POLITECNICO DI TORINO

Collegio di Ingegneria Chimica e dei Materiali

**Corso di Laurea Magistrale
in Ingegneria Chimica e dei Processi Sostenibili**
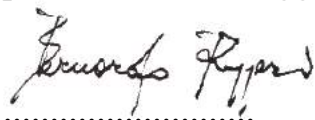
Tesi di Laurea Magistrale

Hydrogenase maturation enzymes

Relatori

firma dei relatori
prof. Bernardo Ruggeri

....

prof. Marc Fontecave

Candidato

firma del candidato
Carlotta Panzone

Ottobre 2017

The project has been carried out at 'Laboratoire de Chimie des Processus Biologiques' in Collège de France, during an Erasmus project at Université Pierre et Marie Curie of Paris.

Table of contents

1. Riassunto in italiano

1.1 Introduzione

Il lavoro è parte di un progetto più ampio che coinvolge una ricerca di base sulle idrogenasi al fine di comprendere il meccanismo di reazione, la struttura del sito attivo e il coinvolgimento dell'ossigeno nella disattivazione del sito attivo.

Le idrogenasi sono enzimi in grado di catalizzare la reazione reversibile di produzione dell'idrogeno, attraverso la riduzione dei protoni. Tale reazione comporta l'uso di energia chimica che coinvolge reazioni redox. La reazione di riduzione dei protoni (il cui potenziale è pari a -0.42 V) deve quindi essere accoppiata con una semi-reazione di ossidazione che deve fornire gli elettroni e che ha un potenziale redox più negativo.[1] Le idrogenasi sono attive soltanto a valori di potenziale redox corrispondente a condizioni anaerobiche. Aumentando la concentrazione di ossigeno il potenziale redox aumenta e, in queste condizioni, l'idrogenasi non è più attiva perché da un lato l'ossigeno diventa un accettore di elettroni migliore dell'idrogenasi e dall'altro esso ha un'azione ossidativa molto forte che porta alla degradazione del sito attivo. Infatti, in ambienti fortemente ossidanti, l'ossigeno forma dei composti altamente reattivi (ROS) che attaccano il sito attivo dell'idrogenasi, degradandolo. Pertanto, è importante capire come l'ambiente esterno e il modo in cui gli elettroni vengono forniti influenzano l'attività dell'enzima, in modo da creare nuovi sistemi bioispirati in grado di operare a differenti condizioni ambientali. Inoltre, è noto che l'ossigeno reagisce con l'H-cluster ma rimane sconosciuto il meccanismo, ancora molto discusso. Si vuole quindi capire in che modo l'ossigeno reagisce e quale parte dell'H-cluster è coinvolta nella degradazione.

L'obiettivo di questi studi è l'ottenimento di un'idrogenasi che possa essere usata in differenti applicazioni biotecnologiche. L'interesse per questi enzimi è dovuto al fatto che essi sono i migliori catalizzatori per la produzione di idrogeno ad elevata attività e basso sovrapotenziale. Inoltre, sono facili da immobilizzare, perciò possono avere diverse applicazioni come bioelettrodi.[2] Una delle principali applicazioni è la produzione fotoelettrochimica di idrogeno, basata sullo splitting dell'acqua attraverso fotocatalisi con un elettrodo contenente un'idrogenasi immobilizzata; questo sistema potrebbe sostituire la tradizionale produzione di idrogeno a partire da gas naturale e la produzione tramite water splitting con catalizzatore al platino.[3] Altre possibili applicazioni delle idrogenasi sono la prevenzione della biocorrosione, dal momento che questi enzimi sono in grado di favorire la riduzione dello ione cromato,[4] e l'utilizzo come biosensori di H_2[5]. Esistono possibilità di applicazione anche come anodo delle biofuels cells, sfruttando la loro abilità di catalizzare l'ossidazione dell'idrogeno.

Inoltre, la conoscenza del meccanismo e della struttura di questi enzimi può portare al design di complessi sintetici che li imitano, applicabili su larga scala.[6] Il problema principale da affrontare rimane la tolleranza all'ossigeno. Lo sviluppo di catalizzatori che operano in presenza di ossigeno è la direzione attualmente più investigata, ma alcuni recenti studi hanno dimostrato la possibilità di creare dei catalizzatori resistenti all'ossigeno, in grado di ridurre l'ossigeno e di evitare la formazione di ROS. Queste reazioni entrano in competizione con la riduzione dei protoni, perciò l'idrogeno viene prodotto a velocità e rese minori.[7]

1.1.1 Idrogenasi

Le idrogenasi sono una classe di enzimi di grande interesse, grazie alla loro capacità di catalizzare la produzione di idrogeno, secondo la seguente reazione reversibile:

$$H_2 \ (\rightleftharpoons H^+ + H^-) \rightleftharpoons 2H^+ + 2e^- \tag{1.1}$$

L'uso massiccio di combustibili fossili sta iniziando a porre seri problemi di cambiamento climatico e inquinamento ambientale e questo ha dato inizio allo sviluppo di energie da fonti rinnovabili e di sistemi di stoccaggio dell'energia che possano fornire energia quando essa è richiesta.[8] Grazie alla loro capacità di produrre idrogeno, le idrogenasi sono studiate per possibili applicazioni nello sviluppo di

nuove tecnologie di produzione di energia rinnovabile a base di idrogeno. L'idrogeno, infatti, è una molecola a elevato contenuto energetico e non contiene carbonio. Diversi sono i vantaggi di tale elemento: la sua combustione produce soltanto energia e acqua, secondo la reazione $2H_2 + O_2 \rightarrow 2H_2O$, può essere trasportato e stoccato in molte forme, come gas compresso, liquido criogenico o idruro solido. Tuttavia, esistono anche degli svantaggi, per esempio la scarsa disponibilità in natura di idrogeno molecolare e la necessità di estrarlo da composti come acqua o idrocarburi, attraverso metodi costosi e inquinanti (riassunti in Figure 2.1).[10]

