray absorption spectroscopic characterization of *Helicobacter pylori* UreE bound to  $Ni^{2+}$  and  $Zn^{2+}$  reveals a role for the disordered C-terminal arm in metal trafficking. *Biochem. J.* **441**, 1017–1026 [CrossRef](#) [Medline](#)
30. Pérard, J., Covès, J., Castellan, M., Solard, C., Savard, M., Miras, R., Galop, S., Signor, L., Crouzy, S., Michaud-Soret, I., and de Rosny, E. (2016) Quaternary structure of Fur proteins, a new subfamily of tetrameric proteins. *Biochemistry* **55**, 1503–1515 [CrossRef](#) [Medline](#)
31. Mueller-Dieckmann, C., Bowler, M. W., Carpentier, P., Flot, D., McCarthy, A. A., Nanao, M. H., Nurizzo, D., Pernot, P., Popov, A., Round, A., Royant, A., de Sanctis, D., von Stetten, D., and Leonard, G. A. (2015) The status of the macromolecular crystallography beamlines at the European Synchrotron Radiation Facility. *Eur. Phys. J. Plus* **130**, 70 [CrossRef](#)
32. Kabsch, W. (2010) XDS. *Acta Crystallogr. D* **66**, 125–132 [CrossRef](#) [Medline](#)
33. Winn, M. D., Ballard, C. C., Cowtan, K. D., Dodson, E. J., Emsley, P., Evans, P. R., Keegan, R. M., Krissinel, E. B., Leslie, A. G., McCoy, A., McNicholas, S. J., Murshudov, G. N., Pannu, N. S., Potterton, E. A., Powell, H. R., et al. (2011) Overview of the CCP4 suite and current developments. *Acta Crystallogr. D* **67**, 235–242 [CrossRef](#) [Medline](#)
34. Sheldrick, G. M. (2010) Experimental phasing with SHELXC/D/E: combining chain tracing with density modification. *Acta Crystallogr. D* **66**, 479–485 [CrossRef](#) [Medline](#)
35. Adams, P. D., Afonine, P. V., Bunkóczi, G., Chen, V. B., Davis, I. W., Echols, N., Headd, J. J., Hung, L.-W., Kapral, G. J., Grosse-Kunstleve, R. W., McCoy, A. J., Moriarty, N. W., Oeffner, R., Read, R. J., Richardson, D. C., et al. (2010) PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr. D* **66**, 213–221 [CrossRef](#) [Medline](#)
36. Emsley, P., Lohkamp, B., Scott, W. G., and Cowtan, K. (2010) Features and development of Coot. *Acta Crystallogr. D Biol. Crystallogr.* **66**, 486–501 [CrossRef](#) [Medline](#)
37. Afonine, P. V., Grosse-Kunstleve, R. W., Echols, N., Headd, J. J., Moriarty, N. W., Mustyakimov, M., Terwilliger, T. C., Urzhumtsev, A., Zwart, P. H., and Adams, P. D. (2012) Towards automated crystallographic structure refinement with phenix.refine. *Acta Crystallogr. D* **68**, 352–367 [CrossRef](#) [Medline](#)
38. Petoukhov, M. V., Franke, D., Shkumatov, A. V., Tria, G., Kikhney, A. G., Gajda, M., Gorba, C., Mertens, H. D. T., Konarev, P. V., and Svergun, D. I. (2012) New developments in the ATSAS program package for small-angle scattering data analysis. *J. Appl. Crystallogr.* **45**, 342–350 [CrossRef](#) [Medline](#)
39. Franke, D., and Svergun, D. I. (2009) DAMMIF, a program for rapid *ab initio* shape determination in small-angle scattering. *J. Appl. Crystallogr.* **42**, 342–346 [CrossRef](#) [Medline](#)
40. Konarev, P. V., Volkov, V. V., Sokolova, A. V., Koch, M. H. J., and Svergun, D. I. (2003) PRIMUS: a Windows PC-based system for small-angle scattering data analysis. *J. Appl. Crystallogr.* **36**, 1277–1282 [CrossRef](#)
41. Svergun, D. I. (2007) Determination of the regularization parameter in indirect-transform methods using perceptual criteria. *J. Appl. Crystallogr.* **25**, 495–503 [CrossRef](#)
42. Volkov, V. V., and Svergun, D. I. (2003) Uniqueness of *ab initio* shape determination in small-angle scattering. *J. Appl. Crystallogr.* **36**, 860–864 [CrossRef](#)
43. Svergun, D., Barberato, C., and Koch, M. H. J. (2007) CRYSOLE: a program to evaluate X-ray solution scattering of biological macromolecules from atomic coordinates. *J. Appl. Crystallogr.* **28**, 768–773 [CrossRef](#)
44. Brooks, B. R., Brucoleri, R. E., Olafson, B. D., States, D. J., Swaminathan, S., and Karplus, M. (1983) CHARMM: a program for macromolecular energy, minimization, and dynamics calculations. *J. Comput. Chem.* **4**, 187–217 [CrossRef](#)
45. Mackerell, A. D., Jr., Feig, M., and Brooks, C. L., 3rd (2004) Extending the treatment of backbone energetics in protein force fields: Limitations of gas-phase quantum mechanics in reproducing protein conformational distributions in molecular dynamics simulations. *J. Comput. Chem.* **25**, 1400–1415 [CrossRef](#) [Medline](#)
46. MacKerell, A. D., Bashford, D., Bellott, M., Dunbrack, R. L., Evanseck, J. D., Field, M. J., Fischer, S., Gao, J., Guo, H., Ha, S., Joseph-McCarthy, D., Kuchnir, L., Kucera, K., Lau, F. T., Mattos, C., et al. (1998) All-atom empirical potential for molecular modeling and dynamics studies of proteins. *J. Phys. Chem. B* **102**, 3586–3616 [CrossRef](#) [Medline](#)

## A novel Ni(II)-binding site in the nickel chaperone *CooJ*

47. Wu, X., Hodoscek, M., and Brooks, B. R. (2012) Replica exchanging self-guided Langevin dynamics for efficient and accurate conformational sampling. *J. Chem. Phys.* **137**, 044106 [CrossRef](#) [Medline](#)
48. Jiang, W., Hodoscek, M., and Roux, B. (2009) Computation of absolute hydration and binding free energy with free energy perturbation distributed replica-exchange molecular dynamics. *J. Chem. Theory Comput.* **5**, 2583–2588 [CrossRef](#) [Medline](#)
49. Hassan, S. A., Mehler, E. L., Zhang, D., and Weinstein, H. (2003) Molecular dynamics simulations of peptides and proteins with a continuum electrostatic model based on screened Coulomb potentials. *Proteins* **51**, 109–125 [CrossRef](#) [Medline](#)
50. Crooks, G. E., Hon, G., Chandonia, J.-M., and Brenner, S. E. (2004) WebLogo: a sequence logo generator. *Genome Res.* **14**, 1188–1190 [CrossRef](#) [Medline](#)
51. Humphrey, W., Dalke, A., and Schulten, K. (1996) VMD: Visual molecular dynamics. *J. Mol. Graph.* **14**, 33–38 [CrossRef](#) [Medline](#)
52. Yachdav, G., Kloppmann, E., Kajan, L., Hecht, M., Goldberg, T., Hamp, T., Hönigsmid, P., Schafferhans, A., Roos, M., Bernhofer, M., Richter, L., Ashkenazy, H., Punta, M., Schlessinger, A., Bromberg, Y., *et al.* (2014) PredictProtein—an open resource for online prediction of protein structural and functional features. *Nucleic Acids Res.* **42**, W337–W343 [CrossRef](#) [Medline](#)

**The carbon monoxide dehydrogenase accessory protein CooJ is a histidine-rich multidomain dimer containing an unexpected Ni(II)-binding site**

Marila Alfano, Julien Pérard, Philippe Carpentier, Christian Basset, Barbara Zambelli, Jennifer Timm, Serge Crouzy, Stefano Ciurli and Christine Cavazza

*J. Biol. Chem.* 2019, 294:7601-7614.

doi: 10.1074/jbc.RA119.008011 originally published online March 11, 2019

---

Access the most updated version of this article at doi: [10.1074/jbc.RA119.008011](https://doi.org/10.1074/jbc.RA119.008011)

Alerts:

- [When this article is cited](#)
- [When a correction for this article is posted](#)

[Click here](#) to choose from all of JBC's e-mail alerts

This article cites 52 references, 18 of which can be accessed free at <http://www.jbc.org/content/294/19/7601.full.html#ref-list-1>

**Supplementary information for**

**The CO-dehydrogenase accessory protein CooJ is a histidine-rich multidomain dimer with  
an unexpected Ni(II)-binding site**

Marila Alfano, Julien Pérard, Philippe Carpentier, Christian Basset, Barbara Zambelli, Jennifer  
Timm, Serge Crouzy, Stefano Ciurli and Christine Cavazza\*



## Cloning and expression of *RrCooJ* and its mutants

The *Rhodospirillum rubrum* *CooJ* gene (*RrcooJ*) was ordered from DNA 2.0 in pET15b as codon-optimized gene for expression in *E. coli*. A previously described protocol (1) was used with the primers described in Table S1 for generation of the different mutants: the truncated protein lacking the C-terminal disordered region from Asp69 to Pro115 containing the histidine tail (*RrCooJ*-Δ), the triple H18A/H22A/H26A mutants (*RrCooJ*-3HA & *RrCooJ*-Δ3HA), and the *RrCooJ*-Δ-E29A mutant. All protein sequences contain an additional glycine at the N-terminus from the cloning procedure.

PCR was carried out in the presence of the appropriate primers using Phusion polymerase-GC (NEB) at the recommended TM. PCR samples were incubated with a reaction buffer containing 2UI of DpnI, 10 UI of T4 DNA Ligase, 1mM of ATP and 2 UI PNK in PNK buffer from NEB for 15 min at RT before transformation in Top10 ultra competent cells. Each mutant was DNA-sequenced (MWG).

*E. coli* BL21 (DE3) cells harboring *RrCooJ* or its mutants were cultured in LB medium, supplemented with appropriate antibiotics, at 37 °C and 180 rpm. The gene expression was induced with 0.5 mM IPTG at OD<sub>600</sub> ~0.6 and left overnight at 16 °C and 180 rpm before cell harvest by centrifugation. The cell pellets were resuspended in 50 mM Tris-HCl pH 8.5 containing 1 mM EDTA, and frozen at -80 °C.

For the production of selenomethionine-labeled *RrCooJ*-Δ, *E. coli* BL21 (DE3) cells harboring *RrCooJ*-Δ were cultured in M9 medium, supplemented with appropriate antibiotics, at 37 °C and 180 rpm. When the OD<sub>600</sub> reached ~0.6, 100 mg/L of phenylalanine, lysine and threonine and 50 mg/L of leucine, isoleucine, valine and selenomethionine were added to the culture. After 30 minutes at 37 °C and 180 rpm, the gene expression was induced with 0.5 mM IPTG and left overnight at 18 °C and 180 rpm before cell harvest by centrifugation.

## Molecular Dynamics simulations details

A dimer of *RrCooJ* was modelled from Gly1(NH<sub>3</sub><sup>+</sup>) to Pro115(COO<sup>-</sup>) with the all-atom force field *all22* for proteins (2)(3). All histidines were protonated on their Nδ atom, whereas aspartate and glutamate residues were kept unprotonated (negative charge). All arginine and lysine residues were positively charged. Initial coordinates were taken from the X-ray structure of the Ni-*RrCooJ*-Δ dimer solved from E3 to L68 for chain A and P5 to L68 for chain B. All missing atom coordinates were built with CHARMM using internal coordinates and the structure was energy minimized down to a gradient of 0.1 kcal/mol/Å with the ABNR algorithm and subject to harmonic restraints

on heavy atoms resolved in the X-ray structure with force constant of 5 kcal/mol/Å<sup>2</sup>. For RXSGLD, temperatures of replicas and their self-guiding temperatures were exponentially spaced from 300 to 400 K and from 300 to 600K, respectively. The guiding factor increment was set to 0.2 and exchange was attempted every 50 steps (4). Coordinates of backbone atoms resolved in the X-ray structure were subject to harmonic restraints with force constant of 5 kcal/mol/Å<sup>2</sup>. Additionally, NOE-type distance restraints using a bi-harmonic potential between 1.8 and 2.2 Å with a force constant of 80 kcal/mol/Å<sup>2</sup> were applied between atoms O of residue *i* and H of residue *i*+4 (*i*=4 to 6 and 60 to 62) for both chains, to prevent fraying of α-helices at their N- and C-terminal during high temperature dynamics. The Screened Coulomb Potentials Implicit Solvent Model (SCPISM) (5), a continuum model of solvation with implicit waters was used in this work in conjunction with the all-atom force field *all22* for proteins. 16 replicas were created and RXGLD simulations were run for 8 ns with a time step of 1 fs on 16 processors.



**Table S1. Primers sequence used to generate *RrCooJ* mutants.**

| Name    | Sequence (5' – 3')                   |
|---------|--------------------------------------|
| CooJ-ΔF | TAAGGATCCGGCTGCTAAC                  |
| CooJ-ΔR | CAGATCGCGCAGAACGGC                   |
| H18AF   | GGTCGAGGGTGCGATGGGCGAGATCGGTGTT      |
| H18AR   | GCGTCCAGAAAGTGTGCCAAATAAATACCCAAACG  |
| H22AF   | TGTCGAGGGTGCGATGGGCGAGATCGGTGTT      |
| H22AR   | TGGTCCAGAAACGCTGCCAAATAAATACCCAAACG  |
| H26AF   | GGTCGAGGGTCACATGGGCGAGATCGGTGTT      |
| H26AR   | GCGTCCAGAAACGCTGCCAAATAAATACCCAAACG  |
| E29AF   | TCACATGGGC <sub>gct</sub> ATCGGTGTTC |
| E29AR   | CCCTCGACATGGTCCAGA                   |

**Table S2: Elements detected by ICP-AES in *RrCooJ* and mutants samples.**

| Sample name      | Element concentration in µg/L |             |              |                |                |
|------------------|-------------------------------|-------------|--------------|----------------|----------------|
|                  | Sodium (Na)                   | Sulphur (S) | Calcium (Ca) | Magnesium (Mg) | Phosphorus (P) |
| apoRrCooJ        | 13                            | 2.3         | 4.4          | 3.3            | 13             |
| apoRrCooJ-Δ      | 6.9                           | 2.0         | 3.8          | 3.2            | 14             |
| apoRrCooJ C2S    | 11                            | 2.1         | 4.0          | 2.7            | 11             |
| apoRrCooJ 3HA    | 11                            | 2.2         | 2.4          | 2.9            | 15             |
| apoRrCooJ-Δ 3HA  | 7.2                           | 2.1         | 2.3          | 2.9            | 10             |
| apoRrCooJ-Δ E29A | 7.0                           | 2.1         | 2.4          | 2.6            | 14             |

Protein concentration is 0.41µM dimer for each sample. This table reports the detected elements by ICP AES after the standard mineralization procedure with HNO<sub>3</sub>. The other elements are undetectable under the experimental conditions because they are below the ICP AES detection limit.

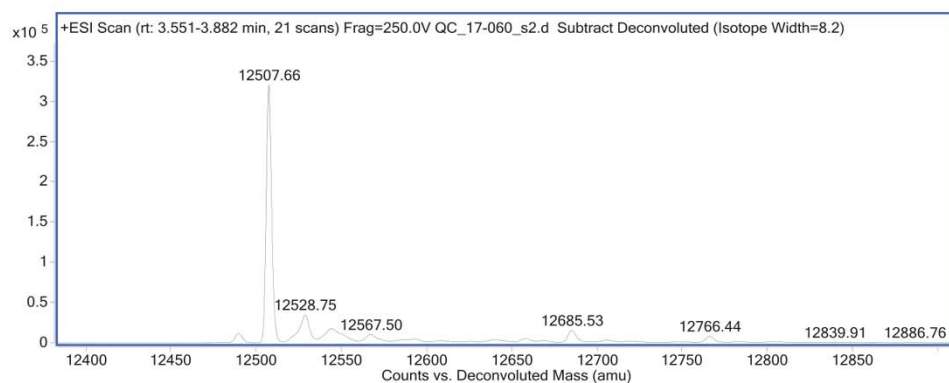
**Table S3: Data collection and refinement statistics for the data set used to solve the Ni-RrCooJ- $\Delta$  structure.**

| Data collection                          | Ni-RrCooJ- $\Delta$  |
|--|--|
| Beamline                                 | ID30B (ESRF)   |
| Wavelength (Å)                           | 0.9786   |
| Space group                              | P2 <sub>1</sub>  |
| Unit cell parameters                     | a = 46.80 Å<br>b = 74.28 Å<br>c = 70.64 Å<br>$\alpha=\gamma= 90^\circ$<br>$\beta= 92.74^\circ$ |
| Resolution range (Å)                     | 46.75-2.04 [2.16-2.04]   |
| $R_{\text{sym}}$ (%)                     | 14 [88.3]  |
| Mean I/ $\sigma$ (I)                     | 9.4 [1.6]  |
| Completeness (%)                         | 99.2 [96]  |
| Redundancy                               | 6.4 [6.2]  |
| $N_{\text{measured}}$                    | 198360 [29572]   |
| $N_{\text{unique}}$                      | 30586 [4749]   |
| CC1/2 (%)                                | 99.7 [69.1]  |
| <b>Refinement</b>                        |  |
| $R$ factor/ $R_{\text{free}}$ factor (%) | 20.0 / 25.9  |
| No. atoms                                | 4161   |
| No. water molecules                      | 300  |
| No. Nickel ions [Ni <sup>2+</sup> ]      | 4  |
| No. SeMet                                | 16   |
| Wilson B-factor (Å <sup>2</sup> )        | 31.4   |
| Rmsd bonds (Å)                           | 0.007  |
| Rmsd angles (°)                          | 1.0  |
| Ramachandran plot (%)                    |  |
| Favored                                  | 98.2   |
| Outliers                                 | 0.2  |
| Allowed                                  | 1.6  |
| Other ions and ligands                   | Cl <sup>-</sup> , Ca <sup>2+</sup> , PEG, oxidized TCEP  |
| <b>PDB DEPOSITION CODE</b>               | 6HK5   |

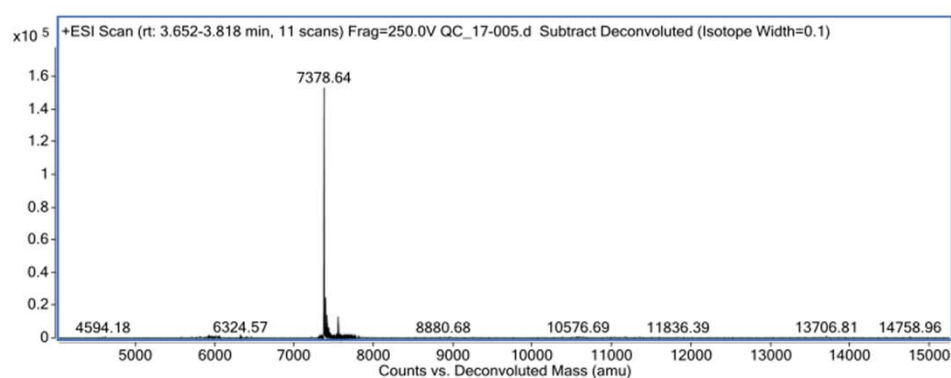
Values in brackets are for the highest resolution shell.

**Figure S1. ESI-MS spectra of *RrCooJ* and mutants.**

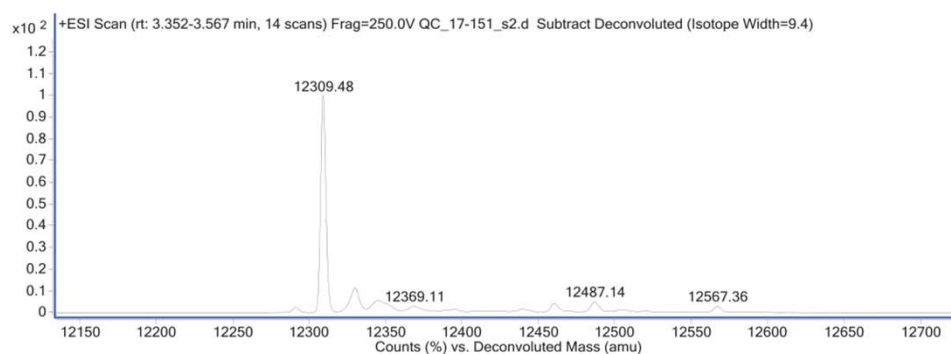
**A *RrCooJ***



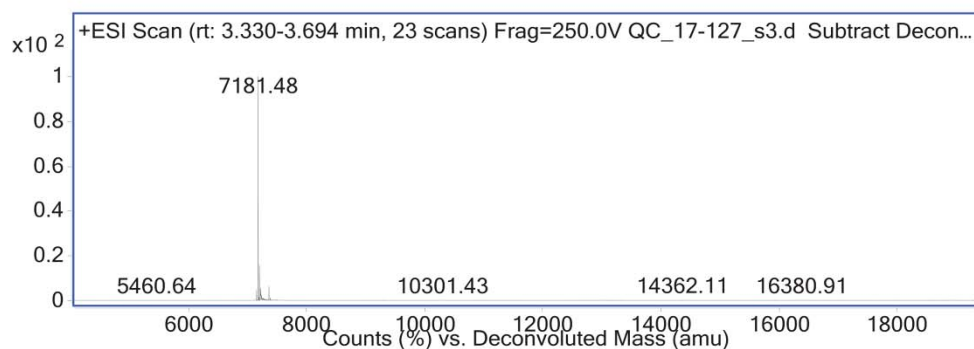
**B *RrCooJ*-Δ**



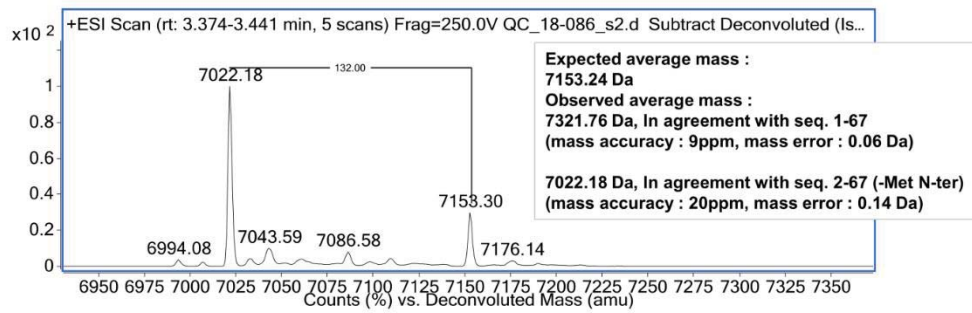
**C *RrCooJ*-3HA**



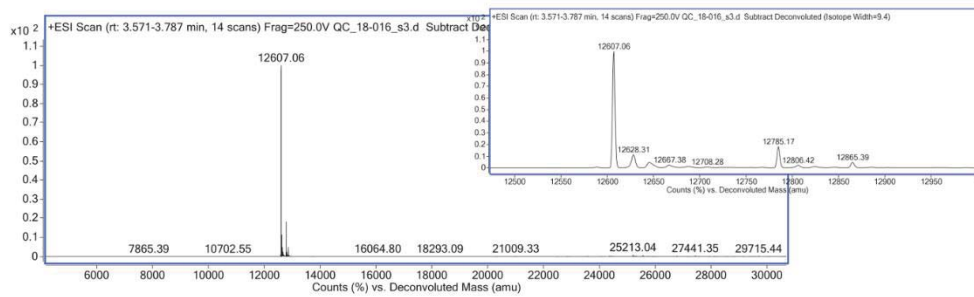
**D *RrCooJ*-Δ3HA**



## E *RrCooJ*- $\Delta$ E29A

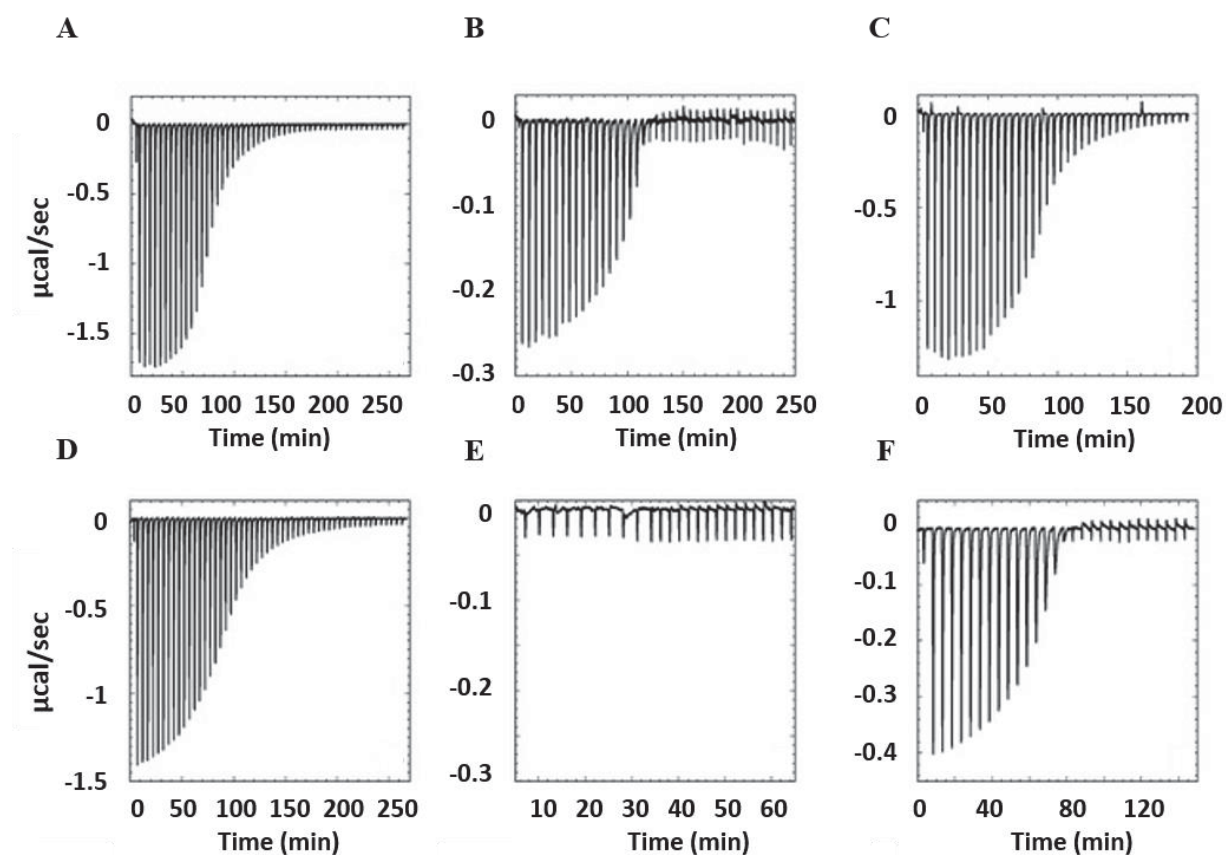


## F *RrCooJ*-2CS



(A) *RrCooJ* (B) *RrCooJ*- $\Delta$  (C) *RrCooJ*-3HA (D) *RrCooJ*- $\Delta$ 3HA (E) *RrCooJ*- $\Delta$ E29A (F) *RrCooJ*-C2S.

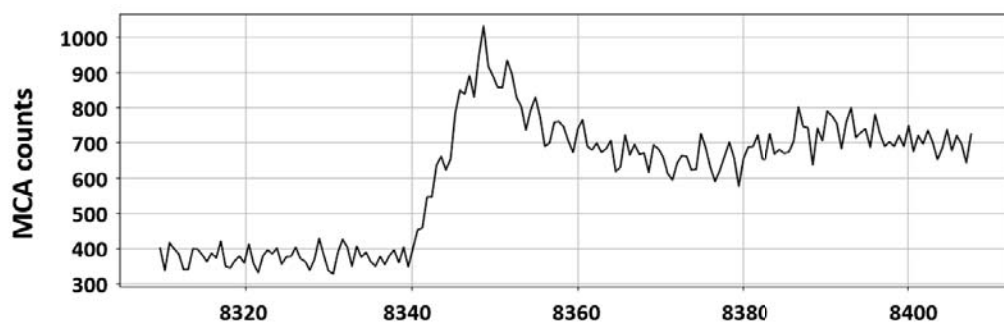
**Figure S2. Raw titration data representing the heat released upon Ni(II) injection over RrCooJ and its variants, determined using ITC.**



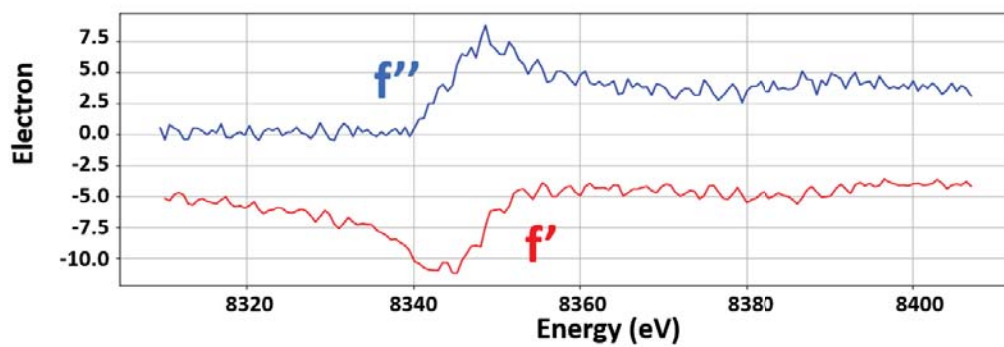
(A) RrCooJ (B) RrCooJ- $\Delta$  (C) RrCooJ-C2S (D) RrCooJ-3HA (E) RrCooJ- $\Delta$ 3HA (F) RrCooJ- $\Delta$ E29A. The integrated heat data, as well as the best fits to derive thermodynamic parameters, are shown in Figure 4.

Figure S3. X-ray fluorescence spectrum at the nickel edge of Ni-RrCooJ- $\Delta$  crystal.

**A**



**B**



(A) Experimental X-ray fluorescence spectrum at Ni K edge measured on a crystal of Ni-CooJ- $\Delta$  on beamline ID29 at the ESRF. (B) Calculated anomalous scattering factors  $f'$  and  $f''$ .



**Figure S4: Alignment of 46 putative CooJ sequences using ClustalW 2.1.**

```

A0A2N2PV90 -----MSMTDIEKLRVLLPHWLEHNTAHADDYRAWIERARAAGEPHV
A0A1V5RSM0 -----MSMSDIEKLRLLLLPHWLEHNASHADDYHGWIERTLAAGEPHA
A0A1W9V5U2 -----MNDQEKLRVLI PHWIEHNHEHADEFRRWAEQAG-----DA
A0A0S8ICL2 -----MTNISNIDKLRVLLPHWIRHNAEHAAEFRRWAEKAG-----DA
A0A1J5AM16 -----MMEQIEKLRVLLPHWIEHNGEHAAEFRDYAEHSG-----KA
A0A1E7J266 -----MLMKDQEKLRVLLVHWVKHNREHAAEEFGQWARRAVELGFNEA
A0A1H0TP87 -----MDNVEKLRVLLQHWIDHNKGHVVEEFETWRKLMSEENRESL
A0A2G2DTS5 -----MDNVEKLRVLLQHWIDHNKGHVVEEFENWRKLMADENQEKT
A0A0F2N9P2 -----MDNVEKLRVMLQHWIEHNKGHMEEFQKQSVITDEGKPAL
K1ZCM0 -----MDNVEKLRVMLQHWIEHNKGHMEEFQKQAVMTDEGKPAL
K1Y6L6 -----MDNVEKLRVMLQHWIEHNKGHMEEFQKQETMAGEGKTSL
A0A1F9NPL5 -----MTGEGMDNVEKLRVMLQHWIEHNKGHMEEFQKQETMAGEGKTSL
A0A1E7GJ61 -----MENIEKLRVLLQHWIDHNKGHAAEFQKQKLMADDAKGVV
A0A1M7Y3C3 -----MENKEKLRVLLQHWIDHNKSHAAEFERWRQIADNDGESAV
A0A0A2HUV8 -----MDNTEKIRILLQHWIDHNKEHAAEYERWQETLVDDGLAPI
A0A1M5TKJ2 -----MKNREKLRVLLQHWIEHNGGHVAFENKWRATMADEHQDRL
A0A2G6MZ09 -----MDNIEKLRVILQHWIDHNGGHVDEFQKWRQLMNNEGKTEI
A0A1V4MI89 -----MTDLEKLRILLPHWVAHNHEHMAEIDRWASLPEISNNVEI
A0A1W9LUV0 -----MNDIEKIRTLPHWIDHTKSHGEEFAHWAGHIRTAGQQES
A0A1W9M2N4 -----MDDIAKIRTLPHWIEHTKSHENEFTHWVTHVREAGQQDC
M1NER2 -----MQQLEKIRVLLPHWIEHNKGHAAEERKWAQAEDK---DV
A0A286TWP4 -----MTEQDTIKLRVLLPHWIEHNNSHMAEFKRWEGEAKKESGEEV
A0A0B0EGJ4 -----MAEQDTIKLRVLLPHWIEHNISHIAEFKRWEGEAKKESGEEV
A0A1G3B468 -----MSEQNTTKKLQVLLPHWIEHNNNHITFRKWESEASTGPGQEV
A0A0F9F8E7 -----MSEQDTTKKLRVLLPHWIEHNNNHIAEFKRWEREAREASGQEV
Q1NTE8 -----MTEQQVVEKLRVLLPHWIEHNQGHAAEFKRWKSATARAEGCQEP
D6Z5H4 -----MTEKQTEKLRVLLPHWIEHNQGHATEFRRWKSATARTENLNEV
A0A1F9L257 -----MDQKTTREKLQILLPHWIEHNNHAEAFKKWADLVRTEGQGGGL
A0A2N2GNQ0 -----MDQKTTMEKLQILLPHWIEHNNHAEAFKKWADLVRSEGKGNL
A0A2N2INF5 -----MDQKTTMEKLRIILLPHWIEHNNHAEAFKKWVELVSGEGHAVL
A0A1F9M2B0 -----MDQKTTREKLQILLPHWVEHNNHETEFKKWAALAQSEGQAL
A0A1G1A614 -----MEPTTIIDKLRILLPHWITHRNHEAEFGKWAALARTEGAENL
A0A2N6C0K0 -----MSEKTNLEKLRILLPHWIEHSQSHQDEFKRWVEVARTEGQAEA
A0A1Q6DHA7 -----MIPHWIEHSRSHQHEFAKWLDVAKKEGAET
A0A2N6C4W8 -----MTEKQYSEIEKLRLLLLTHWLQHNEDHGKEYAKWAAVAREAGESTA
A0A0S8AJX8 -----MTDKQYSDTEKLQMLLTHWLQHNESHGKEYAKWAAAAAREAGHSTT
A0A0S8D9J8 -----MTEKQYSDSEKLQMLITHWLKHNEHSHGREYAKWAAVARQTGHPPAA
A0A0S8CVL2 -----MAEKEFTDIEKLQMLLNHWLQHNQSHGREYAKWAAVARQAGHSA
A0A0S8AGL1 -----MAEKQYSDTEKLQTLLDHWLQHNRGHGAEYQKWAEEVARQAGLSE
A0A2H0AGC5 -----MSIPDIDKIRVLLPHWLEHNQSHRAEFGRWAEVARGAGLAEV
A0A1F9M304 -----MTDLDKLRVLLPHWIEHNRGHLEAFKRWATMAGSSATPQA
A0A1J5DFD8 -----MEGNELKKLAVLLNHWIKHNEDHVKEYREWAEKVYKQGLKEV
P72321 -----MTESPERGRKRLGIYLAHFLDHVEGHMGEIG---VQDALADEARL
A0A212LFB2 -----MTARKTPASTGDDRLGIYLAHFADHLEDHLEDLT--QHRIRETESAEL
Q20XW8 -----MTPQDQAKARLAYLAHFIDHAQDHVAEIE---AQRGAI GASARL
A0A2N3PZM2 MAAAPAWKKTCKDERCAMNEPGVADKLVLLPHWIEHNDRHVAELRRWRRALHDQAP-LA
: *: * * :

A0A2N2PV90 VKHLAAAEKLEGVNRDLEAALEHLGG-----PTEAAGHD----
A0A1V5RSM0 AEHLAAAVEKLTGVNHDLEALLQRLGG-----PAAHADHEGAAA
A0A1W9V5U2 APDIQAAAEAITRANESLAAALKKLGG-----AASHHHHKA
A0A0S8ICL2 QEDIHAAAEQMVLANHKLESAKQLGG-----ALEGKDHRHE--
A0A1J5AM16 KDSLLAAQLMEEANARLSEALDKLGG-----PLEHHHA----
A0A1E7J266 ASELEAAVQEMKVVNSRLMTALDILGG-----PLEHHHCDHKN-
A0A1H0TP87 ADHITAAIEKMATVNGELGKALHELGGPQKSS-----DHHHDDHHHHDDHHHH
A0A2G2DTS5 ASHITAIRQMAMVNAELGKALHQVGGPNAAG-----DKGHEHGHSHSHDHN
A0A0F2N9P2 AEHISRAVKLMATMNEELGKALQEAGG-----TKHDDEHHHHDDHHHH
K1ZCM0 ADHIGRAVKLMVTMNEELGKALHEAGG-----HHHDDDDDDHHHHHDHG
K1Y6L6 AGHIAKAIEMMVGMMNEELHQALHEAGG-----HGHDHDDHHHHHHHHH
A0A1F9NPL5 AAHIAKAIEMMVGMMNEELHQALHEAGG-----HGHDHDDHHHHHHHHH
A0A1E7GJ61 ADHIGGAIEKEMEQA NEHLAKALHDAGGPKE-----GGGGDGGHHHHGDBGH
A0A1M7Y3C3 ATHIAEAI SDMPKANEALQNALNKAGG-----PAEHSDDHHHH
A0A0A2HUV8 AGYISAAVGEMQKVNA YLDKALEEAGG-----RDEASGHHHHNDHH
A0A1M5TKJ2 VAALDRAAAQMDAVSETLQQALDES GG-----PVADHAHHHHHHHGD
A0A2G6MZ09 AASLEEAKTQMNKISD LLA AVLKDIGE-----PVAGDHHHH----
A0A1V4MI89 KKALQKAITATEKVNEELQHAMD MAGG-----PIENPETHGNQQRHG
A0A1W9LUV0 ATALAQAAAALREVEEHLRRALTLAGGP-----LQATQPHDEHHHHHD
A0A1W9M2N4 GDALAQAAAASLRQVEEHLRHALTLAGGP-----LAAEHEHHHHHHHDH
M1NER2 NLHLNAAALTAMEVVTNHLERALAAAGG-----AKTDDHHDDHHHHHH
A0A286TWP4 AQLLDKAI RDMEEGAGKSLSEALEKVGG-----PLEGAGHHHH--
A0A0B0EGJ4 AKLLDKAISDM EKAGKSLSEALEKVGG-----PLESGGGHHHHH-
A0A1G3B468 ALLLGKAI SNMEEAGKSLSEALEKIGG-----PLEGGEGQH HHH
A0A0F9F8E7 ALLLEKAI SDMEKTGKSLSEALEKVGG-----PLEGGGVHHHPHH
Q1NTE8 ANLIDRAAELLDEAERLLSEALTKAGG-----PAQGGHRHHHHHH

```

|            |  |
|------------|--|
| D6Z5H4     | AKLIDRAADLMEDADRLLGEALIKAGG-----PTHRGHSHHHHHQ            |
| A0A1F9L257 | AELLDKAVASMGETDSILQKVLAEIGG-----PGEGHCHHHHPH             |
| A0A2N2GNQ0 | AELLDKAVASMGETDGLVKKVLAIEIGG-----PGESHGHGHHHH            |
| A0A2N2INF5 | ADLLAKAVVSMETDITLKKVLAIEIGG-----PGEGHHHGHSHH             |
| A0A1F9M2B0 | AELLNKAVTSMAETDGLLEKALAEIGG-----PGEGHHHHHHHY             |
| A0A1G1A614 | ATLLDKAAAAMVATDITLQVQGAVGA-----PGEIQGHHRHD               |
| A0A2N6C0K0 | AARIDRALALMADTDQALEEAELELGG-----KLGEHHHHGHHH             |
| A0A1Q6DHA7 | AEAIGKAMEKMAKTDKYLQKALESLGG-----AAEGHHHHHHHH             |
| A0A2N6C4W8 | AECIEQAVDLLAKADRAFEKALTSIGG-----PGKGHQHHHHHD-            |
| A0A0S8AJX8 | AEYIEQAIKLLAQADKAFAEQALESVGG-----PGKAHKHHHHHH            |
| A0A0S8D9J8 | ADYIEEAAGLLAKADKAFAEKALESVGG-----PHQGHQHGHSHH            |
| A0A0S8CVL2 | AQYIEQAVDLLTEADKVFEEKALAAVGG-----PGNEDQRHHHHHH           |
| A0A0S8AGL1 | AELIERAVAFLLKADDTALAKALASAGG-----PPKKHPHHHHQH            |
| A0A2H0AGC5 | AELIEQAMAKMAAVDVLLARALEKAGG-----KVAGHEHHH---             |
| A0A1F9M304 | AMLLEQAAARLAEADVALAQALVQLGG-----AVPDSPPHHHD              |
| A0A1J5DFD8 | AHNLKTAADLILQSNERFIEAQKGIPAS-----CAEEDHKHDPHHD           |
| P72321     | GALIDRALADMAVARASLNAVLRDLGEAPAPASPEAVHSPFHSASHDHDHAGHSHD |
| A0A212LFB2 | ARLLDEASASIEASRRIGRVLHLGAETGA-----GHGHGHDHGSHHH          |
| Q20XW8     | DQLLDQAVADVQLARRSLEAALEAIGGSPAR-----DVFSQHAHSHQPHLH      |
| A0A2N3PZM2 | ASMMDMAITQMEDAGRALAATARELAAG-----ETCRGVHPRAV-            |
|            | : * :  |
| A0A2N2PV90 | ----HPHEQHDHVR-----                                      |
| A0A1V5RSM0 | HHHEHPHEHGEHGHSHH-----                                   |
| A0A1W9V5U2 | -----  |
| A0A0S8ICL2 | -----  |
| A0A1J5AM16 | -----  |
| A0A1E7J266 | -----  |
| A0A1H0TP87 | HHHDHHDHHDH-----   |
| A0A2G2DTS5 | HTHDHHDHKKH-----   |
| A0A0F2N9P2 | HHH-----   |
| K1ZCM0     | HHHHHH-----  |
| K1Y6L6     | HHHD-----  |
| A0A1F9NPL5 | HHHD-----  |
| A0A1E7GJ61 | HHH-----   |
| A0A1M7Y3C3 | HHHH-----  |
| A0A0A2HUV8 | HHHH-----  |
| A0A1M5TKJ2 | GHHHH-----   |
| A0A2G6MZ09 | -----  |
| A0A1V4MI89 | HIHQKYGTD-----   |
| A0A1W9LUV0 | HHHDHGLHHHH-----   |
| A0A1W9M2N4 | HHGHDDHHKKH-----   |
| MLNER2     | HKH-----   |
| A0A286TWP4 | -----  |
| A0A0B0EGJ4 | -----  |
| A0A1G3B468 | -----  |
| A0A0F9F8E7 | -----  |
| Q1NTE8     | D-----   |
| D6Z5H4     | D-----   |
| A0A1F9L257 | HHD-----   |
| A0A2N2GNQ0 | HHYD-----  |
| A0A2N2INF5 | QHYD-----  |
| A0A1F9M2B0 | D-----   |
| A0A1G1A614 | HEPGTASPDHGHAG-----                                      |
| A0A2N6C0K0 | H-----   |
| A0A1Q6DHA7 | -----  |
| A0A2N6C4W8 | -----  |
| A0A0S8AJX8 | HHHD-----  |
| A0A0S8D9J8 | HD-----  |
| A0A0S8CVL2 | D-----   |
| A0A0S8AGL1 | HQHD-----  |
| A0A2H0AGC5 | -----  |
| A0A1F9M304 | HDDHHHGS-----  |
| A0A1J5DFD8 | GLH-----   |
| P72321     | HAHDHCHCHDHP-----  |
| A0A212LFB2 | HHGDGTHARD-----  |
| Q20XW8     | HSHEVVTGPGADVTEHHKANIGGPRPTA                             |
| A0A2N3PZM2 | -----  |

Alignment of 46 CooJ homologues. Each sequence is identified by its UniProtKB ID. ID sequence is colored according to its phylum: *Proteobacteria* (red), *Planctomycetes* (green), *Lentisphaerae* (blue), *Chloroflexi* (orange) and *Nitrospirae* (purple). Other CooJ homologues are unclassified sequences found in metagenomes.