Le idrogenasi sono delle metallo-proteine, cioè contengono un cofattore metallico senza il quale la proteina non riuscirebbe a svolgere le sue funzioni. Lo ione metallico è quindi parte integrante del sito attivo ed ha un ruolo chiave. Le idrogenasi, in particolare, contengono come cofattore metallico il cluster ferro-zolfo, molto comune in diversi tipi di proteine. Esso può intervenire in diverse reazioni, ma è principalmente coinvolto in quelle di ossido-riduzione e nel trasporto di elettroni. Le configurazioni più comuni sono la disposizione rombica [2Fe-2S] e quella cubica [4Fe-4S] (vedere Figure 2.2).[12]

Le idrogenasi possono essere classificate in tre classi: idrogenasi [Ni-Fe], [Fe-Fe] e solo [Fe].[3] Le più importanti sono le prime due: le idrogenasi [Ni-Fe] sono coinvolte nel consumo di idrogeno, mentre quelle [Fe-Fe] sono coinvolte nella produzione di idrogeno, perciò più interessanti per questo studio. Le idrogenasi [Ni-Fe] contengono nel sito attivo un gruppo etero-atomico, costituito da un atomo di ferro e uno di zinco, connessi tramite due cisteine.[23] Le idrogenasi [Fe-Fe], note come HydA, invece, hanno un sito attivo costituito da un cluster [4Fe-4S] connesso a un sub-cluster [2Fe], attraverso il gruppo tiolo di una cisteina. Tale complesso è chiamato H-cluster. I due atomi di ferro sono connessi da un gruppo azaditiolato e ogni atomo è legato ai gruppi CO e CN. La sfera di coordinazione è completata da un ponte carbonilico (come mostrato in Figura 1.1). Le differenti idrogenasi [Fe-Fe] differiscono nel numero di cluster accessori. Alcuni microrganismi, infatti, hanno soltanto il cluster [4Fe-4S] dell'H-cluster, altri invece presentano cluster aggiuntivi che sono coinvolti nel trasferimento di elettroni.[24]
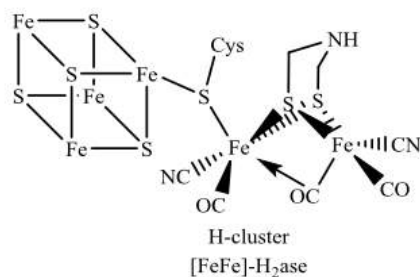
Figura 1.1 Sito attivo delle idrogenasi [Fe-Fe].

1.1.2 Maturazione delle idrogenasi [Fe-Fe]

In natura, la maturazione delle idrogenasi è un processo post-traduzionale molto complesso che coinvolge reazioni difficili. L'ottenimento della forma attiva della proteina, infatti, coinvolge l'assemblaggio dei cluster [Fe-S], la sintesi dei leganti (CO, CN e azaditiolato), l'assemblaggio del complesso [2Fe] e la sua incorporazione nell'enzima.

Il processo di maturazione non è ancora del tutto noto e coinvolge tre proteine, dette maturasi (HydE, HydF, HydG) che insieme costituiscono la macchina di maturazione delle idrogenasi, chiamata HYD. Secondo le ipotesi più accreditate, il processo di maturazione richiede l'assemblaggio del cluster [4Fe-4S] dell'H-cluster prima dell'inserimento del sub-cluster [2Fe] all'interno della proteina. Perciò, le maturasi HydE, HydF e HydG sintetizzano il sub-cluster [2Fe] e lo trasferiscono alla apo-idrogenasi (che contiene già il cluster [4Fe-4S]).[28]

Le due proteine HydE e HydG appartengono alla classe radical SAM (*S*-adenosil metionina) che catalizza reazioni chimiche complesse, svolgendo svariate funzioni biologiche. HydG contiene due cluster [4Fe-4S], entrambi necessari per l'attività e aventi ognuno il proprio ruolo: il cluster N-

terminale è richiesto per la produzione di p-cresolo, mentre quello C-terminale è richiesto per la formazione dei gruppi biatomici CO e CN. È stato inoltre dimostrato che HydG utilizza come substrato la tirosina, dalla cui scissione derivano gli atomi di ferro del gruppo binucleare dell'H-cluster. Studi spettroscopici hanno dimostrato che HydG è in grado di fornire il sintone [Fe(CO$_2^-$)CN$^-$], un complesso contenente Fe e i due gruppi biatomici CO e CN$^-$ e considerato l'effettivo precursore dell'H-cluster.[31-38] Anche HydE contiene due cluster [4Fe-4S], ma quello C-terminale sembra non essere coinvolto nella biosintesi dell'H-cluster. È stato ipotizzato che HydE fornisca il ponte azapropanditiolato che unisce i due atomi di ferro nel sub-cluster biatomico e che utilizzi cisteina come substrato.[39-41]

La terza proteina coinvolta, HydF, appartiene al gruppo delle GTPasi. Anch'essa contiene un cluster [4Fe-4S], legato da tre cisteine e da un legante non cisteinico, e un sub-cluster [2Fe], legato da CO e CN$^-$. Diversi esperimenti hanno dimostrato che HydF agisce come proteina scaffold ed è in grado di interagire prima con HydE e HydG, per assemblare il sub-cluster biatomico e poi con l'apo-idrogenasi per trasferirle quest'ultimo. Probabilmente, HydF interagisce con HydE e HydG attraverso il dominio GTP che viene poi dissociato per consentire l'interazione con apo-HydA.[42-45]