## References

1. Pérard, J., Nader, S., Levert, M., Arnaud, L., Carpentier, P., Siebert, C., Blanquet, F., Cavazza, C., Renesto, P., Schneider, D., Maurin, M., Coves, J., Crouzy, S., and Michaud-Soret, I. (2018) Structural and functional studies of the metalloregulator Fur identify a promoter-binding mechanism and its role in *Francisella tularensis* virulence. *Commun. Biol.* **1**, 93
2. Mackerell, A. D., Feig, M., and Brooks, C. L. (2004) Extending the treatment of backbone energetics in protein force fields: Limitations of gas-phase quantum mechanics in reproducing protein conformational distributions in molecular dynamics simulations. *J. Comput. Chem.* **25**, 1400–1415
3. MacKerell, A. D., Bashford, D., Bellott, M., Dunbrack, R. L., Evanseck, J. D., Field, M. J., Fischer, S., Gao, J., Guo, H., Ha, S., Joseph-McCarthy, D., Kuchnir, L., Kuczera, K., Lau, F. T. K., Mattos, C., Michnick, S., Ngo, T., Nguyen, D. T., Prodhom, B., Reiher, W. E., Roux, B., Schlenkrich, M., Smith, J. C., Stote, R., Straub, J., Watanabe, M., Wiórkiewicz-Kuczera, J., Yin, D., and Karplus, M. (1998) All-Atom Empirical Potential for Molecular Modeling and Dynamics Studies of Proteins. *J. Phys. Chem. B.* **102**, 3586–3616
4. Jiang, W., Hodosek, M., and Roux, B. (2009) Computation of Absolute Hydration and Binding Free Energy with Free Energy Perturbation Distributed Replica-Exchange Molecular Dynamics. *J. Chem. Theory Comput.* **5**, 2583–2588
5. Hassan, S. A., Mehler, E. L., Zhang, D., and Weinstein, H. (2003) Molecular dynamics simulations of peptides and proteins with a continuum electrostatic model based on screened Coulomb potentials. *Proteins Struct. Funct. Bioinforma.* **51**, 109–125

### 3.3.2 Paper II



Communication

## Nickel-Induced Oligomerization of the Histidine-Rich Metallochaperone CooJ from *Rhodospirillum Rubrum*

Marila Alfano , Julien Pérard and Christine Cavazza \*

Univ. Grenoble Alpes, CEA, CNRS, CBM, F-38000 Grenoble, France

\* Correspondence: christine.cavazza@cea.fr

### Paper highlights

- Nickel binding to *RrCooJ* increases its stability, as demonstrated via chemical denaturation and secondary structural studies of apo vs holo proteins.
- *RrCooJ* forms dynamic and reversible higher-ordered oligomers through the histidine tail interaction with nickel.
- The ability to form higher-ordered oligomers was proven *in vitro* via MALLS and the presence of homogeneous particles was observed via TEM microscopy.
- *In cellulo* the overproduction of both *RrCooJ* and *RrCooJ*-Δ protects *E. coli* from increasing amount of nickel in the culture media.
- *RrCooJ* has a tendency to precipitate in the presence of high nickel concentrations both *in vitro* and *in cellulo*.
- The precipitation process is fully reversible upon addition of EDTA.



Communication

# Nickel-Induced Oligomerization of the Histidine-Rich Metallochaperone CooJ from *Rhodospirillum Rubrum*

Marila Alfano , Julien Pérard and Christine Cavazza \*

Univ. Grenoble Alpes, CEA, CNRS, CBM, F-38000 Grenoble, France

\* Correspondence: christine.cavazza@cea.fr

Received: 27 May 2019; Accepted: 25 June 2019; Published: 1 July 2019

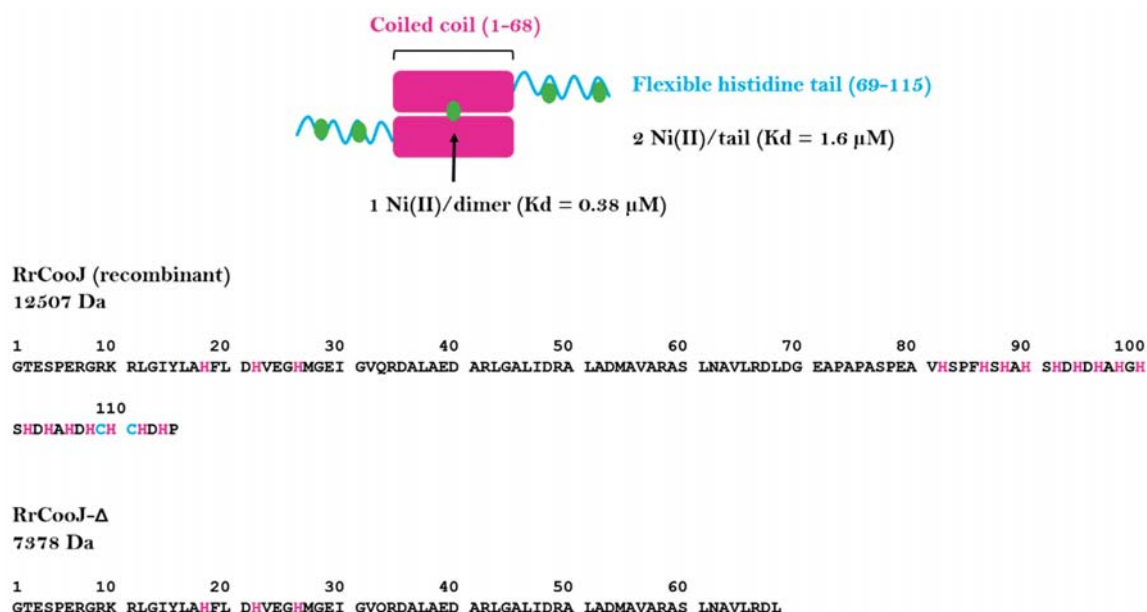


**Abstract:** [NiFe]-carbon monoxide dehydrogenase reversibly catalyzes the oxidation of CO to CO<sub>2</sub>. Its active site is a unique NiFe<sub>4</sub>S<sub>4</sub> cluster, known as C-cluster. In *Rhodospirillum rubrum*, three nickel-dependent proteins, CooC, CooT and CooJ are required for Ni insertion into the active site. Among them, CooJ is a histidine-rich protein, containing two distinct and spatially separated Ni(II)-binding sites: a strictly conserved N-terminal site and a variable histidine tail at the C-terminus. Here, using biophysical techniques, we study the behavior of the protein upon Ni(II) addition. Using circular dichroism and chemical denaturation, we show that the binding of Ni(II) to the protein increases its stability. Moreover, high-order oligomers are formed through nickel–histidine tail interactions, both in vitro and in cellulo, via a dynamical and reversible process.

**Keywords:** histidine-rich protein; carbon monoxide dehydrogenase; nickel chaperone; nickel-induced oligomerization

## 1. Introduction

Monofunctional nickel-dependent carbon monoxide dehydrogenases (CODH) reversibly catalyzes the oxidation of CO to CO<sub>2</sub> [1]. CODH's active site, called C-cluster, is a NiFe<sub>4</sub>S<sub>4</sub> cluster, unique in biology, whose atypical nature was revealed by X-ray crystallography [2,3]. In the hydrogenogenic carboxydrotroph *Rhodospirillum rubrum*, CODH plays an essential role in the energy metabolism when CO is the sole energy source [4]. In this bacterium, the structural gene of CODH, called *cooS*, is found in the *cooFSCTJ* operon [5]. Three of the proteins encoded by the *coo* operon, namely CooC, CooT and CooJ, are nickel chaperones dedicated to nickel insertion into CODH, while CooF is a ferredoxin [6]. In vivo studies have shown that the ATPase CooC and the Ni(II)-binding proteins CooT and CooJ play a significant role in the maturation pathway leading to a fully active enzyme. Moreover, Ni insertion is a key step in the enzyme activation process [5,7]. Among these chaperones, CooJ from *R. rubrum* (RrCooJ) is a histidine-rich protein, with a histidine tail comprising 16 histidines and 2 cysteines at its C-terminus [8]. Initially thought to be only present in *R. rubrum*, our recent study revealed the existence of at least 46 CooJ homologues in bacteria [9]. In solution, RrCooJ forms a 25-kDa homodimer with a central coiled coil and two independent C-terminal his-tails and possesses several Ni(II) binding sites. The first one is present in the N-terminal region and binds one Ni(II) per dimer via a "H-X3-H-X3-H" motif, strictly conserved in the CooJ family. In addition, the two Histidine-rich regions bind at least 2 Ni(II) ions each (Figure 1), via both histidine and cysteine residues, as shown by site-directed mutagenesis of cysteines 109 and 111 that led to decreased stoichiometry [9].



**Figure 1.** Schematic representation of *RrCooJ* and its amino acid sequence. Ni(II) ions are depicted as green spheres. Histidine residues are depicted in magenta and cysteines 109 and 111 are in blue.

Histidine-rich proteins, such as histatin 5 or  $\alpha\beta$  peptides [10], are predicted to be intrinsically disordered with a tendency to form oligomers driven either by self-association or through interaction with multivalent ions. In nickel chaperones, the presence of histidine-rich regions is often proposed to be related to nickel storage, and/or detoxification, due to their ability to quickly bind and release nickel ions. Among them, one remarkable example is Hpn from *Helicobacter pylori*, a small protein of 7 kDa composed of histidines for about half of the total amino-acid sequence, which is expected to form high-order oligomers, with 20-mers as predominant species, even in the absence of metal [11].

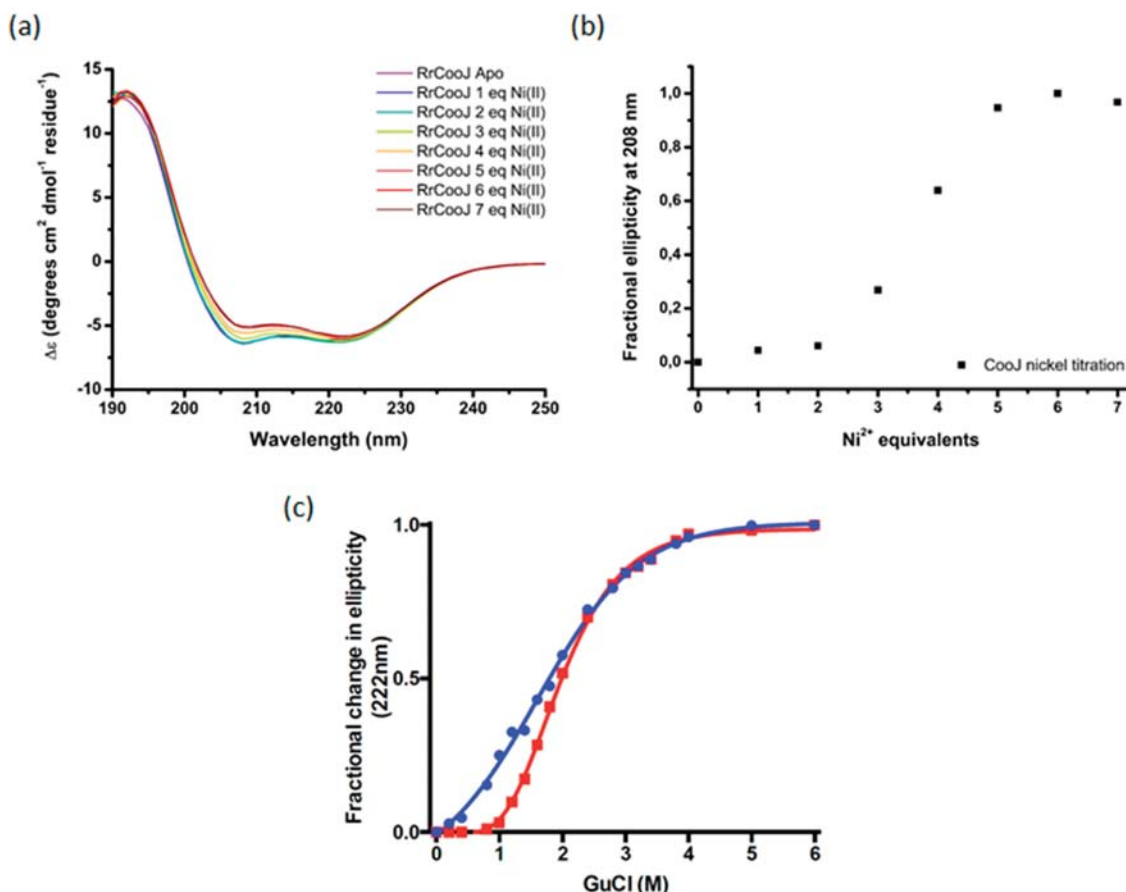
Here, we report the behavior of *RrCooJ* towards nickel in vitro and in cellulo, using a combination of biophysical techniques. *RrCooJ* exists as an equilibrium of oligomeric states in solution in the presence of Ni(II) ions, while the apo-protein forms a stable dimer. We reveal that the His-rich region is responsible for the oligomeric process.

## 2. Results

Recently, we showed that the two histidine tails present in *RrCooJ* are highly flexible and predicted to be partly disordered [9]. Moreover, Ni addition to the apo-dimer led to its conformational change, as observed by size-exclusion chromatography coupled to multi-angle laser light scattering (SEC-MALLS) experiments. Here we used circular dichroism (CD) in the far UV region (190–250 nm) and chemical denaturation of *RrCooJ*, either in its apo- or holo-form (Ni-*RrCooJ*), to get more information about the impact of Ni(II) ions on *RrCooJ* conformation. The CD spectrum of apo-*RrCooJ* is characterized by a positive peak at 193 nm and two negative peaks at 220 nm and 208 nm, attributed to the presence of a high  $\alpha$ -helix content in its secondary structures. Upon the addition of Ni(II) ions to the apo-protein solution, the negative peaks increased gradually, as an index of a conformational change (Figure 2a). Quantitative analysis, performed using the web server BeStSel [12], showed that the increase in ellipticity was correlated to a slight increase in the helical content (from 83% for apo-*RrCooJ* to 85% for Ni-*RrCooJ*) and a decrease in the turns content (from 6% for apo-*RrCooJ* to 0.4% for Ni-*RrCooJ*). Interestingly, the CD signal at 208 nm saturated when the Ni(II) concentration reached 20  $\mu$ M, corresponding to 5 molar eq. of Ni(II) per dimer, the same stoichiometry as the one previously determined by ITC (Figure 2b). Considering the predominance of  $\alpha$ -helices in the protein, the  $\Theta_{222/208}$  ratio, which increases if the protein gains in stability [13], can be used to get information about *RrCooJ*'s behavior towards nickel. For a Ni(II) concentration greater than 12  $\mu$ M



(corresponding to 3 molar eq. of Ni(II)), the ratio is  $>1$ , reflecting an increase in the  $\alpha$ -helix content and likely the stabilization of the protein. To further investigate the difference in stability between the apo- and the holo-forms, chemical denaturation using guanidine hydrochloride (GuCl) was performed. Their stability was compared by monitoring the variation of the 222 nm CD ellipticity signal after the addition of increasing concentrations of GuCl. Interestingly, at 1.2 M of GuCl, Ni-*RrCooJ* was still perfectly folded, whereas apo-*RrCooJ* started to denature, as shown by an ellipticity fractional change of 0.09 vs. 0.4, respectively. The midpoint GuCl concentration was determined to be 1.6 M for apo-*RrCooJ* and 2.0 M for Ni-*RrCooJ*. This result reveals a higher stability induced by Ni(II)-binding to the protein (Figure 2c).



**Figure 2.** Far-UV circular dichroism (CD) and chemical denaturation of *RrCooJ* at  $4\ \mu\text{M}$  of dimer upon Ni(II) addition. (a) Evolution of far-UV signal upon the addition of increasing Ni(II) concentrations. (b) Fractional values of the ellipticity registered at 208 nm plotted versus Ni(II) molar equivalents. (c) Chemical denaturation of apo-*RrCooJ* (blue) and Ni-*RrCooJ* (red) using different concentrations of GuCl at room temperature.

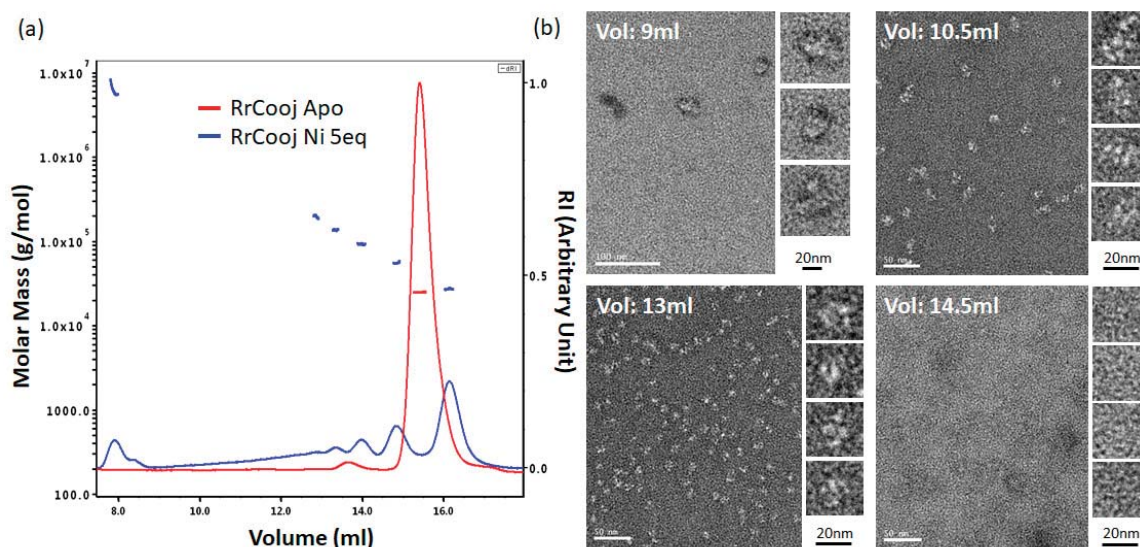
SEC-MALLS experiments were conducted to investigate the oligomeric state of *RrCooJ* in the presence of Ni(II). We previously showed that a truncated version of *RrCooJ* (*RrCooJ*- $\Delta$ ), lacking the histidine tail (Figure 1), is a stable dimer even in the presence of an excess of nickel [9]. Here, the column was equilibrated with a buffer supplemented with  $10\ \mu\text{M}$  of  $\text{NiSO}_4$ , corresponding to a concentration about six times higher than the affinity of the His tails for Ni(II) ( $K_d = 1.6\ \mu\text{M}$ ). The injection of  $50\ \mu\text{M}$  of *RrCooJ* dimer pre-incubated with 5 molar eq. of Ni(II) led to the appearance of a mixture of oligomeric forms ranging from dimers to high-order oligomers (Figure 3a).

To characterize these high-order oligomers, single particle analysis was performed using Transmission Electron Microscopy (TEM). In order to separate the different oligomeric states, size exclusion chromatography was used. Apo-*RrCooJ* dimer pre-incubated with 5 molar eq. of Ni(II)



was injected onto the column and a series of fractions was selected, based on the SEC-MALLS results. Four samples were collected at different elution volumes (9, 10.5, 13 and 14.5 mL) and subsequently analyzed, revealing the existence of high-order oligomers with sizes ranging from  $11 \pm 2$  nm to  $28 \pm 5$  nm in the three first peaks. In the latter one, corresponding to the elution of *RrCooJ* tetramers as shown by SEC-MALLS, high-order oligomers could not be detected as expected. Neither aggregation nor fibrils were observed under any of the tested conditions (Figure 3b). The size of high-order oligomers decreased along the elution with an optimum distribution of particles at an elution peak of 13 mL corresponding to a homogenous particle size of about 11 nm.

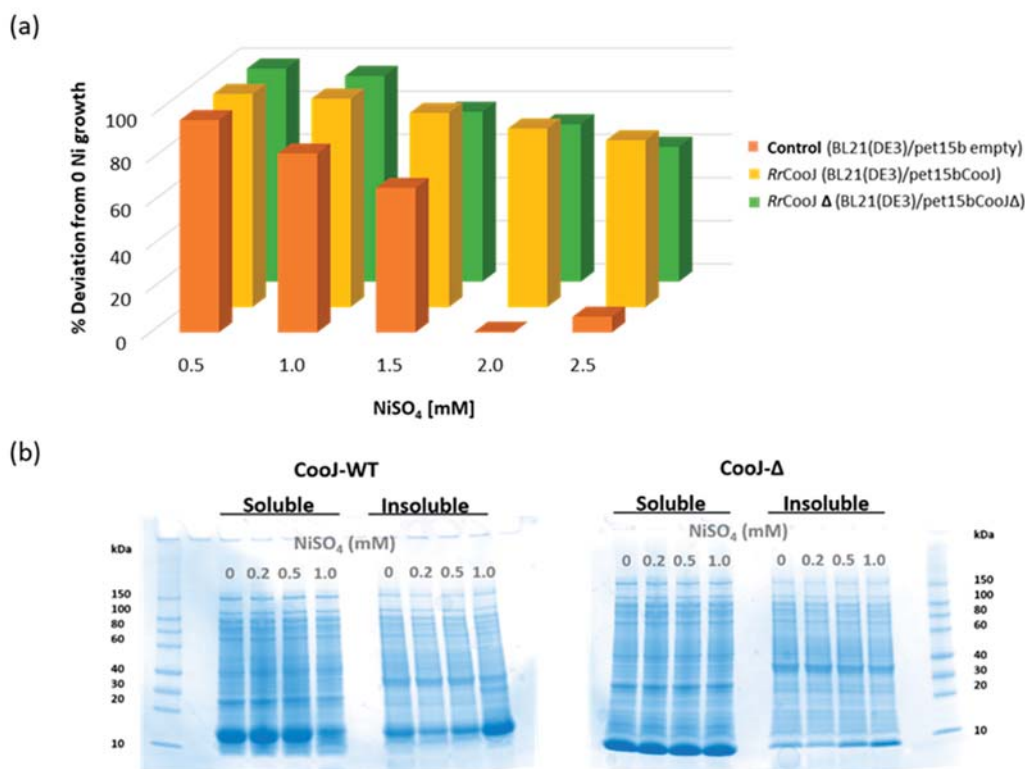
The oligomerization process is both protein and Ni(II) concentration dependent. Indeed, for a protein concentration of 50  $\mu$ M of dimer, *RrCooJ* started to form insoluble aggregates when the Ni(II) concentration reached 500  $\mu$ M. Moreover, the incubation of 100  $\mu$ M *RrCooJ* dimer with 10 molar eq. of  $\text{NiSO}_4$  (1.0 mM) led to its complete precipitation. However, the addition of 10 mM EDTA can fully redissolve the precipitate within seconds (as shown by protein assay) revealing the easy reversibility of the metal-dependent oligomerization process.



**Figure 3.** Nickel-dependent oligomerization of *RrCooJ* by Size Exclusion Chromatography and Multi Angle Laser Light Scattering (SEC-MALLS) and TEM. (a) 50  $\mu$ M of apo-*RrCooJ* (red) were injected onto a column equilibrated with buffer A (96% of dimer); The injection of 50  $\mu$ M of *RrCooJ* dimer preincubated with 250  $\mu$ M of  $\text{NiSO}_4$  (blue) in buffer A supplemented with 10  $\mu$ M of  $\text{NiSO}_4$  shows multiple peaks corresponding to a mixture of oligomers: dimer (35%), tetramer (16%), hexamer (10%), octamer (6.2%) and high-order oligomers (32.8%). (b) Nickel-dependent oligomerization analysis of *RrCooJ* by negative stain after size exclusion chromatography. Image of negatively stained particles (23,000 $\times$  and 30,000 $\times$  of magnification) of high-order oligomers of Ni-*RrCooJ* with highlights of particles measured for the following elution volumes: 9 mL ( $28 \pm 3$  nm), 10.5 mL ( $15 \pm 2$  nm), 13 mL ( $11 \pm 2$  nm) and 14.5 mL (below 5 nm). Oligomer size was determined after merging 20 independent particles by “ImageJ v 1.51” software. Vol: Elution volume.

In order to study their in-cell Ni(II)-binding capability and behavior, *RrCooJ* and its truncated version *RrCooJ*- $\Delta$  were overproduced in *E. coli* in growth medium supplemented with a range of Ni(II) concentrations from 0 to 2.5 mM  $\text{NiSO}_4$ . As a control, the impact of nickel was evaluated on a strain of *E. coli* transformed with an empty vector. The growth of the control culture was strongly affected by nickel, decreasing remarkably between 0.5 mM and 1.5 mM, and stopping for concentrations higher than 1.5 mM. The presence of *RrCooJ* or *RrCooJ*- $\Delta$  increases the nickel resistance of *E. coli*, resulting in a constant growth even at high concentration of metal (Figure 4a). Although *RrCooJ* and *RrCooJ*- $\Delta$  are both able to bind Ni(II) in-cell, we observed that *RrCooJ* tends to form insoluble oligomers upon addition of increasing concentrations of Ni(II) in the growth medium. At 1.0 mM of  $\text{NiSO}_4$ , despite

a bacterial growth rate similar to the one in the absence of nickel, the overproduced *RrCooJ* was mainly found in the insoluble fraction, as shown by SDS gel (Figure 4b). This behavior was not observed with the overproduced *RrCooJ*- $\Delta$ , mainly found in the soluble fractions independently on the nickel concentration (Figure 4b). This is a strong indication that the His-tail triggers the oligomerization process of *RrCooJ* in the presence of nickel, in agreement with the in vitro studies, without drastically affecting the bacterial growth.



**Figure 4.** Growth response to nickel content in *E. coli* strains overproducing *RrCooJ* or *RrCooJ*- $\Delta$ . (a) The resistance towards nickel for the control (orange), *RrCooJ* (yellow) and *RrCooJ*- $\Delta$  (green) cultures is plotted for each different nickel concentration after 16 h from the induction. The resistance is reported as percentile deviation of the OD<sub>600</sub> measured at 16 h considering as maximum value the OD<sub>600</sub>-16h measured for the cultures at 0 mM NiSO<sub>4</sub> versus the ones measured for the cultures at different nickel concentrations. The deviation is calculated using the formula  $\sum_{n=1}^3 \left( \frac{OD_{600 \text{ at } x \text{ mM } 16 \text{ h}} - OD_{600 \text{ before induction}}}{OD_{600 \text{ at } 16 \text{ h for } 0 \text{ mM Ni}}} \right) * 100$ , using the OD<sub>600</sub>-16h measured for each different nickel concentration ( $x \text{ mM} = 0.5, 1.0, 1.5, 2.0, 2.5 \text{ mM NiSO}_4$ ). (b) SDS-PAGE of soluble and insoluble extracts of *E. coli* cultures overproducing *RrCooJ* or *RrCooJ*- $\Delta$  in the presence of Ni(II) in the growth media (0 to 1.0 mM NiSO<sub>4</sub>).

### 3. Materials and Methods

Recombinant *RrCooJ* and *RrCooJ*- $\Delta$  were overproduced in *E. Coli* and purified as previously described [9]. Protein concentration was determined either by SEC-MALLS or by Rose Bengal protein assay to determine the protein concentration for the EDTA solubilization experiment.

#### 3.1. Circular Dichroism (CD) Spectroscopy

CD spectra were recorded using a J-1500 circular dichroism spectrometer (JASCO Analytical Instruments, Easton, PA, USA). A stock solution of 10 mM NiSO<sub>4</sub> was used to monitor metal-dependent secondary structure content upon Ni(II) titration (from 0 to 7 molar equivalents of NiSO<sub>4</sub>). Spectra were recorded from 190 to 250 nm using a 1 mm cuvette, with ten accumulations to increase the signal-to-noise ratio. Proteins were thawed and diluted to 4  $\mu$ M dimer in CD buffer (5 mM potassium phosphate pH

7.0, containing 0.5 mM TCEP). For the determination of the secondary structure elements, the spectra were analyzed using the BeStSel online software [12]. The guanidine hydrochloride denaturation experiments were performed via additions of different concentrations of GuCl on 4  $\mu$ M dimer samples (apo- and Ni-RrCooJ), followed by 2 h of incubation. The Ni-RrCooJ sample was prepared by adding 5 molar equivalents of Ni(II) per protein dimer (20  $\mu$ M final concentration) and incubated for 30 min prior to its denaturation. The fractional change in ellipticity at 222 nm was used to determine the variation of the secondary structure elements due to the denaturation of the protein. Empiric fit was done via a sigmoidal equation.

### 3.2. Size Exclusion Chromatography and Multi Angle Laser Light Scattering (SEC-MALLS)

Purified and frozen RrCooJ was thawed and diluted to 50  $\mu$ M of dimer in 50 mM HEPES pH 7.5, 150 mM NaCl, 1 mM TCEP (buffer A). Ni-RrCooJ was incubated in the presence of 250  $\mu$ M NiSO<sub>4</sub> for 30 min at room temperature prior to injection onto the SEC-MALLS system (Wyatt Dawn HELEOS-II 18-angle light scattering detector and Wyatt Optilab rEX refractive index monitor linked to a Shimadzu HPLC system comprising a LC-20AD pump, a SPD20A UV/Vis detector, and a Superdex 200 10/300 increase column (GE Healthcare Life Sciences, Pittsburg, KS, USA). Injections were carried out using a 20  $\mu$ L loop. The size exclusion column was equilibrated using buffer A  $\pm$  10  $\mu$ M NiSO<sub>4</sub>. Protein concentration in all samples was determined by integration of the differential refractive index (dRI) peak ( $dn/dc = 0.185$ ). It is important to note that all used samples came from the same stock protein. The data were analyzed using the ASTRA software (version 6) (WYATT Technology Corporation, Santa Barbara, CA, USA).

### 3.3. Electron Microscopy

100  $\mu$ L of Ni-RrCooJ dimer at 50  $\mu$ M were injected onto a Superdex 200 10/300 column (GE Healthcare) equilibrated with buffer A + 10  $\mu$ M NiSO<sub>4</sub>. 200  $\mu$ L-fractions were collected and 4 fractions corresponding to elution volumes of 9, 10.5, 13 and 14.5 mL were selected. The negative stain Mica-carbon Flotation Technique (MFT) was used to prepare the samples. Briefly, 3–5  $\mu$ L of protein sample (0.1 to 0.01 mg/mL) were applied to a clean side of a carbon layer (between a carbon and a mica layer). The carbon was then floated on stain (2% *w/v* uranyl acetate (AcU) pH 4.5) and covered by a copper grid. The images were captured under low-dose conditions ( $<10\text{ e}^- \text{ \AA}^{-2}$ ) at a magnification of 23K $\times$  and 30K $\times$  with defocus values between 1.2 and 2.5  $\mu$ m on a Tecnai 12 LaB6 electron microscope at 120 kV accelerating voltage using a CCD Camera Gatan Orius 1000 (Gatan, Inc., Pleasanton, CA, USA).

### 3.4. Growth Response to Nickel in *E. coli* Strains Overproducing RrCooJ or RrCooJ- $\Delta$

*E. coli* BL21 (DE3) strain was transformed with pET15b, pET15b-RrCooJ or pET15b-RrCooJ- $\Delta$ . The three different cultures (100 mL each) were grown in LB medium, with the appropriate antibiotic, at 25  $^{\circ}$ C and 180 rpm to an OD<sub>600</sub> of  $\sim$ 0.6 and induced with 0.25 mM IPTG. At this point, each culture was divided into five 50 mL-Falcon tubes (10 mL-culture each) and a range of NiSO<sub>4</sub> concentrations was added to the culture medium (0, 0.5, 1.0, 1.5, 2.0 and 2.5 mM). The cells were grown overnight at 25  $^{\circ}$ C and 180 rpm, and the OD<sub>600</sub> was measured after 16 h for all the cultures. The cultures were done in triplicate for each different condition. The effect of the overproduction of RrCooJ or RrCooJ- $\Delta$  according to the amount of Ni(II) was evaluated by normalizing the culture optical densities, measured at the different nickel concentrations, with the values obtained from the control. To study the solubility of RrCooJ in the presence of Ni(II) in-cell, *E. coli* BL21 (DE3) strain transformed with either pET15b-RrCooJ or pET15b-RrCooJ- $\Delta$  were grown as described above, except that 25 mL-cultures were grown in 50 mL-flasks with 0, 0.2, 0.5 and 1.0 mM NiSO<sub>4</sub> concentration. After overnight growth, the OD<sub>600</sub> was measured ( $5.5 \pm 0.2$  for BL21(DE3)/pET15b-RrCooJ- $\Delta$  cultures and  $5.8 \pm 0.3$  for BL21(DE3)/pET15b-RrCooJ cultures). The cell pellets were collected by centrifugation and then resuspended in 4 mL of 50 mM Tris-HCl, pH 8.0 containing a complete Protease Inhibitor cocktail tablet (Roche Life Science, Indianapolis, IN, USA). After three freeze-thaw cycles, the cultures were sonicated

and centrifuged at 14,000 rpm for 20 min. The 16 pellets were resuspended in 50 mM Tris-HCl pH 8.0, containing 4 M urea and 0.1% SDS. Ten microliters of each sample were examined using SDS-PAGE (Bio-Rad Corporate, Hercules, CA, USA).

#### 4. Conclusions

Recently, structural studies of apo-RrCooJ revealed its atypical topology with two independent histidine tails flanking a central coiled coil region. Here, we focused our attention on the histidine-rich region behavior. Two phenomena were observed upon nickel addition. Firstly, Ni(II)-binding directly impacts the secondary structure content of the protein, with a gain in  $\alpha$ -helices related to higher stability. Secondly, nickel induces a dynamic and reversible oligomerization process in vitro, leading to the protein precipitation observed both in vitro and in cellulo, in the presence of elevated Ni(II) levels. Further studies are necessary to investigate the functional role of RrCooJ oligomerisation in nickel homeostasis under physiological conditions.

**Author Contributions:** Conceptualization, C.C. and M.A.; formal analysis, C.C., M.A., J.P.; investigation, C.C., M.A.; writing—original draft preparation, C.C.; writing—review and editing, M.A.

**Funding:** This work was supported by “the ITERLIS PhD program, CEA Life sciences” for MA’s PhD funding, the “FUNBIOCO” project (IDEX-UGA, Initiatives de Recherche stratégiques) and the “COSYNBIO” project (Projets exploratoires, Cellule énergie-CNRS). This work has been partially supported by Labex ARCANE and CBH-EUR-GS (ANR-17-EURE-0003). The research leading to these results has received funding from the networking support from the COST Action FeSBioNet (Contract CA15133).

**Acknowledgments:** This work used the platforms of the Grenoble Instruct Centre (ISBG; UMS 3518 CNRS-CEA-UJF-EMBL) with support from FRISBI (ANR-10-INSB-05-02) and GRAL (ANR-10-LABX-49-01) within the Grenoble Partnership for Structural Biology (PSB). The electron microscope facility is supported by the Rhône-Alpes Region, the Fondation Recherche Médicale (FRM), the fonds FEDER, the Centre National de la Recherche Scientifique (CNRS), the CEA, the University of Grenoble, EMBL, and the GIS-Infrastructures en Biologie Santé et Agronomie (IBISA). We thank Daphna FENNEL and Dr Guy SCHOEHN, from the Electron Microscopy platform of the Integrated Structural Biology of Grenoble (ISBG, UMI3265).

**Conflicts of Interest:** The authors declare no conflicts of interest.

#### References

1. Kung, Y.; Drennan, C.L. A role for nickel–iron cofactors in biological carbon monoxide and carbon dioxide utilization. *Curr. Opin. Chem. Biol.* **2011**, *15*, 276–283. [[CrossRef](#)] [[PubMed](#)]
2. Dobbek, H.; Svetlitchnyi, V.; Gremer, L.; Huber, R.; Meyer, O. Crystal structure of a carbon monoxide dehydrogenase reveals a [Ni–4Fe–5S] cluster. *Science* **2001**, *293*, 1281–1285. [[CrossRef](#)] [[PubMed](#)]
3. Drennan, C.L.; Heo, J.; Sintchak, M.D.; Schreiter, E.; Ludden, P.W. Life on carbon monoxide: X-ray structure of *Rhodospirillum rubrum* Ni–Fe–S carbon monoxide dehydrogenase. *Proc. Natl. Acad. Sci. USA* **2001**, *98*, 11973–11978. [[CrossRef](#)] [[PubMed](#)]
4. Alfano, M.; Cavazza, C. The biologically mediated water–gas shift reaction: Structure, function and biosynthesis of monofunctional [NiFe]–carbon monoxide dehydrogenases. *Sustain. Energy Fuels* **2018**, *2*, 1653–1670. [[CrossRef](#)]
5. Kerby, R.L.; Ludden, P.W. In vivo nickel insertion into the carbon monoxide dehydrogenase of *Rhodospirillum rubrum*: Molecular and physiological characterization of cooCTJ. *J. Bacteriol.* **1997**, *179*, 2259–2266. [[CrossRef](#)] [[PubMed](#)]
6. Singer, S.W.; Hirst, M.B.; Ludden, P.W. CO-dependent H<sub>2</sub> evolution by *Rhodospirillum rubrum*: Role of CODH: CooF complex. *Biochim. Biophys. Acta Bioenerg.* **2006**, *1757*, 1582–1591. [[CrossRef](#)] [[PubMed](#)]
7. Jeon, W.B.; Cheng, J.; Ludden, P.W. Purification and Characterization of Membrane-associated CooC Protein and Its Functional Role in the Insertion of Nickel into Carbon Monoxide Dehydrogenase from *Rhodospirillum rubrum*. *J. Biol. Chem.* **2001**, *276*, 38602–38609. [[CrossRef](#)] [[PubMed](#)]
8. Watt, R.K.; Ludden, P.W. The Identification, Purification, and Characterization of CooJ. *J. Biol. Chem.* **1998**, *273*, 10019–10025. [[CrossRef](#)] [[PubMed](#)]



9. Alfano, M.; Pérard, J.; Carpentier, P.; Basset, C.; Zambelli, B.; Timm, J.; Crouzy, S.; Ciurli, S.; Cavazza, C. The carbon monoxide dehydrogenase accessory protein CooJ is a histidine-rich multidomain dimer containing an unexpected Ni(II)-binding site. *J. Biol. Chem.* **2019**, *294*, 7601–7614. [[CrossRef](#)] [[PubMed](#)]
10. Iadanza, M.G.; Jackson, M.P.; Hewitt, E.W.; Ranson, N.A.; Radford, S.E. A new era for understanding amyloid structures and disease. *Nat. Rev. Mol. Cell Biol.* **2018**, *19*, 755–773. [[CrossRef](#)] [[PubMed](#)]
11. Ge, R.; Zhang, Y.; Sun, X.; Watt, R.M.; He, Q.-Y.; Huang, J.-D.; Wilcox, D.E.; Sun, H. Thermodynamic and Kinetic Aspects of Metal Binding to the Histidine-rich Protein, Hpn. *J. Am. Chem. Soc.* **2006**, *128*, 11330–11331. [[CrossRef](#)] [[PubMed](#)]
12. Micsonai, A.; Wien, F.; Bulyáki, É.; Kun, J.; Moussong, É.; Lee, Y.-H.; Goto, Y.; Réfrégiers, M.; Kardos, J. BeStSel: A web server for accurate protein secondary structure prediction and fold recognition from the circular dichroism spectra. *Nucleic Acids Res.* **2018**, *46*, W315–W322. [[CrossRef](#)] [[PubMed](#)]
13. Kwok, S.C.; Hodges, R.S. Stabilizing and destabilizing clusters in the hydrophobic core of long two-stranded  $\alpha$ -helical coiled-coils. *J. Biol. Chem.* **2004**, *279*, 21576–21588. [[CrossRef](#)] [[PubMed](#)]



© 2019 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).



# Chapter 4





## CHAPTER 4

### Heterologous production of CODH in *E. coli*

The heterologous expression of *RrCODH* has never been reported in literature, so far. From the growth of native *RrCODH* in a nickel-depleted medium, Spangler et al. proved that the enzyme can be purified in a “Ni-deficient” state, still possessing the 4Fe-4S clusters.<sup>172,113,173,174</sup> Bonam et al. demonstrated that this form could be partially re-activated upon nickel incubation under reducing conditions.<sup>172,118</sup> Within the framework of my project, the characterization of how the nickel insertion activates the enzyme and how this enzyme interacts with its nickel accessory proteins are some of the final aims. The possibility to obtain high amount of enzyme, to co-express it with the partners, to use different growth conditions (such as nickel-depleted medium), are all reasons why the *RrCODH* heterologous expression in *E. coli* is advocated.

Hence I report the results obtained during the last months of my PhD, showing reliable and reproducible protocols for the overproduction of recombinant *RrCODH* in *E. coli*. The purification, characterization of the nickel-deficient *RrCODH* and the *in vitro* nickel insertion and re-activation are presented. The results obtained by its co-expression with the nickel chaperone protein *CooJ* will also be discussed, please consider them as preliminary data that may be used in the future to go further with this project.

*(Experimental procedures reported in the annex session)*

## 4.1 RESULTS AND DISCUSSIONS

### 4.1.1 Heterologous expression and purification

The *RrcooS* gene was cloned in the *prsF* vector (with or without the TEV cleavable his-tag) and was expressed in *E. coli* BL21(DE3). Firstly, an expression protocol similar to the one published for the production of RecChCODH-II was tested to see whether it could be also applied for the recombinant product of *RrCODH*. Following this as a guideline, the cells were grown in TB medium supplemented with metals under anaerobic conditions. The protein expression, after IPTG induction, was checked using an SDS-PAGE gel and resulted in a band that corresponds to the calculated molecular weight of *RrCODH* monomer (~68 KDa). Confident of this result I then proceeded with the purification, which presented however an inconvenience. In fact, after cell lysis and centrifugation, the protein was mainly insoluble and thus present in the pellet, as it has been previously observed with the native enzyme purification.<sup>175</sup> To check if the expression in different growth media would increase the protein solubility, the cells were cultured at the same time in TB, LB and M9ZB, always supplemented with the metal mixture (*described in annex*). The protein production did not succeed in M9ZB and it worked for TB and LB, nevertheless the CODH was still present in the insoluble fraction.

At this point, to control whether the protein solubilisation could be obtained from the pellet, different detergents were tried on both the pellets from the LB and TB cultures. The choice was based on the ones used to solubilize the native enzyme when purified from *R. rubrum* cultures. It is worth mentioning that even if CODH is not a membrane associated protein, it needs a solubilisation step in order to be released into the soluble fraction. Different detergents were chosen to maximize the likelihood of success, covering the three main groups identified by the type of hydrophilic group that they contain: Triton, a non-ionic detergent (no charge); sodium deoxycholate and sodium lauroyl sarcosinate, both ionic (stable charge); CHAPS, a zwitterionic detergent (positive and negative charges but the net charge is equal to zero). (*Procedure described in annex*) The results obtained from the LB and TB cultures are shown in Figure 1.

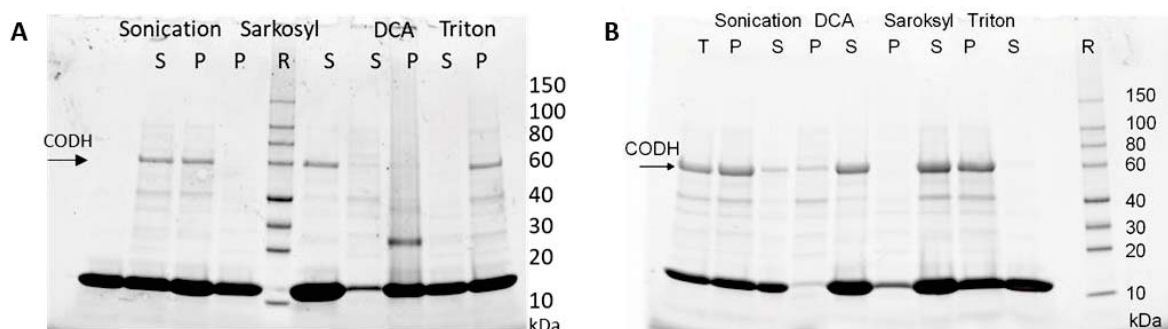


Figure 1. SDS-PAGE gel on the detergent solubility analysis of *RrCODH* grown in LB medium (A) and in TB medium (B). The results obtained after sonication are checked against the sodium deoxycholate (DCA) treatment, sodium lauroyl sarcosinate (sarkosyl) treatment and Triton exposure for both the pellet (P) and supernatant (S). As it is possible to see, the best results are obtained from the DCA and sarkosyl treatment. A small part of the enzyme is also soluble after sonication. (T=total before lysis and R=molecular weight reference)

From the testing of the different detergents, it is evident that the use of ionic ones helps the protein solubilisation, resulting in a complete solubilisation in the case of sodium lauroyl sarcosinate. Considering that the CODH is better over-produced in TB (Figure 1 B), it was decided to perform the purification from this culture, using sodium lauroyl sarcosinate as detergent to solubilize the protein. However, the protein started to behave differently during the purification and its color shifted from a typical brownish observed for [4Fe4S]-containing proteins to green. Worried that the change in color could be related to the presence of the detergent with a possible denaturation effect on the metal-clusters, it was decided to step back and try to improve the solubilisation via a different culture method.

The auto-induction medium was then tested under aerobic condition, giving positive results concerning the protein production but still the same solubility problem was encountered. Considering that a higher amount of cells extract was now available (10 gr per L vs the 2 gr per L obtained with TB under anaerobic conditions) it was decided to purify the soluble CODH fraction released after the sonication.

As a proof of concept, it was decided to purify the Nickel-deficient *RrCODH*-HisTag protein produced in auto-induction medium. The protein was purified on a cobalt column, to avoid the interaction of the enzyme with nickel. The His-tag was then removed by adding a TEV protease and recovering the cleaved CODH. The SDS-PAGE highlighted the presence of minor

contaminants, thus it was passed through a size exclusion column and the purest fractions were recovered. The quality of the purified enzyme was examined via ESI-MS, confirming the nature of the enzyme and the cleavage of the tag (Figure 2). The metal content was checked via ICP, the nickel was not present (as expected) and a value of 9.9 Fe (out of 10 theoretically for a monomer) was found. The ratio 420/280 nm, supposed to be 0.4 native, was 0.38 thus in good agreement. It is possible to conclude that heterologous expression in aerobically auto induced medium is feasible and that the addition of Fe salts and Lcys as S donor is sufficient to correctly insert the Fe and S atoms, most probably thanks to the Isc machinery naturally present in *E. coli*.

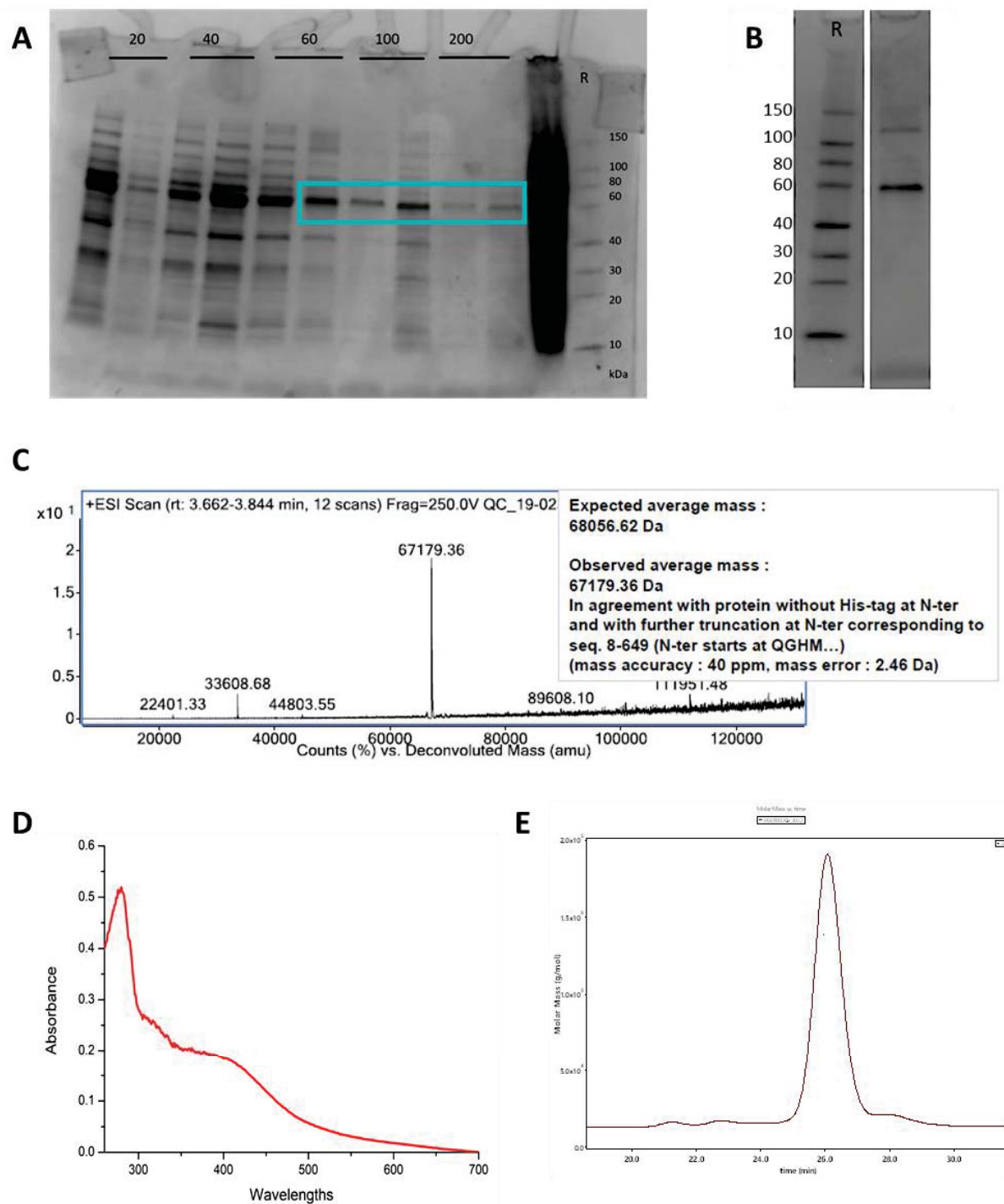


Figure 2. SDS-gel of *RrCODH*-TEV purification. A) Imidazole elution gradient (20, 40, 60, 100 and 200 mM) the chosen fractions for TEV cleavage are shown in the green box. B) Pure protein after the Superdex75. C) ESI-Mass of the purified protein. D) Uv-Vis spectra of *RrCODH* showing the 420 nm absorption band for the iron-sulfur cluster. E) SEC-MALLS profiles for *RrCODH*. The protein is a dimer at 140 kDa.

#### 4.1.2 *In cellulo* activity of *RrCODH* vs *RrCODH-CooJ* vs *RrCODH-CooJA*

Starting from the consideration that native CooJ co-purified with CODH<sup>125</sup> and that two independent Ni-binding sites are present in this chaperone, we decided to evaluate the putative role of CooJ in the CODH maturation, with a closer look to its high-affinity site N-terminal site. In fact, as you may remember from chapter 3, our theory for the presence of a double nickel-binding site on CooJ is a possible double function for this chaperone. On one hand, the his-rich tails could serve as a nickel storage site, while on the other hand the N-terminal site could insert the nickel and activate the CODH. Thus, our idea relies on the concept that if the co-production of *RrCODH* and CooJ or *RrCODH* and CooJA present at least the same activity between each other we could actually have a good starting point to investigate further this theory.

To evaluate if the co-production with CooJ or CooJA in the recombinant system would affect the CODH activity, preliminary experiments were performed: different cultures were grown co-expressing *cooJ* or *cooJ-Δ* genes with the *cooS* gene. In addition, we also wanted to evaluate if the nickel concentration supplied during the bacterial growth could have an impact on the enzyme activity. For this reason, three different nickel concentrations (1, 10 and 100  $\mu\text{M}$   $\text{NiSO}_4$ ) were tested (Figure 5). The activity tests were performed on the whole cell-extracts and the general trend was evaluated crossing the date one against each other.

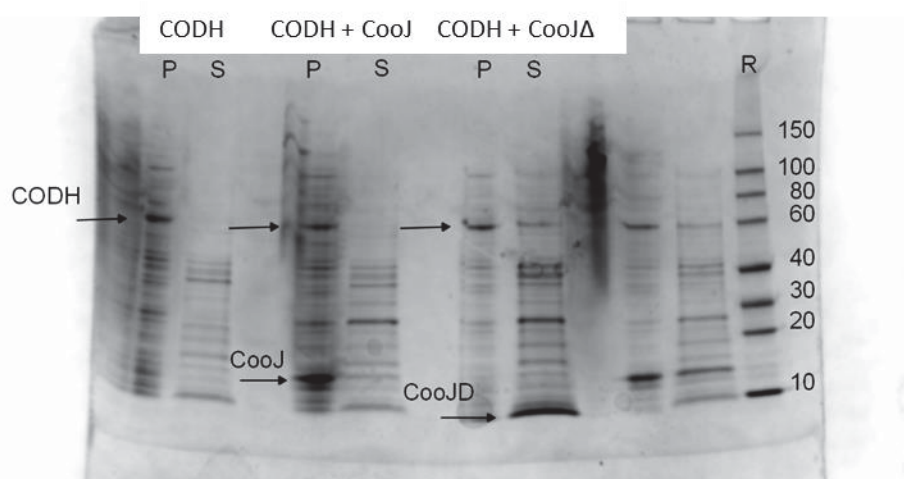


Figure 5. Expression gel to check the presence of *RrCODH*, *RrCODH+CooJ* and *RrCODH+CooJA*. The gel was performed on the cell pellets and soluble fractions. As it is possible to see CooJ is mainly present in the insoluble fraction with the CODH, whereas CooJA is present in the soluble one. This result is in agreement with the hypothesis that the his-tail triggers the formation of insoluble CooJ high-oligomers in the presence of nickel. Instead, CooΔ, lacking the his-tail, is always present in the soluble fraction. These cultures were grown with 100  $\mu\text{M}$  of nickel in the growth medium.

From this test, it seems that the CODH co-expression with the CooJ/CooJΔ chaperones does not influence its activity (Table 1). Different reasons could be found to justify this result; the first one is that the other chaperones, CooT and CooC, are actually needed to increase the activity of the enzyme. It is not surprising that, if they actively cooperate for the delivery of the metal in the active site, the simultaneous co-expression of the full operon is required to restore the native activity.

An interesting point is the possibility to re-activate the inactive cellular extract (obtained with 1  $\mu$ M NiSO<sub>4</sub>) via overnight nickel exposure (1 mM NiSO<sub>4</sub>, 5 mM MV) even if RecRrCODH is mainly present in the insoluble fraction after the sonication, which in my opinion could have been a major obstacle for the correct nickel insertion in the active site (Table 1).

In the previous section we described the feasibility of *in vitro* nickel insertion in a purified enzyme, here the possibility to insert nickel *in vitro* directly in cell extracts by addition of Ni salts is an encouraging news for the use of RrCODH in future applications.

I would like to underline one more time that all the data presented for RecRrCODH are preliminary and cannot be considered as a final proof.

| Whole cell<br>extracts | 1 $\mu$ M NiSO <sub>4</sub><br>U/ml | 10 $\mu$ M NiSO <sub>4</sub><br>U/ml | 100 $\mu$ M NiSO <sub>4</sub><br>U/ml | Reactivation *<br>U/ml |
|------------------------|-------------------------------------|--------------------------------------|---------------------------------------|------------------------|
| <b>RrCODH</b>          | No activity                         | 99.2                                 | 229.41                                | 5900                   |
| <b>RrCODH + CooJ</b>   | No activity                         | 110.2                                | 163.2                                 | 551.47                 |
| <b>RrCODH + CooJΔ</b>  | No activity                         | 112.41                               | 79.41                                 | 694.85                 |

\*the inactive 1  $\mu$ M NiSO<sub>4</sub> samples are re-activated via O/N incubation with 1 mM NiSO<sub>4</sub>, 5 mM MV.

Table 1. CO oxidation activity ( $\mu$ mole CO min<sup>-1</sup> ml<sup>-1</sup>) on the cell extract. Please note that it's an activity per ml ( $\mu$ mole CO min<sup>-1</sup> ml<sup>-1</sup> solution) and not a specific activity calculated per mg of protein ( $\mu$ mole CO min<sup>-1</sup> mg<sup>-1</sup> protein).





# Chapter 5



## CHAPTER 5

### Conclusions and perspectives for further research

To sum up our contribution to the community working on deciphering the maturation of CODH, the biochemical, structural and biophysical characterization of the two chaperone proteins CooT and CooJ has been added to the scientific knowledge.

In the **CooT** family, two distinct nickel-binding coordination modes are present: the “Cys2” only motif and the mixed “Cys2–His55” one. The representative of the first class is CooT from *R. rubrum*, which coordinates the Ni<sup>2+</sup> involving the two cys2, brought close in space by the dimer formation. The metal is coordinated in a square-planar geometry, hold by the 2-amidates and the 2-thiolates of both cys2. Concerning the second motif, we found and characterized a putative CooT from *C. hydrogenoformans*, which will coordinate the Ni<sup>2+</sup> using the cys2 and the his55, also in this case closed-by in space due to the dimer formation. For both CooTs, the nickel site is solvent exposed, probably to facilitate its physiological role as a nickel chaperone. Interestingly, the nickel binding affinity differs from motif to motif, going from the nanomolar range for *RrCooT* to the micromolar one for *ChCooT*, suggesting their possible different role *in vivo*.

**CooJ** in its apo-form is a dimer, created by the close proximity of the structured coiled coil N-terminal regions, which are arranged antiparallel to each other causing the independence of the two histidine tails. It possesses two spatially separated nickel binding sites: one is a “high-affinity site” in the N-terminal part and a second one in the histidine-rich region. Based on the comparison with other described nickel chaperones, a double function for this protein can be envisioned as: the his-tail region is needed for the nickel storage, while the N-terminal site is devoted to the CODH maturation. The *in vitro* studies showed that the nickel exposure triggers CooJ’s stabilization and the nickel-binding induces a dynamical reversible oligomerisation. The tendency to form high-order oligomers in the presence of increasing amount of nickel could be related to a possible physiological function in nickel-homeostasis.

We demonstrated for the first time the ability to overproduce and purify the **recombinant *R. rubrum* CODH** in *E. coli*. The ability to activate the nickel-deficient enzyme has been demonstrated via *in vitro* nickel reconstitution.

The capability to heterogeneously produce high amounts of RrCODH was one of the steps needed in order to study the maturation mechanism, and in our particular case gain more information about the nickel insertion and activation of this enzyme.

For what concerns the study of CODH maturation, specifically for the nickel ion insertion and C-cluster activation, the following reported perspectives are most likely achievable in the imminent future (months/years).

Considering that we succeeded in purifying large quantities of RrCooT, RrCooJ and RrCODH (in both nickel bound and nickel depleted states), an evaluation of the possible interactions between these different partners (CooT-CooJ, CooT-CODH, CooJ-CODH) will be the first aim. Recently the laboratory acquired a system for direct protein-protein interaction, (BLItz system for direct binding assays on a biosensor), which has been installed in a glove-box allowing for the study of anaerobic systems. Other biochemical techniques are going to be employed, as well as co-crystallization assays.

In addition, remember that we are missing one of main players of the *R. rubrum* CODH nickel-maturation story: RrCooC. Its purification is crucial to verify the interactions among the three-nickel chaperones and the enzyme, and to study how the nickel insertion occurs and activates the C-cluster.

Further steps could also be made towards the *in cellulo* co-production of the enzyme with different partners (CooFCTJ) to see if and how the final activity can be modulated directly in the whole cell extracts. The preliminary tests presented in Chapter 4 suggest that the co-production of CODH with only CooJ/CooJ-Δ is not sufficient to enhance the enzyme activity compared to the enzyme by itself in *E. coli*. In particular, I would personally be interested to check if the co-production CooJ-Δ with the other partners (CooC and CooT) is sufficient to enhance the enzyme activity.

Diverting now from the perspective strictly related to the enzyme maturation, from the introduction, I hope that I managed to persuade you that the possibility of using this biological system for the production of biofuels and chemicals is one of the reasons why the community is interested in our field. The biological water gas shift (WGS) reaction promoted by the two enzymes CODH and hydrogenase has an appealing industrial outlook. However, major limitations are imposed by the lack of knowledge of this biological system. Taking advantage of the expertise that we are acquiring about CODH purification and maturation, the next idea would be to focus the effort towards the purification of the *R. rubrum* CODH enzyme in complex with its physiological partners, the ferredoxin CooF and hydrogenase (Figure 1).

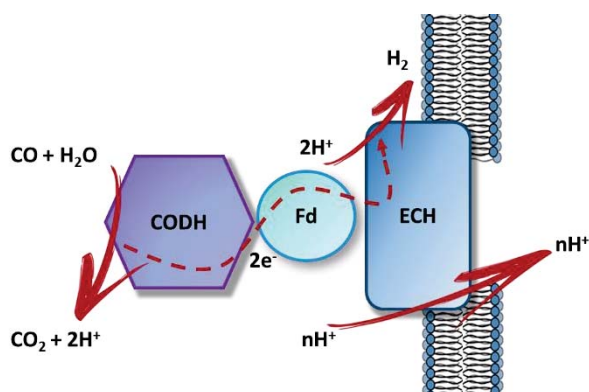


Figure 1. The biological WGS is catalyzed by a multiprotein complex constituted of a monofunctional NiFe-CODH (CODH), a ferredoxin CooF (Fd) and a CO-induced energy-conserving NiFe-H<sub>2</sub>ase (Ech). CO is oxidized by CODH, releasing electrons used for the reduction of two protons into dihydrogen, catalysed by the hydrogenase. CooF transfers electrons from CODH to hydrogenase.

The feasibility of this idea has already been proven by purification of the whole CODH:CooF:Hydrogenase complex from *C. hydrogenoformans*,<sup>176</sup> supporting our assumption of a tight link between the three partners also in *R. rubrum*. Achieving this goal, we could have the chance to study to a molecular level how the complete WGS reaction occurs, determine the limiting parameters and predict the possible implementable factors to increase the system performances. From the study of the purified complex, we can now think about two different approaches to develop a full performing biological WGS, either *in vivo* by the direct use of the living microorganism or *in vitro* by combining nature and nanotechnology.

Linked to this last possibility, the feasibility of an *in vitro* enzyme-promoted system is currently under evaluation in another branch of this CODH project. The idea is to use already purified enzymes grafted on an inorganic-electrode support, substituting in this way the role of the ferredoxin in the electron transfer. U. Contaldo, in the context of his PhD thesis, is working on the grafting of purified CODH and [NiFeSe]-hydrogenase using nanotubes as electron carriers. However, the use of purified fragile enzymes can be a challenging point during the experimental applications.<sup>177</sup> To overcome this problem, our prediction is that the substitution of the *in solution* enzyme by the use of cross-linked enzyme crystals (CLEC) technology would increase the robustness of the system. The CLEC approach, mastered in the laboratory, provides the enzyme stabilization over a vast range of temperatures, pH and solvents thanks to the direct cross-linking of the proteins inside the crystal. The CLEC can be then immobilized on the electrode surface and their performances can be compared to the *in solution* grafted enzymes. Both approaches would be implemented in a fluidic cell, to gain a better control of the gas fluxes and the system temperature and pressure, in order to maximize the biological-WGSR efficiency (Figure 2).

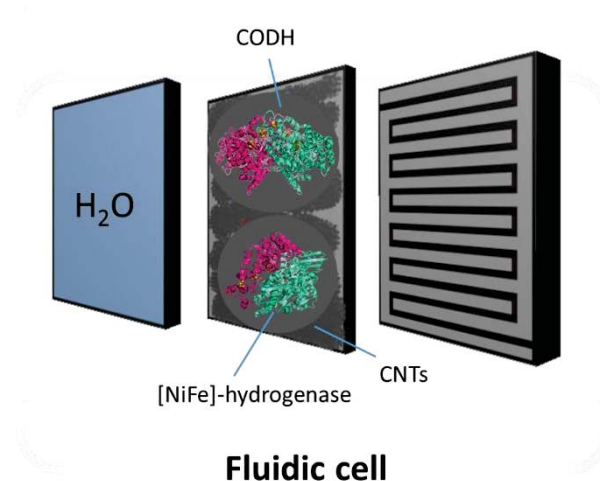


Figure 2. Development of water gas shift reaction in a fluidic cell system using the grafting of either the purified enzymes on a nanotube surface (CNTs) or the cross-linked CLEC crystals. The feasibility of the WGS reaction will be evaluated for both approaches.

Looking now at the concept of a bacterial *in vivo* water gas shift reaction, their use to replace the current industrial inorganic catalyst has proven valid and small-scale bioreactors already



shown. However, the employed bacteria were the ones that naturally perform the bio-WGS reaction for survival, thereby requiring restrictive growth conditions (anaerobic atmosphere, specific media, temperature etc.). The possibility opened from the utilization of the heterologous expression in *E. coli* is creating new industrial perspectives. This microorganism, widely used in current biotechnology, has already proven to be a robust and reliable bacterium, able to support different growth conditions in different culture media and with known tools for its metabolic engineering. Ideally, the heterologous expression of the *cooFSCTJ* operon in *E. coli* could create an easy to grow bacterium, able to reduce the CO, transfer the electrons to its natural hydrogenases thanks to the ferredoxin CooF and produce in this way bio-hydrogen (Figure 3). The success of the eventual CODH coupling with the naturally *E. coli* Hydrogenase 3 is going to be evaluated by monitoring the culture hydrogen production via gas chromatography technique, correlating it the volume of the supplied CO (Figure 3).

With this idea in mind, I tried some primary tests to see if we could co-express the *cooFSCTJ* gene in an ideal growth condition for the expression of *E. coli* Hydrogenase 3 (anaerobic fermentation<sup>130,155</sup>). Additional experiments are going to be done after the writing of this manuscript, but the preliminary once shown the ability to express the CODH (CO oxidation activity tested).

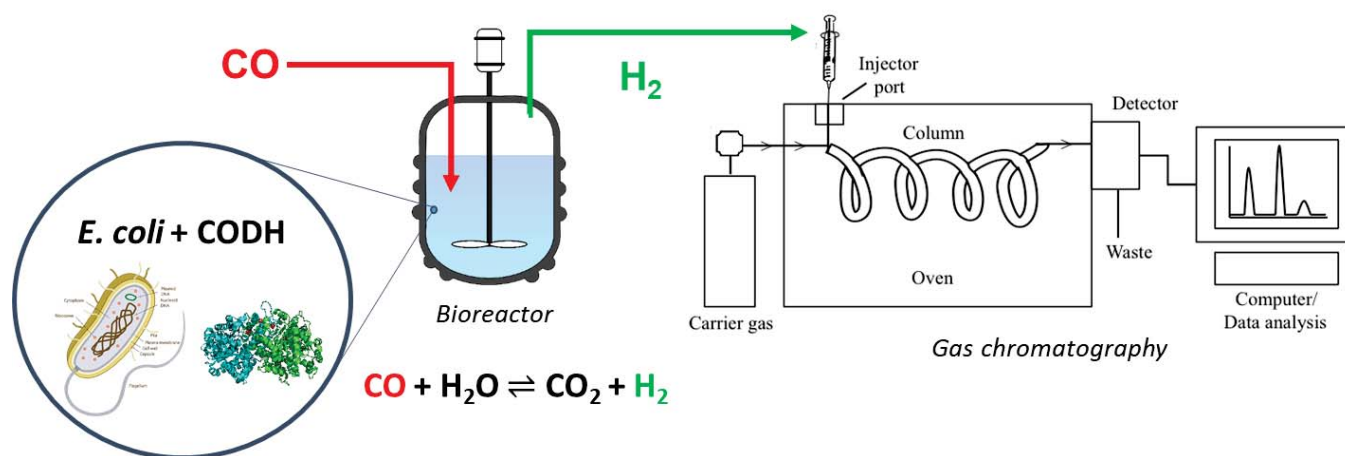


Figure 3. Development of water gas shift reaction in *E. coli* via heterologous production of CODH. The idea behind this strategy is that the protons and the electrons produced from the CODH's CO oxidation will be directly reduced by native *E. coli* hydrogenase. The produced hydrogen will be quantified via gas chromatography.

As a final reflection, nowadays one of the major concerns about the industrialization of microorganisms-based systems is the unpredictability related to the use of living systems. Most of the time, these technologies are developed using a consortium of microorganisms naturally present in specific niches (such as wastewater plants) and different experimental conditions are tried only evaluating the final effects on the energy efficiency of the working systems and not the metabolic impact caused on the microorganisms. This implies an enormous variability of the inner nature of these kind of technologies, an unappealing factor for the economical industrial applications. The ability to control and engineer a well-characterized bacterial strain, modulating its metabolism and supervising its lifetime growth, may be a valid option to address this problem. In this context, the use of modified *E. coli* to perform a complete WGS reaction *in vivo* becomes a valuable alternative to overcome all the drawbacks of the microorganisms that employ it naturally.

For this reason, I believe that a better scientific dialogue is required between fundamental and applied sciences. Just like the cross-talk between different nickel-insertion pathways could be the key to achieving a more efficient enzyme activation, so a better communication between our most specialized experts can be the catalyst to reach our full potential. Great ideas and solutions often result from the encounter of different knowledge fields.

# Chapter 6



## CHAPTER 6

### Methods

The study of metal-dependent macromolecules is a special subject requiring the use of specific methods in order to achieve it. Among the techniques used during my PhD, I will introduce the basis of the physical properties that govern their functioning in order to understand why and how some specific methods are used at the detriment of others.

The first main factor used to differentiate between different techniques is the physical state of the sample required during the experimental data collection, which split them in “**solution-state**” analyses, such as UV-Visible spectroscopy, Circular Dichroism (CD), Nuclear Magnetic Resonance (NMR), Isothermal titration calorimetry (ITC), Multi-Angle Laser Light Scattering (MALLS), Extended X-Ray Absorption Fine Structure (EXAFS) and Small-Angle X-ray Scattering (SAXS) and “**solid-state**” analysis, such as X-ray crystallography.

#### 6.1 X-Ray Crystallography

In protein crystallography, an ordered crystalline protein sample is used to investigate its three-dimensional structure using X-rays. This radiation is chosen due to its wavelength, which ranges between 1 Å to 10 nm (the same order of magnitude of an interatomic bond) and thus designates X-rays as the proper tool to unveil structure’s atomic resolution. The physical principle behind is based on the concept that a crystal exposed to the x-ray radiation will diffract, producing a diffraction pattern which contains information about the structure of the molecule. In order to achieve this condition, mainly two things need to occur: the protein needs to crystallize in a perfect ordered crystal and the X-ray needs to be monochromatic (single wavelength).<sup>178</sup>

##### 6.1.1 Crystallization

In order to obtain a perfectly ordered protein crystal, the first step is to produce a homogenous, high quality and concentrated protein. Once the purified protein is obtained, the focus is shifted in finding the best protein crystallization condition, the most important and delicate step in order to gain a diffracting crystal, and for which different protocols are available to screen multiple crystallization options. The principle of crystallization relies on finding the nucleation phase, achieved by the correct balance between the protein

concentration and the different precipitant conditions used (as pH, precipitant nature, buffer, concentrations etc.). As depicted in figure 1A, if the protein is in the unsaturated region (light orange session), neither crystal nor precipitant is formed and the drop will appear clear. In order to form a crystal, the protein needs to be in the nucleation zone (dark blue), a specific portion of the supersaturated region. Once the crystal nuclei are formed, the crystallization growth will occur in the metastable zone (light blue), lowering the protein concentration until the protein in solution will reach the solubility limit.<sup>179</sup> It is thus clear that many different conditions can lead to the protein crystallization and to screen all of then the vapor diffusion method is generally used (Figure 1B). When and if a promising condition is found, it is the possible to increase the crystal quality and size using multiple techniques, such as manual optimization, micro-seeding, additive screens and temperature dependent growth. It is crucial to favor a regular periodic crystal lattice, any anomaly introduced during the crystal nucleation and growth will have a negative impact during the diffraction collections (*see Cool-twinning problem chapter*).

It is important to remember that not all the proteins are able to crystallize. Generally speaking, highly flexible proteins or proteins with highly disordered regions represent a main challenge for the crystallographer community.

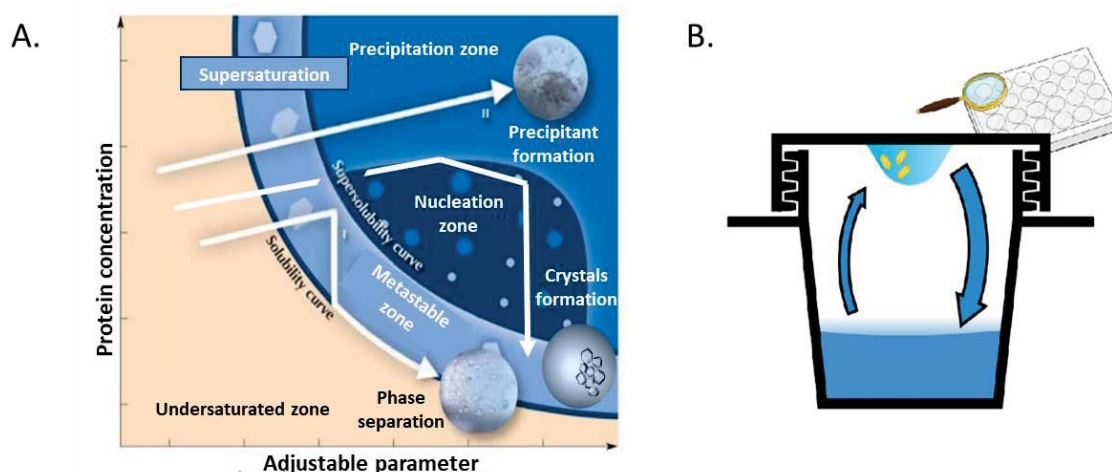


Figure 1. A) Crystallization diagram phase. B) Hanging drop vapor diffusion method is prepared using a drop composed of a mixture of protein and crystallizing solution suspended over a reservoir. As the drop contains a lower concentration of crystallizing solution, the vapor pressure of the drop solution is bigger than the one of the reservoir. This creates an overall gradient that leads to a water loss from the drop towards the reservoir,

increasing the protein and precipitant concentrations in the drop until supersaturation, allowing nucleation and crystal growth.

### 6.1.2 Synchrotron radiation

Briefly, X-ray radiation is generated by accelerating electrons in a synchrotron ring. A single wavelength is selected using a monochromator, focused on a beam and collimate to ensure its parallelism. The crystal is mounted on a goniometer and its position is adjusted to ensure a perfect beam-sample alignment. The collision between X-rays on the sample will cause the scattering of some of the X-ray in various directions. The collection of the diffracting X-rays arriving on the detector gives a diffraction pattern, which is related to the atoms spatial position accordingly to the Bragg's Law as:

$$2d_{hkl} \sin\theta = n \cdot \lambda$$

where  $d$  is the distance between two electron regions of the electron density,  $\theta$  is the angle of diffraction and  $\lambda$  is the wavelength of the diffracted X-ray. Considering that different angles of diffraction  $\theta$  are going to be present at the same time, the diffracted X-rays will be recorded at different intensities, which are related to the atomic 3D arrangement of the electrons in the crystal's atoms ( $d$ ). These information will later allow the re-construction of the protein structure.

### 6.1.3 Data Analysis

The collected images are the diffraction pattern, pictured as spots with different intensities. The data analysis aim is to convert these 2D pictures on a 3D electron density map. In order to do this, a Fourier Transformation is applied to the geometrical spatial arrangement detected (diffraction spots on the images) to obtain the spatial frequency of the atoms in the crystal (electron density). The electron density will allow the model building of the macromolecule in the crystal (see pipeline Figure 2).

Nowadays, many software and program suits are available to process almost automatically the diffraction data, which is a difficult mathematical task. Here I will present the main steps that need to be understood by a neophyte in the domain of data analysis.

#### 1) Determination of crystal space group and unit cell dimension.

Each space group is defined by a specific set of symmetry operations. A good choice of the space group is essential in order to correctly predict where the diffraction spots



will appear according to the symmetry law imposed. The registered images are indexed, meaning that to each spot three spatial coordinates ( $h,k,l$ ) will be assigned. A scale factor is donated to correlate the intensity of all the images of the full data set. The result of this process is a scaled file that contains the index of each spot and its relative intensity. During this process, in order to assess the quality of the diffraction data, the residual  $R_{merge}$  (or  $R_{sym}$ ,  $R_{meas}$  is similar but more elaborated version) is used to compare the intensities of all symmetrical related equivalent reflections. If a reflection does not match the identified unique symmetrical once, it is defined as an outlier and rejected. A high-quality data set should have a  $R_{merge}$  lower than 4-5% but a possible solution can be found with values as high as 20%. Where to cut the data without losing information is usually decided considering the resolution limit, which is generally reached when the value of  $I/\sigma(I)$  (average ratio of intensity reflection over its estimated error) drops below 2.0. However, this cutoff is arbitrary and recently a new indicator for data accuracy has been introduced as  $CC_{1/2}$ , the correlation between the reflection intensity of two half of a data set randomly split and defined as insignificant data  $<0.15$ . The completeness of the data set, defined as the coverage of all the possible unique reflections within the data set, should be around 100% in all resolution shells in the best possible scenario.<sup>180 181</sup>

## 2) Amplitudes determination from the intensity.

As mentioned above, the spots will have a different intensity depending on the angle (related to phase) and the amplitude (related to intensity) of the diffracted X-rays. Theoretically, from the knowledge of the amplitude and the phase angle between the diffracted waves is possible to determine the structure factors that describe the atom positions in the unit cell. The determination of the structure amplitude  $|F|$  is relatively easy considering its relation to the square root of the intensity, the main problem concerns the determination of the phase.

## 3) Phase problem.

The phase information is lost during the data collection, which means that indirect ways need to be used to be able to determine it. The most used methods are by molecular replacement and anomalous diffraction. In molecular replacement, the atomic coordinates of a protein homolog already solved is needed. In this case, the phase are “stolen” from the already solved homolog, placed in the new data set and

used to calculate the relative structure factor. In order to use these phases, the model structure needs to be placed correctly in the unit cell. The phases can be used just after this step to calculate a new set of structure factors that are going to be refined in the final model. If a structural homolog is not present, anomalous diffraction is the way to go. In order to achieve this, the crystal needs to be either soaked with heavy metals or the protein can be purified using seleno-methionine enrichment (*as in our case for RrCool*). The wavelength will then be tuned to match the external added metal. The position of these heavy atoms can be easily recognized in the diffraction patterns, and thus it can be used as spatial reference in order to determine the phases.

#### 4) Electron density map.

Knowing the phase and the amplitude, the structure factors can be calculated, which gives the electron density map of our structure. Higher is the collection resolution, more defined is going to be this map, which is needed to place our model.

#### 5) Model building and refinement.

When the electron density reaches a sufficient quality, the placement of the amino acids can start. The model is checked and refined in the electron density to reach a good match among the observed structure factors ( $F_{\text{obs}}$ ) and the calculated ones ( $F_{\text{cal}}$ ) using the formula  $2F_{\text{obs}} - F_{\text{cal}}$ . These maps are calculated with statistically weighted coefficient ( $\sigma$ ), which is the root-mean-square deviations (rmsd) of the mean electron density level. Usually the  $2F_{\text{obs}} - F_{\text{cal}}$  map should be countered at  $1\sigma$ , where else the difference map ( $F_{\text{obs}} - F_{\text{cal}}$ ) around  $2.5-3\sigma$ . Lowering these contours level results in highlighting phantom features, which is always counter-productive.