Il meccanismo di funzionamento più accreditato per la macchina di maturazione HYD è che esso sintetizzi e trasferisca il sub-cluster [2Fe] e i suoi leganti all'apo-idrogenasi, già contenente il cluster [4Fe-4S]. Tutte e tre le maturasi sono essenziali: HydG effettua la sintesi del sintone [Fe(CO$_2$)CN$^-$] da tirosina, HydE probabilmente fornisce l'anello azapropanditiolato da un substrato ancora non noto e HydF interagisce con le altre due maturasi per assemblare tutti i componenti in un complesso bi-ferrico che viene poi trasferito all'apo-idrogenasi, portando all'attivazione della proteina, come mostrato in Figura 1.2.[28]
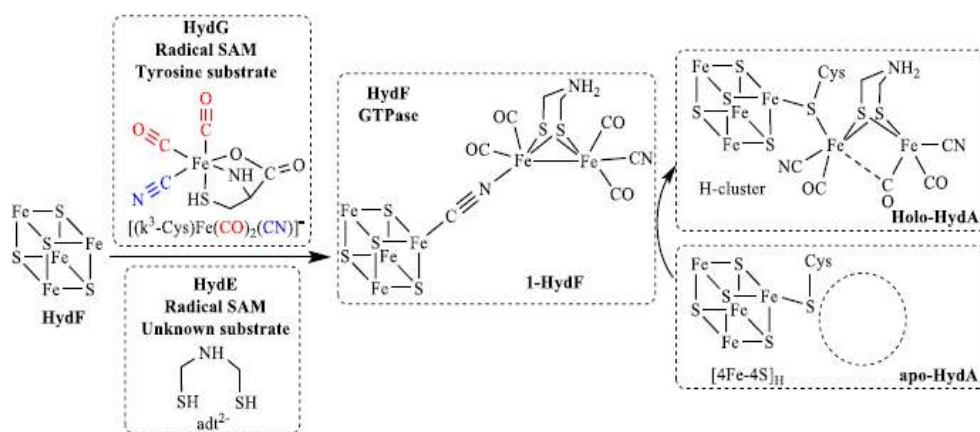
Figura 1.2 Rappresentazione schematica della maturazione di un'idrogenasi.[48]

1.1.3 Maturazione chimica

Diversi studi hanno dimostrato la possibilità di attivare le idrogenasi chimicamente, attraverso un processo che prende ispirazione dalla natura senza coinvolgere l'espressione delle tre maturasi. Infatti, è stato dimostrato che complessi bi-ferrici possono essere trasferiti, attraverso la maturasi HydF, all'idrogenasi [4Fe-4S]. Questi complessi da soli non sono attivi ma, inseriti nell'idrogenasi, hanno mostrato elevate attività catalitiche.[52] Diversi complessi sintetici che imitano il sub-cluster bi-ferrico sono stati realizzati (Figura 1.3) e inglobati in [4Fe-4S]-TmHydF, dando origine a proteine ibride del tipo x-HydF. È stato dimostrato che solo il complesso **2** (complesso adt) può essere efficacemente trasferito da HydF a HydA ed è in grado di attivare l'apo-idrogenasi, indicando che il precursore naturale è molto simile a tale complesso. Questo mostra anche che il complesso sintetico subisce le dovute trasformazioni che lo convertono nel substrato naturale attivo.[51]
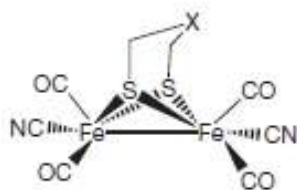
Figura 1.3 Struttura dei biomimic sintetici analizzati. X=CH$_2$ nel complesso **1**; X=NH nel complesso **2**; X=O nel complesso **3**.[51]

È stata anche dimostrata la possibilità di attivare HydA incubando in maniera diretta [4Fe-4S]-HydA con il complesso **2**, senza l'aiuto di HydF. Le due possibili vie di maturazione chimica sono mostrate in Figura 1.4.
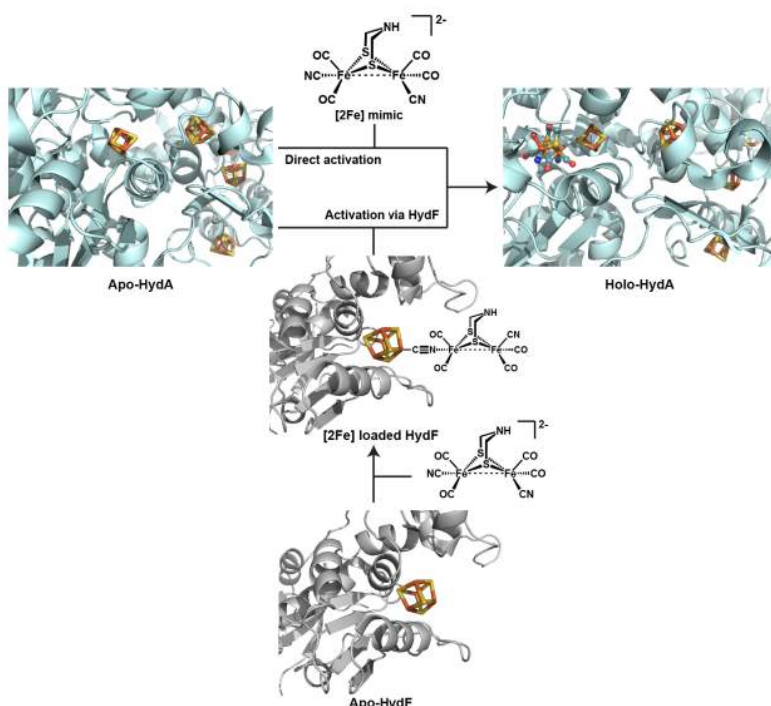
Figura 1.4 Due possibili vie di maturazione delle idrogenasi: i) attivazione diretta incubando apo-HydA con il complesso **2**; ii) attivazione via HydF, attraverso l'incubazione di apo-HydF con il complesso **2** e poi il trasferimento del subcluster [2Fe] da [2Fe]-HydF a apo-HydA.[62]

Recentemente, il laboratorio ha dimostrato la possibilità di evitare non soltanto la macchina HYD ma anche il by-pass della procedura di purificazione completamente anaerobica. E' possibile, infatti, purificare le idrogenasi in condizioni aerobiche e, successivamente, ricostituire i cluster metallici in condizioni anaerobiche.[53]

Nonostante questi studi abbiano portato a interessanti applicazioni che altrimenti sarebbero state impossibili da ottenere (identificazione dei siti attivi e produzione di idrogenasi artificiali), diversi sono i problemi che devono ancora essere affrontati per poter usare le idrogenasi in campo biotecnologico: i) la necessità di manipolare le proteine in condizioni strettamente anaerobiche per evitarne la disattivazione causata dall'ossigeno; ii) la difficoltà a produrli in forma attiva, parzialmente superata con la maturazione chimica.

IV

1.1.4 Idrogenasi artificiali

Negli ultimi anni, diversi biomimic metallici sono stati sintetizzati, consentendo di riprodurre strutture e funzioni del sito attivo metallico naturale attraverso composti sintetici organometallici a basso peso molecolare.[54-57] Tuttavia, la maggior parte di questi complessi ha mostrato un'attività non comparabile con quella delle idrogenasi, probabilmente a causa della mancanza di interazioni con la catena polipeptidica. Pertanto, è nata l'idea di migliorare le prestazioni di questi complessi, incorporandoli in una proteina e creando un'idrogenasi artificiale.[58]

Il concetto di enzimi artificiali fonda le sue origini nel lavoro di Wilson e Whitesides[59] e si basa sull'idea di combinare uno scaffold biomolecolare e un complesso sintetico come sito attivo. Il caso delle idrogenasi è molto interessante perché offre la possibilità di ottimizzare lo scambio di elettroni e protoni. Il concetto di idrogenasi artificiali implica la combinazione di una proteina ospite, ben selezionata e facilmente accessibile, con un complesso sintetico in grado di catalizzare la riduzione di protoni o l'ossidazione di idrogeno. Il fine è l'ottenimento di un sistema ibrido economico, stabile, solubile in acqua e attivo. Tra le idrogenasi, HydF è stata considerata come la proteina più adatta per questo tipo di complessi sintetici bioispirati, a causa della sua abilità di legare e trasferire il sub-cluster [2Fe]. È stato ottenuto un sistema sintetico con una piccola attività, incubando TmHydF con il complesso pdt. Nonostante l'attività misurata fosse molto bassa, questo è un punto di partenza per lo sviluppo di idrogenasi artificiali più efficienti.[58]

Attualmente, questi sistemi artificiali sono molto meno attivi degli enzimi naturali, ma c'è speranza che in futuro possa essere generato un metalloenzima attivo e stabile, con potenziale applicazione in dispositivi tecnologici.

1.1.5 Obiettivi del progetto

HydF da Thermosipho melanesiensis

Diversi studi sono stati condotti sulla struttura della maturasi HydF, studiata per la sua interessante capacità di assemblare e trasferire il complesso bi-ferrico a HydA. Studi spettroscopici e cristallografici su TnHydF da *Thermotoga neapolitana*[44] e TmeHydF da *Thermosipho melanesiensis*[45] hanno dimostrato che la proteina ha normalmente configurazione dimerica e ogni monomero è costituito da tre domini differenti (Figura 1.5 a): (I) il dominio I lega il gruppo GTP; (II) il dominio II è connesso al I attraverso una lunga catena di amminoacidi ed è responsabile della dimerizzazione di HydF; (III) il dominio III lega il cluster [4Fe-4S] tramite tre residui cisteinici altamente conservati. In TmeHydF è stato identificato anche un legante non cisteinico, il gruppo carbossilato di un glutammato. Questi residui amminoacidici formano una cavità carica positivamente che è responsabile dell'interazione di HydF con molecole cariche negativamente ($[4Fe-4S]^{1-/2-}$ e sub-cluster [2Fe]).