To guarantee a correct match between the model and the electron density map, the well-known 'R factors' are used to evaluate the structure quality based on individual residues. They measure the global relative discrepancy between the experimental obtained structure factor amplitude ( $F_{\text{obs}}$ ) and the calculated structure factor amplitude obtained from the model ( $F_{\text{cal}}$ ) as  $\sum |F_{\text{obs}} - F_{\text{cal}}| / \sum F_{\text{obs}}$ . A good matching R-value is typical around 20 %. R<sub>free</sub> is the analogue of R-factor but considering just 1000 random reflections that did not enter in the model refinement. It is an important factor to avoid data over-interpretation caused by possible inclusion of superfluous data in the model during the refinement cycles. It should not be higher than 7 % compared to the R-factor, a likely indication of data over fitting or in general

some problem with the model. Another quality criteria is the root-mean-square deviations (Rmsd) is an index of the reliability of the model compared to the typical geometrical parameters based on chemical knowledge. Different structural parameters can be considered but, generally, the bond length is the most used. For a good quality model, the rmsd(bond) is around 0.02 Å (values >0.03 Å are index of a problematic model). The Ramachandran plot is graphical way to visualize the torsion angles of the backbone versus the expected contour and shows if the individual residues fall in the allowed regions.<sup>180</sup>

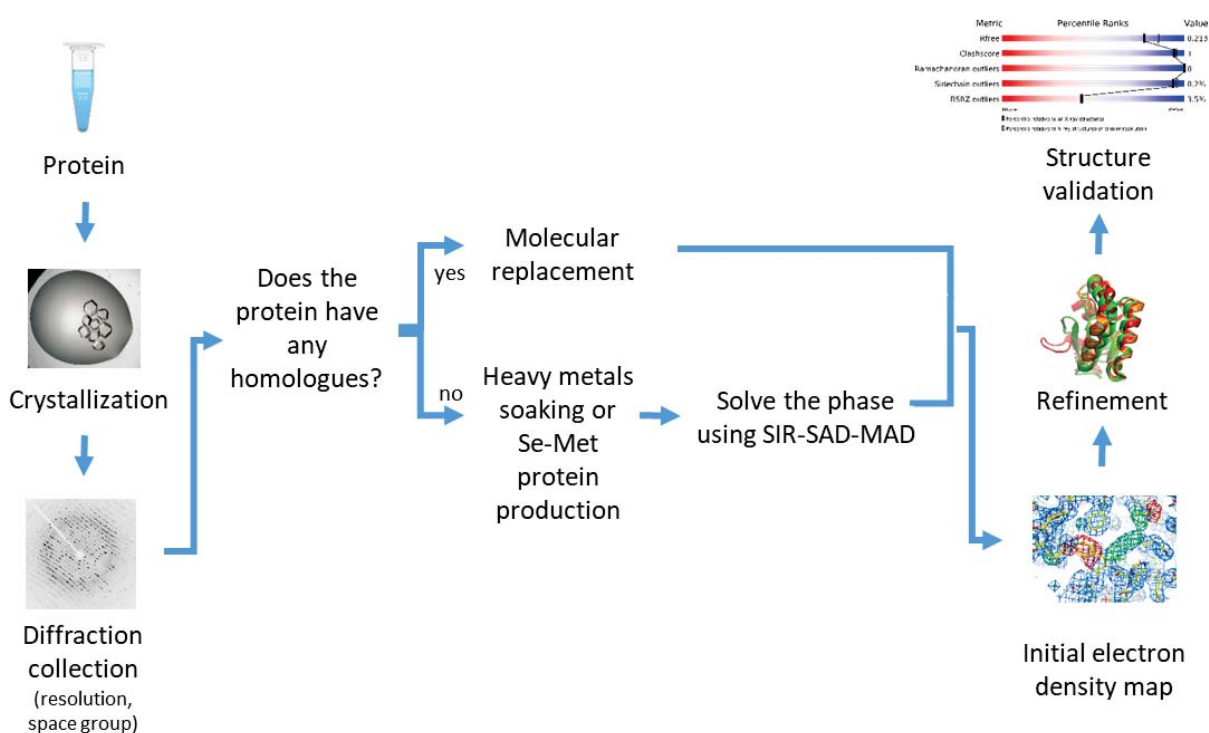


Figure 2. Pipeline for crystal structure data processing.<sup>182</sup>

## 6.2 Circular Dichroism

Circular dichroism (CD) uses circular polarized light and registers how a molecule differs in left- and right-handed light absorption.<sup>183</sup> Polarization is an intrinsic property of all transverse waves, of which the main exponent is the well-known electromagnetic wave *light*, and it specifies the geometrical orientation of the oscillation. Electromagnetic waves are synchronized oscillation of perpendicular electric and magnetic fields and the polarization always refers to the direction of the oscillating electric field. In circular polarization, the electric field has a constant magnitude while it rotates always perpendicular to the wave's direction. This rotation can have two possible verses:

- Right hand sense (RHC) with the respect to the traveling wave direction, named right circular polarization;
- Left hand sense (LHC) named left circular polarization.

When this polarized light hits a molecule that contains a chiral light-absorbing group, the result is a difference in the absorption of the right and left polarization light ( $\lambda_L \neq \lambda_R$ ), which affects the extent to which they are absorbed ( $\epsilon_L \neq \epsilon_R$ ). The circular dichroism is defined as this difference  $\Delta\epsilon \equiv \epsilon_L - \epsilon_R$ , which is measured over a range of wavelengths. The CD measurement are usually carried out in the visible (700-250 nm) and in the ultra-violet region (250-190 nm) and a chiral chromophore is needed in order to have a signal different from 0, due to a measurable imbalance between  $\epsilon_L$  and  $\epsilon_R$ . The signal registered can be either positive (+) or negative (-) depending if the LHC is absorbed to a greater extent compared to the RHC or vice versa. In proteins, all amino acids are left hand chiral, which makes them the perfect object of study for CD spectroscopy (with the exception of the achiral glycine residue, silent in a CD measurement).

In order to gain as much information as possible, the CD can be used to study macromolecules in two different wavelength ranges: the ultra-violet and the visible regions.

In the far ultra-violet region (190-250 nm), it is possible to investigate the protein secondary structural elements (such as  $\alpha$ -helix and  $\beta$ -sheet) and predict their possible percentage composition in the final structure, normalized by amino acids composition and protein concentration. Generally,  $\alpha$ -helical proteins are characterized by negative bands at 222 nm and 208 nm and a positive band at 193 nm,  $\beta$ -sheet proteins have a negative bands at 218 nm

and a positive one at 195 nm, while a random coil have really low values above 210 nm and a negative band near 195 nm (Figure 3).<sup>184</sup> These standard values are used by the secondary structure prediction software to determine the secondary structure composition of the registered spectrum.<sup>185</sup>

In the near UV-visible region (250-700 nm), it is possible to investigate the tertiary structure and the absorptions of aromatic groups. Moreover, it can be used to determine how a metal is interacting with a protein. In fact, a free metal ion in solution will not give a CD signal, meaning that only the ions bound to the protein are going to be detected. The CD spectroscopy can resolve individual electronic d-d transitions, usually not fully resolved in absorption spectra. In the case of a transition metal, this optical activity is mainly attributed to the configuration effect, caused by the asymmetrical distribution of the coordinating donor atoms towards the metal. Thus, it is possible to have an insight into the geometrical disposition of the protein bound metal.

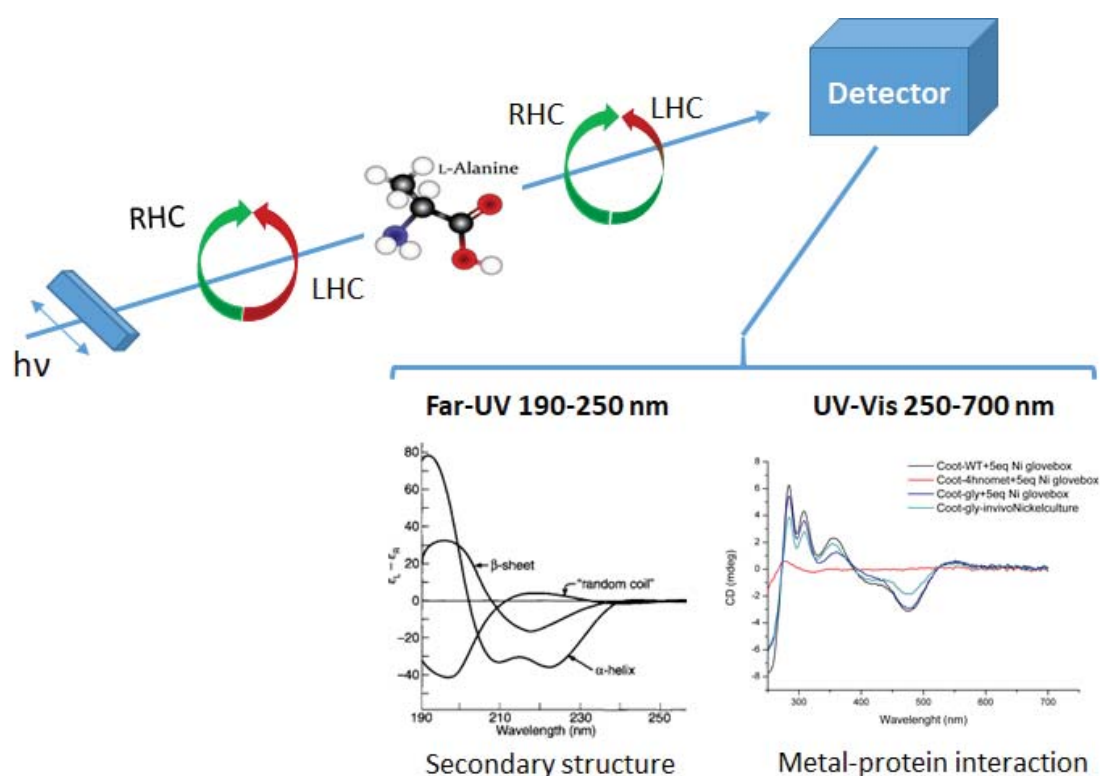


Figure 3. Monochromatised light passes through a modulator, which converts it into left (LHC) and right (RHC) handed polarized light. The two polarizations are differently absorbed from a chiral center, and the difference in LHC-RHC absorption is detected.

Metal titration on the protein can be performed to obtain a preliminary idea of the stoichiometry. Based on the knowledge that the metal-protein CD signal will evolve depending on the amount of added metal, an invariable CD signal indicates the saturation of the protein-binding site and the possible number of bound ions.

Analysis of CD spectra as a function of temperature can be used to determine the midpoint of the unfolding transition ( $T_M$ ) of a protein and its related structural stability. Another technique is to use a chemical denaturant agent, as guanidine hydrochloride, and check the evolution of the  $\alpha$ -helix signature to detect its folding stability (*see RrCooJ chapter*).<sup>186</sup>

CD by itself gives information about structure folding, metal binding geometry, affinity and stoichiometry but it has less resolution power than other techniques that are going to be described later on (as NMR and X-ray crystallography, and ITC). However, it is a fast spectroscopic analysis (one spectra can be collected in less than a minute), it can be performed in different solvents, temperature, protein concentration, pH and in presence of different co-factors. It is, also, a perfect method to verify the quality of the protein sample before proceeding with more time demanding and consuming techniques.

A series of papers tackling CD spectroscopy as a concrete tool to investigate protein structure and metal interactions can be find here <sup>187–190</sup>.

### 6.3 Isothermal titration calorimetry

Isothermal titration calorimetry (ITC) is used to determine the thermodynamic parameters of biomolecules interactions in solution, like association ( $K_a$ ) and dissociation constants ( $K_d$ ), enthalpy ( $\Delta H$ ) and entropy ( $\Delta S$ ) changes and binding stoichiometry ( $N$ ).<sup>191</sup> A typical experiment is performed via the addition of a reactant in the cell containing the reagent, which is constantly compared to a reference one to register the resulting modifications in temperatures. The chemical reaction created by the injection of the reagent in the reactant can releases (exothermic) or absorbs (endothermic) a defined amount of heat ( $q$ ) which is proportional to the amount of bound ligand to the protein.

$$q_i = \Delta H \times v \times \Delta L_i$$

With  $v$  volume of the injected ligand,  $\Delta L_i$  related to the ligand concentration and  $\Delta H$  linked to the characteristic binding enthalpy. The working principle of the instrument is to maintain constant the temperature between the reference cell and the sample cell. This amount of temperature is recorded and converted in the heat ( $q$ ) after each injection using the area of the peak ( $\mu\text{cal/s}$ ) vs time (Figure 4). The heat will decrease proportionally to the binding accessibility of the protein until the saturation is achieved. The final formula used to analyses the data depends by the number of binding events happening and it differs substantially between 'one set of sites' and 'two set of sites' (see *RrCool chapter*). From this measure the Gibbs free energy ( $\Delta G$ ) and the entropy ( $\Delta S$ ) can be derived as:

$$\Delta G = -RT\ln K = \Delta H - T\Delta S.$$

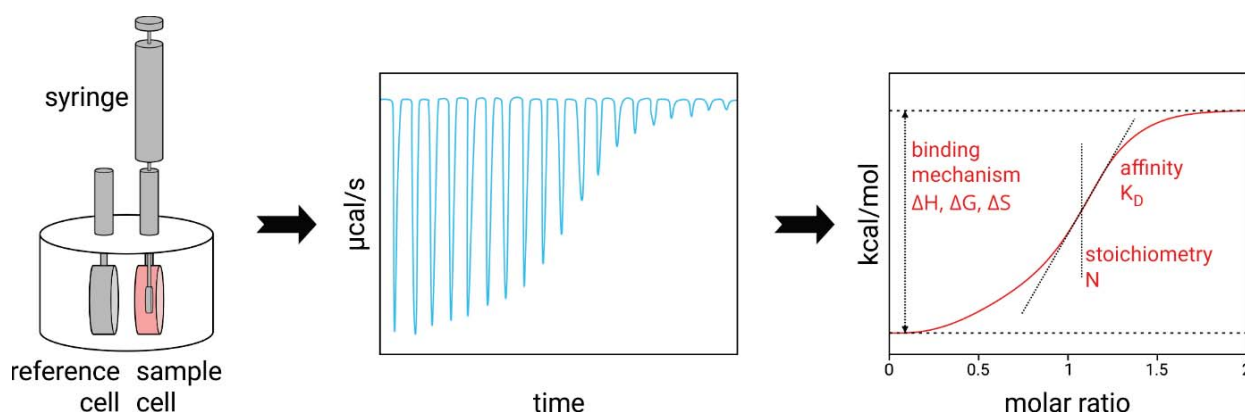


Figure 4. ITC instrument. The resulting spectra is used to integrate the area of each peak ( $q$ ), resulting in a new plot versus the molar ratio of ligand to protein. The resulting isotherm can be fitted to a binding model from which the affinity ( $K_D$ ) is derived. The molar ratio at the center of the binding isotherm give us the reaction stoichiometry. The plot shown below is an example of a 1:1 binding reaction. [Adapted from <https://2bind.com/itc>]

Also for this technique is possible to easily modulate the protein and ligand concentrations and the set up buffers and the temperatures, considering that we are working in solution. A measurement does not need a high protein concentration ( $\sim 5\text{-}10\ \mu\text{M}$ ), making this analysis particularly interesting when we are working with difficult to purify proteins.



Concerning the protein-metal binding study, the binding stoichiometry and dissociation constant can be confidently determined, conferring to this approach a commonly recognized “queen” role in  $K_D$  and stoichiometry determination.

Problems may arise in the data interpretation when multiple binding sites are present; a careful and pondered data interpretation is always advocated.

## 6.4 Multi angle laser light scattering

Multi-angle laser light scattering (MALLS) experiment allows the determination of some key features of proteins in solution, as protein oligomeric state, concentration and molecular weight. The main concept to keep in mind is that the detection of light scattered from a sample at several angles can be used to determine its absolute molecular weight (related to the scattering at  $0^\circ$  scattering angle) and the sample concentration at the same time. In a MALLS set up a laser source is used to illuminate a sample holder and the scattering light is collected in each different detector (this system is usually combined to a size-exclusion chromatography (SEC) in order to be confidently sure about the sample state and to isolate single species from a mixture (Figure 5). The scattering is produced when a photon collides with a molecule and it is remitted at a certain energy as light. This phenomena can be analyzed with different theory and in the simplest case of elastic scattering approximation (object  $\ll$  than the incident wavelength) the Rayleigh scattering theory can be used. In this approximation, the scattered light intensity will be proportional to the number of molecule in each elution volume, thus the molecular weight (Mw), and the molecular size of the sample.

$$\frac{KC}{R_{(\theta,C)}} = \left( \frac{1}{Mw} + 2A_2C \right) \frac{1}{P_\theta}$$

With  $R_\theta$  defined as the Rayleigh ratio,  $C$  the sample concentration,  $\theta$  the measured angle,  $Mw$  the molecular weight,  $P_\theta$  the form factor,  $A$  as the second virial coefficient and  $C$  as coefficients dependent by the system.  $K$  an optical constant, related to the refractive index increment  $dn/dc$ , which is plotted versus the concentration to the molecular weight of the samples ( $K = (2\pi^2 n^2)/(\lambda^4 N) * (dn/dc)^2$ ).

In the case of small molecules, the scattered light is going to be independent from the scattering angle (isotropic scatter), where else for big molecules the scattered light will reach the detectors with different phases and for this reason we need to register it at different angles (anisotropic scatter). At  $\theta = 0$  the form factor is going to be equal to 1 ( $P_\theta = 1$ ), thus independent by the interference. Multiple collection angles are registered and extrapolated back to  $0^\circ$ , in order to eliminate the phase problem. In addition, the size of the molecules can be determine, as radius of gyration  $R_g$ , giving an idea about the sample conformation.

MALLS is a powerful technique to determine protein conformational changes as oligomers or aggregates formation, size and shape modification in native solution conditions. It requires a low amount of protein in order to make measurements (20  $\mu\text{L}$  at least 1 mg/ml) and it does not required the use of calibrating samples during the experiments.

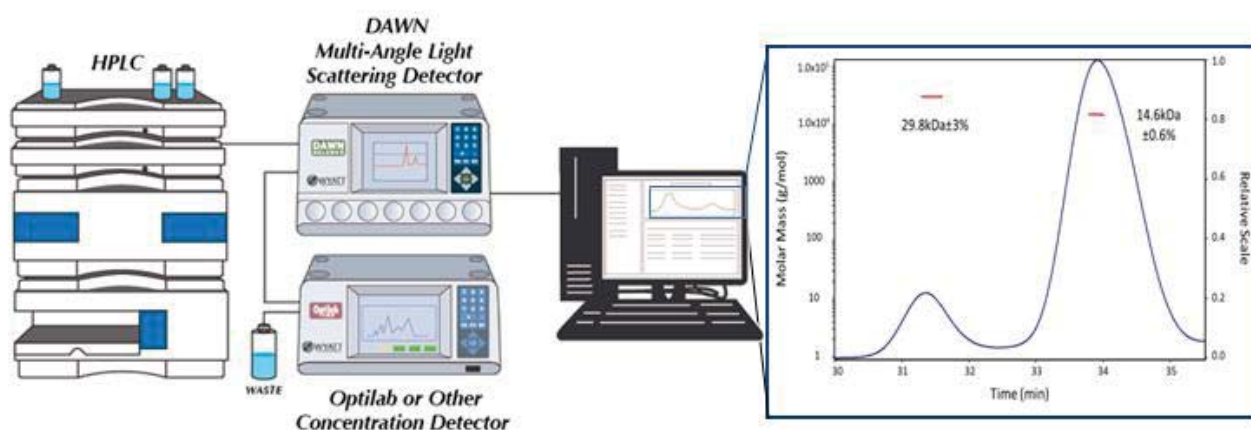


Figure 5. Schematic representation of SEC-MALLS set up.

## 6.5 Nuclear magnetic resonance

Nuclear magnetic resonance (NMR) spectroscopy is currently the only technique capable of determining the 3D atomic resolution structures of macromolecules in solution, combined to the possibility of studying time dependent phenomena such as protein folding and interaction. The physical principle behind this technique is based on the interaction between the atomic nuclei and a magnetic field, which would cause a splitting in the spin energy levels of the atomic nuclei spin resulting in at least two separate states as in the case of spin  $\pm 1/2$  (Figure 6).

This phenomenon rises from the fact that atomic nuclei possess a non-zero spin ( $s$ ) associated to a non-zero magnetic moment ( $\bar{\mu}$ ), which can interact with an external magnetic field orienting itself accordingly. (For this reason, nuclei with an even number of protons and neutrons ( $\text{spin}(S)=0$ ) do not have NMR signal because they have a zero magnetic moment, which would not be perturbed by the applied magnetic field.) This interaction splits the nuclear spin levels in parallel and antiparallel polarization due to the presence of a constant magnetic field ( $B_0$ ), defined as alignment. The time dependent perturbation of this alignment via a perpendicular magnetic field ( $B_1$ ) generates the wanted NMR signals when the system goes back to the equilibrium.

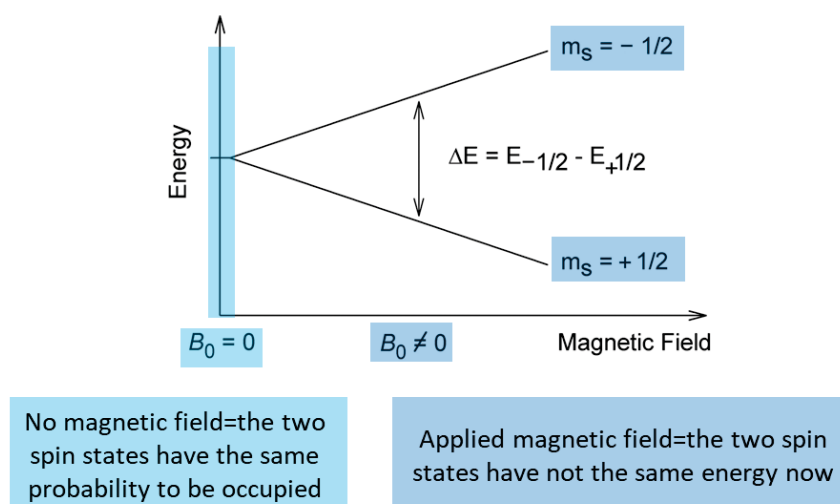


Figure 6. Interaction of nuclei  $s = \pm 1/2$  with an external magnetic field. The splitting of nuclei energy correspond to the applied magnetic field and can be parallel to it  $m = 1/2$ , the most stable level, or antiparallel  $m = -1/2$ . During an NMR experiment, an oscillating magnetic field ( $B_1$ ) is applied perpendicular to  $B_0$  with a frequency equal to the energy-splitting  $\Delta E$ . In this way a promotion from the state  $m_s = +1/2$  to  $m_s = -1/2$  is induced and the resulting relaxation is recorded.

The precise resonant frequency is strictly dependent by the nature of the nucleus and by its surrounding shells of electron, which cause the 'shielding' effect. In fact, the electrons are also polarized by the applied magnetic field, producing an opposite magnetic field that lowers the nucleus one as a reflection of its chemical environment. This shift, proportional to the electron density distribution, shields the local magnetic nuclear resonance and results in a chemical shift that can be analyzed compared to a reference frequency.

In 1D experiments, the proton chemical shifts are investigated and it is largely used in chemistry. From a "protein point of view" this spectra gives information about the nature of the proton ( $H^N$ ,  $H^{\alpha}$ , aromatic etc..) but also about the secondary structure folding. However, 1D spectra are too complex to determine the protein structure and another dimension is needed for this purpose. 2D spectra can either detect the same isotope at two different frequency (homonuclear experiment) or use the same frequency but different isotopes (heteronuclear). This method works well for the determination of small macromolecule, however for the determination of large protein (>20 kDa) an additional dimension is once again required in order to spatially resolve the signals. For the 3D spectra, the proteins are isotopically double labelled ( $N^{15}$  and  $C^{13}$ ) and the resonance of these nuclei is correlated and solved in a single experiment. In this case, the spectra overlap is reduced and the information arises from the transfer magnetization over the peptide bond and the connectivity among the amino acid side chains. It is essential to assign each chemical shift to each atom in order to have the resonance assignment of the protein.

From the experimental point of view, in the case of a 3D spectra, the sample needs to be prepared in a double labelled fashion using  $N^{15}$  and  $C^{13}$  ( $^{15}NH_4Cl$  and  $^{13}C$ -Glucose in minimal medium) during the protein production, an expensive procedure if many liters of medium are needed. Once the protein is purified, it will be placed in a magnetic field and the detected resonant frequency will be proportional to the chosen instrument magnetic field. It is important to remember that the solvent protons are going to contribute to the proton signals of the protein; therefore the use of a deuterated water is essential in order to distinguish it from the protein signature. The duration of the spectra acquisition depends by the signal-to-noise ratio and different acquisitions are needed in order to obtain a defined spectra (a range that goes from a 10 min to a couple of hours for a double labeled protein). In all these time laps, the protein needs to be stable in solution at high concentration (average protein

concentration of 1 mM) and stable at the experimental temperature (generally 20-25°C), conditions that can be challenging for proteins in solution. Protein size is another limitation for this technique and usually proteins smaller than 40 kDa are the best candidates for NMR structure resolution due to the fact that the peaks attribution is still mainly done by hand. On the other side, NMR does not present the limitations encountered for the x-ray structure determination, as the need to have the protein in crystal state, and also intrinsically unstructured proteins can be investigated. Moreover, the interaction of proteins with co-factors in solution can be studied, giving response about the conformational dynamics and flexibility of different protein portions (*as RrCooT in our case*).

## 6.6 X-ray absorption spectroscopy (XAS)

X-ray absorption spectroscopy (XAS) is a powerful method to determine the local geometry and the electronic structure of a specific element. As for X-ray crystallography, this technique needs x-rays that are generally produced by a synchrotron. The goal is to analyze the modulations (i.e. the fine structure) of the X-ray absorption coefficient around the absorption energy of the core electron of a selected element (photoelectric effect), which are related to the coordination geometry and electronic properties of the absorber. In order to do so, the core electrons are excited using monochromatic energetic photons that scan a region of about 800-1000 eV around the core electron's ionization energy. The x-ray absorption phenomenon is described by the Beer's law as:

$$I_t = I_0 e^{-\mu x}$$

Where  $I_0$  is the incoming x-ray intensity,  $\mu$  is the absorption coefficient of the element,  $x$  is the sample thickness and  $I_t$  is the transmitted intensity. We are particularly interested in the absorption coefficient  $\mu$ , because it is proportional to the photoelectric cross section, i.e. the probability of absorption of a photon at a given energy. For a given element the absorption coefficient  $\mu$  can be written as:

$$\mu = \frac{-\ln\left(\frac{I_t}{I_0}\right)}{x}$$

When the x-ray has an energy equal to the binding energy of a core-level electron in the sample, a drop in the transmitted x-ray intensity appears corresponding to its absorption. This is called absorption edge and it is specific for each element. What we do in a XAS experiment is to register the absorption coefficient, in our case of Nickel, near and above its absorption edge. (Figure 7).<sup>192</sup>

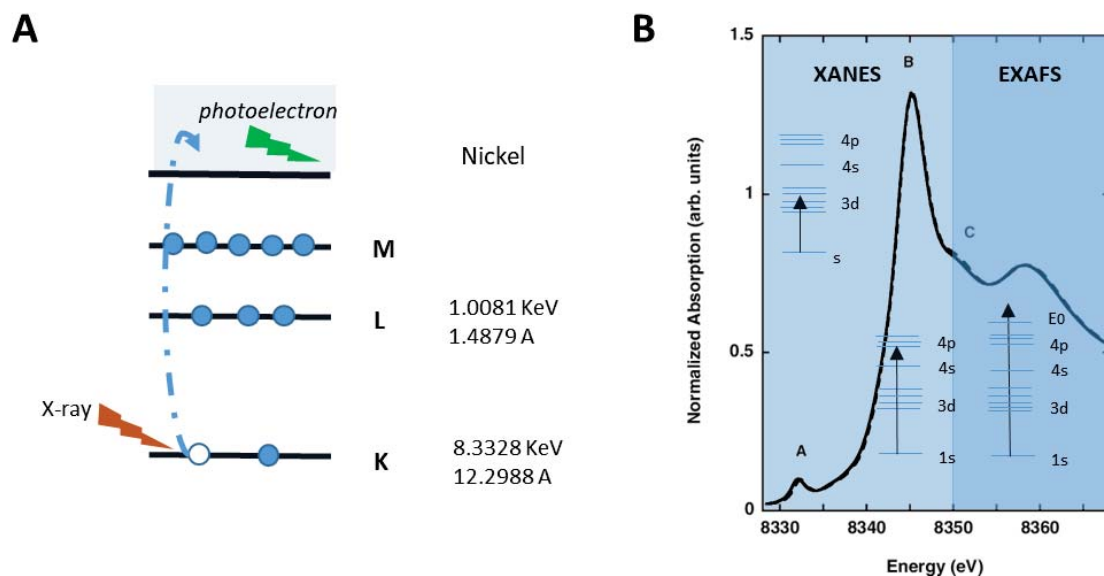


Figure 7. A) Representation of the photoelectric effect, in which an x-ray is absorbed and a core level electron is promoted to a vacant level. The electron binding energies for nickel (our selected element) are reported. B) An example of Ni K-edge spectrum is shown highlighting two main XAS regions with the relative energies.<sup>193</sup>

The two main regions in a XAS experiment are:

- X-ray absorption near edge structure (XANES). From some eV before the edge to ~ 50-100 eV above. It is due to transitions of the excited 1s electron to bound and quasi-bound states within the atomic potential. It shows in general pre-edge features that are indicative of the coordination geometry of the absorber. For a 3d transition metal, the 1s electron is promoted to 3d and 4p states.
- Extended X-ray Absorption Fine Structure (EXAFS), from ~50-100 eV above the edge. The photoelectron is promoted to the continuum and is backscattered by the neighboring atoms. This phenomenon produces an interference of the outgoing photoelectron wave with the backscattered wave, which is correlated to the

geometrical configuration around the absorbing atom. Thus, this region provides information about interatomic distances, coordination number and nature of neighboring atoms of the absorber. The oscillatory portion of the spectrum can be extracted and Fourier transformed to determine these structural parameters.

Concerning the advantages of this technique, it is a powerful tool to determine the coordination and the geometrical environments of a metal binding protein in solution, since it does not require crystallization. The measurements are performed at liquid nitrogen temperature and the sample can be fully recovered after its analysis. On the other hand, the computational analyses behind the spectra interpretation is challenging and not straightforward. The comparison of the spectrum of a metalloprotein with reference spectra of compounds with known geometry is desirable to drive the analysis.

## **6.7 Biological small-angle scattering (SAXS)**

The small-angle scattering of x-rays (SAXS) is used to investigate macromolecules in solution, differing in this way from x-ray crystallography. In this technique, the sample is exposed to a small angle x-ray beam at 12.5 KeV and the resulting scattered radiation is registered on a detector. The small-angle scattering is generated by the coherent secondary wave scattered by a single molecule's atoms, resulting in an isotropic intensity distribution due to the random positions and orientations of the particles in solution. The intensity of the resulting x-ray scattering curve ( $I(s)$ ) is plotted against the scattering angle ( $s$ ) and it can be used to create a low-resolution model of the protein in solution (blue region Figure 8) but it is not suitable for reaching the atomic resolution.<sup>194–196</sup>



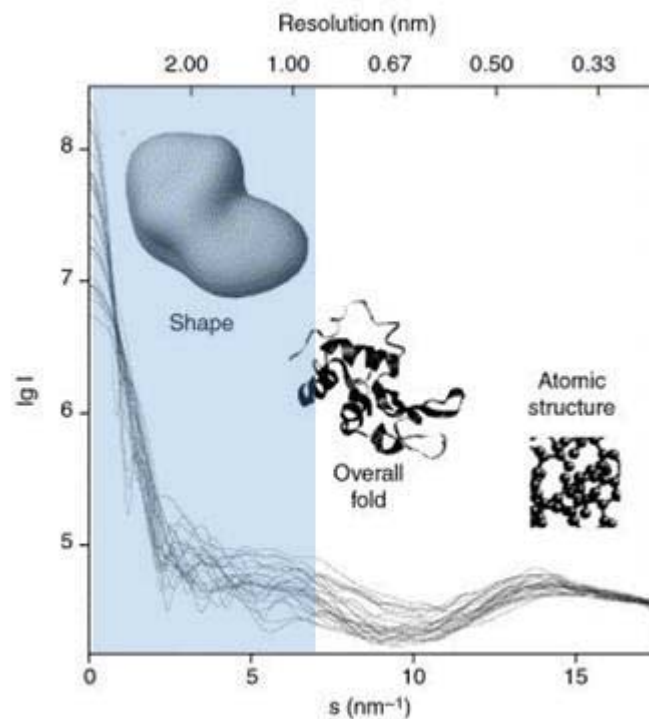


Figure 8. X-ray computed scattering curves for 25 different proteins. It is clear that at low angles (blue region ) the curves are really different between each other, while above this region it is difficult to distinguish the curves between each other. The atomic resolution cannot be resolved considering that all the curves basically match in the same spot. Adapted from <sup>197</sup>.

In order to extrapolate the highest amount of information from this curve, the particles need to be considered as mono-disperse and non-interacting with each other. With this approximation, the signal is going to be considered as the scattering of a single spherical particle averaged on all the possible orientations. In order to apply this approximation, the starting region of the scattering curve ( $s=0$ ,  $I_0$ ) needs to follow the Guinier approximation in order to be considered not aggregated (Figure 9 Guinier plot). Considering this step, it is clear that the sample preparation has a critical and essential role for the success of the experiment. The Fourier transformation of the scattering profile ( $I(q)$ ) is what is generally used to define the interatomic distances ( $P(r)$ ) in the real space. This term is the one used to describe the space and the volume occupied by a macromolecule, the final envelope of the protein sample (Figure 9).

The goal of a SAXS experiment is thus to determine a pool of 3D structural models starting from a 1D scattering plot. In order to achieve this, *ab initio* models are generated from the scattering curve and the best solution able to confidently match the experimental data is chosen and used as the macromolecule envelope in solution (remember that we are always in sphere approximation).

Considering the approximations required during the data analysis, a valuable data interpretation, as for most of the techniques, needs to be supported on other analysis methods as x-ray crystallography and NMR analysis (*see RrCool proposed structural model data analysis*). Compared to other techniques, SAXS is a powerful tool an evaluation of shape determination, flexibility and oligomer mixture of the proteins in solution.

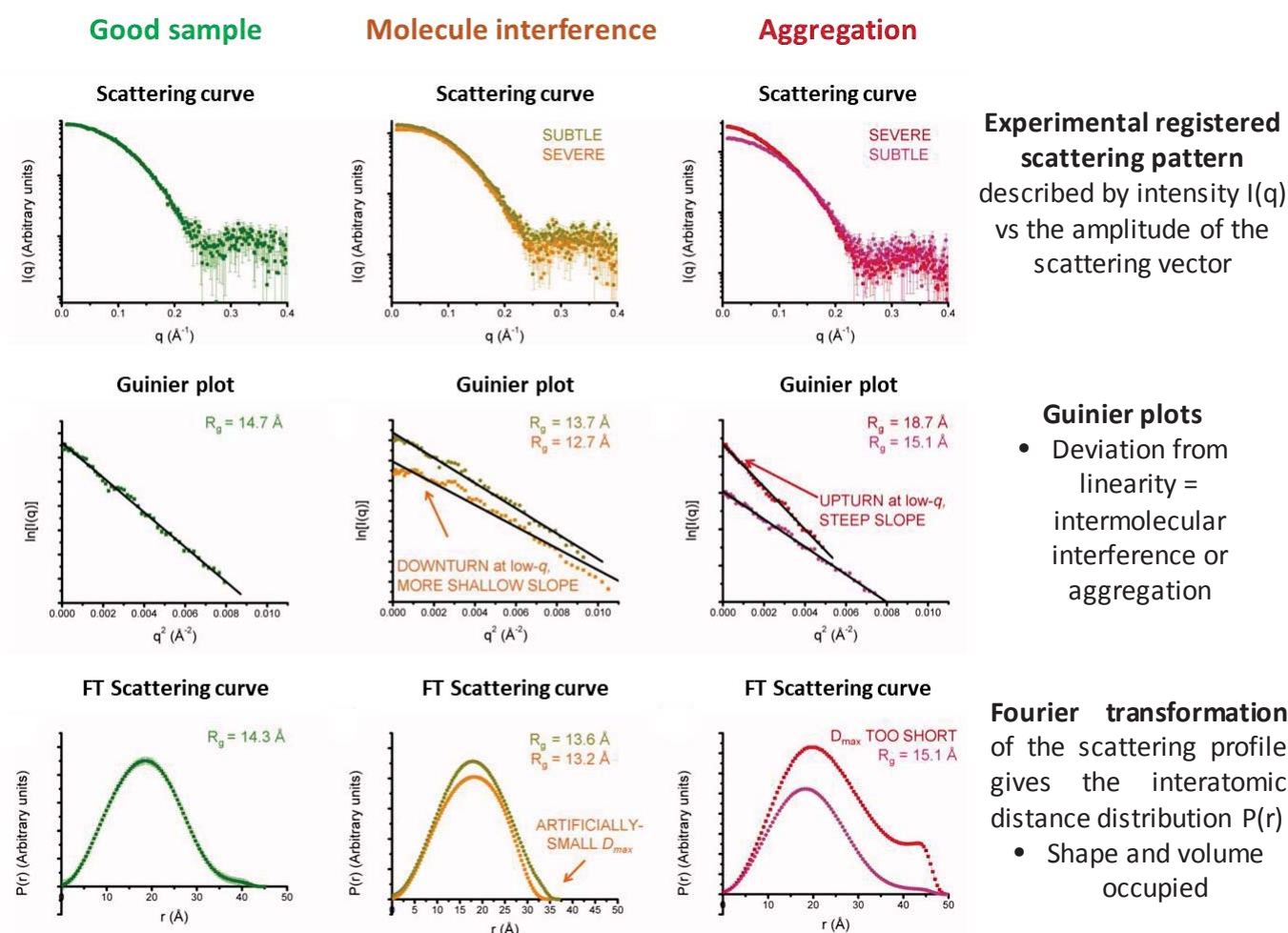


Figure 9. Essential steps required for data interpretation. Adapted from <sup>197</sup>.

|   | Physical state  | Information   | Limitation   |
|---|-----------------|---|--|
| <b>X-ray absorption spectroscopy</b>  | Solution        | Geometry and oxidation states (XANES)<br>Distances , coordination numbers and nature of the ligands (EXAFS)   | Complex computational work to determine the ligands, specific for one element.   |
| <b>Small Angle X-ray Scattering (SAXS)</b>                                    | Solution        | Protein envelope in solution.   | No conclusive by itself (multiple models proposed to fit the surface), limited by sample quality.  |
| <b>Size Exclusion Chromatography-Multi Angle Light Scattering (SEC-MALLS)</b> | Solution        | Protein Mw and concentration as well as oligomerization/aggregation caused by protein/protein, or protein/metal interaction or by environmental conditions modification (salt, pH, ....).       | Protein concentration needs to be higher than 50µM in order to be detectable. Limited by the size exclusion of the gel filtration (high oligomers or elongated molecules will be excluded) |
| <b>X-crystal structure</b>  | Solid (crystal) | 3D structure in solid<br>Possibility to evaluate protein/protein-protein/cofactors interactions. Mechanistic studies by time-resolved crystallography.<br>Identification of metal-binding sites | Crystals needed.<br>Artifact due to solid-state effects.<br>X-ray damages and X-ray reduction of metal sites.  |
| <b>Nuclear Magnetic Resonance (NMR)</b>                                       | Solution        | 3D structure in solution<br>Possibility to evaluate protein/protein-protein/cofactors interactions  | High protein conc.<br>Stable over time sample at room temperature needed (recording depending by the instrument)<br>Limited to small proteins (<40 Kda)                                    |
| <b>Circular Dichroism (CD)</b>  | Solution        | 190-250 nm secondary structure determination.<br>700-250 nm protein-metal interaction.  | Protein concentration in solution, homogeneity of the samples.<br>Buffer composition is critical.  |
| <b>Isothermal titration calorimetry (ITC)</b>                                 | Solution        | Dissociation constants with metals-partners, stoichiometry determination.   | Protein concentration in solution, homogeneity of the samples.   |

## Annex

### MATERIALS AND METHODS OF CHAPTER 4

#### Heterologous gene expression

The plasmids harboring the *RrcooS* gene (prsF and prsF-TEV) were transformed into *E. coli* BL21(DE3). From this transformation, a colony was picked and grown in 3 mL LB medium supplied with the right antibiotic at 37 °C for 8 h under agitation. One mL of this culture is then used to inoculate the 100 mL O/N pre-culture grown at 37 °C under agitation. Using this pre-culture, different growth media and protocols were performed. In the case of the anaerobic growth, the pre-culture was added to 1L of either TB or LB or M9 supplied with 8.7 mL glycerol, 0.1 mM Na<sub>2</sub>S, 0.5 mM NiSO<sub>4</sub>, 0.1 mM FeSO<sub>4</sub> and 1% glucose to reach an optical density at 600 nm around 0.1 and left at 30 °C under agitation until OD<sub>600</sub> reached approx. 0.5. In the case of “defficient-Nickel” enzyme production NiSO<sub>4</sub> was not added. At this point, the cultures were introduced in the glovebox and left to degas for 30 minutes before their induction with 0.2 mM IPTG. Harvesting took place 20-23 hours after induction. The cultures were centrifuged for 30 minutes under anaerobic conditions. To recover the cell’s pellet, the centrifugation pot was re-entered in the glovebox, the supernatant was removed and the cells pellet was stored at -80 °C until purification. The CODH activity of the cultures was checked before purification. Concerning the aerobic growth, auto-induction medium has been chosen: 15 mL of O/N pre-culture was used to inoculate 1 L of 5 gr yeast extract, 10 gr casamino acid, 20 mL 5052, 50 mL NPS and 1 mL MgSO<sub>4</sub>, left at 37 °C at 180 rpm. After 3 hours, 1 mM of FeSO<sub>4</sub> and 2 mM of L-cysteine were added and the temperature was switched to 15 °C at 180 rpm. After 21 hours, the cells were harvested, centrifuged and the pellet was recovered under argon and stored at -80 °C until purification.

#### Purification

The purifications were carried out in the glovebox at room temperature. For the *RrCODH*-TEV, frozen pellet was re-suspended in buffer A (0.05 M Tris-HCl pH 8.5, 2 mM DTH, 1 mM TCEP, 0.150 M NaCl) and one complete Protease Inhibitor cocktail tablet (Roche) per 50 mL buffer. The cell suspension was sonicated 8 minutes using a cycle of 50 sec OFF/10 sec ON with 50%

amplitude at 4 °C. The lysate was then centrifuged for 20 minutes at 40,000 rpm. The 50 mL of supernatant were loaded on 20 mL of HisPur Cobalt Resin, previously equilibrated with buffer A, and left under stirring for 1 hour. After washing the column with 4 column volumes of buffer A, the protein was eluted with a step gradient of 20 mM, 40 mM, 60 mM, 100 mM, 200 mM imidazole in the same buffer (2x10 mL each elution). The fractions containing the enzyme were checked via SDS-PAGE, pulled together for a total volume of 50 mL and concentrated to 3 mL via amicon cell 30 kDa cutoff. The his-tag was cleaved via incubation with His-taggedTEV protease, which was added to a sample left in dialysis in 500 mL of buffer A O/N. The next day, the protein solution was passed through the HisPur Cobalt resin column, the not cleaved portion and the His-tagged TEV remained on the resin. A. The flow-through was collected, concentrated, and loaded on a Superdex75 16/600 column equilibrated in buffer A. Fractions corresponding to *RrCODH* were pooled, concentrated via amicon cell 30kDa cutoff, frozen in liquid nitrogen and stored at -80 °C.

For *RrCODH*, the lysis procedure was carried out as described above in a buffer A' (0.1 M Tris-HCl pH 8.5, 2 mM DTH, 2 mM DTT) and one complete Protease Inhibitor cocktail tablet (Roche) per 50 mL buffer. The supernatant was then loaded at 0.5 mL min<sup>-1</sup> onto a Q-sepharose HP 16/10 column (GE Healthcare) equilibrated with buffer A'. After loading, the column was washed with buffer A' and the protein was eluted with a linear gradient (0-50% in six column volumes, 50-100% in six column volumes) of buffer B (0.1 M Tris-HCl pH 8.5, 2 mM DT, 2 mM DTT, 1 M NaCl). The fractions of the main elution peak were analyzed by SDS-PAGE and those containing *RrCODH* were pooled and concentrated via vivaspin 100 kDa cutoff to reach a final volume of 0.5 mL. The protein was then loaded onto a Superdex75 16/600 size exclusion column (GE Healthcare) equilibrated in buffer C (50 mM HEPES pH 7.5, 300 mM NaCl, 1 mM TCEP). The fractions containing *RrCODH* were pooled and concentrated before flash freezing in liquid nitrogen and stored at -80 °C.

### **Solubilisation assay of *RrCODH***

The pellets coming from the 50 mL cultures were re-suspended in 15 mL of 0.1 M TRIS pH 8.0, 10 % sucrose, 2 mM DTT, 0.1 gr lysozyme and one complete Protease Inhibitor cocktail tablet (Roche) per 50 mL buffer. These 15 mL are then split in 3 falcon tube of 5 mL each and to each separate volume were added 1% sodium deoxycholate (DCA), 1% Triton or 0.1% sodium lauroyl sarcosinate (Sarkosyl). The samples were then sonicated for 2min30sec 2 sec ON/4 sec

OFF with an amplitude of 30%. The sonicated cells were stirred in the cold room for 30 min. From each volume, 1 mL was centrifuged at 170000g for 30 min. The supernatant was separated from the pellet, which was re-suspended before SDS-PAGE.

### Activity tests

The CO-oxidation activities of CODH was performed by monitoring the reduction of methyl viologen (MV) over time at 604 nm ( $\epsilon = 13.6 \text{ mM}^{-1}\cdot\text{cm}^{-1}$ ). Activity tests were done in a buffer made of 0.1 mM MOPS pH 7.5, 1 mM TCEP and 5 mM MV. 800  $\mu\text{L}$  of this mixture were injected in a 1 mL quartz cuvette, which was then sealed with a septum to be pressure and airtight. 200  $\mu\text{L}$  of saturated CO solution were added prior to start the reaction. The reaction was initiated by injecting the enzyme using a Hamilton syringe. Specific activities are expressed as  $\mu\text{mol}$  of CO oxidized per min per mg enzyme ( $\text{units}\cdot\text{mg}^{-1}$ ).

### Metal content analysis

Inductively coupled plasma atomic emission spectroscopy (ICP-AES) (Shimadzu ICP 9000 instrument with Mini plasma Torch in axial reading mode) was used to determine the metal content. Samples were incubated in 65%  $\text{HNO}_3$  overnight at  $95^\circ\text{C}$  before measurement. The samples were then diluted to 1%  $\text{HNO}_3$ .

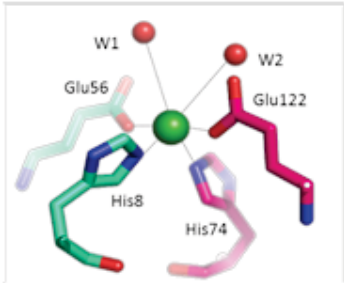
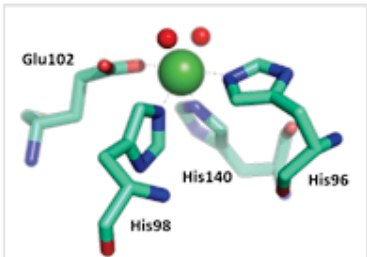
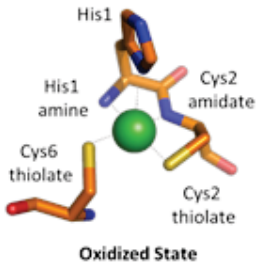
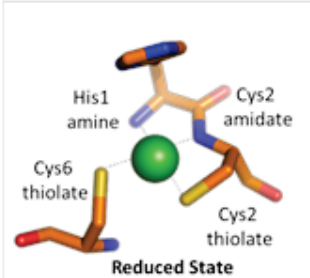
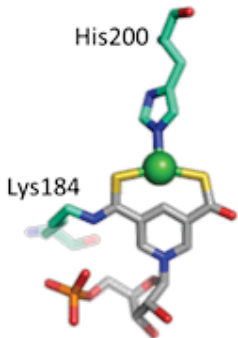
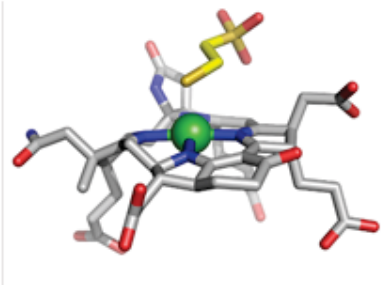
### *In cellulo* activity of RrCODH vs RrCODH-CooJ vs RrCODH-CooJA

Concerning the co-expression experiments with *cooJ* and *cooJA*, the *pet15bcooJ/cooJA* and *prsFRrcooS* plasmids were co-transformed into *E. coli* BL21(DE3). From these transformations, three different colonies were picked for each one and grown in three different falcon supplied with 3 mL LB medium plus the right antibiotics at  $37^\circ\text{C}$  for 8 h under agitation. One mL of these cultures was then used to inoculate the 100 mL O/N pre-culture grown at  $37^\circ\text{C}$  under agitation. 1 mL of these pre-cultures were added to 45 mL of TB medium supplied with 8.7 mL glycerol, 0.1 mM  $\text{Na}_2\text{S}$ , 1-10-100  $\mu\text{M}$   $\text{NiSO}_4$ , 0.1 mM  $\text{FeSO}_4$  and 1% glucose and left at  $37^\circ\text{C}$  under agitation until the optical density reached approx. 0.5. These cultures are done in 50 mL falcon tubes. At this point, all the falcon tubes are sealed and left for 30 minutes on the bench before their induction with 0.2 mM IPTG. Harvesting took place 20 hours after induction. The cultures were centrifuged for 30 min in anaerobic conditions, the supernatant was removed,

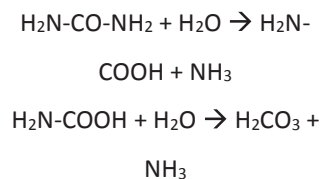
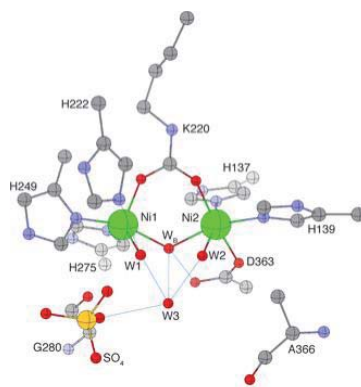
and the cells pellets were fluxed under argon and stored at -80 °C. To check their activity, the pellets are entered in the glovebox and 6 mL of previously degassed buffer (0.1 M MOPS pH 7.5, 200 mM NaCl, 2 mM DTT, 3 Mm DT and a one complete Protease Inhibitor cocktail tablet (Roche) per 50 mL buffer) are added to each. The solubilized cultures are then sonicated for 4 minutes with a cycle 20 sec ON/40 sec OFF 40% amplitude. The activity of the total lysate is checked injecting 50 µL in the activity buffer (0.1 mM MOPS pH 7.5, 200 mM NaCl, 2 mM DTT and 10 mM MV) using the same method reported in point 4.2.3.



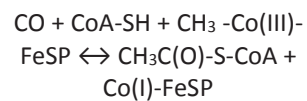
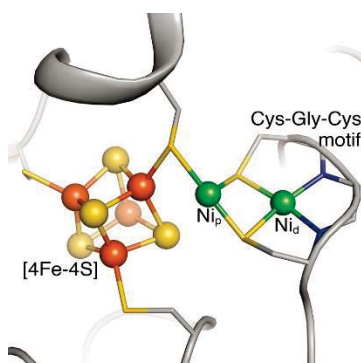
**Table 1 Nickel enzymes comparison**

| Enzyme                            | Active site   | Reaction   |
|-----------------------------------|---|--|
| Glyoxalase I                      |    | $\text{CH}_3\text{-CO-C(OH)-SG} \rightarrow \text{CH}_3\text{-CH(OH)-CO-SG}$   |
| Acireductone Dioxygenase (ARD)    |    | Acireductone $\rightarrow$ 3-methylthiopropionate + CO + Formic acid   |
| Superoxide dismutase (Ni-SOD)     |   | $2 \text{O}_2^{\cdot -} + 2\text{H}^+ \rightarrow \text{H}_2\text{O}_2 + \text{O}_2$   |
| Lactate racemase (LarA)           |    | L-lactate $\leftrightarrow$ D-lactate  |
| Methyl Coenzyme M Reductase (MCR) |    | $\text{CH}_3\text{-S-CoM (methyl-S-coenzyme M)} + \text{CoB-SH (coenzyme B)} \rightarrow \text{CH}_4 + \text{CoB-S-S-CoM (heterodisulfide)}$ |

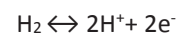
Urease



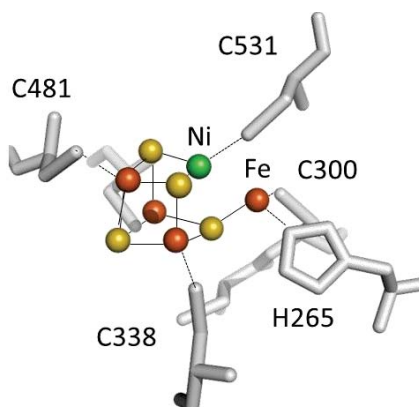
Acetyl-CoA  
synthase (ACS)



[Ni-Fe]  
hydrogenase



Carbon  
Monoxide  
Dehydrogenase  
(CODH)



## Abbreviations

ARD: Acireductone Dioxygenase

CD: Circular Dichroism

CODH: Carbon Monoxide Dehydrogenase

DTH: Sodium hydrosulfite

DTT: Dithiothreitol

EDTA: Ethylenediaminetetraacetic acid

EXAFS: Extended X-ray Absorption Fine Structure

HEPES: 4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid

HRC : Histidine rich cluster

ITC: Isothermal titration calorimetry

LarA: Lactate racemase

LB: Lysogeny broth

MCR: Methyl Coenzyme M Reductase

NMR: Nuclear magnetic resonance

SAXS: Small-angle X-ray scattering

SOD: Superoxide Dismutase

TB: Terrific Broth

TCEP: Tris(2-carboxyethyl)phosphine

TEM: Transition Electron Microscopy

TRIS: Tris(hydroxymethyl)aminomethane

UV/vis : Ultraviolet–visible spectroscopy

WGSR : Water Gas Shift reaction

XANES: X-ray Absorption Near-Edge Structure

XAS: X-ray absorption spectroscopy



# Review



Cite this: *Sustainable Energy Fuels*,  
2018, 2, 1653

Received 23rd February 2018  
Accepted 4th May 2018

DOI: 10.1039/c8se00085a  
rsc.li/sustainable-energy

# The biologically mediated water–gas shift reaction: structure, function and biosynthesis of monofunctional [NiFe]-carbon monoxide dehydrogenases

Marila Alfano and Christine Cavazza \*

Among the possible renewable energies sources, biomass gasification has been considered one promising alternative to meet the daily growing energy demand. The outcome of this process is a synthetic gas (syngas) mainly composed of carbon monoxide and hydrogen. Syngas can be upgraded by a group of anaerobic micro-organisms, thanks to the biologically mediated water–gas shift (WGS) reaction. In this process, the conversion of CO and H<sub>2</sub>O into CO<sub>2</sub> and H<sub>2</sub> is catalyzed by two enzymes: carbon monoxide dehydrogenase (CODH) and hydrogenase. In order to efficiently use micro-organisms as a cost effective and environmentally friendly technology, it is fundamental to deeply understand how this process occurs. In this review paper, an overview on the possible biotechnological uses of the WGS reaction is presented, focusing mainly on the fundamental characterization of the CODH enzyme.

## 1. Introduction

The intense exploitation of fossil fuels, continuous demographic growth, expanding industrialization and living standard improvements require new sustainable strategies to meet the future demand for commodity products and energy sources. For example, the demand for *n*-butanol is expected to reach 5.61 Mton per year by 2030 in the manufacture of pharmaceuticals and plastics.<sup>1</sup> In the energy sector, liquid, gaseous or solid biofuels hold great promise to deliver an increasing share of the energy required to power a new global green economy. Among them, dihydrogen is an attractive fuel since it can be converted to energy without the production of CO<sub>2</sub>. Nowadays, its production comes mainly from fossil fuels, with an annual production of more than 60 million metric tons worldwide.<sup>2</sup> Biohydrogen production processes through an economically acceptable method are therefore promising alternatives for a possible future H<sub>2</sub>-based economy.

Among available renewable sources, biomass conversion from organic waste is becoming increasingly popular. It is also likely to evolve into the main source of primary energy feedstock for developing countries. Besides bioconversion processes to transform biomass into biofuel or different valuable chemicals, an interesting alternative is the thermal conversion (gasification) of poorly degradable biomass sources like straw and wood.<sup>3</sup> Biomass gasification occurs through a combination of complex reactions, including drying, pyrolysis, combustion and

reduction, resulting in the production of synthesis gas (syngas). Syngas is mainly composed of a mixture of H<sub>2</sub>, CO and CO<sub>2</sub> and minor amounts of other gasses. It is an inexpensive and versatile substrate generated from any hydrocarbon feedstock, from fossil fuels to biomass. Therefore, a wider application of routes starting from syngas is more appealing since it enables a gradual transition to more sustainable energy. Biomass gasification to provide motor fuel has been in use since the early 1930s. Over the decades, syngas was successfully used in the USA and Europe for heat, electricity and the synthesis of higher alcohols (Fig. 1). This has also prompted the utilization of new isolates and some known anaerobic microorganisms capable of growing with CO and H<sub>2</sub> as substrates. CO can be metabolized by four main trophic groups of microorganisms: methanogenic archaea, hydrogenogenic bacteria, sulfate-reducing bacteria and acetogenic bacteria.<sup>4</sup> Among the latter group, some *Clostridia* produce, besides acetate, significant amounts of ethanol, butyrate and butanol.<sup>5</sup> Specifically, syngas fermentation to ethanol and isobutanol using *Clostridium ljungdahlii* is exploited on a relatively small scale by companies like Lanza-Tech or INEOS Bio. Hydrogenogenic carboxydotrophs directly produce H<sub>2</sub> from CO, according to the reaction CO + H<sub>2</sub>O → H<sub>2</sub> + CO<sub>2</sub> known as the water–gas shift (WGS) reaction. The conversion of CO into H<sub>2</sub> and CO<sub>2</sub> involves two key enzymes, a [NiFe]-CO dehydrogenase (CODH), which oxidizes CO to CO<sub>2</sub>, and an energy-conserving [NiFe]-hydrogenase, which produces dihydrogen.<sup>6</sup> Nowadays, the water–gas shift reaction plays a significant role in industrial applications to balance H<sub>2</sub>/CO ratios of syngas, via a H<sub>2</sub>-enrichment process.<sup>7</sup> The discovery of bacteria capable of growth employing a biological analogue of

Univ. Grenoble Alpes, CEA, CNRS, BIG, CBM, F-38000 Grenoble, France. E-mail: christine.cavazza@cea.fr

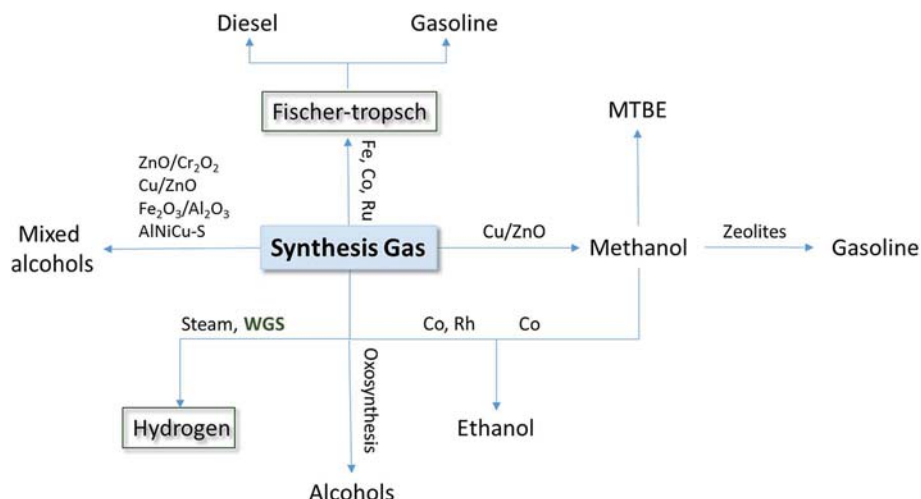


Fig. 1 Fuel products obtained from synthesis gas transformation. The green squares highlight the process described in this review (adapted from ref. 103).

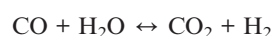
the WGS reaction holds promise for the design of a biological alternative to this chemical material. In this review, we first discuss the industrial applications of the WGS reaction. In the second part, we describe at the molecular level [NiFe]-CODH, a key enzyme of the WGS reaction.

## 2. Industrial applications: water–gas shift reaction & Fischer–Tropsch process

Since the beginning of the 20th century, the development and exploitation of the water–gas shift (WGS) reaction have followed the increasing industrial demands for hydrogen.<sup>8</sup> This has been achieved progressively, evolving the feedstock and the materials used to meet the production requirement for either pure hydrogen or CO to H<sub>2</sub> ratio modulation, the latter being required in the Fischer–Tropsch process or the synthesis of platform molecules such as methanol.<sup>7,9</sup> For this reason, WGS facilities can present major differences in the equipment and materials employed and are strongly bounded to the feedstock used for the syngas production, for both economical and practical purposes. In the next paragraphs the different requirements will be described.<sup>10</sup>

### 2.1 The water–gas shift reaction

The WGS reaction, discovered in 1780 by F. Fontana, is a reversible exothermic reaction, described as:



$$\Delta H_{298}^0 = -41.09 \text{ kJ mol}^{-1}$$

In this reaction, the equilibrium constant decreases with increasing temperature. Carbon monoxide and hydrogen conversions are favoured at low temperature but the reaction is kinetically favoured at high temperature.<sup>9</sup> Concretely, we face

a trade-off between the speed and purity of the products. In this framework, catalysts, specifically their composition and working temperatures, play a predominant role.

When hydrogen production is the main goal, industries traditionally rely on the use of two stage catalytic steps.<sup>11</sup> The first step, carried out at high temperatures (HT), between 400 and 500 °C, uses iron oxide stabilized in chromium oxide as a catalyst (Fe<sub>2</sub>O<sub>3</sub>–Cr<sub>2</sub>O<sub>3</sub>).<sup>12</sup> This high temperature reaction is generally extremely fast, but it has the drawback of being thermodynamically limited, resulting in an incomplete CO conversion of the initial syngas composition (almost 90% conversion). To refine the final product, a slower low temperature (LT) step is required, typically between 200 and 220 °C, in order to move the equilibrium towards the hydrogen product. Copper–zinc oxide supported on alumina (Cu–ZnO/Al<sub>2</sub>O<sub>3</sub>) catalysts,<sup>13</sup> characteristic of this reaction, are able to finally generate the desired product, lowering the CO content to less than 1% in the final mixture.<sup>8,9,14</sup> The combination of two steps is also necessary since LT catalysts can easily suffer from sulphate poisoning, a substance commonly present in pre-high-temperature mixtures.

Currently, the methods commonly available to perform the WGS reaction are distinguished on the basis of the specific type of reactor used: the traditional reactor, the membrane reactor and the photocatalytic reactor. Concerning the traditional reactor, the WGS reaction can be conducted in a homogeneous state (catalyst and reactant are in the same state) or a heterogeneous state (catalyst and reactant in different states).<sup>8</sup> This latter method is by far the most commonly used due to the scarcity of gaseous catalysts. A membrane reactor is able to selectively remove one of the products of a given reaction through a porous membrane.<sup>15,16</sup> This allows for a modulation of the outlet gas composition, achieving a complete CO conversion and H<sub>2</sub>/CO<sub>2</sub> separation. In photo-catalytic reactors, Pt/TiO<sub>2</sub>-based catalysts are irradiated and activated by visible or UV-light.<sup>17,18</sup> Although this technology allows for processes to occur at room temperature, its industrialization has been impeded by its many disadvantages for the commercialization.



Comprehensively, in WGS reaction facilities, their construction and their operation are tightly dependent on the syngas: a more controlled syngas composition, having for example CO and H<sub>2</sub> as predominant gases, leads to strong advantages in the catalytic process. In addition, the syngas composition strongly depends on the feedstock source and gasification process, resulting in a mixture of gases in which the percentage of each component is strongly variable (Table 1). As mentioned above, the presence of contaminant species, such as sulfur and nitrate compounds, in the inlet gas can lead to catalyst poisoning and a decrease in the global process efficiency. Another important limitation for the HT shift reaction is the H<sub>2</sub>/CO ratio, which needs to be relatively high to avoid side reactions, such as the formation of metallic iron, methanation, carbon deposition and Fischer–Tropsch reactions, due to the over reduction of the iron oxide.<sup>9</sup> In some cases the products of these side reactions can be the real objective, conferring to the WGS reaction a contradistinctive role.

## 2.2 The use of the WGS reaction in the Fischer–Tropsch process

The Fischer–Tropsch (FT) process is an assembly of different reactions with the aim of converting a CO and H<sub>2</sub> mixture into liquid hydrocarbons, using the general reactions shown in Table 2. The process uses syngas as the main source converting it catalytically with a particular inorganic catalyst, working in a range of temperature between 150 and 300 °C and depending upon the syngas nature and desired product.<sup>10,19</sup>

In the FT synthesis the overall process from syngas is commonly named XTL (X-to-liquid), with X depending on the carbon source. Four main classes of carbon sources are the most popular: coal (CTL), natural gas (GTL), biomass (BTL) and waste (WST) (Fig. 2).<sup>10</sup>

The synthesis is, therefore, intrinsically different for the specific carbon sources and an appropriate selection of syngas feed conditions is essential to obtain optimal activity, stability and selectivity in industrial reactors. In general, the H<sub>2</sub>/CO ratio most suitable for a FT synthesis is 2 : 1, which changes according to the desired product (Table 2). In a GTL process, the syngas is characterised by a high optimal H<sub>2</sub>/CO (1.9–2.2) whereas in the CTL and BTL processes, the syngas is characterised by a lower H<sub>2</sub>/CO (0.6–1.7), which therefore needs to be adjusted.<sup>10</sup> This highlights one of the problems that the

Table 2 Main reactions used during Fischer–Tropsch process<sup>10</sup>

|                          |  |  | H <sub>2</sub> /CO ratio |
|--------------------------|--|--|--------------------------|
| <b>Main reactions</b>    |  |  |                          |
| Alkanes                  | $n\text{CO} + (2n + 1)\text{H}_2 \rightarrow \text{C}_n\text{H}_{2n+2} + n\text{H}_2\text{O}$        |  | $(2n + 1)/n$             |
| Alkenes                  | $n\text{CO} + 2n\text{H}_2 \rightarrow \text{C}_n\text{H}_{2n} + n\text{H}_2\text{O}$                |  | 2                        |
| Water–gas shift reaction | $\text{CO} + \text{H}_2\text{O} \rightarrow \text{CO}_2 + \text{H}_2$                                |  | —                        |
| <b>Side reactions</b>    |  |  |                          |
| Alcohols                 | $n\text{CO} + 2n\text{H}_2 \rightarrow \text{H}(-\text{CH}_2-)\text{OH} + (n - 1)\text{H}_2\text{O}$ |  | 2                        |
| Boudouard reaction       | $2\text{CO} \rightarrow \text{CO}_2 + \text{C}$  |  | —                        |

industry is currently facing: syngas obtained from wood and straw, which is becoming more and more popular due to the low economic and environmental costs, has a H<sub>2</sub>/CO ratio which is often too low (0.7) for optimal industrial use.<sup>20</sup> This is also related to the most currently used industrial catalysts, iron unsupported metal and cobalt deposited on an inert oxide support, which need a specific H<sub>2</sub>/CO ratio in order to perform efficiently.<sup>7</sup> In recent years, new catalysts have been designed to work with a lower H<sub>2</sub>/CO ratio.<sup>21</sup> However, the practicality of this approach is limited by its different drawbacks, which arise from the materials used (working temperatures, fouling, *etc.*).<sup>21</sup> Today, the main route followed is to exploit the WGS reaction that, used to balance the H<sub>2</sub>/CO ratio, can actively modulate hydrogen concentration at the expense of carbon monoxide reaching the desired feed ratio.<sup>7</sup> The employment of such a strategy on an industrial scale is extremely interesting, and its versatility is expanded by the possibility of microbial processes, where the use of inorganic catalysts is set apart in favour of more bio-compatible alternatives.

## 3. The biologically mediated water–gas shift reaction

An alternative to these inorganic industrial processes is the use of bacteria able to perform the WGS reaction at room temperature and pressure. This biological process has the promise to become a favourable and cost-effective technology for bio-

Table 1 Composition of synthesis gas derived from various carbon sources

| Source                             | Composition (vol%) |       |                 |                |                 |       | Ref. |
|------------------------------------|--------------------|-------|-----------------|----------------|-----------------|-------|------|
|                                    | H <sub>2</sub>     | CO    | CO <sub>2</sub> | N <sub>2</sub> | CH <sub>4</sub> | Other |      |
| Natural gas steam reforming by air | 56–57              | 10–15 | 7–12            | 22–23          | 0.2–0.3         | —     | 22   |
| Water gas                          | 31.7               | 30    | 3.4             | 13.1           | 12.2            | 9.6   | 24   |
| Coal gasification                  | 29.4               | 59.4  | 10              | 0.6            | —               | 0.6   | 24   |
| Coke oven gas                      | 55.4               | 5.6   | 1.4             | 4.3            | 28.4            | 4.9   | 24   |
| Charcoal                           | 29                 | 48    | —               | —              | 48              | —     | 20   |
| Grass straw                        | 2.6                | 12.9  | 17.4            | 64.2           | 2.1             | 0.8   | 37   |
| Demolition wood/paper residue      | 6.1                | 9.2   | 16.1            | 63.2           | 2.8             | 2.6   | 37   |

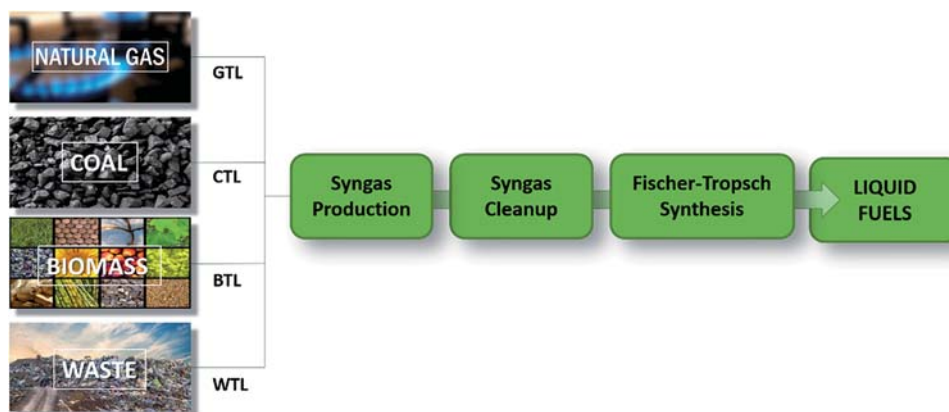
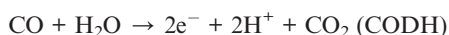


Fig. 2 General scheme for describing the XTL process, starting from different carbon sources. X = gas, coal, biomass, and waste (adapted from ref. 10).

hydrogen production and  $H_2$ -enriched syngas for the Fischer-Tropsch process.<sup>22,23</sup> The key for both these applications is the existence of microorganisms able to convert CO to worthwhile compounds. Among them, anaerobic hydrogenogenic carboxydutrophs use CO as the sole energy source thanks to the biological water-gas shift reaction, fundamental for their growth.<sup>24,25</sup> In the following pages, we will focus on hydrogenogenic CO metabolism.

In the biologically mediated WGS reaction, the overall reaction is catalyzed by two different metalloenzymes, a monofunctional [NiFe]-carbon monoxide dehydrogenase ([NiFe]-CODH) and an energy-conserving [NiFe]-hydrogenase,<sup>6</sup> according to:



CO is oxidized by CODH, releasing electrons used for the reduction of two protons into bio-hydrogen, catalyzed by an energy-conserving hydrogenase (Fig. 3). CO has a strong tendency to inhibit hydrogenase activity by binding to the Ni ion

present in the active site.<sup>23</sup> Therefore, the hydrogenase coupled with CODH needs to present a high CO tolerance in order to avoid being inactivated along the catalytic cycle.

Among hydrogenogenic carboxydutrophs, we find thermophilic and mesophilic species (Table 3). Thermophilic species seem to be more abundant than the mesophilic ones, suggesting that high temperature might help hydrogenogenic CO metabolism, by increasing gas diffusion rates.<sup>25</sup> The physiology and the utilization for biotechnological applications of two model organisms, *Rhodospirillum rubrum* and *Carboxydotherrmus hydrogenoformans*, are detailed below.

### 3.1 *Rhodospirillum rubrum*

*R. rubrum* is a purple non-sulfur photosynthetic mesophilic bacterium.<sup>26</sup> It is a versatile facultative anaerobe, able to obtain energy through a variety of mechanisms. *R. rubrum* is capable of aerobic or anaerobic respiration, photosynthesis and acid-mixed fermentation, depending on the growth conditions.<sup>27</sup> Moreover, it can use CO as the sole energy source during anaerobic growth in the dark, *via* the WGS reaction.<sup>28</sup> It has a higher rate of CO uptake and conversion yield than other similar organisms and can tolerate a small amount of oxygen in the culture medium, making it a particularly attractive strain to investigate for the biohydrogen production from syngas.<sup>28,29</sup>

Focusing on the growth studies made for *R. rubrum* under syngas, it has been demonstrated that mild-light exposure (1000 m cd) has a slight positive effect on the growth rate and hydrogen production.<sup>30</sup> *R. rubrum* also requires an additional carbon source to maintain constant cell growth and efficiently convert CO.<sup>31,32</sup> Najafpour *et al.*<sup>33</sup> investigated hydrogen production using different carbon sources, such as malate, formate and acetate, showing that not all the substrates are suitable for hydrogen production. The maximum CO conversion was reached using acetate, resulting in a maximum hydrogen yield of 98% from a syngas composed of 55% CO, 10% CO<sub>2</sub>, 20% H<sub>2</sub> and 15% Ar.<sup>30</sup>

*R. rubrum* is thus capable to directly convert syngas into biohydrogen catalysing the WGS reaction. As a proof of concept, Younesi *et al.*<sup>34</sup> studied *R. rubrum* growth using acetate under

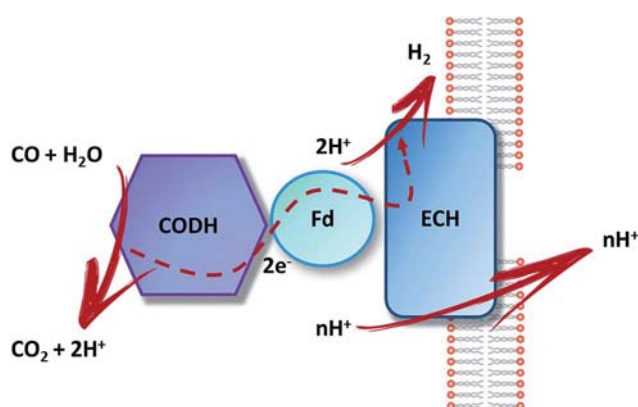


Fig. 3 Schematic representation of CO oxidation, electron transfer,  $H_2$  production and proton translocation in hydrogenogenic carboxydutrophs. CODH: CO dehydrogenase; Fd: ferredoxin (Coof); ECH: energy-conserving hydrogenase.

Table 3 Overview of hydrogenogenic CO converting microorganisms

|                       | Species   | Origin  | Growth conditions |           | Substrate                                  | Product                                     | H <sub>2</sub> Production rate       |         | Ref. |
|-----------------------|---|---|-------------------|-----------|--|---|--------------------------------------|---------|------|
|                       |   |   | pH                | Temp (°C) |  |   | mmol g <sup>-1</sup> h <sup>-1</sup> |         |      |
| Mesophilic bacteria   | <i>Rhodospirillum rubrum</i>                    | Brackish ditch                                | 6.5               | 30        | CO-H <sub>2</sub> /CO <sub>2</sub> acetate | H <sub>2</sub>                              | 16                                   | 28,45   |      |
|                       | <i>Rubrivivax gelatinosus</i>                   | Lake and pond sediments                       | 6.7               | 34        | CO-H <sub>2</sub> /CO <sub>2</sub>         | H <sub>2</sub>                              | 1.3–33                               | 105     |      |
| Thermophilic bacteria | <i>Rhodopseudomonas palustris</i>               | Anaerobic sludge digester                     | n.r.              | 30        | CO-H <sub>2</sub> /CO <sub>2</sub>         | H <sub>2</sub>                              | 41                                   | 106     |      |
|                       | <i>Carboxydotherrnus hydrogeniformans</i>       | Hydrothermal springs on Kunashir island       | 6.9–7.8           | 70        | CO   | H <sub>2</sub>                              | 83.3 <sup>a</sup>                    | 38,42   |      |
|                       | <i>Desulfotomaculum carboxydivorans</i>         | Anaerobic bioreactor sludge                   | 6.8–7             | 55        | CO-H <sub>2</sub> /CO <sub>2</sub>         | H <sub>2</sub>                              | —                                    | 107     |      |
|                       | <i>Carboxydocella pertinax</i>                  | Thermal spring of the volcanic Karymskoe lake | 6–6.5             | 65        | CO   | H <sub>2</sub>                              | —                                    | 108     |      |
|                       | <i>Caldanaerobacter subterraneus pacificus</i>  | Submarine hot vent                            | 6.8–7.1           | 70        | CO-H <sub>2</sub> /CO <sub>2</sub>         | Acetate, CO <sub>2</sub> and H <sub>2</sub> | —                                    | 109     |      |
|                       | <i>Carboxydocella sporoproducens</i>            | Hot spring Karymskoe lake                     | 6.8               | 60        | CO-H <sub>2</sub> /CO <sub>2</sub>         | H <sub>2</sub>                              | —                                    | 110     |      |
| Archaea               | <i>Carboxydocella thermoautotrophica</i>        | Terrestrial hot vent                          | 7                 | 58        | CO-H <sub>2</sub> /CO <sub>2</sub>         | H <sub>2</sub>                              | —                                    | 111     |      |
|                       | <i>Thermoanaerobacter thermohydrosulfuricus</i> | Geothermal spring in Ayas                     | 6.3–6.8           | 70        | CO   | H <sub>2</sub>                              | —                                    | 112     |      |
|                       | <i>Moorella stansii</i>                         | Digester sludge                               | 7.5               | 65        | CO   | CO <sub>2</sub> , H <sub>2</sub>            | —                                    | 113     |      |
|                       | <i>Thermincola carboxydiphila</i>               | Mud, water and cyanobacterial mat             | 7                 | 73        | CO-H <sub>2</sub> /CO <sub>2</sub>         | H <sub>2</sub>                              | —                                    | 114     |      |
|                       | <i>Thermincola ferriacetica</i>                 | Hydrothermal spring                           | 7–7.2             | 60        | CO-H <sub>2</sub> /CO <sub>2</sub>         | H <sub>2</sub>                              | —                                    | 115     |      |
|                       | <i>Calderihabitans maritimus</i>                | Sediment from a marine caldera                | 7–7.5             | 65        | CO   | H <sub>2</sub>                              | —                                    | 116     |      |
|                       | <i>Thermococcus onnurineus</i>                  | Deep-sea hydrothermal vent                    | 6.1–6.2           | 80        | CO   | H <sub>2</sub>                              | 151.3                                | 117,118 |      |
|                       | <i>Thermococcus strain AM4</i>                  | Hydrothermal vent                             | 6.8               | 82        | CO   | H <sub>2</sub> , H <sub>2</sub> S           | —                                    | 119     |      |

<sup>a</sup> (%mol mol<sup>-1</sup> CO).

a continuous syngas flow in a stirred bioreactor. The study demonstrated a hydrogen production rate of 16 mmol per g of substrate per cell per h and a global conversion yield of  $87 \pm 2.4\%$ , confirming the practicality of the WGS reaction on a small scale. The feasibility of this method from a more industrial and economical point of view was described by Dadak *et al.*,<sup>35</sup> who studied the photobiological hydrogen production from syngas in industrial continuous bioreactors. The effect of acetate concentration on CO consumption as well as H<sub>2</sub> and CO<sub>2</sub> production was evaluated, based on both conventional exergy and eco-exergy concepts. This work showed that the eco-exergy concept seems more suitable as a tool for large-scale bio-hydrogen production using living organisms, highlighting the environmental feedback which could be a sufficient incentive for pushing forward this technology. Concerning the production of high value added products, *R. rubrum* has shown the ability to produce biopolymers such as polyhydroxyalkanoates (PHA) *via* syngas fermentation. Revelles *et al.*<sup>36</sup> demonstrated that in the presence of acetate as a carbon source, PHA accumulation in the bacteria was almost 30% of the dried cell weight. As previously described, syngas cannot be used as a sole carbon source for *R. rubrum*. Moreover, the Calvin–Benson–Bassham cycle plays a minor role in carbon assimilation during syngas fermentation. Instead, CO<sub>2</sub> produced from CO *via* CODH is assimilated into biomass and PHA *via* carboxylases such as crotonyl-coA carboxylase or pyruvate synthase. This study shed light on the effective utilization of *R. rubrum* for the production of bioplastics from syngas.

### 3.2 Carboxydotherrmus hydrogenoformans

Considering that biomass gasification results in high temperature gas, strictly anaerobic thermophilic bacteria could be better suited for syngas industrial applications.<sup>37</sup> Among them, *C. hydrogenoformans*, isolated from a hot spring on the volcanic island of Kunashir, is a strictly anaerobic gram-positive thermophilic bacterium capable of CO-growth at an optimum temperature of 70 °C.<sup>38</sup> It grows fast (doubling time 2 h) in the dark using CO as the sole carbon and energy source by catalysing the WGS reaction, making it an extremely interesting candidate to test as a possible bio-hydrogen production source from syngas.<sup>39</sup> However, no data have been published on the ability of this bacterium to convert CO into H<sub>2</sub> from syngas, all the studies considering only the conversion from pure CO gas.<sup>37</sup> Due to the low aqueous solubility of CO at 70 °C, an optimum substrate/biomass ratio must be used to avoid gas–liquid mass transfer limitation.<sup>40</sup> *C. hydrogenoformans* displays a maximal specific activity of 3.5 mol per CO per g VSS (volatile suspended solid) per day at 0.55 mM CO, which compares favourably with those of other hydrogenogenic bacteria. For comparison, the WGS activity of *Rubrivivax gelatinosus* is optimal at a dissolved CO concentration of 90–100 µM and starts to be inhibited at concentrations greater than 150 µM.<sup>40</sup> Considering this remarkable stability, Zhao *et al.*<sup>41</sup> built a working setup using a hollow fibre membrane bioreactor for bio-hydrogen production from CO. *C. hydrogenoformans* was grown for four months with a hydrogen yield of 92%. Despite the high efficiency, major

problems arise from the system used (membrane fouling and low CO transfer rate), which limit the prospective long-term use. Haddad *et al.*<sup>42</sup> obtained a production yield of 95% using a gas-lift reactor with a continuous supply of gas for three months and addition of bacto-peptone in the medium, leading to a remarkable increase of growth, yield and conversion rate regarding CO conversion. However, the cell density remained low, probably due to some deficiency in essential minerals, which restricts the CO feeding-rate and overall CO conversion performance. Another complementary application is the use of a biological WGS reaction to reduce the CO level in outlet gases, which remains the limiting step for some technologies due to catalyst CO poisoning. Henstra *et al.*<sup>43</sup> proved that *C. hydrogenoformans* can reach levels lower than 2 ppm of CO in outlet gas, making the H<sub>2</sub> suitable for CO-sensitive low temperature fuel cells.