I cristalli ottenuti da tali proteine contenevano soltanto il cluster [4Fe-4S] e non mostravano prove della presenza del subcluster [2Fe], nonostante diversi tentativi siano stati provati per ottenere cristalli della proteina completamente attiva. Tra questi tentativi, il soaking di cristalli di [4Fe-4S]-TmeHydF, in un cocktail di cristallizzazione contenente i leganti [2Fe], ha portato all'osservazione di una rottura progressiva dei cristalli a seguito della diffusione dei leganti. Diverse ipotesi sono state formulate per spiegare questo fenomeno: (i) la presenza del residuo glutammico rende difficile l'ingresso del complesso bi-ferrico; (ii) nell'impaccamento dei cristalli di [4Fe-4S]-TmeHydF, il contatto tra i dimeri ha luogo nell'area in cui si legano i cluster ferro-zolfo e bi-ferro, portando alla rottura del cristallo, dovuta probabilmente all'introduzione della carica negativa del sub-cluster [2Fe]; (iii) i domini GTP hanno fattori B molto più elevati rispetto agli altri due domini e ciò implica che il dominio GTP è spesso non ben ordinato e causa di disordine nel cristallo, pertanto porta alla formazione di cristalli a bassa risoluzione (Figura 1.5 b).
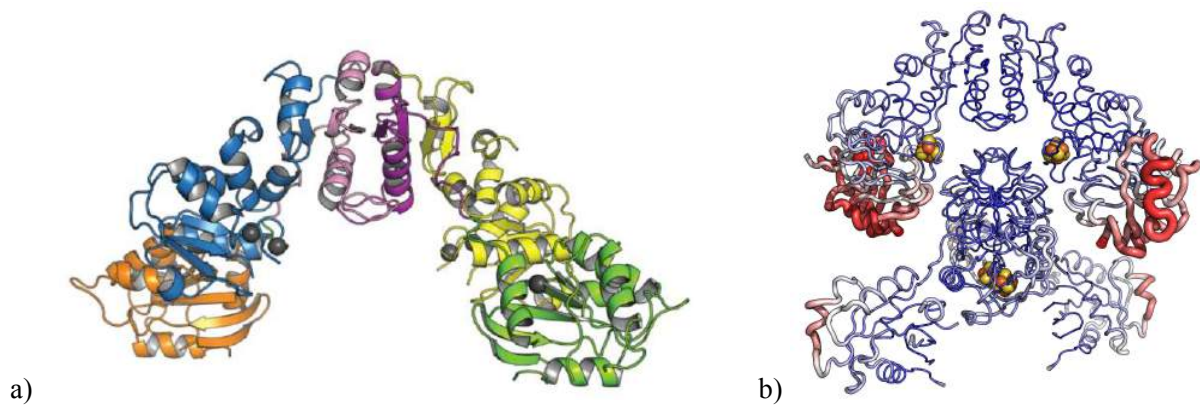
a)   b)

Figura 1.5 a) Rappresentazione della struttura di [4Fe-4S]-TmeHydF che si comporta come un dimero. Il dominio GTP è rappresentato in verde e arancione, il dominio di dimerizzazione in magenta e rosa e il dominio che lega il cluster metallico in blu e giallo. **b)** Rappresentazione dei fattori B in [4Fe-4S]-TmeHydF. I fattori B sono colorati da blu a rosso in un range [40:170] Å.[45]

Pertanto, il primo obiettivo del progetto è focalizzato sull'ottenimento della struttura di HydF che contenga sia il cluster [4Fe-4S], sia il sub-cluster [2Fe]. Per fare questo, è stata creata una forma troncata della proteina HydF da *Thermosipho melanesiensis,* eliminando il dominio GTP dalla proteina wild-type. L'idea è basata sul fatto che è stato osservato che questo dominio è spesso mal-orientato e disordinato, causa una diffrazione dei cristalli a bassa risoluzione e ostacola l'inserimento del complesso [2Fe] all'interno della struttura cristallina.

HydA from Megasphaera elsdenii

Tre sono le strutture di HydA attualmente disponibili, ottenute da tre diversi microrganismi: *Clostridium pasteurianum* (CpI)[65], *Desulfovibrio desulfuricans* (DdHydA)[66] e *Chlamydomonas reinhardtii* (CrHydA1)[67].

Le prime due strutture presentano delle analogie: il sito attivo contiene sei atomi di Fe che formano un cluster [4Fe-4S] legato ad un sub-cluster [2Fe] e che insieme costituiscono l'H-cluster. Il cluster [4Fe-4S] è legato alla proteina da quattro cisteine e il sub-cluster [2Fe] è legato al precedente da una cisteina. Inoltre, presentano dei cluster ferro-zolfo aggiuntivi che non fanno parte del sito attivo. Alcune differenze sono state osservate nel numero e tipo di cluster aggiuntivi e negli elementi che compongono la sfera di coordinazione del gruppo bi-ferrico. La struttura di CrHydA1, invece, manca del sub-cluster [2Fe] e di eventuali cluster accessori. Proprio per le sue caratteristiche di semplicità strutturale, questa proteina è stata molto sfruttata per approfondire gli studi sulla maturazione dell'H-cluster. Le tre strutture sono riportate e descritte in dettaglio in Figura 1.6.

Nonostante diverse strutture siano già disponibili e abbiano fornito informazioni importanti riguardo al comportamento delle idrogenasi [Fe-Fe], un'altra forma di HydA viene studiata. L'idrogenasi isolata da *Megasphaera elsdenii,* infatti, ha mostrato comportamenti molto interessanti dal punto di vista della sensibilità all'ossigeno. Infatti, è stato dimostrato dal laboratorio che essa ha una resistenza all'ossigeno maggiore di quella delle altre idrogenasi studiate. Perciò, siamo interessati a studiarne la struttura per capire la causa di questo rilevante comportamento della proteina in presenza di ossigeno.

Durante il mio progetto, quindi, mi sono concentrata anche sullo studio delle condizioni di cristallizzazione di una forma troncata di HydA da *M. elsdenii,* detta MeH HydA, dove tutti i cluster accessori sono stati eliminati, in modo da ottenere una proteina più semplice, in analogia con CrHydA1.
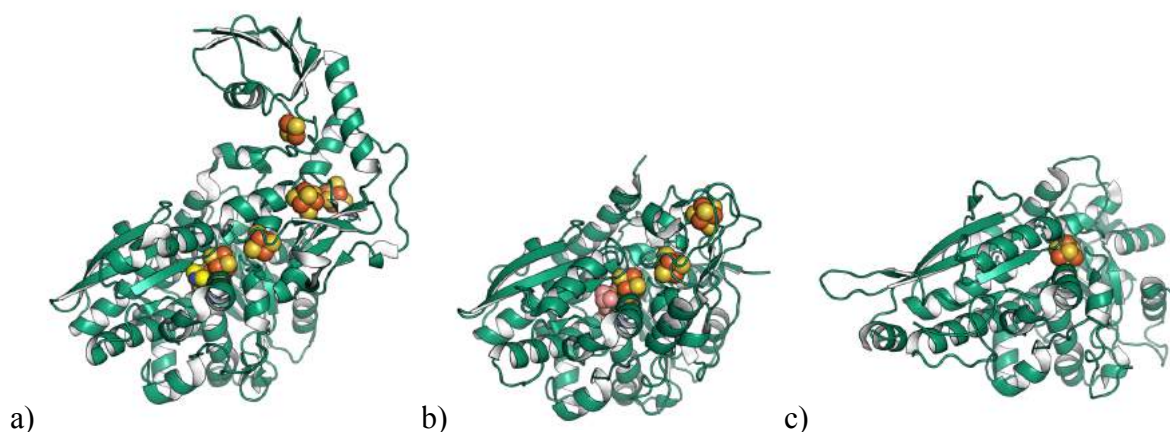
Figura 1.6 a) Struttura di CpI. Si distinguono il primo dominio, contenente l'H-cluster, composto dal cluster [4Fe-4S] e dal subcluster [2Fe], e i tre domini contenenti i quattro cluster [Fe-S] accessori. **b)** Struttura di DdHydA. Si distinguono l'H-cluster, composto dal cluster [4Fe-4S] e dal subcluster [2Fe], e i domini accessori contenenti i due cluster [Fe-S] accessori. **c)** Struttura di CrHydA1. Essa contiene solo un cluster [4Fe-4S] e manca del subcluster [2Fe] e dei cluster accessori. In ogni immagine, gli atomi di ferro sono rappresentati in giallo, quelli di zolfo in arancione.

1.2 Materiali e Metodi

1.2.1 Espressione eterologa delle proteine

Per espressione eterologa si intende l'espressione di un gene in un organismo ospite che non è in grado di produrlo naturalmente. Il gene viene inserito all'interno del sistema di espressione con la tecnica del DNA ricombinante. Il sistema di espressione deve avere una proprietà, nota come competenza, con cui si intende la capacità di una cellula di acquisire DNA extracellulare dal suo ambiente.

La trasformazione è l'operazione che consente alle cellule competenti di acquisire il DNA extracellulare, attraverso la membrana plasmatica. La sequenza di DNA che deve essere amplificata è contenuta in un plasmide, cioè una piccola molecola di DNA circolare che contiene diversi elementi: l'origine di replicazione dove ha inizio la duplicazione del DNA, il gene che codifica per la resistenza all'antibiotico (necessaria per distinguere le cellule che hanno acquisito il plasmide dalle altre) e il gene di interesse preceduto da un promotore (Figure 3.1). La trasformazione è stata usata per produrre tutte le proteine usate in questo studio. 1 μL di DNA plasmidico è stato trasferito in 100 μL di cellule competenti e le cellule sono state lasciate 30 minuti nel ghiaccio. Le cellule sono poi state sottoposte a uno shock termico a 42°C per 45 secondi e incubate 2 minuti a 0°C. Sono poi state unite a 400 μL di Lisogeny Broth (LB) caldo a 37°C e incubate a 37°C per 50 minuti, mantenendo un'agitazione continua. LB è un terreno complesso contenente triptone come fonte di peptidi, estratto di lievito come fonte di vitamine e cloruro di sodio come fonte di ioni sodio per il trasporto e il bilancio osmotico. Le cellule poi sono state centrifugate per 5 minuti a 2000 g e parte dei surnatanti è stata eliminata per concentrare le cellule. Il precipitato è stato risospeso all'interno dei surnatanti e il 20 e l'80% del prodotto di trasformazione sono stati caricati su due piastre agar contenti 100 mg/mL di ampicillina e incubati a 37°C per tutta la notte. Nel caso di proteine prodotte con l'utilizzo di chaperons molecolari, alle cellule competenti e al plasmide contenente il DNA della proteina da esprimere, è stata aggiunta anche una pari quantità di plasmide contenente il DNA della proteina chaperon. Inoltre, le piastre agar usate contenevano ampicillina e cloramfenicolo, antibiotico a cui il plasmide per la proteina chaperon è resistente.

Il prodotto di trasformazione è stato utilizzato come primo inoculo per la precoltura. Dalle piastre sono state estratte tre colonie, inserite in 130 mL di medium di coltura contenente antibiotico e lasciate a 37°C tutta la notte (maggiori informazioni sul terreno di coltura usato per ciascuna proteina sono riportate in Table 3.1).