Besides the use of bioreactors, an alternative consists of the utilization of the two key enzymes involved in the WGS, *i.e.*, CODH and hydrogenase. Armstrong and collaborators<sup>44</sup> successfully co-attached hydrogenase 2 from *E. coli* and CODH from *C. hydrogenoformans* to graphite platelets to produce a heterogeneous catalyst. They showed that the amount of H<sub>2</sub> produced is equal to the amount of CO consumed and that the enzymes constituting their catalyst react in concert with an efficiency as high as that of the industrial WGS reaction. However, the utilization of this catalyst in applied technology is hindered at the moment because of the difficulty to obtain large amounts of stable active enzymes, which are therefore not competitive with robust synthetic catalysts.

### 3.3 Prospects of the biologically mediated WGS reaction

The production of hydrogen and chemicals from syngas *via* the biological WGS reaction and Fischer–Tropsch process offers different advantages compared to conventional chemical metal-based catalysis.<sup>4</sup> As already mentioned, biological processes occur at ambient temperature and pressure, lowering energy requirements and costs. In contrast to inorganic catalysts, which can easily be irreversibly poisoned or fouled during utilization, living microorganisms can automatically regenerate *via* newly replicated cells, leading to a longer lifetime. Furthermore, biological wastes are considered as by-products that can be treated to produce fuel and/or other organic compounds.<sup>4</sup> Concerning the potential development for the commercialization of this technology, one of the areas of improvement is bioreactor design and utilization,<sup>45</sup> through the optimization of culture media, metabolic or promoter engineering and biodiversity screening. However, several drawbacks need to be overcome for the implement of industrial-scale bioreactors. First, bacterial activity potential is limited by the CO availability to the microorganisms. It is thus necessary to optimize the CO concentration in the feedstock, the gas to liquid mass transfer and the concentration and distribution of bacteria. Important modifications have been done in order to enhance the syngas mass transfer from the gas phase to the liquid phase. Without mass transfer limitation, the biochemical reaction rate should be close to the maximal intrinsic reaction rate. Pressure and mixing conditions also play a role: high



pressure (1 MPa) increases mass transfer rates and thus the rate of CO conversion. Second, CO toxicity towards bacteria is another important limiting step. Indeed, CO is a potent inhibitor of growth and catabolic activity due to its high affinity for metalloenzymes, even if intact cells may offer more resistance than purified CO-sensitive enzymes. A better understanding of syngas metabolism would help the design or the selection of CO-resistant strains.<sup>45</sup> Finally, the scientific community has not yet established appropriate tools to genetically modify thermophilic carboxydrotrophs.<sup>37</sup> Despite the latest advances in bio-hydrogen making, the production is still limited to the laboratory scale and improvements in system productivity and hydrogen yield are needed.

An alternative consists in the “*in vitro*” utilization of the two isolated key enzymes involved in the WGS reaction, *i.e.*, CODH and hydrogenase.

#### 4. CO-induced energy-conserving [NiFe]-hydrogenase

Hydrogenases have been classified into three phylogenetically unrelated classes, based on the metal composition of their active site: [Fe]-, [FeFe]- and [NiFe]-hydrogenases. Among them, the two latter types catalyze the reversible oxidation of molecular hydrogen to protons and electrons.<sup>46</sup> [NiFe]-hydrogenases, found in bacteria and archaea, have been classified in four main groups. Membrane-associated H<sub>2</sub> uptake hydrogenases belong to group 1. Group 2 contains soluble uptake hydrogenases and sensory hydrogenases; group 3 contains heteromultimeric cytoplasmic hydrogenases harbouring a reducible cofactor (F420 or NAD(P)) and group 4, energy-converting hydrogenases.<sup>47</sup> CO-induced hydrogenases involved in the WGS reaction belong to the latter group with the Ech hydrogenase from *Methanosarcina barkeri* and hydrogenase 3 from *E. coli* (Fig. 4). In *E. coli*, hydrogenase 3 is part of the formate hydrogen-lyase (FHL) complex and its expression is formate-induced.<sup>48</sup>

Standard [NiFe]-hydrogenases are heterodimers of 60- and 30-kDa subunits. The dinuclear [NiFe] active site contains additional ligands (1 CO and 2 CN<sup>−</sup>) terminally bound to the iron ion and is buried in the large subunit, while the small subunit harbors three FeS clusters (named proximal, medial and distal clusters) forming the electron transport chain. Depending on the different groups, additional subunits are present.

In hydrogenogenic carboxydrotrophs, the CO-induced hydrogenase is encoded by the *cooMKLXUH* operon, identified by Ludden and collaborators in *R. rubrum*,<sup>49</sup> present upstream the CODH operon. These membrane-bound hydrogenases contain six subunits: two integral membrane-bound proteins, CooM and CooK corresponding to HycC and HycD in *E. coli* hydrogenase 3, and four hydrophilic subunits cooLXUH. CooU in *R. rubrum* has no homologue in *E. coli* hydrogenase 3. CooX has been postulated to be an Fe<sub>4</sub>S<sub>4</sub> cluster-containing ferredoxin-like protein. CooH and CooL, corresponding to HycE and HycG in *E. coli* hydrogenase 3, are related to the large and small subunits of standard [NiFe]-hydrogenases, respectively. However, the small subunit is considerably smaller and contains only the cysteine ligands for the proximal FeS cluster. The two subunits are more closely related to the energy-conserving NADH:quinone oxidoreductase (complex I) found in the respiratory chain in mitochondria and aerobic bacteria, and the sequence similarity is also valid for the hydrophobic subunits and CooU.<sup>49</sup>

Most of the members of this hydrogenase family have been found to be too labile to be purified intact, except the Ech enzyme from *M. barkeri* which was purified to homogeneity in a high yield.<sup>50</sup> However, the characterization of the partly purified *R. rubrum* hydrogenase highlighted its unique resistance towards CO as well as a very high ratio of H<sub>2</sub> evolution to H<sub>2</sub> uptake activity.<sup>51</sup> In 2002 Hedderich and collaborators successfully purified the whole “CO-oxidizing:H<sub>2</sub> evolving” complex from *C. hydrogenoformans*, containing the monofunctional CODH, the ferredoxin CooF and the CO-induced hydrogenase. They showed that the three partners form a tight complex, accessible to further biochemical studies.<sup>52</sup>

#### 5. Monofunctional [NiFe]-CODH: structure and mechanism

Ni-dependent CODHs reversibly catalyze the oxidation of CO to CO<sub>2</sub>, playing a central role in the global carbon cycle in anaerobic microorganisms.<sup>6</sup> CODH can form either a monofunctional enzyme, as in the case of *R. rubrum*, or a bifunctional complex with acetyl-CoA synthase (ACS) such as the one involved in the Wood–Ljungdahl pathway used by acetogens. In this case, the CODH catalyzes the production of CO from CO<sub>2</sub> to form acetyl-CoA at the ACS's active site.<sup>53</sup> The most commonly described CODHs are the CODH/ACS complex from *Moorella thermoacetica* (MtCODH/ACS)<sup>54</sup> and the monofunctional CODHs from

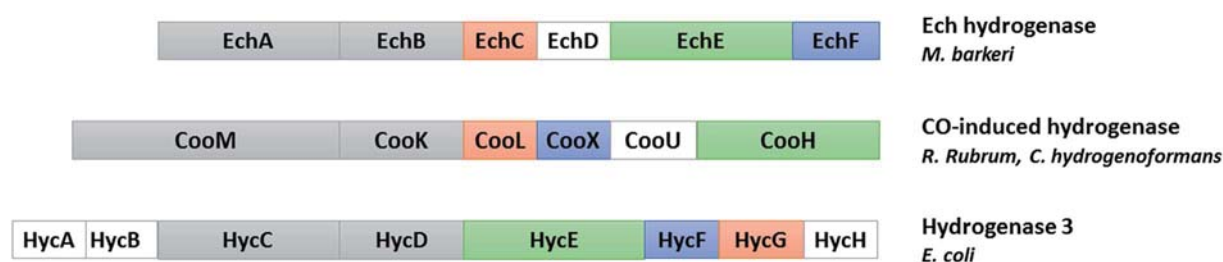


Fig. 4 Comparison of the subunits encoded by the *cooMKLXUH* operon from *R. rubrum* or *C. hydrogenoformans*, the *Ech* operon from *Methanosarcina barkeri* and the *hyc* operon from *E. coli*. Proteins without sequence similarity are shown in white (adapted from ref. 104).

*Rhodospirillum rubrum* and *Carboxydotherrmus hydrogenoformans*. In this chapter, we will focus on monofunctional CODHs. Five genes encoding CODHs have been identified in *C. hydrogenoformans* with different proposed functions, depending on their genomic context: CODH-I, involved in the WGS reaction; CODH-II, linked to NADH generation; CODH-III, coupled to ACS in the acetyl CoA pathway; and CODH-IV supposed to play a role in the oxidative stress response.<sup>55</sup> No function has been found for CODH-V. On the other hand, *R. rubrum* possesses one CODH involved in the WGS reaction.<sup>56</sup> Recently, the monofunctional CODH from *Desulfovibrio vulgaris* (DvCODH) was described, although its physiological function has not been clearly identified.<sup>57</sup>

### 5.1 Biochemical, structural and spectroscopic characterization

Ni-dependent CODHs are extremely oxygen-sensitive and their purification and manipulation require strictly anaerobic conditions in glove boxes. *Rr*CODH was purified for the first time in 1984,<sup>58</sup> followed by the 1st purification of *Ch*CODH-I and *Ch*CODH-II in 2001.<sup>59</sup> *Rr*CODH and *Ch*CODH-II represent today the best-studied monofunctional enzymes. The three enzymes are membrane-associated proteins that became solubilized upon cell disruption. The two *Ch*CODHs are described as homodimeric NiFeS proteins in solution, with one atom of Ni per molecule of monomer. All described monofunctional CODHs share high specific activity for CO oxidation and high affinity for CO (Table 4). Three X-ray structures of monofunctional CODHs are available in the literature. *Rr*CODH and *Ch*CODH-II structures were solved in 2001 at 2.8 Å (ref. 60) and 1.63 Å (ref. 61) resolution, respectively, while the *Ch*CODH-IV structure was determined at 2.52 Å resolution in 2017.<sup>55</sup> The three enzymes display highly similar overall structures (root-mean-square deviation values of 1.16 Å for the *Ch*CODH-IV/*Rr*CODH dimer, considering C $\alpha$ -atoms of 1206 residues, and 0.82 for the *Ch*CODH-II/*Rr*CODH monomer, considering C $\alpha$ -atoms of 602 residues) with two identical intertwined subunits covalently bound (Fig. 5A). Each monomer comprises three domains: an N-terminal helical domain, and a middle and a C-terminal  $\alpha/\beta$  (Rossmann-like) domains (Fig. 5B). Despite gel filtration experiments suggesting that *Rr*CODH is a monomer,<sup>62</sup>

structural analyses evidenced that the enzyme is a homodimer, comparable to *Ch*CODH (Fig. 5C). With an  $\alpha_2$  conformation, monofunctional CODHs are considered as prototypes and good models for  $\beta_2$  CODH subunits of the CODH/ACS ( $\alpha\beta$ )<sub>2</sub> tetramer found in acetogens. In 2001, the two structures revealed the unique nature of the active site (called the C-cluster), constituted of a [NiFe<sub>3</sub>S<sub>4</sub>] cluster, coordinated to a mononuclear Fe(II) site (Fe1) (Fig. 5D). The X-ray structures highlighted that the Ni ion is an integral constituent of the C-cluster. Initially, a [NiFe<sub>3</sub>S<sub>5</sub>] state was observed in the *Ch*CODH-II structure<sup>61</sup> but is no longer thought to be physiologically relevant.<sup>6</sup> Additional Fe<sub>4</sub>S<sub>4</sub> clusters are present to mediate electron transfer, generated during CO oxidation, from the C-cluster to the protein surface and subsequently to its electron transfer partner. In the WGS reaction, the electrons are mediated to a ferredoxin, CooF, which transfers electrons to the [NiFe]-hydrogenase. A total of five FeS clusters per dimer are present in CODH. In addition to the two C-clusters, the dimer harbors two other types of FeS clusters: two [Fe<sub>4</sub>-S<sub>4</sub>] clusters (B-clusters) and a [Fe<sub>4</sub>-S<sub>4</sub>] cluster bridging the two subunits (D-cluster) (Fig. 5C). The B-cluster position is in agreement with a role in electron transfer from the C-cluster to the protein surface, towards CooF. In contrast, the role of the D-cluster has not been well established: this cluster is not reducible at potentials as low as –530 mV and may play a structural or oxygen-protective role.<sup>63</sup>

Before the identification of the C-cluster revealed by the first crystal structures of CODH in 2001, extensive spectroscopic studies have been conducted on this enzyme.<sup>62,64–67</sup> The presence of five metal centers in the protein dimer rendered the EPR analysis not straightforward. In *Rr*CODH, the B-clusters have usual properties of the [4Fe–4S]<sup>2+/1+</sup> cluster with midpoint potential close to –400 mV and  $g_{av} \sim 1.94$ . For the C-cluster, four different oxidation states have been proposed: a catalytically inactive and EPR-silent C<sub>ox</sub> state; a one-electron reduced C<sub>red1</sub> state, which binds CO and has an EPR signal with  $g$ -values at 2.03, 1.88, and 1.71; a two-electron-reduced EPR-silent C<sub>int</sub> state; and a three-electron-reduced form, C<sub>red2</sub>, which binds CO<sub>2</sub> and has a distinct EPR signature with  $g$ -values of 1.97, 1.87, and 1.75.<sup>68</sup> The electronic structure of these redox states is not clear yet; however, the majority of unpaired electron spin density is localized on Fe in both C<sub>red1</sub> and C<sub>red2</sub>, which exhibit large <sup>57</sup>Fe and small <sup>61</sup>Ni hyperfine values.<sup>69</sup>

Table 4 Catalytic properties of monofunctional CODH<sup>a</sup>

|                               | Molecular weight, kDa | Specific activity (CO oxidation), U mg <sup>–1</sup> | $K_m$ (CO), $\mu$ M | $k_{cat}$ (CO), s <sup>–1</sup> | $\frac{k_{cat}}{K_m}$ (CO), M <sup>–1</sup> s <sup>–1</sup> | CO <sub>2</sub> reduction activity, U mg <sup>–1</sup> | Ref. |
|-------------------------------|-----------------------|--|---------------------|---------------------------------|---|--|------|
| <i>Ch</i> CODH-I native       | 62.5                  | 15 756(70 °C), 1300(20 °C)                           | 30(70 °C)           | 39 000(70 °C)                   | $1.3 \times 10^9$ (70 °C)                                   | NR   | 59   |
| <i>Ch</i> CODH-I recombinant  |                       | 8060(70 °C)  | NR                  | NR                              | NR  | NR   | 83   |
| <i>Ch</i> CODH-II native      | 64.5                  | 13 828(70 °C), 1000(20 °C)                           | 18(70 °C)           | 31 000(70 °C)                   | $1.7 \times 10^9$ (70 °C)                                   | NR   | 59   |
| <i>Ch</i> CODH-II recombinant |                       | 9600/12 229(70 °C)                                   | NR                  | NR                              | NR  | 16, 9(25 °C)   | 120  |
| <i>Ch</i> CODH-IV recombinant | 68.2                  | NR   | 0.047(25 °C)        | 14(25 °C)                       | $3 \times 10^8$ (25 °C)                                     | NR   | 55   |
| <i>Rr</i> CODH native         | 63.5                  | 7700(25 °C)  | 32(25 °C)           | 8000                            | $2.5 \times 10^8$   | 30, 1  | 121  |
| <i>Dv</i> CODH recombinant    | 67                    | 1660(37 °C)  | NR                  | NR                              | NR  | 1, 4   | 57   |

<sup>a</sup> NR = non-reported, 1 U: 1  $\mu$ mol of CO oxidized per min.

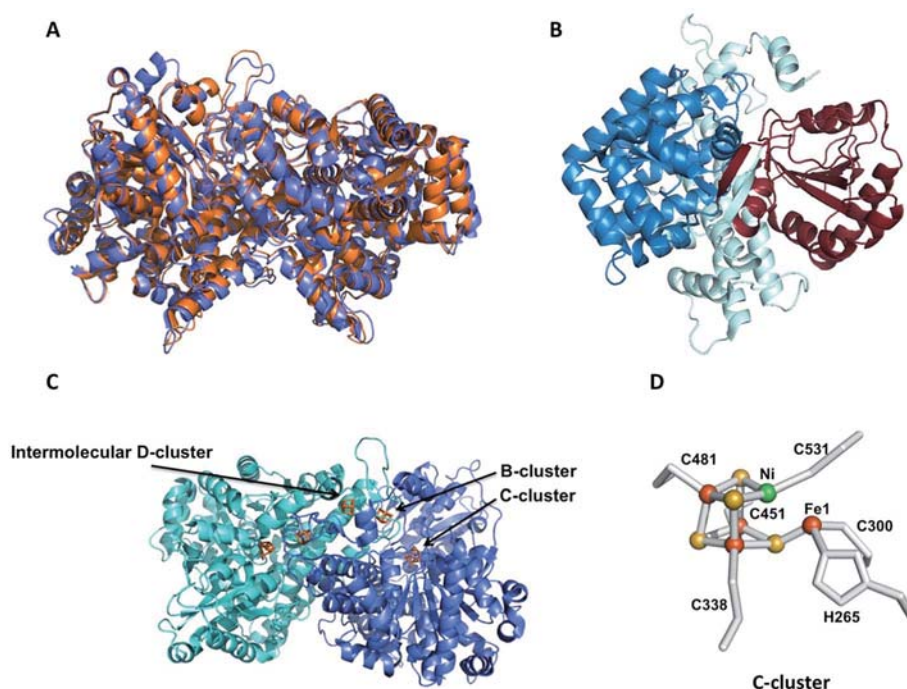


Fig. 5 Crystal structures of monofunctional CODHs. (A) Superposition of the crystal structures of *RrCODH* (PDB code: 1JQK) in orange and *ChCODH-IV* (PDB code: 6ELQ) in blue; (B) ribbon drawing of the *RrCODH* monomer: the N-terminal domain is in pale cyan, and the Rossman domains are in blue and brown; (C) homodimer of *RrCODH*: the two monomers are in blue and cyan, and the  $\text{Fe}_4\text{S}_4$  clusters (1 B-cluster per monomer and 1 intermolecular D-cluster) and C-clusters are depicted in orange; (D) C-cluster: nickel is depicted in green, iron in red and sulfur in yellow.

## 5.2 Reaction mechanism

The reaction mechanism of CODH has been extensively studied in the past decade, thanks to a combination of kinetic, theoretical and spectroscopic approaches. Great support was also given by the determination of several X-ray structures of *ChCODH-II*,<sup>70–73</sup> *MtCODH/ACS*<sup>74</sup> and the CODH component of acetyl-CoA decarbonylase/synthase (ACDS)<sup>75</sup> from *Methanosarcina barkeri* (*MbCODH*), in complex with substrates ( $\text{CO}$  or  $\text{CO}_2$ ) and inhibitors or analogues ( $\text{CN}^-$  and  $\text{NCO}^-$ ). They permitted to propose a basic mechanistic model as follows. Several reaction intermediates have been described, postulating that  $\text{CO}$  binds to nickel and  $\text{H}_2\text{O}$  binds to  $\text{Fe1}$ . After deprotonation of water, a nucleophilic attack of  $\text{OH}^-$  on  $\text{CO}$  gives a  $\text{Fe1-Ni}$ -bridging  $\text{COOH}$  intermediate. A second deprotonation yields an  $\text{Fe1-Ni}$ -bridging  $\text{COO}^-$  species, before the release of  $\text{CO}_2$  and the reduction of the cluster by two electrons.<sup>6</sup> Electrons are then transferred to an external ferredoxin *via*  $\text{Fe}_4\text{S}_4$  clusters. CODHs are mostly inactive at redox potentials higher than  $-300$  mV. The  $\text{C}_{\text{ox}}/\text{C}_{\text{red1}}$  redox couple is  $-200$  mV and the  $\text{C}_{\text{red1}}/\text{C}_{\text{red2}}$  redox couple potential was reported to be  $-530$  mV and matches well with the  $\text{CO}/\text{CO}_2$  redox potential.  $\text{C}_{\text{red1}}$  (one-electron reduced), competent for  $\text{CO}$  oxidation, and  $\text{C}_{\text{red2}}$  (three-electron reduced) for  $\text{CO}_2$  reduction are the two catalytically important oxidation states of the C-cluster (Fig. 6). However, some important mechanistic details remain unclear. The inhibitor study showed that  $\text{CN}^-$  inhibits  $\text{CO}$  oxidation by binding to  $\text{C}_{\text{red1}}$ , but not  $\text{CO}_2$  reduction and that  $\text{NCO}^-$  inhibits  $\text{CO}_2$  reduction by binding to

$\text{C}_{\text{red2}}$ , but not  $\text{CO}$  oxidation. The structure of  $\text{CO}^-$  bound C-cluster showed that it binds Ni with bent geometry, completing a distorted tetrahedral geometry, and that the C atom is too far away from water-bound  $\text{Fe1}$ . Therefore, a “carbon shift” may occur to bring CO close to the  $\text{Fe1}$ -bound hydroxy group to allow the nucleophilic attack step.<sup>6</sup> Moreover, the nature of the  $\text{CO}_2$  ligand bridged between Ni and  $\text{Fe1}$ , a key intermediate of the catalysis has not been yet clearly identified. In 2014, Ragsdale and collaborators published a thorough description of the reaction mechanism of CODH.<sup>68</sup> Therefore, in this review, we detail only recently published data. In 2015, the determination of *ChCODH-II* structures in complex with  $\text{CO}_2$  or its analogue,  $\text{NCO}^-$ , at atomic resolution gave new insights into the reaction mechanism.<sup>71</sup> The improvement of the resolution to the atomic level permitted the description of molecular details of the C-cluster and structural analyses with unbiased determination of geometric parameters. To obtain the crystal structures of reduced  $\text{CO}_2^-$  and  $\text{NCO}^-$ -bound states, *ChCODH-II* crystals were anaerobically poised to a redox potential of  $-600$  mV (equivalent to the  $\text{C}_{\text{red2}}$  state) by 5 mM titanium(III) citrate soaking, in the presence of  $\text{NaHCO}_3$  or  $\text{KNCO}$ , respectively. The  $1.03$  Å structure of  $\text{CO}_2$ -bound and the  $1.09$  Å structure of  $\text{NCO}^-$ -bound *ChCODH-II* showed a similar  $\mu_2$ ,  $\eta^2$  coordination for  $\text{CO}_2$  and  $\text{NCO}^-$  as bent bridging ligands between Ni and  $\text{Fe1}$ .  $\text{CO}_2$  is stabilized in the protein pocket by H bonding with His93 and Lys563. The geometry analysis of  $\text{CO}_2$  binding showed unexpected elongated and identical C–O bond lengths of  $1.32$  Å (bridging O atom) and  $1.30$  Å, while in  $\mu_2$ ,  $\eta^2$



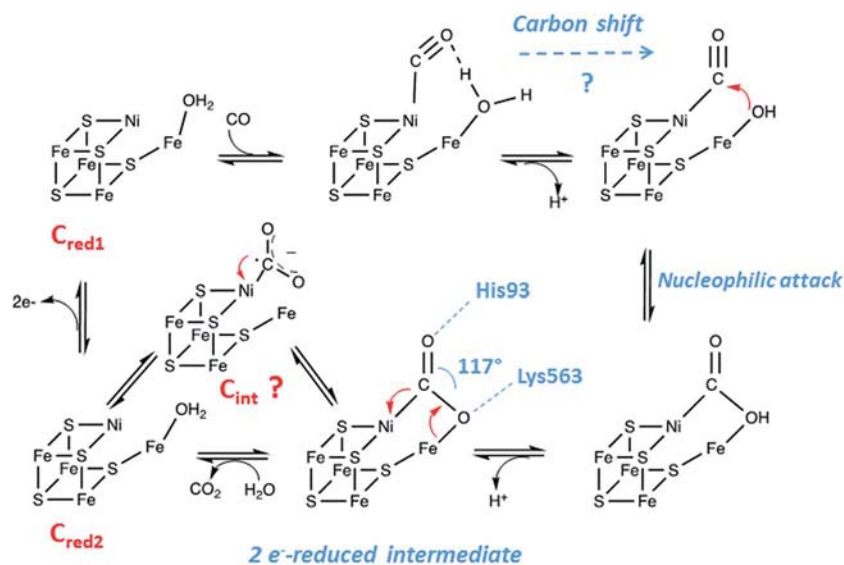


Fig. 6 Proposed reaction mechanism of [NiFe]-CODH (adapted from ref. 6).

complexes, CO<sub>2</sub> has generally one shorter and one longer C–O bond. The major breakthrough in this study was the observation of a change in the position of the C atom involving a change in the O–C–O angle, 15° smaller than the one previously reported at 130°. The 117° angle suggests a more reduced state of CO<sub>2</sub>, corresponding to a two-electron, rather than a one-electron, reduced state. Moreover, a short Ni–C bond of 1.81 Å was observed, indicating an unusual electronic structure, likely due to its integration in an FeS scaffold. It also indicates that the CO<sub>2</sub> ligand is not protonated. The reaction intermediate trapped in the X-ray structure was then proposed to correspond to a two-electron reduced carboxylate, waiting to be protonated. From a bifunctional attack of a nucleophilic Ni center and an electrophilic Fe center, a flow of electrons from the C-cluster to CO<sub>2</sub> occurs, favoring a two-electron transfer process. However, the existence of a transient anionic radical CO<sub>2</sub><sup>•−</sup> intermediate along the catalysis, reported previously, cannot be ruled out (Fig. 6).<sup>76</sup> NCO<sup>−</sup> also undergoes a 2-electron reduction to yield a carbamoyl fragment (H<sub>2</sub>NCO<sup>+</sup>). A shift by 1.36 Å of His93 was observed, compared to free or CO<sub>2</sub>-bound enzymes, giving the required space for the two hydrogen atoms bound to the N atom, suggesting a protonation coupled with reduction process. However, further reaction of the bound carbamoyl did not occur *in crystallo*, while the reduction of cyanate to cyanide by ChCODH-II was evidenced in solution, using IR spectroscopy and isotopically labeled N<sup>13</sup>CO<sup>−</sup>. The same experiment performed on a single ChCODH-II crystal did not yield any signal corresponding to the CN-bound enzyme, showing that the reaction is impaired in the crystalline state.<sup>70</sup> These unprecedented detailed studies are highly challenging in the field of complex multimetallic oxygen-sensitive enzymes and highlight the necessity of combining “in solution” and *in crystallo* approaches to unravel reaction mechanisms.

### 5.3 Oxygen inhibition

Iron and sulfur were used in the formation of FeS clusters within ancient proteins when the environment was anaerobic. Today, most of the FeS-containing proteins are recognized to be oxygen-sensitive by the formation of unstable forms of FeS clusters that quickly decompose upon oxygen exposure. In the case of enzymes containing several multimetallic cofactors, the understanding of their oxygen-sensitivity is often complex which may arise for a number of reasons. For example, B- and D-clusters of [NiFe]-CODHs could be directly oxygen-damaged, but the buried active site is also a potential target, since O<sub>2</sub> can diffuse through the enzyme, *via* internal channels dedicated to substrates and product trafficking.

The nature of the reactivity of DvCODH, ChCODH-II and ChCODH-IV with dioxygen has been investigated by protein film voltammetry (PFV).<sup>55,77</sup> Previous studies have shown that only 10 s-exposure to air is sufficient to inactivate RrCODH and half-lives of 20 min and 60 min were measured for ChCODH-I and ChCODH-II exposed to air, respectively.<sup>59</sup> The most O<sub>2</sub>-resistant enzyme described is ChCODH-IV that retains between 25 and 50% of its original activity after air-exposure. However, while DvCODH and ChCODH-II can be reactivated upon reduction, ChCODH-IV is irreversibly inactivated. Although the mechanism of reactivity of the enzyme towards dioxygen is unknown, three distinct inactive species have been identified thanks to PFV experiments by using a range of O<sub>2</sub> concentration at different potentials: one that reactivates spontaneously once oxygen is removed, one that reactivates at reductive potential and one that is irreversibly inactivated.<sup>77</sup> Thus, the sensitivity of CODH to O<sub>2</sub> is not common to all enzymes: DvCODH reacts more slowly with O<sub>2</sub> than ChCODH-II and can be fully reactivated by reduction, unlike ChCODH-II that is partly reactivated and ChCODH-IV that is not reactivated. Moreover, the use of an inhibitor in CO oxidation, which binds the active site, slows

down the inhibitory effect of O<sub>2</sub>, demonstrating that the C-cluster is directly attacked.

## 6. CODH's active site biosynthesis

Understanding the biosynthesis of complex enzyme-bound heteronuclear metallocusters, such as those found in nitrogenases, hydrogenases, ACS or CODH, evolved as one of the most challenging topics in the field of metalloenzymology. Given their complexity, the assembly and insertion of the active sites into the target enzyme require specific and elaborated maturation machineries. In the case of CODH, the identification of the dedicated accessory proteins, the characterization of the protein complexes involved in metal insertion and the sequence of the different events leading to the final active enzyme remain unclear. However, the development of the high-level production of fully active CODH for biotechnological applications requires a better understanding of the biosynthesis of the C-cluster.

### 6.1 *In vivo* activation of CODH

Maintaining the appropriate intracellular concentration of essential metals while avoiding toxicity caused by overload is absolutely critical for cells. If the concentration becomes too low, the cell will suffer from the inactivation of essential metalloenzymes. However, non-physiological high intracellular concentrations of metals can either replace the native metals in essential enzymes, causing their inactivation, or catalyze the formation of highly toxic reactive oxygen species. Transporters, storage proteins and metalloregulators control the intracellular availability of nickel ions and protect cells against free toxic nickel. They work in concert with accessory proteins and metallochaperones that guide nickel to their specific target sites into enzymes. The study of nickel enzyme maturation systems, such as those of urease and hydrogenase, showed that a “nickel delivery complex”, involving the cooperation of several nickel-chaperones, is responsible for nickel insertion into the enzyme's active site.<sup>78</sup> In *R. rubrum*, an operon named *cooFSCTJ* has been described downstream the hydrogenase operon. The *coo* operon contains five genes encoding the ferredoxin CooF (mentioned above), CooS, the structural gene of CODH, and three additional nickel proteins, CooC, CooJ and CooT.<sup>79</sup> Both operons are regulated by a CO-sensing transcriptional activator, CooA, present downstream of the *coo* operon.<sup>80</sup> CooA co-occurs with CODH operons in most of the CO-oxidizing bacteria. This homodimeric heme protein is a member of the cAMP receptor protein (CRP) family. CooA consists of two heme containing regulatory domains where CO binds and a DNA-binding domain.<sup>81</sup> CO binding induces an allosteric conformational change allowing binding to specific target regions resulting in positive gene regulation. Phylogenetic analyses showed that CooC is widely distributed in microorganisms possessing a [NiFe]-CODH, generally in operons with CooS. In contrast, CooT and CooJ are more difficult to identify in genomes. When present, CooC has been shown to be essential for CODH maturation in the absence of high nickel concentration. For example, in *R. rubrum*, strains containing mutations in the *cooC*

gene displayed minimal CODH activity and required an approximately 1000-fold higher extracellular Ni(II) concentration than the wild-type to restore anaerobic growth under CO and a partial CODH activity.<sup>82</sup> Although not essential, CooJ and CooT are nevertheless relevant for nickel delivery to CODH since strains bearing polar insertions in the *cooJ* gene and polar or non-polar insertions in the *cooT* gene required 10-fold-higher Ni levels to grow, resulting in a reduced CODH activity. The activation of recombinant ChCODH-II or ChCODH-IV heterologously produced in *E. coli* does not require CooC: the supply of an excess of Ni(II) (500 μM) in the growth medium is sufficient to correctly activate CODH (Table 4).<sup>55,73</sup> However, it does not exclude the involvement of accessory proteins under physiological growth conditions, since the high nickel concentration can compensate for the lack of accessory proteins, as previously shown for other Ni-enzymes. In contrast, when ChCODH-I was heterologously produced in *Escherichia coli*, its co-production with CooC3 is required and leads to an approximately 4-fold increase in CO oxidation activity. However, with a specific activity of 8060 U mg<sup>-1</sup>, this value is approximately half the one described for native ChCODH-I which is equal to 15 756 U mg<sup>-1</sup> (Table 4).<sup>83</sup> This could suggest the requirement of other accessory proteins to produce a fully active enzyme. The role of CooC in Ni insertion was also demonstrated in DvCODH, heterologously produced in *Desulfovibrio fructosovorans*.<sup>57</sup> Interestingly, despite similar overall structures, active sites and FeS-cluster organization, the maturation pathways of the different monofunctional CODHs diverge at the molecular level. The respective role of CooC, CooT and CooJ in bacteria needs to be determined (Fig. 7A). Concerning Fe and S insertion into the C-cluster, the co-production of ChCODH-I, II and IV with ISC machinery<sup>84</sup> in *E. coli* leads to an active CODH with the expected 20 Fe and 20 S per dimer. No accessory protein dedicated to the building of the FeS unit of the C-cluster has been identified yet and the classical machinery involved in FeS biogenesis (ISC and/or SUF) seems to participate in the biosynthesis and insertion of the active site (Fig. 7A).

### 6.2 Maturation process

Previous studies showed that the “Fe/S unit” is inserted prior to nickel. Indeed, when *RrCODH* is produced in *R. rubrum* in nickel-depleted media, the purified enzyme corresponds to a nickel-free inactive form.<sup>65</sup> EPR spectroscopy showed that the redox properties of B-clusters are similar in holo- and nickel-deficient enzymes.<sup>66</sup> As expected, the presence of nickel dramatically impacts the midpoint potential of the C-cluster. Mössbauer studies showed that the mononuclear Fe site (Fe1) is not observed if nickel is not present and the spectroscopic properties of the Ni-free C-cluster would be consistent with those of tetracoordinated iron atoms. The addition of nickel salts in the presence of CO under reductive conditions to nickel-deficient CODH is sufficient to convert the Ni-free C-cluster into an active C-cluster. Therefore, nickel incorporation into the C-cluster likely alters the coordination environment of one iron atom. Site-directed mutagenesis of residues coordinating the C-cluster highlighted that the only critical residue required for Ni

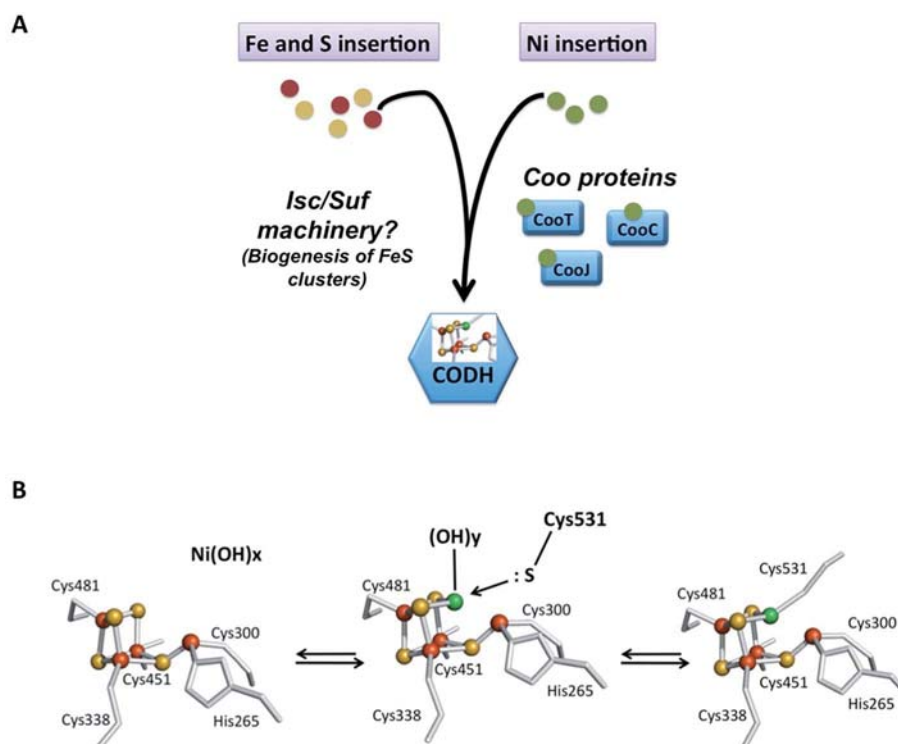


Fig. 7 Metal insertion into CODH. (A) Schematic representation of Fe, S and Ni insertion and metallochaperones; (B) proposed model for the mechanism of Ni insertion into Ni free-CODH in *R. rubrum*, adapted from ref. 85.

binding is His265 since the His265Val variant contains 0.24 Ni atoms per monomer.<sup>67</sup> Surprisingly, all cysteine mutants, even the Cys531Ala variant affecting the residue directly involved in Ni coordination, contain one Ni per monomer.<sup>85</sup> Therefore, cysteine ligands are crucial for CODH activity, but not for C-cluster formation. A model for nickel insertion into the C-cluster has been proposed in the literature but has not been experimentally validated yet (Fig. 7B). This model raised several questions: is an open distorted Fe<sub>4</sub>S<sub>4</sub> cluster (with one iron atom linked with only one inorganic sulfur atom) a real intermediate of the maturation process? How does His265 direct the assembly of the C-cluster? If the present model is wrong with a classical cubic Fe<sub>4</sub>S<sub>4</sub> cluster, nickel atom insertion implicates bond breaking. Whatever the mechanism, nickel insertion is the key step in CODH activation although actual data are insufficient to unravel the molecular mechanism of C-cluster formation, and further deep studies are required.

## 7. Nickel chaperones involved in CODH maturation

### 7.1 CooC

CooC proteins belong to the SIMIBI-class of NTPases, which also contains UreG<sup>86</sup> and HypB,<sup>87</sup> involved in urease and [NiFe]-hydrogenase maturation, respectively. Today, NTPases are recognized as common components of maturation pathways of nickel enzymes.<sup>78</sup> They potentially play a regulatory role by affecting the interaction of partners in protein complexes or

nickel ion affinity, as recently observed in archaea, where the ATPase HypB is described as a metallochaperone enhancer, increasing the affinity of HypA for Ni(II) from 4 μM to 7 nM.<sup>88</sup> The CooC family can be divided in two main groups: the “CooC-type” cluster, regrouping proteins involved in monofunctional CODH maturation and the “AcsF-type” cluster, postulated to be involved in either ACS/CODH or CODH activation.<sup>89</sup> Three *cooC* genes were identified in *C. hydrogenoformans*: *cooC1*, *cooC2* and *cooC3*. CooC1 and CooC2 are located in the same gene cluster as CODH-III and ACS, while CooC3 is co-expressed with CODH-I. Recently, CooC2 was purified and characterized as an AcsF-type enzyme, with a function associated with the maturation of ACS.<sup>89</sup> CooC1 is a nickel-binding enzyme with a stoichiometry of one Ni(II) per dimer, coordinated by two cysteines per dimer within a conserved CXC motif present in all CooC homologues (Fig. 8A). Several X-ray structures of CooC1 were described in the presence or absence of ADP and/or Zn(II) (Fig. 8B).<sup>90</sup> Apo-CooC1 is a monomer and dimerizes upon ADP and metal binding. The metal-binding site is located at the dimer interface (Fig. 8C). As observed for HypB, an intricate link between nucleotide and metal binding, metal-induced dimerization of the protein, and the presence of a mixture of monomeric and dimeric states of the NDP-bound form were observed. The only other CooC described in the literature is CooC from *R. rubrum* (*RrCooC*). *RrCooC* was purified directly from *R. rubrum* cultures.<sup>91</sup> The protein is a homodimer in solution and was reported to associate with the membrane. Thus, its purification required a first step of solubilization from the membrane. However, *RrCooC* is not an integral protein and once released, the protein does not



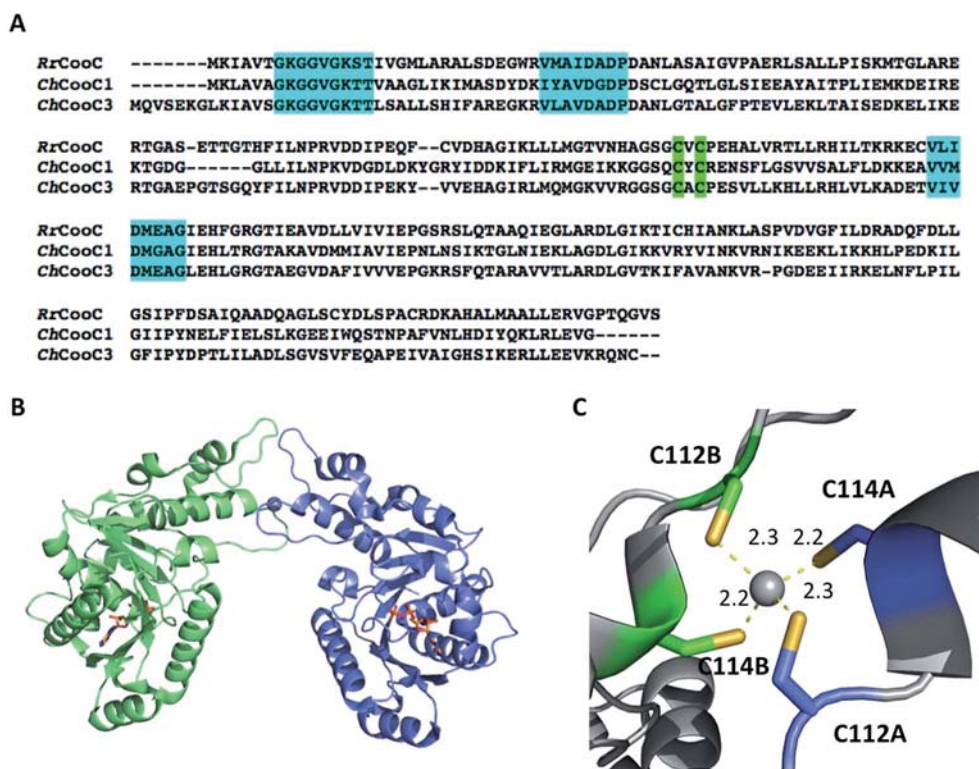


Fig. 8 The NTPase CooC. (A) Sequence alignment of *RrCooC* (accession number: P31897), *ChCooC1* (accession number: Q3ACS5) and *ChCooC3* (accession number: Q3AB30). The nucleotide-binding site is highlighted in cyan, and the metal-binding site is highlighted in green; (B) overall structure of the *ChCoo1* homodimer (PDB code: 3KJ1). The two monomers are in green and blue, ADP molecules are depicted as orange sticks and Zn is depicted as a grey sphere; (C) metal-binding site at the dimer interface.

require the presence of detergent to be stable. The protein is an ATPase and its function is directly linked to nickel insertion into CODH.

## 7.2 CooJ

CooJ is a metallochaperone with only a very few biochemical data and no structural information available.<sup>92</sup> This 12.5 kDa-protein contains a C-terminal nickel-binding domain with 16 histidines and 2 cysteines in the final 34 amino acids (Fig. 9). The protein was reported to bind four Ni(II) per monomer with a  $K_D$  value of 4.3  $\mu$ M, as shown by equilibrium dialysis. When CooJ was purified directly from *R. rubrum* cultures, the protein co-eluted with CODH, as shown by immunoblotting with an anti-CODH antibody, in association with other proteins, suggesting the formation of a multiprotein complex bound to CODH. Interestingly, Kerby *et al.* were able to show the *RrCooJ*'s nickel-binding domain to be partially dispensable for its physiological function: a truncated version of CooJ with only 6 histidines and no cysteine in the C-terminal part is still able to activate CODH *in vivo*, suggesting that the His-tail could play another role in the bacterium, in addition to CODH maturation. Although not widely distributed, CooJ possesses common features with other nickel chaperones involved in Ni-enzyme maturation (Fig. 9). Some HypB and UreE<sup>93</sup> (urease maturation) homologues possess a histidine-rich motif proposed to be involved in nickel storage, conferring them a second function in

addition to their role in enzyme maturation. As for CooJ, the role of histidines in SlyD<sup>94</sup> (hydrogenase maturation) and LarC<sup>95</sup> (lactate racemase maturation) remains to be determined. The two unique proteins Hpn and Hpn-2 that contribute to nickel storage are also related to urease maturation in *Helicobacter pylori*. Both required and essential for *H. pylori* colonization, phylogenetic analyses revealed that they are restricted to gastric strains of *Helicobacter*.<sup>96</sup>

## 7.3 CooT

*RrCooT* is a homodimeric all- $\beta$  protein of 14 kDa with seven strands forming two bent anti-parallel  $\beta$ -sheets made of  $\beta_6$ ,  $\beta_7$ ,  $\beta_1$  and  $\beta_2$  ( $\beta$ -sheet I), and  $\beta_3$ ,  $\beta_4$  and  $\beta_5$  ( $\beta$ -sheet II), respectively, tightly packed on top of each other (Fig. 10A).<sup>97</sup> The protein binds specifically one Ni(II) per dimer with a high affinity in the nanomolar range. In its amino acid sequence, two potential Ni(II)-binding residues are present, a cysteine in position 2 and a histidine in position 41 (Fig. 10B). Site-directed mutagenesis showed that the replacement of cysteine 2 by serine prevents Ni(II)-binding to CooT. Concerning Histidine 41, the X-ray structure of apo-*RrCooT* revealed that the residue is too far away from cysteine 2 to contribute together to a common nickel-binding motif (Fig. 10A). Moreover a conformational change in the  $\beta$ -sheet *RrCooT*'s structure upon Ni(II) binding, to bring His41 close to Cys2, is highly unlikely. In turn, the two cysteines A and B from each monomer are very close to each other,

**RrCooJ**

MTESPERGRKRLGIYLAHFLDHVEGHMGEIGVQORDALAEDARLGALIDRALADMAVARASLN  
 AVLRLDLDGEAPAPASPEAVHSPFSHAHSHDHDHAHGHSHDHAHDHCHCHDHP

**RrHypB**

MCETCGCAGTGPHHVHGDHAHSHSHGDGHTHDHAHDHGHGHSHDHDHDHGHGHSHSHDHDHG  
 HSHSHGSASEGTRTVIVLEDLLAKNDHQAHVRAHFDARGILAVNLMSSPGSGKTSLEATI  
 QALPAGVRVAVIEGDLTENDAERIRRHGVPVAVQITTGTACHLDAHMIHDALHHLDLGDIDI  
 VFIENVGNLVCATFDIGQHRNVLLSVTEGDDKPAKYPVIMRAADRVLITKADLLPHIEEF  
 DVERARASIAGVAGPVPVVCVSSKRGPMDDWIGWLMAEHAARKPAVA

**KaUreE**

MLYLTQRLEIPAAATASVTLPIDVRVKSrvKVTLNDRDAGLLPRGLLLRGGDVLSNEE  
 GTEFVQVIAADEEVSVVRCDDPFMLAKACYHLGNRHVPLQIMPGLRYHHDHVLDDMLRQ  
 FGLTVTFGQLPFEPEAGAYASESHGHHHAHHDHHAHSH

**EcSlyD**

MKVAKDLVVSLEYQVTEDEGLVDESPVSAPLDYLHGHGSLISGLETALEGHEVGDKFDVAV  
 GANDAYGQYDENLVQRPKDVFMGVDELQVGMFLAETDQGPVPEITAVEDDHVVVDGNHM  
 LAGQNLKFNVEVVAIREATEEELAHGHVHGAHDHHDHHDHDCCGGHGHDHGHEHGGECCG  
 KKGNGGCGCH

**LfLarC**

MAYLYLDPFSGLSGNMMLGVFLDLGLDFDQFAELEKLNVSgyHLTLEKTTSAIGGQLFGV  
 VLDDAHHHADESLETAQEVEAHDHHDHDDQLAEHDHHDHDLAEHDHHDHDLAATLHHHHHHH  
 GRNLQEIETIINNSNLSQSVKIQALGVFNEIAQAEAHVHQLSLDEIHFHEVGALDSIVDIVG  
 FFGLEMMGIDGIIISGTLVDGSGTIEVAHGTMPVPPVAVMKMRQNSEVPVRQRTDIHTELVT  
 PTGFAIAKLAFDSFGPIPDMDQVQKVGyGFGTRDTGHLNGLRGLLLTPMASQKEVHSTSDEI  
 VEIHANVDDQSGEGLGYAIERLMTAGAYDAYFTPIFMKKNRPAYQVTVITTSKQLAMVDLL  
 MTHTTTTFGVRWANMKRATLPRTFKTVTTAFGEVTLKIGTFNGQKQVTVEYDSAAALADSHHV  
 SLDAVRRALLAYDKEDGMN

**HpHpn**

MAHHEEQHGGHHHHHHHTHHHHYHGGEHHHHHSSHHEEGCCSTSDSHHQEEGCCHGHHE

**HpHpn-2**

MAHHEQQQQAQQQQQQQANSQHSHHHHAHHHHYHGGEHHHHNAEQHAEQQAQQQAQQQAHQ  
 QQQQKAQQQNQQY

Fig. 9 Histidine-rich proteins related to Ni-enzyme maturation. Histidines and cysteines are depicted in green and blue, respectively. Histidine-rich motifs are underlined. UniprotKB accession numbers: RrCooJ: P72321, RrHypB: Q2RXN8, UreE from *Klebsiella aerogenes*: P18317, SlyD from *Escherichia coli*: P0A9K9, LarC from *Lactobacillus fermentum*: V4XLJ5, Hpn from *Helicobacter pylori*: P0A0V6, and Hpn-2 from *Helicobacter pylori*: A3RDS2.

suggesting the presence of a nickel-binding site formed by the two N-termini of the homodimer. In addition to cysteine 2, the amino terminal and/or the backbone amide group of close residues, such as Met3, are the putative ligands to complete the coordination of Ni(II) in a square-planar geometry, as shown by spectroscopic characterization.

Initially, CooT was only identified in *R. rubrum* and *Rhodospirillum rubrum*. However, the analysis of 2775 complete prokaryotic proteomes performed in 2017 allowed the identification of 111 CooT homologues, mainly present in

anaerobic bacteria, but also in anaerobic archaea such as methanogens and hyperthermophilic species.<sup>97</sup> In 83 proteomes, CooT co-occurs with CooS, suggesting that CooT should be considered as a potential actor with CooC in the maturation pathway of monofunctional CODHs. Moreover, it is worth mentioning that in the remaining 28 proteomes, CooT co-occurs with CooC and ACS, mainly found in methanogens (22 proteomes), or with CooC and ACS/CODH in *Thermodesulfobacterium yellowstonii* and *Butyrivibrio proteoclasticus*. In the four remaining proteomes, the role of CooT needs to be elucidated

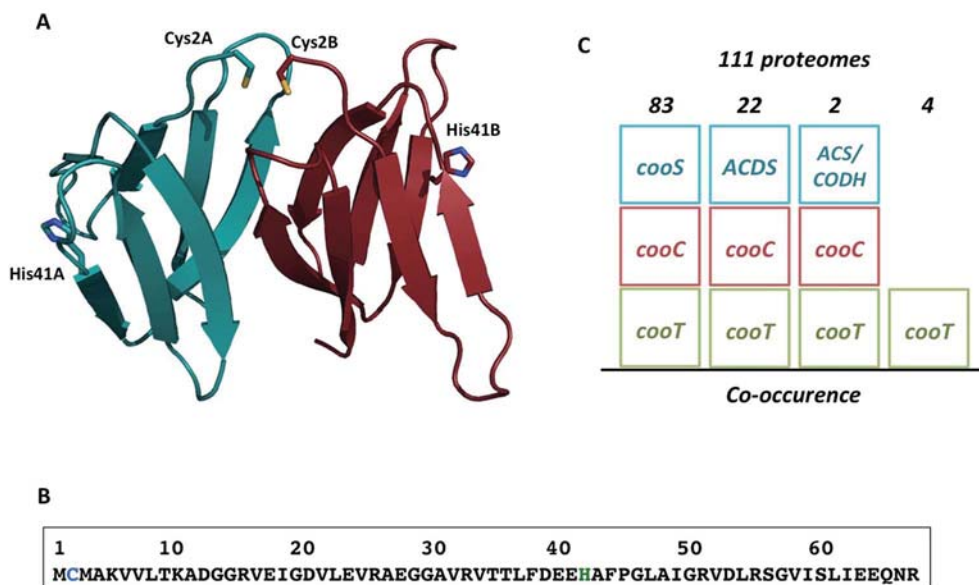


Fig. 10 The nickel-binding protein CooT. (A) Crystal structure of the *RrCooT* homodimer; (B) amino-acid sequence of *RrCooT*; (C) occurrence of CooT in anaerobic archaea and bacteria.

(Fig. 10C). In any case, the presence of CooT is related to an anaerobic life style and to nickel metabolism. Based on their amino acid sequences, all CooT homologues harbor a single conserved RNA\_bind\_2 domain, predicting a role in RNA metabolism. However, electrophoretic mobility shift assay experiments with nucleic acids purified from *R. rubrum* showed that *RrCooT* could bind to neither RNA nor DNA, refuting this hypothesis.

In all CooT homologues, cysteine 2, a key residue in Ni(II)-binding, is strictly conserved, highlighting the existence of a novel Ni(II)-binding protein family linked to anaerobic metabolism in bacteria and archaea.

## 8. Summary and prospects

The development of next generation biofuel and bio-based chemical technologies, respecting green chemistry's principles, is a real challenge. In this field, synthetic biology has the potential to shift biology from a research field centered on the observation of nature to a discipline that highlights the design and construction of novel biological systems and organisms.

In this review, we focused on the biologically mediated WGS reaction, involving two amazing enzymes, CODH and hydrogenase, containing remarkable active sites. These enzymes are involved in energy conversion processes and are often the most highly active and selective catalysts operating under mild conditions with small molecules that are known to be difficult to activate (such as H<sub>2</sub> or CO<sub>2</sub>). Interest in hydrogenases and CODH has dramatically increased in recent years due to the hydrogen fuel initiatives and the development of new methods for CO<sub>2</sub> sequestration. In the field of hydrogenases, these concerns led to the profusion in the literature of deep studies on the deciphering of their reaction mechanism, the synthesis of bio-inspired catalysts, and the assembly of their active site in

the protein scaffold. The combination of all these approaches has been really successful in recent years with the development of new bio-inspired processes, including combining accessory proteins with active site synthetic analogues.<sup>98</sup>

In the case of CODH, the mechanism, the structure and the biosynthesis of its active site, unique in biological systems, intrigue both chemists and biologists. The major difficulties are related to the complexity of its multimetallic cofactor, and the attempts to synthesize bio-inspired models of CODH have failed. However, since CODH catalyzes reversibly the oxidation of CO, the full understanding of the C-cluster synthetic pathway would open new prospects for the synthesis of metal catalysts for both syngas upgrading and CO<sub>2</sub> valorization. Today, the most active catalysts described are based on palladium phosphine complexes<sup>99,100</sup> and there is still a need to explore bio-inspired metallic catalysts to efficiently and economically reduce CO<sub>2</sub>.

Besides the use of inorganic catalysts, the optimization of microorganisms capable of performing the WGS reaction and the engineering of CODH to enhance its activity will represent a great opportunity for biotechnological applications. However, bacteria that naturally produce CODH require complex growth conditions (temperature, strict anaerobic conditions, high CO concentrations, defined growth media, etc), limiting their use in bioreactors on an industrial scale. On the other hand, heterologous production of complex metalloenzymes is often limited because of the requirement of specific maturation machinery. However, recent advances highlighted the possibility to produce several fully active CODH in the easy-to-grow bacterium *Escherichia coli*. *E. coli* is a robust microorganism,<sup>101</sup> able to grow easily and quickly, under aerobic or anaerobic conditions (using aerobic/anaerobic respirations or fermentation<sup>102</sup>), with a wide variety of energy sources, and is widely used in modern biotechnology. Therefore, the use of optimized genetically



modified *E. coli* strains to perform a biological WGS reaction becomes worth considering.

To conclude, the combination of inorganic chemistry, biochemistry and microbiology would have great impacts on the design of new methods to optimize biological WGS reaction efficiency and its integration into biomass gasification devices.

## Conflicts of interest

There are no conflicts to declare.

## References

- H. T. Luk, C. Mondelli, D. C. Ferré, J. A. Stewart and J. Pérez-Ramírez, *Chem. Soc. Rev.*, 2017, **46**, 1358–1426.
- D. Stefan, M. Michael, J. Holger, H. Marta, B. Andreas, P. Regina and W. Peter, *ChemSusChem*, 2017, **10**, 42–47.
- O. Tirado-Acevedo, M. S. Chinn and A. M. Grunden, in *Advances in Applied Microbiology*, Academic Press, 2010, vol. 70, pp. 57–92.
- A. M. Henstra, J. Sipma, A. Rinzema and A. J. Stams, *Curr. Opin. Biotechnol.*, 2007, **18**, 200–206.
- B. Lagoa-Costa, H. N. Abubakar, M. Fernández-Romasanta, C. Kennes and M. C. Veiga, *Bioresour. Technol.*, 2017, **239**, 244–249.
- Y. Kung and C. L. Drennan, *Curr. Opin. Chem. Biol.*, 2011, **15**, 276–283.
- D. B. Bukur, B. Todir and N. Elbashir, *Catal. Today*, 2016, **275**, 66–75.
- G. K. Reddy and P. G. Smirniotis, *Water Gas Shift Reaction*, Elsevier, 2015, vol. 83.
- B. Smith RJ, M. Loganathan and M. S. Shantha, *Int. J. Chem. React. Eng.*, 2010, **8**, 85201.
- P. M. Maitlis and A. de Klerk, *Greener Fischer-Tropsch Processes for Fuels and Feedstocks*, Wiley-VCH Verlag GmbH & Co. KGaA, Weinheim, Germany, 2013.
- C. Ratnasamy and J. P. Wagner, *Catal. Rev.*, 2009, **51**, 325–440.
- G. C. Chinchén, R. H. Logan and M. S. Spencer, *Appl. Catal.*, 1984, **12**, 69–88.
- J. L. Ayastuy, M. A. Gutiérrez-Ortiz, J. A. González-Marcos, A. Aranzabal and J. R. González-Velasco, *Ind. Eng. Chem. Res.*, 2005, **44**, 41–50.
- D. C. Grenoble, M. M. Estadt and D. F. Ollis, *J. Catal.*, 1981, **67**, 90–102.
- D. Mendes, A. Mendes, L. M. Madeira, A. Iulianelli, J. M. Sousa and A. Basile, *Asia-Pac. J. Chem. Eng.*, 2010, **5**, 111–137.
- R. Koc, N. K. Kazantzis and Y. H. Ma, *J. Loss Prev. Process Ind.*, 2011, **24**, 852–869.
- S. Sato and J. M. White, *J. Am. Chem. Soc.*, 1980, **102**, 7206–7210.
- C. Yixuan, W. Zhaobin, C. Yanxin, L. Huaxin, H. Zupei, L. Huiqing, D. Yonglei, Y. Chuying and L. Wenzhao, *J. Mol. Catal.*, 1983, **21**, 275–289.
- D. Selvatico, A. Lanzini and M. Santarelli, *Fuel*, 2016, **186**, 544–560.
- S. K. Sansaniwal, K. Pal, M. A. Rosen and S. K. Tyagi, *Renewable Sustainable Energy Rev.*, 2017, **72**, 363–384.
- Y. Lu, Q. Yan, J. Han, B. Cao, J. Street and F. Yu, *Fuel*, 2017, **193**, 369–384.
- O. Bičáková and P. Straka, *Int. J. Hydrogen Energy*, 2012, **37**, 11563–11578.
- S. K. M. R. Rittmann, H. S. Lee, J. K. Lim, T. W. Kim, J. H. Lee and S. G. Kang, *Biotechnol. Adv.*, 2015, **33**, 165–177.
- J. Sipma, A. M. Henstra, S. M. Parshina, P. N. Lens, G. Lettinga and A. J. M. Stams, *Crit. Rev. Biotechnol.*, 2006, **26**, 41–65.
- M. Diender, A. J. M. Stams and D. Z. Sousa, *Front. Microbiol.*, 2015, **6**, 1–18.
- R. L. Kerby, S. S. Hong, S. A. Ensign, L. J. Coppoc, P. W. Ludden and G. P. Roberts, *J. Bacteriol.*, 1992, **174**, 5284–5294.
- J. E. Schultz and P. F. Weaver, *J. Bacteriol.*, 1982, **149**, 181–190.
- R. L. Kerby, P. W. Ludden and G. P. Roberts, *J. Bacteriol.*, 1995, **177**, 2241–2244.
- H. Zhu, T. Wakayama, Y. Asada and J. Miyake, *Int. J. Hydrogen Energy*, 2001, **26**, 1149–1154.
- G. D. Najafpour and H. Younesi, *World J. Microbiol. Biotechnol.*, 2007, **23**, 275–284.
- K. T. Klasson, K. M. O. Lundbäck, E. C. Clausen and J. L. Gaddy, *J. Biotechnol.*, 1993, **29**, 177–188.
- G. Najafpour, H. Younesi and A. R. Mohamed, *Biochem. Eng. J.*, 2004, **21**, 123–130.
- G. Najafpour, H. Younesi and A. R. Mohamed, *Energy Sources, Part A*, 2006, **28**, 1013–1026.
- H. Younesi, G. Najafpour, K. S. Ku Ismail, A. R. Mohamed and A. H. Kamaruddin, *Bioresour. Technol.*, 2008, **99**, 2612–2619.
- A. H. Y. Ali Dadak, M. Aghbashlo, M. Tabatabaei and G. Najafpour, *Environ. Sci. Technol.*, 2014, **33**, 482–489.
- O. Revelles, N. Tarazona, J. L. García and M. A. Prieto, *Environ. Microbiol.*, 2016, **18**, 708–720.
- S. M. Tiquia-Arashiro, *Thermophilic Carboxydrotrophs and Their Applications in Biotechnology*, 2014, vol. 10.
- V. A. Svetlichny, T. G. Sokolova, M. Gerhardt, M. Ringpfeil, N. A. Kostrikina and G. A. Zavarzin, *Syst. Appl. Microbiol.*, 1991, **14**, 254–260.
- M. Wu, Q. Ren, A. S. Durkin, S. C. Daugherty, L. M. Brinkac, R. J. Dodson, R. Madupu, S. A. Sullivan, J. F. Kolonay, D. H. Haft, W. C. Nelson, L. J. Tallon, K. M. Jones, L. E. Ulrich, J. M. Gonzalez, I. B. Zhulin, F. T. Robb and J. A. Eisen, *PLoS Genet.*, 2005, **1**, e65.
- Y. Zhao, R. Cimpola, Z. Liu and S. R. Guiot, *Appl. Microbiol. Biotechnol.*, 2011, **91**, 1677–1684.
- S. R. G. Ya Zhao, M. Haddad, R. Cimpola and Z. Liu, *Int. J. Hydrogen Energy*, 2013, **38**, 2167–2175.
- M. Haddad, R. Cimpola and S. R. Guiot, *Int. J. Hydrogen Energy*, 2014, **39**, 2543–2548.
- A. M. Henstra and A. J. M. Stams, *Int. J. Microbiol.*, 2011, **2011**, 1–4.
- O. Lazarus, T. W. Woolerton, A. Parkin, M. J. Lukey, E. Reisner, J. Seravalli, E. Pierce, S. W. Ragsdale,



- F. Sargent and F. A. Armstrong, *J. Am. Chem. Soc.*, 2009, **131**, 14154–14155.
- 45 K. S. K. Ismail, G. Najafpour, H. Younesi, A. R. Mohamed and A. H. Kamaruddin, *Biochem. Eng. J.*, 2008, **39**, 468–477.
- 46 J. C. Fontecilla-Camps, P. Amara, C. Cavazza, Y. Nicolet and A. Volbeda, *Nature*, 2009, **460**, 814.
- 47 W. Lubitz, H. Ogata, O. Rüdiger and E. Reijerse, *Chem. Rev.*, 2014, **114**, 4081–4148.
- 48 J. S. McDowall, B. J. Murphy, M. Haumann, T. Palmer, F. A. Armstrong and F. Sargent, *Proc. Natl. Acad. Sci. U. S. A.*, 2014, **111**, E3948–E3956.
- 49 J. D. Fox, H. E. Yiping, D. Shelper, G. P. Roberts and P. W. Ludden, *J. Bacteriol.*, 1996, **178**, 6200–6208.
- 50 M. Jörn, B. Stefan, K. Jürgen, K. Andreas and H. Reiner, *Eur. J. Biochem.*, 2001, **265**, 325–335.
- 51 J. D. Fox, R. L. Kerby, G. P. Roberts and P. W. Ludden, *J. Bacteriol.*, 1996, **178**, 1515–1524.
- 52 B. Soboh, D. Linder and R. Hedderich, *Eur. J. Biochem.*, 2002, **269**, 5712–5721.
- 53 S. W. Ragsdale, L. Yi, G. Bender, N. Gupta, Y. Kung, L. Yan, T. A. Stich, T. Doukov, L. Leichert, P. M. Jenkins, C. M. Bianchetti, S. J. George, S. P. Cramer, R. D. Britt, U. Jakob, J. R. Martens, G. N. Phillips and C. L. Drennan, *Biochem. Soc. Trans.*, 2012, **40**, 501–507.
- 54 T. I. Doukov, T. M. Iverson, J. Seravalli, S. W. Ragsdale and C. L. Drennan, *Science*, 2002, **298**, 567–572.
- 55 L. Domnik, M. Merrouch, S. Goetzl, J. H. Jeoung, C. Léger, S. Dementin, V. Fourmond and H. Dobbek, *Angew. Chem., Int. Ed.*, 2017, **56**, 15466–15469.
- 56 D. Bonam, L. Lehman, G. P. Roberts and P. W. Ludden, *J. Bacteriol.*, 1989, **171**, 3102–3107.
- 57 J. Hadj-Saïd, M. E. Pandelia, C. Léger, V. Fourmond and S. Dementin, *Biochim. Biophys. Acta, Bioenerg.*, 2015, **1847**, 1574–1583.
- 58 D. Bonam, S. A. Murrell and P. W. Ludden, *J. Bacteriol.*, 1984, **159**, 693–699.
- 59 V. Svetlitchnyi, C. Peschel, G. Acker and O. Meyer, *J. Bacteriol.*, 2001, **183**, 5134–5144.
- 60 C. L. Drennan, J. Heo, M. D. Sintchak, E. Schreiter and P. W. Ludden, *Proc. Natl. Acad. Sci. U. S. A.*, 2001, **98**, 11973–11978.
- 61 H. Dobbek, V. Svetlitchnyi, L. Gremer, R. Huber and O. Meyer, *Science*, 2001, **293**, 1281–1285.
- 62 D. Bonam and P. W. Ludden, *J. Biol. Chem.*, 1987, **262**, 2980–2987.
- 63 J. Feng and P. A. Lindahl, *Biochemistry*, 2004, **43**, 1552–1559.
- 64 P. Stephens, M. McKenna, S. A. Ensign, D. Bonam and P. W. Ludden, *J. Biol. Chem.*, 1989, **264**, 16347–16350.
- 65 D. Bonam, M. C. McKennat, P. J. Stephenst and P. W. Ludden, *Biochemistry*, 1988, **85**, 31–35.
- 66 N. J. Spangler, P. A. Lindahl, V. Bandarian and P. W. Ludden, *J. Biol. Chem.*, 1996, **271**, 7973–7977.
- 67 N. J. Spangler, M. R. Meyers, K. L. Gierke, R. L. Kerby, G. P. Roberts and P. W. Ludden, *J. Biol. Chem.*, 1998, **273**, 4059–4064.
- 68 M. Can, F. A. Armstrong and S. W. Ragsdale, *Chem. Rev.*, 2014, **114**, 4149–4174.
- 69 P. A. Lindahl, S. W. Ragsdale and E. Munck, *J. Biol. Chem.*, 1990, **265**, 3880–3888.
- 70 A. Ciaccafava, D. Tombolelli, L. Domnik, J. H. Jeoung, H. Dobbek, M. A. Mroginiski, I. Zebger and P. Hildebrandt, *Angew. Chem., Int. Ed.*, 2017, **56**, 7398–7401.
- 71 J. Fessler, J. H. Jeoung and H. Dobbek, *Angew. Chem., Int. Ed.*, 2015, **54**, 8560–8564.
- 72 J. H. Jeoung and H. Dobbek, *J. Am. Chem. Soc.*, 2009, **131**, 9922–9923.
- 73 J.-H. Jeoung and H. Dobbek, *Science*, 2007, **318**, 1461–1464.
- 74 Y. Kung, T. I. Doukov, J. Seravalli, S. W. Ragsdale and C. L. Drennan, *Biochemistry*, 2009, **48**, 7432–7440.
- 75 W. Gong, B. Hao, Z. Wei, D. J. Ferguson, T. Tallant, J. A. Krzycki and M. K. Chan, *Proc. Natl. Acad. Sci. U. S. A.*, 2008, **105**, 9558–9563.
- 76 M. W. Ribbe, *Angew. Chem., Int. Ed.*, 2015, **54**, 8337–8339.
- 77 M. Merrouch, J. Hadj-Saïd, L. Domnik, H. Dobbek, C. Léger, S. Dementin and V. Fourmond, *Chem.-Eur. J.*, 2015, **21**, 18934–18938.
- 78 C. J. Zeer-Wanklyn and D. B. Zamble, *Curr. Opin. Chem. Biol.*, 2017, **37**, 80–88.
- 79 R. L. Kerby and P. W. Ludden, *J. Bacteriol.*, 1997, **179**, 2259–2266.
- 80 Y. He, D. Shelper, R. L. Kerby and G. P. Roberts, *J. Biol. Chem.*, 1996, **271**, 120–123.
- 81 G. P. Roberts, R. L. Kerby, H. Youn and M. Conrad, *J. Inorg. Biochem.*, 2005, **99**, 280–292.
- 82 W. B. Jeon, J. Cheng and P. W. Ludden, *J. Biol. Chem.*, 2001, **276**, 38602–38609.
- 83 T. Inoue, K. Takao, F. Yuto, T. Yoshida and Y. Sako, *Biosci., Biotechnol., Biochem.*, 2014, **78**, 582–587.
- 84 B. Roche, L. Aussel, B. Ezraty, P. Mandin, B. Py and F. Barras, *Biochim. Biophys. Acta, Bioenerg.*, 2013, **1827**, 923–937.
- 85 W. B. Jeon, S. W. Singer, P. W. Ludden and L. M. Rubio, *J. Biol. Inorg. Chem.*, 2005, **10**, 903–912.
- 86 B. Zambelli, M. Stola, F. Musiani, K. De Vriendt, B. Samyn, B. Devreese, J. Van Beeumen, P. Turano, A. Dikiy, D. A. Bryant and S. Ciurli, *J. Biol. Chem.*, 2005, **280**, 4684–4695.
- 87 A. M. Sydor, H. Lebrette, R. Ariyakumaran, C. Cavazza and D. B. Zamble, *J. Biol. Chem.*, 2014, **289**, 3828–3841.
- 88 S. Watanabe, T. Kawashima, Y. Nishitani, T. Kanai, T. Wada, K. Inaba, H. Atomii, T. Imanaka and K. Miki, *Proc. Natl. Acad. Sci. U. S. A.*, 2015, **112**, 7701–7706.
- 89 C. M. Gregg, S. Goetzl, J. H. Jeoung and H. Dobbek, *J. Biol. Chem.*, 2016, **291**, 18129–18138.
- 90 J. H. Jeoung, T. Giese, M. Grünwald and H. Dobbek, *J. Mol. Biol.*, 2010, **396**, 1165–1179.
- 91 W. B. Jeon, J. Cheng and P. W. Ludden, *J. Biol. Chem.*, 2001, **276**, 38602–38609.
- 92 R. K. Watt and P. W. Ludden, *J. Biol. Chem.*, 1998, **273**, 10019–10025.
- 93 S. B. Mulrooney, S. K. Ward and R. P. Hausinger, *J. Bacteriol.*, 2005, **187**, 3581–3585.

- 94 H. Kaluarachchi, J. W. Zhang and D. B. Zamble, *Biochemistry*, 2011, **50**, 10761–10763.
- 95 B. Desguin, P. Soumillion, P. Hols and R. P. Hausinger, *Proc. Natl. Acad. Sci. U. S. A.*, 2016, **113**, 5598–5603.
- 96 D. Vinella, F. Fischer, E. Vorontsov, J. Gallaud, C. Malosse, V. Michel, C. Cavazza, M. Robbe-Saule, P. Richaud, J. Chamot-Rooke, C. Brochier-Armanet and H. De Reuse, *PLoS Pathog.*, 2015, **11**, e1005312.
- 97 J. Timm, C. Brochier-Armanet, J. Perard, B. Zambelli, S. Ollagnier-de-Choudens, S. Ciurli and C. Cavazza, *Metallomics*, 2017, **9**, 575–583.
- 98 G. Berggren, A. Adamska, C. Lambertz, T. R. Simmons, J. Esselborn, M. Atta, S. Gambarelli, J. M. Mouesca, E. Reijerse, W. Lubitz, T. Happe, V. Artero and M. Fontecave, *Nature*, 2013, **499**, 66–69.
- 99 A. Miedaner, C. J. Curtis, R. M. Barkley and D. L. DuBois, *Inorg. Chem.*, 1994, **33**, 5482–5490.
- 100 J. W. Raebiger, J. W. Turner, B. C. Noll, C. J. Curtis, A. Miedaner, B. Cox and D. L. DuBois, *Organometallics*, 2006, **25**, 3345–3351.
- 101 P. Wang, L. Robert, J. Pelletier, W. L. Dang, F. Taddei, A. Wright and S. Jun, *Curr. Biol.*, 2010, **20**, 1099–1103.
- 102 G. Uden and J. Bongaerts, *Biochim. Biophys. Acta, Bioenerg.*, 1997, **1320**, 217–234.
- 103 P. L. Spath and D. C. Dayton, *Preliminary Screening – Technical and Economic Assessment of Synthesis Gas to Fuels and Chemicals with Emphasis on the Potential for Biomass-Derived Syngas*, Golden, CO, United States, 2003.
- 104 A. Kunkel, J. A. Vorholt, R. K. Thauer and R. Hedderich, *Eur. J. Biochem.*, 1998, **252**, 467–476.
- 105 E. J. Wolfrum and A. S. Watt, *Appl. Biochem. Biotechnol.*, 2002, **98–100**, 611–625.
- 106 Y.-K. Oh, Y.-J. Kim, J.-Y. Park, T. H. Lee, M.-S. Kim and S. Park, *Biotechnol. Bioprocess Eng.*, 2005, **10**, 270.
- 107 S. N. Parshina, J. Sipma, Y. Nakashimada, A. M. Henstra, H. Smidt, A. M. Lysenko, P. N. L. Lens, G. Lettinga and A. J. M. Stams, 2018, 2159–2165.
- 108 Y. Yoneda, T. Yoshida, S. Kawaichi, T. Daifuku, K. Takabe and Y. Sako, *Int. J. Syst. Evol. Microbiol.*, 2012, **62**, 1692–1697.
- 109 T. G. Sokolova, J. M. González, N. A. Kostrikina, N. A. Chernyh, T. P. Tourova, C. Kato, E. A. Bonch-Osmolovskaya and F. T. Robb, *Int. J. Syst. Evol. Microbiol.*, 2001, **51**, 141–149.
- 110 T. V. Slepova, T. G. Sokolova, A. M. Lysenko, T. P. Tourova, T. V. Kolganova, O. V. Kamzolnikina, G. A. Karpov and E. A. Bonch-Osmolovskaya, *Int. J. Syst. Evol. Microbiol.*, 2006, **56**, 797–800.
- 111 T. G. Sokolova, N. A. Kostrikina, N. A. Chernyh, T. P. Tourova, T. V. Kolganova and E. A. Bonch-Osmolovskaya, *Int. J. Syst. Evol. Microbiol.*, 2002, **52**, 1961–1967.
- 112 M. Balk, H. G. H. J. Heilig, M. H. A. van Eekert, A. J. M. Stams, I. C. Rijpstra, J. S. Sinninghe-Damsté, W. M. de Vos and S. W. M. Kengen, *Extremophiles*, 2009, **13**, 885–894.
- 113 J. I. Alves, A. H. van Gelder, M. M. Alves, D. Z. Sousa and C. M. Plugge, *Int. J. Syst. Evol. Microbiol.*, 2013, **63**, 4072–4076.
- 114 T. G. Sokolova, N. A. Kostrikina, N. A. Chernyh, T. V. Kolganova, T. P. Tourova and E. A. Bonch-Osmolovskaya, *Int. J. Syst. Evol. Microbiol.*, 2005, **55**, 2069–2073.
- 115 D. G. Zavarzina, T. G. Sokolova, T. P. Tourova, N. A. Chernyh, N. A. Kostrikina and E. A. Bonch-Osmolovskaya, *Extremophiles*, 2007, **11**, 1–7.
- 116 Y. Yoneda, T. Yoshida, H. Yasuda, C. Imada and Y. Sako, *Int. J. Syst. Evol. Microbiol.*, 2013, **63**, 3602–3608.
- 117 G. J. Schut, G. L. Lipscomb, D. M. N. Nguyen, R. M. Kelly and M. W. W. Adams, *Front. Microbiol.*, 2016, **7**, 1–9.
- 118 M.-S. Kim, S. S. Bae, Y. J. Kim, T. W. Kim, J. K. Lim, S. H. Lee, A. R. Choi, J. H. Jeon, J.-H. Lee, H. S. Lee and S. G. Kang, *Appl. Environ. Microbiol.*, 2013, **79**, 2048–2053.
- 119 T. G. Sokolova, C. Jeanthon, N. A. Kostrikina, N. A. Chernyh, A. V. Lebedinsky, E. Stackebrandt and E. A. Bonch-Osmolovskaya, *Extremophiles*, 2004, **8**, 317–323.
- 120 T. Inoue, T. Yoshida, K. Wada, T. Daifuku, K. Fukuyama and Y. Sako, *Biosci., Biotechnol., Biochem.*, 2011, **75**, 1392–1394.
- 121 S. A. Ensign and P. W. Ludden, *J. Biol. Chem.*, 1991, **266**, 18395–18403.

## Reference

1. Kamber, B. S. *et al.* Oceanic nickel depletion and a methanogen famine before the Great Oxidation Event. *Nature* **458**, 750–753 (2009).
2. Roles, T. U. *Nickel and Its Surprising Impact in Nature. Proteins* (John Wiley & Sons, Ltd, 2007). doi:10.1002/9780470028131
3. Zambelli, B., Uversky, V. N. & Ciurli, S. Nickel impact on human health: An intrinsic disorder perspective. *Biochim. Biophys. Acta - Proteins Proteomics* **1864**, 1714–1731 (2016).
4. Raymond, J. The Effect of Oxygen on Biochemical Networks and the Evolution of Complex Life. *Science*. **311**, 1764–1767 (2006).
5. Falkowski, P. G. EVOLUTION: Tracing Oxygen’s Imprint on Earth’s Metabolic Evolution. *Science*. **311**, 1724–1725 (2006).
6. Wang, S.-J., Rudnick, R. L., Gaschnig, R. M., Wang, H. & Wasylenki, L. E. Methanogenesis sustained by sulfide weathering during the Great Oxidation Event. *Nat. Geosci.* (2019). doi:10.1038/s41561-019-0320-z
7. Fontecilla-Camps, J. C., Amara, P., Cavazza, C., Nicolet, Y. & Volbeda, A. Structure–function relationships of anaerobic gas-processing metalloenzymes. *Nature* **460**, 814 (2009).
8. Khorasani-Motlagh, M., Lacasse, M. J. & Zamble, D. B. High-affinity metal binding by the *Escherichia coli* [NiFe]-hydrogenase accessory protein HypB is selectively modulated by SlyD. *Metallomics* **9**, 482–493 (2017).
9. Boer, J. L., Mulrooney, S. B. & Hausinger, R. P. Nickel-dependent metalloenzymes. *Arch. Biochem. Biophys.* **544**, 142–152 (2014).
10. Macomber, L. & Hausinger, R. P. Mechanisms of nickel toxicity in microorganisms. *Metallomics* **3**, 1153–1162 (2011).
11. Hausinger, R. P. *Biochemistry of Nickel*. (1993). doi: 10.1007/978-1-4757-9435-9\_2
12. Li, Y. & Zamble, D. B. Nickel Homeostasis and Nickel Regulation: An Overview. *Chem. Rev.* **109**, 4617–4643 (2009).
13. Zamble, D., Rowińska-Żyrek, M. & Kozłowski, H. *The Biological Chemistry of Nickel*. (2017).