Per dare inizio alla coltura, per ogni litro di medium è stata aggiunta una quantità di precoltura tale che la densità ottica misurata a 600 nm fosse compresa tra 0.05 e 0.08. Le cellule sono state lasciate a

37°C con agitazione di 200 rpm negli incubatori INFORS e la crescita della biomassa è stata seguita attraverso la densità ottica a 600 nm. Quando questa ha raggiunto il valore di 0.5-0.6, l'espressione della proteina è stata indotta con l'aggiunta di Isopropyl β-D-1-thiogalactopyranoside (IPTG), un metabolita del lattosio, in grado di attivare la trascrizione dell'operone *lac*. Le cellule sono state incubate 5 ore a 20°C, poi separate per centrifugazione a 6000 g e 4°C per 15 minuti e poi conservate a -20°C. La resa è solitamente attorno a 2 g di cellule per litro di coltura. Nel caso della produzione di proteine con chaperons molecolari, quando la OD_{600} ha raggiunto un valore pari a 1.3, è stata indotta l'espressione del chaperon con arabinosio a 2 mg/mL. Quando OD_{600} ha raggiunto 1.8, i campioni sono stati incubati a 4°C per il tempo necessario a raggiungere una temperatura di 16°C nell'incubatore e, a quel punto, è stata indotta anche l'espressione della proteina di interesse con IPTG.

1.2.2 Purificazione aerobica

Poiché il prodotto desiderato è una proteina intracellulare, è stato necessario rompere le cellule per purificarlo. Le cellule, quindi, sono state sospese in un tampone di lisi e sonicate in ghiaccio per 10 minuti, in modo da causare una rottura per cavitazione dovuta alle onde sonore emesse ad elevata frequenza dal sonicatore. Per promuovere la lisi, sono stati aggiunti benzonasi e magnesio. Il prodotto di lisi è stato ultracentrifugato a 190000 g a 4°C per 45 minuti, in modo da separare i detriti cellulari, e il precipitato è stato eliminato.

Per purificare l'enzima sono state usate tecniche cromatografiche. Il primo step di purificazione è stato effettuato attraverso una cromatografia di affinità (HisTrap crude FF) che contiene ioni Ni^{2+} che possono legare selettivamente il tag di 6 istidine (6-his tag) contenuto dalla proteina di interesse, mentre le proteine che non lo possiedono passano attraverso la colonna. La proteina viene poi eluita con un gradiente di imidazolo. Le frazioni contenenti la proteina di interesse, identificate attraverso SDS-PAGE, sono state estratte, stabilizzate con DTT, concentrate tramite ultrafiltrazione e ulteriormente stabilizzate con glicerolo, prima di essere congelate a -80°C. Successivamente, il campione è stato sottoposto a cromatografia di esclusione molecolare (Superdex S200 26/600), in grado di separare le proteine sulla base del loro peso molecolare. Le frazioni contenenti la proteina di interesse sono state isolate, concentrate per ultrafiltrazione e conservate a -80°C. La concentrazione è stata misurata attraverso l'analisi Bradford.

1.2.3 Ricostituzione e purificazione anaerobiche

La ricostituzione del cluster [4Fe-4S] è stata condotta in condizioni strettamente anaerobiche in una glove box M. Brown contenente meno di 0.5 ppm di ossigeno. Un'aliquota di proteina a 100 μM è stata trattata con DTT a 10 mM per 10 minuti, in modo da avere tutte le cisteine in forma ridotta. Successivamente, sono stati aggiunti Fe(II) dal sale di Mohr e L-cisteina in eccesso molare di 6, rispettivamente come fonti di ferro e di zolfo, e una quantità catalitica di cisteina-desulfurasi CsdA, per favorire il trasferimento dello zolfo dalla cisteina alla proteina. La purificazione della proteina ricostituita è stata effettuata in condizioni anaerobiche e con un unico step cromatografico di esclusione molecolare (Superdex S200 10/300), dopo aver centrifugato la proteina per eliminare l'eccesso di reagenti e dopo averla concentrata con un concentratore Vivaspin con cut-off di 10 kDa. Attraverso uno spettro UV-Vis è stato valutato il rapporto tra assorbanze a 400 e a 200 nm.

1.2.4 Maturazione chimica e trasferimento del sub-cluster [2Fe] da HydF a HydA

La proteina [4Fe-4S] è stata sottoposta a maturazione chimica, in modo da incorporare il sub-cluster [2Fe]. Come precursore del complesso bi-ferrico sono stati usati $(Et_4N)_2[Fe_2(adt)(CO)_4(CN)_2]$ (complesso adt) o $(Et_4N)_2[Fe_2(pdt)(CO)_4(CN)_2]$ (complesso pdt). La proteina è stata lavata con un tampone privo di DTT (fosfato di potassio a 100 mM a pH 6.8) per evitare che l'agente riducente chelasse i metalli. Poi, la proteina è stata incubata con il complesso adt o pdt in eccesso molare di 10, in condizioni anaerobiche e, dopo un'ora, l'eccesso di reagenti è stato rimosso attraverso una colonna desalinizzante NAP-10. L'incorporazione del sub-cluster [2Fe] è stata provata paragonando gli spettri

UV-Vis dell'idrogenasi [4Fe-4S] e dell'idrogenasi incubata con adt/pdt. La proteina ricostituita [2Fe]-ΔTmeHydF è stata usata per testarne l'attività catalitica o l'abilità a trasferire il sub-cluster [2Fe] a HydA. Per quest'ultimo esperimento, [2Fe]-ΔTmeHydF in eccesso molare di 10 è stato incubato con [4Fe-4S]-MeHydA e l'attività è stata testata dopo 15 minuti per verificare il corretto trasferimento del sub-cluster.