14. James Sumner, B. B. *THE ISOLATION AND CRYSTALLIZATION OF THE ENZYME UREASE*. (1926)
15. Nicholas E. Dixon, Carlo Gazzola, James J. Watters Robert L. Blakeley, B. Z. A Metalloenzyme. A Simple Biological Role for Nickel? *J. Am. Chem. Soc.* / **97**, (1975).
16. Kappaun, K., Piovesan, A. R., Carlini, C. R. & Ligabue-Braun, R. Ureases: Historical aspects, catalytic, and non-catalytic properties – A review. *J. Adv. Res.* **13**, 3–17 (2018).
17. Benini, S. *et al.* The complex of *Bacillus pasteurii* urease with acetohydroxamate anion from X-ray data at 1.55 Å resolution. *JBIC* **5**, (2000).
18. Benini, S., Cianci, M., Mazzei, L. & Ciurli, S. Fluoride inhibition of *Sporosarcina pasteurii* urease: Structure and thermodynamics. *J. Biol. Inorg. Chem.* **19**, 1243–1261 (2014).
19. He, M. M., Clugston, S. L., Honek, J. F. & Matthews, B. W. Determination of the structure of *Escherichia coli* glyoxalase I suggests a structural basis for differential metal activation. *Biochemistry* **39**, 8719–8727 (2000).
20. Dai, Y., Wensink, P. C. & Abeles, R. H. One protein, two enzymes. *J. Biol. Chem.* **274**, 1193–5 (1999).
21. Deshpande, A. R., Wagenpfeil, K., Pochapsky, T. C., Petsko, G. A. & Ringe, D. Metal-Dependent Function of a Mammalian Acireductone Dioxygenase. *Biochemistry* **55**, 1398–1407 (2016).
22. Pochapsky, T. C., Pochapsky, S. S., Ju, T., Hoefler, C. & Liang, J. A refined model for the structure of acireductone dioxygenase from *Klebsiella* ATCC 8724 incorporating residual dipolar couplings. *J. Biomol. NMR* **34**, 117–27 (2006).
23. Hausinger, R. P. New metal cofactors and recent metallocofactor insights. *Curr. Opin. Struct. Biol.* **59**, 1–8 (2019).
24. Miłaczewska, A. *et al.* On the Structure and Reaction Mechanism of Human Acireductone Dioxygenase. *Chem. - A Eur. J.* **24**, 5225–5237 (2018).
25. Desguin, B. *et al.* Lactate racemase is a nickel-dependent enzyme activated by a widespread maturation system. *Nat. Commun.* **5**, 3615 (2014).
26. Desguin, B. Lactate Racemization and Beyond An Intriguing Cofactor in Lactate Racemase. (2018). doi:10.4172/2155-9597.1000335
27. Desguin, B., Soumillon, P., Hols, P. & Hausinger, R. P. Nickel-pincer cofactor biosynthesis

- involves LarB-catalyzed pyridinium carboxylation and LarE-dependent sacrificial sulfur insertion. *Proc. Natl. Acad. Sci.* **113**, 5598–5603 (2016).
28. Fellner, M., Desguin, B., Hausinger, R. P. & Hu, J. Structural insights into the catalytic mechanism of a sacrificial sulfur insertase of the N-type ATP pyrophosphatase family, LarE. *Proc. Natl. Acad. Sci.* **114**, 9074–9079 (2017).
  29. Sheng, Y. *et al.* Superoxide Dismutases and Superoxide Reductases. *Chem. Rev.* **114**, 3854–3918 (2014).
  30. Shearer, J. Insight into the Structure and Mechanism of Nickel-Containing Superoxide Dismutase Derived from Peptide-Based Mimics. *Acc. Chem. Res.* **47**, 2332–2341 (2014).
  31. Wuerges, J. *et al.* Crystal structure of nickel-containing superoxide dismutase reveals another type of active site. (2004).
  32. Huang, H.-T. *et al.* The Role of Mixed Amine/Amide Ligation in Nickel Superoxide Dismutase. *Inorg. Chem.* **57**, 12521–12535 (2018).
  33. Herbst, R. W. *et al.* Role of Conserved Tyrosine Residues in NiSOD Catalysis: A Case of Convergent Evolution. *Biochemistry* **48**, 3354–3369 (2009).
  34. Repeta, D. J. *et al.* Marine methane paradox explained by bacterial degradation of dissolved organic matter. *Nat. Geosci.* **9**, 884–887 (2016).
  35. Vinson, V. Methanogenic archaea metabolism. *Science*. **357**, 656.17-658 (2017).
  36. Cedervall, P. E. *et al.* Structural analysis of a Ni-methyl species in methyl-coenzyme M reductase from *Methanothermobacter marburgensis*. *J. Am. Chem. Soc.* **133**, 5626–8 (2011).
  37. Moore, S. J. *et al.* Elucidation of the biosynthesis of the methane catalyst coenzyme F 430. *Nat. Publ. Gr.* **543**, (2017).
  38. Hausinger, R. P. New metal cofactors and recent metallocofactor insights. *Current Opinion in Structural Biology* (2019). doi:10.1016/j.sbi.2018.12.008
  39. Zheng, K., Ngo, P. D., Owens, V. L., Yang, X. & Mansoorabadi, S. O. The biosynthetic pathway of coenzyme F430 in methanogenic and methanotrophic archaea. *Science*. **354**, 339–342 (2016).
  40. Lubitz, W., Ogata, H., Rüdiger, O. & Reijerse, E. Hydrogenases. *Chem. Rev.* **114**, 4081–4148 (2014).

41. Fontecilla-Camps, J. C., Volbeda, A., Cavazza, C. & Nicolet, Y. Structure/Function Relationships of [NiFe]- and [FeFe]-Hydrogenases. *Chem. Rev.* **107**, 4273–4303 (2007).
42. Ogata, H., Lubitz, W. & Higuchi, Y. Structure and function of [NiFe] hydrogenases. *J. Biochem.* **160**, 251–258 (2016).
43. Alfano, M. & Cavazza, C. The biologically mediated water–gas shift reaction: structure, function and biosynthesis of monofunctional [NiFe]-carbon monoxide dehydrogenases. *Sustain. Energy Fuels* **2**, 1653–1670 (2018).
44. Lacasse, M. J. & Zamble, D. B. [NiFe]-Hydrogenase Maturation. *Biochemistry* **55**, 1689–1701 (2016).
45. Pinske, C. & Sargent, F. Exploring the directionality of Escherichia coli formate hydrogenlyase: a membrane-bound enzyme capable of fixing carbon dioxide to organic acid. *Microbiologyopen* **5**, 721–737 (2016).
46. Fox, J. D., Yiping, H. E., Shelver, D., Roberts, G. P. & Ludden, P. W. Characterization of the region encoding the CO-induced hydrogenase of Rhodospirillum rubrum. *J. Bacteriol.* **178**, 6200–6208 (1996).
47. Correction for Adam et al., Evolutionary history of carbon monoxide dehydrogenase/acetyl-CoA synthase, one of the oldest enzymatic complexes. *Proc. Natl. Acad. Sci.* **115**, E5836–E5837 (2018).
48. Robb, F. T. & Techtman, S. M. Life on the fringe: microbial adaptation to growth on carbon monoxide. *F1000Research* **7**, (2018).
49. Darnault, C. et al. Ni-Zn-[Fe<sub>4</sub>-S<sub>4</sub>] and Ni-Ni-[Fe<sub>4</sub>-S<sub>4</sub>] clusters in closed and open  $\alpha$  subunits of acetyl-CoA synthase/carbon monoxide dehydrogenase. *Nat. Struct. Mol. Biol.* **10**, 271–279 (2003).
50. Gencic, S. & Grahame, D. A. Nickel in subunit beta of the acetyl-CoA decarbonylase/synthase multienzyme complex in methanogens. Catalytic properties and evidence for a binuclear Ni-Ni site. *J. Biol. Chem.* **278**, 6101–10 (2003).
51. Amara, P., Volbeda, A., Carlos Fontecilla-Camps, J. & Field, M. J. A Quantum Chemical Study of the Reaction Mechanism of Acetyl-Coenzyme A Synthase. (2005). doi:10.1021/ja0439221
52. Golden, M. L., Rampersad, M. V., Reibenspies, J. H. & Darensbourg, M. Y. Capture of Ni<sup>II</sup>, Cu<sup>I</sup> and Zn<sup>II</sup> by thiolate sulfurs of an N<sub>2</sub>S<sub>2</sub>Ni complex: A role for a metallocythiolate ligand in the



- acetyl-coenzyme A synthase active site. *Chem. Commun.* 1824–1825 (2003).  
doi:10.1039/B304884P
53. Can, M., Armstrong, F. A. & Ragsdale, S. W. Structure, Function, and Mechanism of the Nickel Metalloenzymes, CO Dehydrogenase, and Acetyl-CoA Synthase. *Chem. Rev.* **114**, 4149–4174 (2014).
  54. Inoue, M. *et al.* Structural and Phylogenetic Diversity of Anaerobic Carbon-Monoxide Dehydrogenases. *Front. Microbiol.* **9**, 3353 (2018).
  55. Fessler, J. M. Biochemische und strukturelle Untersuchungen der Kohlenmonoxid-Dehydrogenasen CODH-II und CODH-V aus Carboxydotherrnus hydrogenoformans.
  56. Svetlitchnyi, V., Peschel, C., Acker, G. & Meyer, O. Two membrane-associated NiFeS-carbon monoxide dehydrogenases from the anaerobic carbon-monoxide-utilizing eubacterium Carboxydotherrnus hydrogenoformans. *J. Bacteriol.* **183**, 5134–5144 (2001).
  57. Domnik, L. *et al.* CODH-IV: A High-Efficiency CO-Scavenging CO Dehydrogenase with Resistance to O<sub>2</sub>. *Angew. Chemie - Int. Ed.* **56**, 15466–15469 (2017).
  58. Wang, J. *et al.* The roles of the hybrid cluster protein, Hcp and its reductase, Hcr, in high affinity nitric oxide reduction that protects anaerobic cultures of *E. coli* against nitrosative stress. *Mol. Microbiol.* **100**, 877–892 (2016).
  59. Drennan, C. L., Heo, J., Sintchak, M. D., Schreiter, E. & Ludden, P. W. Life on carbon monoxide: X-ray structure of *Rhodospirillum rubrum* Ni-Fe-S carbon monoxide dehydrogenase. *Proc. Natl. Acad. Sci.* **98**, 11973–11978 (2001).
  60. Dobbek, H., Svetlitchnyi, V., Gremer, L., Huber, R. & Meyer, O. Crystal structure of a carbon monoxide dehydrogenase reveals a [Ni-4Fe-5S] cluster. *Science* **293**, 1281–1285 (2001).
  61. Feng, J. & Lindahl, P. A. Carbon Monoxide Dehydrogenase from *Rhodospirillum rubrum*: Effect of Redox Potential on Catalysis. *Biochemistry* **43**, 1552–1559 (2004).
  62. Can, M., Armstrong, F. A. & Ragsdale, S. W. Structure, Function, and Mechanism of the Nickel Metalloenzymes, CO Dehydrogenase, and Acetyl-CoA Synthase. *Chem. Rev.* **114**, 4149–4174 (2014).
  63. Kroneck, P. M. H. & Torres, M. E. S. *The Metal-Driven Biogeochemistry of Gaseous Compounds in the Environment.* **14**, (2014).
  64. Lindahl, P. A. The Ni-containing carbon monoxide dehydrogenase family: light at the end of



- the tunnel? *Biochemistry* **41**, 2097–2105 (2002).
65. Amara, P., Mouesca, J., Volbeda, A. & Fontecilla-Camps, J. Carbon monoxide dehydrogenase reaction mechanism: a likely case of abnormal CO<sub>2</sub> insertion to a Ni-H(-) bond. *Inorg. Chem.* **50**, 1868–1878 (2011).
  66. Lindahl, P. A., Ragsdale, S. W. & Munck, E. Mossbauer Study of CO Dehydrogenase from *Clostridium thermoaceticum*. *J. Biol. Chem.* **265**, 3880–3888 (1990).
  67. Jeoung, J., Fessler, J., Goetzl, S. & Dobbek, H. *The Metal-Driven Biogeochemistry of Gaseous Compounds in the Environment*. **14**, (Springer Netherlands, 2014).
  68. Kung, Y. & Drennan, C. L. A role for nickel–iron cofactors in biological carbon monoxide and carbon dioxide utilization. *Curr. Opin. Chem. Biol.* **15**, 276–283 (2011).
  69. Fessler, J., Jeoung, J. H. & Dobbek, H. How the [NiFe<sub>4</sub>S<sub>4</sub>] Cluster of CO Dehydrogenase Activates CO<sub>2</sub> and NCO<sup>-</sup>. *Angew. Chemie - Int. Ed.* **54**, 8560–8564 (2015).
  70. Ciaccavava, A. *et al.* Carbon Monoxide Dehydrogenase Reduces Cyanate to Cyanide. *Angew. Chemie - Int. Ed.* **56**, 7398–7401 (2017).
  71. Ciaccavava, A. *et al.* When the inhibitor tells more than the substrate: The cyanide-bound state of a carbon monoxide dehydrogenase. *Chem. Sci.* (2016). doi:10.1039/c5sc04554a
  72. Zeer-Wanklyn, C. J. & Zamble, D. B. Microbial nickel: cellular uptake and delivery to enzyme centers. *Curr. Opin. Chem. Biol.* **37**, 80–88 (2017).
  73. Soriano, A. *et al.* GTP-dependent activation of urease apoprotein in complex with the UreD, UreF, and UreG accessory proteins. *Proc. Natl. Acad. Sci. U. S. A.* **96**, 11140–4 (1999).
  74. Zambelli, B. *et al.* Nickel binding properties of *Helicobacter pylori* UreF, an accessory protein in the nickel-based activation of urease. *J. Biol. Inorg. Chem.* **19**, 319–34 (2014).
  75. Fong, Y. H. *et al.* Structure of UreG/UreF/UreH Complex Reveals How Urease Accessory Proteins Facilitate Maturation of *Helicobacter pylori* Urease. *PLoS Biol.* **11**, e1001678 (2013).
  76. Rain, J. C. *et al.* The protein-protein interaction map of *Helicobacter pylori*. *Nature* **409**, 211–5 (2001).
  77. Volland, P. *et al.* Interactions among the seven *Helicobacter pylori* proteins encoded by the urease gene cluster. *Am. J. Physiol. Gastrointest. Liver Physiol.* **284**, G96–G106 (2003).
  78. Bellucci, M., Zambelli, B., Musiani, F., Turano, P. & Ciurli, S. *Helicobacter pylori* UreE, a urease

- accessory protein: specific Ni(2+)- and Zn(2+)-binding properties and interaction with its cognate UreG. *Biochem. J.* **422**, 91–100 (2009).
79. Yang, X., Li, H., Lai, T.-P. & Sun, H. UreE-UreG Complex Facilitates Nickel Transfer and Preactivates GTPase of UreG in *Helicobacter pylori*. *J. Biol. Chem.* **290**, 12474–12485 (2015).
  80. Brayman, T. G. & Hausinger, R. P. Purification, characterization, and functional analysis of a truncated *Klebsiella aerogenes* UreE urease accessory protein lacking the histidine-rich carboxyl terminus. *J. Bacteriol.* **178**, 5410–6 (1996).
  81. Song, H. K., Mulrooney, S. B., Huber, R. & Hausinger, R. P. Crystal structure of *Klebsiella aerogenes* UreE, a nickel-binding metallochaperone for urease activation. *J. Biol. Chem.* **276**, 49359–49364 (2001).
  82. Mulrooney, S. B., Ward, S. K. & Hausinger, R. P. Purification and Properties of the *Klebsiella aerogenes* UreE Metal-Binding Domain, a Functional Metallochaperone of Urease. *J. Bacteriol.* **187**, 3581–3585 (2005).
  83. Hu, H. Q., Johnson, R. C., Merrell, D. S. & Maroney, M. J. Nickel Ligation of the N-Terminal Amine of HypA Is Required for Urease Maturation in *Helicobacter pylori*. *Biochemistry* **56**, 1105–1116 (2017).
  84. Banaszak, K. *et al.* Crystallographic and X-ray absorption spectroscopic characterization of *Helicobacter pylori* UreE bound to Ni<sup>2+</sup> and Zn<sup>2+</sup> reveals a role for the disordered C-terminal arm in metal trafficking. *Biochem. J.* **441**, 1017–26 (2012).
  85. 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.* **280**, 4684–4695 (2005).
  86. Yuen, M. H., Fong, Y. H., Nim, Y. S., Ho Lau, P. & Wong, K.-B. Structural insights into how GTP-dependent conformational changes in a metallochaperone UreG facilitate urease maturation. doi:10.1073/pnas.1712658114
  87. Yang, X., Li, H., Lai, T.-P. & Sun, H. UreE-UreG complex facilitates nickel transfer and preactivates GTPase of UreG in *Helicobacter pylori*. *J. Biol. Chem.* **290**, 12474–85 (2015).
  88. Lacasse, M. J. & Zamble, D. B. [NiFe]-Hydrogenase Maturation. *Biochemistry* **55**, 1689–1701 (2016).
  89. Olson, J. W., Mehta, N. S. & Maier, R. J. Requirement of nickel metabolism proteins HypA and HypB for full activity of both hydrogenase and urease in *Helicobacter pylori*. *Mol. Microbiol.*

- 39**, 176–182 (2001).
90. Rowinska-Zyrek, M., Potocki, S., Witkowska, D., Valensin, D. & Kozlowski, H. The zinc-binding fragment of HypA from *Helicobacter pylori*: a tempting site also for nickel ions. *Dalt. Trans.* **42**, 6012 (2013).
  91. Hu, H. Q., Johnson, R. C., Merrell, D. S. & Maroney, M. J. Nickel Ligation of the N-Terminal Amine of HypA Is Required for Urease Maturation in *Helicobacter pylori*. *Biochemistry* **56**, 1105–1116 (2017).
  92. Cheng, T., Li, H., Yang, X., Xia, W. & Sun, H. Interaction of SlyD with HypB of *Helicobacter pylori* facilitates nickel trafficking. *Metallomics* **5**, 804–7 (2013).
  93. Spronk, C. A. E. M. *et al.* Structure and dynamics of *Helicobacter pylori* nickel-chaperone HypA: an integrated approach using NMR spectroscopy, functional assays and computational tools. *JBIC J. Biol. Inorg. Chem.* **23**, 1309–1330 (2018).
  94. Watanabe, S. *et al.* Structural basis of a Ni acquisition cycle for [NiFe] hydrogenase by Ni-metallochaperone HypA and its enhancer. *Proc. Natl. Acad. Sci.* **112**, 7701–7706 (2015).
  95. Kennedy, D. C., Herbst, R. W., Iwig, J. S., Chivers, P. T. & Maroney, M. J. A Dynamic Zn Site in *Helicobacter pylori* HypA: A Potential Mechanism for Metal-Specific Protein Activity. *J. Am. Chem. Soc.* **129**, 16–17 (2007).
  96. Herbst, R. W. *et al.* Communication between the Zinc and Nickel Sites in Dimeric HypA: Metal Recognition and pH Sensing. *J. Am. Chem. Soc.* **132**, 10338–10351 (2010).
  97. Chan Chung, K. C. *et al.* A High-Affinity Metal-Binding Peptide from *Escherichia coli* HypB. *J. Am. Chem. Soc.* **130**, 14056–14057 (2008).
  98. Xia, W., Li, H., Yang, X., Wong, K.-B. & Sun, H. Metallo-GTPase HypB from *Helicobacter pylori* and Its Interaction with Nickel Chaperone Protein HypA. *J. Biol. Chem.* **287**, 6753–6763 (2012).
  99. Fu, C., Olson, J. W. & Maier, R. J. HypB protein of *Bradyrhizobium japonicum* is a metal-binding GTPase capable of binding 18 divalent nickel ions per dimer. *Proc. Natl. Acad. Sci.* **92**, 2333–2337 (1995).
  100. Sydor, A. M., Lebrette, H., Ariyakumaran, R., Cavazza, C. & Zamble, D. B. Relationship between Ni(II) and Zn(II) coordination and nucleotide binding by the *Helicobacter pylori* [NiFe]-hydrogenase and urease maturation factor HypB. *J. Biol. Chem.* **289**, 3828–3841 (2014).
  101. Zhang, J. W., Butland, G., Greenblatt, J. F., Emili, A. & Zamble, D. B. A role for SlyD in the

- Escherichia coli hydrogenase biosynthetic pathway. *J. Biol. Chem.* **280**, 4360–4366 (2005).
102. Pinske, C., Sargent, F. & Sawers, R. G. SlyD-dependent nickel delivery limits maturation of [NiFe]-hydrogenases in late-stationary phase Escherichia coli cells. *Metallomics* **7**, 683–690 (2015).
  103. Weininger, U. *et al.* NMR solution structure of SlyD from Escherichia coli: spatial separation of prolyl isomerase and chaperone function. *J. Mol. Biol.* **387**, 295–305 (2009).
  104. Kovermann, M., Schmid, F. X. & Balbach, J. Molecular function of the prolyl cis/trans isomerase and metallochaperone SlyD. *Biological Chemistry* (2013). doi:10.1515/hsz-2013-0137
  105. Kaluarachchi, H., Zhang, J. W. & Zamble, D. B. Escherichia coli SlyD, More Than a Ni(II) Reservoir. *Biochemistry* **50**, 10761–10763 (2011).
  106. Khorasani-Motlagh, M., Noroozifar, M., Kerman, K. & Zamble, D. B. Complex formation between the Escherichia coli [NiFe]-hydrogenase nickel maturation factors. *BioMetals* 1–12 (2019). doi:10.1007/s10534-019-00173-9
  107. Lacasse, M. J., Douglas, C. D. & Zamble, D. B. Mechanism of Selective Nickel Transfer from HypB to HypA, Escherichia coli [NiFe]-Hydrogenase Accessory Proteins. *Biochemistry* **55**, 6821–6831 (2016).
  108. Leach, M. R., Zhang, J. W. & Zamble, D. B. The role of complex formation between the Escherichia coli hydrogenase accessory factors HypB and SlyD. *J. Biol. Chem.* **282**, 16177–86 (2007).
  109. Kerby, R. L. & Ludden, P. W. In vivo nickel insertion into the carbon monoxide dehydrogenase of Rhodospirillum rubrum : molecular and physiological characterization of cooCTJ . *J. Bacteriol.* **179**, 2259–2266 (1997).
  110. He, Y., Shelver, D., Kerby, R. L. & Roberts, G. P. Characterization of a CO-responsive transcriptional activator from Rhodospirillum rubrum. *J. Biol. Chem.* **271**, 120–123 (1996).
  111. Roberts, G. P., Kerby, R. L., Youn, H. & Conrad, M. CooA, a paradigm for gas sensing regulatory proteins. *J. Inorg. Biochem.* **99**, 280–292 (2005).
  112. Jeon, W. B., Cheng, J. & Ludden, P. W. Purification and Characterization of Membrane-associated CooC Protein and Its Functional Role in the Insertion of Nickel into Carbon Monoxide Dehydrogenase from Rhodospirillum rubrum. *J. Biol. Chem.* **276**, 38602–38609

- (2001).
113. Spangler, N. J., Lindahl, P. A., Bandarian, V. & Ludden, P. W. Spectroelectrochemical characterization of the metal centers in carbon monoxide dehydrogenase (CODH) and nickel-deficient CODH from *Rhodospirillum rubrum*. *J. Biol. Chem.* **271**, 7973–7977 (1996).
  114. Spangler, N. J. *et al.* Substitution of valine for histidine 265 in carbon monoxide dehydrogenase from *Rhodospirillum rubrum* affects activity and spectroscopic states. *J. Biol. Chem.* **273**, 4059–4064 (1998).
  115. Jeon, W. B., Singer, S. W., Ludden, P. W. & Rubio, L. M. New insights into the mechanism of nickel insertion into carbon monoxide dehydrogenase: analysis of *Rhodospirillum rubrum* carbon monoxide dehydrogenase variants with substituted ligands to the [Fe<sub>3</sub>S<sub>4</sub>] portion of the active-site C-cluster. *JBIC J. Biol. Inorg. Chem.* **10**, 903–912 (2005).
  116. Wittenborn, E. C. *et al.* Redox-dependent rearrangements of the NiFeS cluster of carbon monoxide dehydrogenase. *Elife* **7**, (2018).
  117. Inoue, T., Takao, K., Yuto, F., Yoshida, T. & Sako, Y. Over-expression of carbon monoxide dehydrogenase-I with an accessory protein co-expression: a key enzyme for carbon dioxide reduction. *Biosci. Biotechnol. Biochem.* **78**, 582–587 (2014).
  118. Jeon, W. B., Cheng, J. & Ludden, P. W. Purification and Characterization of Membrane-associated CooC Protein and Its Functional Role in the Insertion of Nickel into Carbon Monoxide Dehydrogenase from *Rhodospirillum rubrum*. *J. Biol. Chem.* **276**, 38602–38609 (2001).
  119. Miraula, M., Ciurli, S. & Zambelli, B. Intrinsic disorder and metal binding in UreG proteins from Archae hyperthermophiles: GTPase enzymes involved in the activation of Ni(II) dependent urease. *J. Biol. Inorg. Chem.* (2015). doi:10.1007/s00775-015-1261-7
  120. Gasper, R., Scrima, A. & Wittinghofer, A. Structural Insights into HypB, a GTP-binding Protein That Regulates Metal Binding. *J. Biol. Chem.* **281**, 27492–27502 (2006).
  121. Jeoung, J.-H. H., Giese, T., Grünwald, M. & Dobbek, H. Crystal structure of the ATP-dependent maturation factor of Ni,Fe-containing carbon monoxide dehydrogenases. *J. Mol. Biol.* **396**, 1165–1179 (2010).
  122. Gregg, C. M., Goetzl, S., Jeoung, J. H. & Dobbek, H. AcsF catalyzes the ATP-dependent insertion of Nickel into the Ni, Ni-[4Fe<sub>4</sub>S] cluster of Acetyl-CoA synthase. *J. Biol. Chem.* **291**, 18129–18138 (2016).

123. Gregg, C. M., Goetzl, S., Jeoung, J.-H. & Dobbek, X. H. AcsF Catalyzes the ATP-dependent Insertion of Nickel into the Ni<sub>2</sub>[4Fe4S] Cluster of Acetyl-CoA Synthase \* Downloaded from. *NUMBER 35 J. Biol. Chem.* **291**, 18129–18138 (2016).
124. Jeoung, J.-H. H. *et al.* CooC1 from Carboxydotherrmus hydrogenoformans is a nickel-binding ATPase. *Biochemistry* **48**, 11505–11513 (2009).
125. Watt, R. K. & Ludden, P. W. The Identification, Purification, and Characterization of CooJ. *J. Biol. Chem.* **273**, 10019–10025 (1998).
126. Song, H. K., Mulrooney, S. B., Huber, R. & Hausinger, R. P. Crystal structure of Klebsiella aerogenes UreE, a nickel-binding metallochaperone for urease activation. *J. Biol. Chem.* **276**, 49359–64 (2001).
127. Timm, J. *et al.* The CO dehydrogenase accessory protein CooT is a novel nickel-binding protein. *Metallomics* **9**, 575–583 (2017).
128. Inoue, H., Nojima, H. & Okayama, H. High efficiency transformation of Escherichia coli with plasmids. *Gene* **96**, 23–28 (1990).
129. Wang, P. *et al.* Robust growth of Escherichia coli. *Curr. Biol.* **20**, 1099–1103 (2010).
130. Maeda, T., Tran, K. T., Yamasaki, R. & Wood, T. K. Current state and perspectives in hydrogen production by Escherichia coli: roles of hydrogenases in glucose or glycerol metabolism. *Applied Microbiology and Biotechnology* **102**, 2041–2050 (2018).
131. Windsor, L. J. & Steele, D. L. Expression of Recombinant Matrix Metalloproteinases in Escherichia coli. in *Methods in molecular biology (Clifton, N.J.)* **622**, 67–81 (2010).
132. Jeoung, J.-H. & Dobbek, H. Carbon Dioxide Activation at the Ni<sub>2</sub>Fe-Cluster of Anaerobic Carbon Monoxide Dehydrogenase. *Science (80-. )*. **318**, 1461–1464 (2007).
133. Inoue, T., Takao, K., Fukuyama, Y., Yoshida, T. & Sako, Y. Over-expression of carbon monoxide dehydrogenase-I with an accessory protein co-expression: A key enzyme for carbon dioxide reduction. *Biosci. Biotechnol. Biochem.* (2014). doi:10.1080/09168451.2014.890027
134. Bar-Even, A. *et al.* The Moderately Efficient Enzyme: Evolutionary and Physicochemical Trends Shaping Enzyme Parameters. **50**, 4402–4410 (2011).
135. Wise, D. L. & Houghton, G. Diffusion coefficients of neon, krypton, xenon, carbon monoxide and nitric oxide in water at 10–60°C. *Chem. Eng. Sci.* **23**, 1211–1216 (1968).

136. Svetlitchnyi, V. *et al.* A functional Ni-Ni-[4Fe-4S] cluster in the monomeric acetyl-CoA synthase from *Carboxydotherrmus hydrogenoformans*. *Proc. Natl. Acad. Sci. U. S. A.* **101**, 446–51 (2004).
137. Hadj-Saïd, J., Pandelia, M. E., Léger, C., Fourmond, V. & Dementin, S. The Carbon Monoxide Dehydrogenase from *Desulfovibrio vulgaris*. *Biochim. Biophys. Acta - Bioenerg.* **1847**, 1574–1583 (2015).
138. Wittenborn, E. C. *et al.* Structural insight into metallocofactor maturation in carbon monoxide dehydrogenase. *J. Biol. Chem.* jbc.RA119.009610 (2019). doi:10.1074/jbc.RA119.009610
139. Henstra, A. M., Sipma, J., Rinzema, A. & Stams, A. J. Microbiology of synthesis gas fermentation for biofuel production. *Curr. Opin. Biotechnol.* **18**, 200–206 (2007).
140. Roberts, G. P., Youn, H. & Kerby, R. L. CO-sensing mechanisms. *Microbiol. Mol. Biol. Rev.* **68**, 453–73, table of contents (2004).
141. Hädicke, O., Grammel, H. & Klamt, S. Metabolic network modeling of redox balancing and biohydrogen production in purple nonsulfur bacteria. *BMC Syst. Biol.* **5**, 150 (2011).
142. Dubbs, J. M. & Tabita, F. R. Regulators of nonsulfur purple phototrophic bacteria and the interactive control of CO<sub>2</sub> assimilation, nitrogen fixation, hydrogen metabolism and energy generation. *FEMS Microbiology Reviews* **28**, 353–376 (2004).
143. Fox, J. D., Kerby, R. L., Roberts, G. P. & Ludden, P. W. Characterization of the CO-induced, CO-tolerant hydrogenase from *Rhodospirillum rubrum* and the gene encoding the large subunit of the enzyme. *J. Bacteriol.* **178**, 1515–24 (1996).
144. Revelles, O. *et al.* Syngas obtained by microwave pyrolysis of household wastes as feedstock for polyhydroxyalkanoate production in *Rhodospirillum rubrum*. *Microb. Biotechnol.* **10**, 1412–1417
145. Revelles, O., Tarazona, N., García, J. L. & Prieto, M. A. Carbon roadmap from syngas to polyhydroxyalkanoates in *Rhodospirillum rubrum*. *Environ. Microbiol.* **18**, 708–720 (2016).
146. Sansaniwal, S. K., Pal, K., Rosen, M. A. & Tyagi, S. K. Recent advances in the development of biomass gasification technology: A comprehensive review. *Renew. Sustain. Energy Rev.* **72**, 363–384 (2017).
147. Chinchén, G. C., Logan, R. H. & Spencer, M. S. Water-gas shift reaction over an iron oxide/chromium oxide catalyst.: I: Mass transport effects. *Appl. Catal.* **12**, 69–88 (1984).
148. Najafpour, G. D. & Younesi, H. Bioconversion of synthesis gas to hydrogen using a light-



- dependent photosynthetic bacterium, *Rhodospirillum rubrum*. *World J. Microbiol. Biotechnol.* **23**, 275–284 (2007).
149. Klasson, K. T., Lundbäck, K. M. O., Clausen, E. C. & Gaddy, J. L. Kinetics of light limited growth and biological hydrogen production from carbon monoxide and water by *Rhodospirillum rubrum*. *J. Biotechnol.* **29**, 177–188 (1993).
  150. Najafpour, G., Younesi, H. & Mohamed, A. R. A Survey on Various Carbon Sources for Biological Hydrogen Production via the Water-Gas Reaction Using a Photosynthetic Bacterium (*Rhodospirillum rubrum*). *Energy Sources, Part A Recover. Util. Environ. Eff.* **28**, 1013–1026 (2006).
  151. Najafpour, G., Younesi, H. & Mohamed, A. R. Effect of organic substrate on hydrogen production from synthesis gas using *Rhodospirillum rubrum*, in batch culture. *Biochem. Eng. J.* **21**, 123–130 (2004).
  152. Younesi, H., Najafpour, G., Ku Ismail, K. S., Mohamed, A. R. & Kamaruddin, A. H. Biohydrogen production in a continuous stirred tank bioreactor from synthesis gas by anaerobic photosynthetic bacterium: *Rhodospirillum rubrum*. *Bioresour. Technol.* **99**, 2612–2619 (2008).
  153. Ismail, K. S. K., Najafpour, G., Younesi, H., Mohamed, A. R. & Kamaruddin, A. H. Biological hydrogen production from CO: Bioreactor performance. *Biochem. Eng. J.* **39**, 468–477 (2008).
  154. Ali Dadak, Mortaza Aghbashlo, Meisam Tabatabaei, Ghasem Najafpour, A. H. Y. Sustainability Assessment of Photobiological Hydrogen Production Using Anaerobic Bacteria (*Rhodospirillum rubrum*) via Exergy Concept: Effect of Substrate Concentrations. *Environ. Sci. Technol.* **33**, 482–489 (2014).
  155. Valle, A., Cantero, D. & Bolívar, J. Metabolic engineering for the optimization of hydrogen production in *Escherichia coli*: A review. *Biotechnol. Adv.* (2019).  
doi:10.1016/j.biotechadv.2019.03.006
  156. Stanyon, H. F. *et al.* Developing predictive rules for coordination geometry from visible circular dichroism of copper(II) and nickel(II) ions in histidine and amide main-chain complexes. *FEBS J.* **281**, 3945–3954 (2014).
  157. Artur Krężel, Rafał Latajka, Grzegorz D. Bujacz and Wojciech Bal, Coordination Properties of Tris(2-carboxyethyl)phosphine, a Newly Introduced Thiol Reductant, and Its Oxide. (2003).  
doi:10.1021/IC025969Y
  158. Colpas, G. J. *et al.* X-ray spectroscopic studies of nickel complexes, with application to the

- structure of nickel sites in hydrogenases. *Inorg. Chem.* **30**, 920–928 (1991).
159. Poole, L. B. The basics of thiols and cysteines in redox biology and chemistry. *Free Radic. Biol. Med.* **80**, 148–57 (2015).
  160. Marino, S. M. & Gladyshev, V. N. Cysteine Function Governs Its Conservation and Degeneration and Restricts Its Utilization on Protein Surfaces. *J. Mol. Biol.* **404**, 902–916 (2010).
  161. Gilmore, J. M., Scheck, R. A., Esser-Kahn, A. P., Joshi, N. S. & Francis, M. B. N-Terminal Protein Modification through a Biomimetic Transamination Reaction. *Angew. Chemie Int. Ed.* **45**, 5307–5311 (2006).
  162. Witus, L. S. & Francis, M. Site-Specific Protein Bioconjugation via a Pyridoxal 5'-Phosphate-Mediated N-Terminal Transamination Reaction. *Curr. Protoc. Chem. Biol.* **2**, 125–34 (2010).
  163. Magalon, A. & Böck, A. 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.* **275**, 21114–21120 (2000).
  164. Wang, L., Xia, B. & Jin, C. Solution structure of *Escherichia coli* HypC. *Biochem. Biophys. Res. Commun.* **361**, 665–669 (2007).
  165. Soboh, B. *et al.* [NiFe]-hydrogenase maturation: Isolation of a HypC–HypD complex carrying diatomic CO and CN– ligands. *FEBS Lett.* **586**, 3882–3887 (2012).
  166. Watanabe, S., Matsumi, R., Atomi, H., Imanaka, T. & Miki, K. Crystal Structures of the HypCD Complex and the HypCDE Ternary Complex: Transient Intermediate Complexes during [NiFe] Hydrogenase Maturation. *Structure* **20**, 2124–2137 (2012).
  167. Thomas, C., Waclawek, M., Nutschan, K., Pinske, C. & Sawers, R. G. The Extended C-Terminal  $\alpha$ -Helix of the HypC Chaperone Restricts Recognition of Large Subunit Precursors by the Hyp-Scaffold Machinery during [NiFe]-Hydrogenase Maturation in *Escherichia coli*. *J. Mol. Microbiol. Biotechnol.* **28**, 87–97 (2018).
  168. Hagemans, D., van Belzen, I. A. E. M., Morán Luengo, T. & Rüdiger, S. G. D. A script to highlight hydrophobicity and charge on protein surfaces. *Front. Mol. Biosci.* **2**, 56 (2015).
  169. Chayen, N. E. The role of oil in macromolecular crystallization. *Ways Means* **5**, 1269–1274 (1269).
  170. Engilberge, S. *et al.* Crystallophore: a versatile lanthanide complex for protein crystallography

- combining nucleating effects, phasing properties, and luminescence. *Chem. Sci.* **8**, 5909–5917 (2017).
171. Kaluarachchi, H. *et al.* The Ni(II)-Binding Properties of the Metallochaperone SlyD. doi:10.1021/ja9081765
  172. Bonam, D., McKennat, M. C., Stephenst, P. J. & Ludden, P. W. Nickel-deficient carbon monoxide dehydrogenase from *Rhodospirillum rubrum*: In vivo and in vitro activation by exogenous nickel (hydrogenase/iron-sulfur protein/electron paramagnetic resonance). *Biochemistry* **85**, 31–35 (1988).
  173. Stephens, P., McKenna, M., Ensign, S. A., Bonam, D. & Ludden<sup>1</sup>, P. W. Identification of a Ni- and Fe-containing cluster in *Rhodospirillum rubrum* carbon monoxide dehydrogenase. *J. Biol. Chem.* **264**, 16347–16350 (1989).
  174. Ensign, S. A., Campbell, M. J. & Ludden, P. W. Activation of the Nickel-Deficient Carbon Monoxide Dehydrogenase from *Rhodospirillum rubrum*: Kinetic Characterization and Reductant Requirement1&quot; *Biochemistry* **29**, 2162–2168 (1990).
  175. Ensign, S. A. & Ludden, P. W. Characterization of the CO oxidation/H<sub>2</sub> evolution system of *Rhodospirillum rubrum*: Role of a 22-kDa iron-sulfur protein in mediating electron transfer between carbon monoxide dehydrogenase and hydrogenase. *J. Biol. Chem.* **266**, 18395–18403 (1991).
  176. Soboh, B. *et al.* [NiFe]-hydrogenase maturation: Isolation of a HypC-HypD complex carrying diatomic CO and CN<sup>-</sup> ligands. *FEBS Lett.* **586**, 3882–3887 (2012).
  177. Lazarus, O. *et al.* Water–Gas Shift Reaction Catalyzed by Redox Enzymes on Conducting Graphite Platelets. *J. Am. Chem. Soc.* **131**, 14154–14155 (2009).
  178. Smyth, M. S. & Martin, J. H. x ray crystallography. *Mol. Pathol.* **53**, 8–14 (2000).
  179. McPherson, A. & Gavira, J. A. Introduction to protein crystallization. *Acta Crystallogr. Sect. F, Struct. Biol. Commun.* **70**, 2–20 (2014).
  180. Karplus, P. A. & Diederichs, K. Assessing and maximizing data quality in macromolecular crystallography. *Curr. Opin. Struct. Biol.* **34**, 60–8 (2015).
  181. Evans, P. R. & Murshudov, G. N. How good are my data and what is the resolution? *Acta Crystallogr. Sect. D Biol. Crystallogr.* **69**, 1204–1214 (2013).
  182. Garman, E. F. Developments in x-ray crystallographic structure determination of biological

- macromolecules. *Science* **343**, 1102–8 (2014).
183. Martin, S. R. & Schilstra, M. J. Circular Dichroism and Its Application to the Study of Biomolecules. in *Methods in cell biology* **84**, 263–293 (2008).
  184. Greenfield, N. J. Using circular dichroism spectra to estimate protein secondary structure. *Nat. Protoc.* (2007). doi:10.1038/nprot.2006.202
  185. Micsonai, A. *et al.* BeStSel: a web server for accurate protein secondary structure prediction and fold recognition from the circular dichroism spectra. *Nucleic Acids Res.* **46**, W315–W322 (2018).
  186. Klewpatinond, M. & Viles, J. H. Empirical rules for rationalising visible circular dichroism of Cu<sup>2+</sup> and Ni<sup>2+</sup> histidine complexes: Applications to the prion protein. *FEBS Lett.* **581**, 1430–1434 (2007).
  187. Greenfield, N. J. Determination of the folding of proteins as a function of denaturants, osmolytes or ligands using circular dichroism. *Nat. Protoc.* **1**, 2733–2741 (2006).
  188. Smith, L., Greenfield, N. J. & Hitchcock-DeGregori, S. E. The effects of deletion of the amino-terminal helix on troponin C function and stability. *J. Biol. Chem.* **269**, 9857–63 (1994).
  189. Greenfield, N. J. & Fasman, G. D. Computed circular dichroism spectra for the evaluation of protein conformation. *Biochemistry* **8**, 4108–4116 (1969).
  190. Greenfield, N. J. Using circular dichroism collected as a function of temperature to determine the thermodynamics of protein unfolding and binding interactions. *Nat. Protoc.* **1**, 2527–35 (2006).
  191. Leavitt, S. & Freire, E. Direct measurement of protein binding energetics by isothermal titration calorimetry. *Curr. Opin. Struct. Biol.* **11**, 560–566 (2001).
  192. van Oversteeg, C. H. M., Doan, H. Q., de Groot, F. M. F. & Cuk, T. In situ X-ray absorption spectroscopy of transition metal based water oxidation catalysts. *Chem. Soc. Rev.* **46**, 102–125 (2017).
  193. Acosta-Alejandro, M. *et al.* Local lattice structure change in PrNiO<sub>3</sub> across the metal-insulator transition: X-ray absorption near-edge structure spectroscopy and *ab initio* calculations. *Phys. Rev. B* **77**, 085107 (2008).
  194. Svergun, D. I. & Koch, M. H. J. Small-angle scattering studies of biological macromolecules in solution. *Reports Prog. Phys.* **66**, 1735–1782 (2003).

195. Jacques, D. A. & Trewhella, J. Small-angle scattering for structural biology-Expanding the frontier while avoiding the pitfalls. *Protein Sci.* **19**, 642–657 (2010).
196. Tria, G., Mertens, H. D. T., Kachala, M. & Svergun, D. I. Advanced ensemble modelling of flexible macromolecules using X-ray solution scattering. *IUCrJ* **2**, 207–217 (2015).
197. Svergun, D. I. & Koch, M. H. . Advances in structure analysis using small-angle scattering in solution. *Curr. Opin. Struct. Biol.* **12**, 654–660 (2002).