1.2.5 Biologia molecolare

Tecniche di biologia molecolare è stata utilizzata per eliminare il dominio GTP dalla proteina TmeHydF. Per farlo, è stato necessario eliminare la parte del gene che codifica per il dominio GTP. Quindi, per prima cosa sono stati progettati i primer necessari, ordinati da Eurofins (le sequenze di DNA ordinate sono riportate in Table 3.2), e poi il gene è stato modificato attraverso una Polymerase Chain Reaction (PCR) che consente di amplificare una regione di una molecola di DNA. Per la PCR occorre fornire i primer che forniscono le due estremità in cui ha inizio la replicazione del DNA, il DNA che deve essere amplificato, la DNA polimerasi e i nucleotidi. La PCR è basata su cicli ripetuti di denaturazione del DNA, ibridazione dei primer e sintesi del DNA grazie alla polimerasi (vedere Table 3.3). La presenza del gene è stata verificata attraverso elettroforesi su gel di agarosio.

Le operazioni che consentono di ottenere il gene desiderato sono schematizzate in Figura 1.7. Il prodotto della PCR è un doppio filamento di DNA lineare ed è quindi necessario ricreare la forma circolare del plasmide, per poterlo inserire in un sistema di espressione. Questa operazione è effettuata incubando il prodotto della PCR con ligasi (che unisce filamenti di DNA), chinasi (che effettua la fosforilazione) e l'enzima di restrizione DpnI (che elimina il DNA stampo). Il plasmide così ottenuto può quindi essere incorporato in un sistema di espressione (*E. coli)*, attraverso una trasformazione.
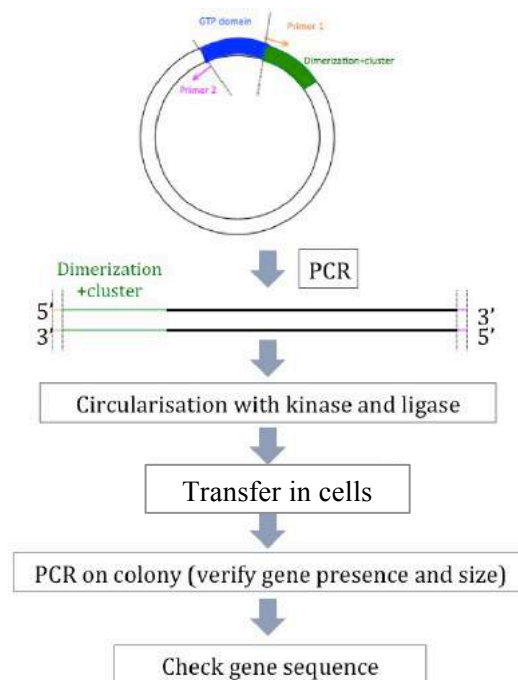
Figura 1.7 Rappresentazione schematica del processo di modifica del gene di TmeHydF attraverso biologia molecolare.

Successivamente, occorre verificare quali colonie batteriche hanno assunto il nuovo plasmide e contengono, quindi, il gene per la proteina troncata. La verifica è stata effettuata con una PCR sulle colonie, unendo la colonia selezionata ai primer, alla DNA polimerasi e ai nucleotidi. Il prodotto della PCR è stato verificato tramite elettroforesi su gel di agarosio e, dalle colonie contenenti un filamento di DNA della giusta dimensione, è stato purificato il plasmide, per verificarne la sequenza.

1.2.6 Test di espressione, solubilità e termostabilità

150 mL di precoltura sono stati diluiti in 100 mL di LB contenente ampicillina, in modo che la OD_{600} iniziale fosse 0.1. Dopo l'induzione dell'espressione della proteina con IPTG, sono state prelevate aliquote da 500 µL ogni ora per 4 ore e ciascuna è stata centrifugata a 6000 g per 5 minuti a 4°C. Il precipitato è stato dissolto in 50 µL di SDS-PAGE e riscaldato a 96°C per 5 minuti. Questi campioni sono stati usati per il test di espressione. Dopo 4 ore, sono state prelevate due aliquote da 2 mL e centrifugate nelle stesse condizioni precedenti. Il precipitato è stato dissolto in 600 µL di tampone di lisi a cui sono stati aggiunti $MgCl_2$ e benzonasi. Dopo 15 minuti di incubazione a temperatura ambiente, i campioni sono stati centrifugati a 18000 g per 15 minuti. Il primo campione è stato usato per il test di solubilità: il precipitato è stato dissolto nel tampone di lisi e sonicato 5 minuti; è stato poi aggiunto SDS-PAGE sia ai surnatanti che al prodotto sonicato. Il secondo campione, invece, è stato utilizzato per testare la termostabilità della proteina: tutti i surnatanti sono stati incubati a 50°C e sono state prelevate aliquote da 50 µL ogni 5 minuti per 20 minuti. Ogni aliquota è stata centrifugata a 18000 g per 5 minuti. I surnatanti sono stati uniti a SDS-PAGE, mentre il precipitato è stato risospeso nel buffer di lisi e poi aggiunto all'SDS-PAGE. Tutti i campioni sono stati scaldati a 96°C per 5 minuti e caricati su un gel elettroforetico con SDS-PAGE.

1.2.7 Dosaggio del ferro con il metodo di Fish

Per determinare la quantità di ferro contenuta nelle proteine, è stato usato il metodo di Fish. Prima di effettuare la misura, è stato necessario costruire una curva di calibrazione, partendo da una soluzione commerciale standard di ferro a 1000 mg/mL, diluita 36 volte per ottenere una soluzione a circa 0.5 mM. Tale soluzione è stata diluita in acqua in modo da avere diversi campioni nell'intervallo 0-10 nmol. Il test è stato effettuato con diverse quantità di proteina, tra 0.5 e 4 nmol, a seconda della quantità di ferro stimata. La proteina è stata incubata 15 minuti a temperatura ambiente con 45 µL di una soluzione 1 M di acido perclorico (che denatura le proteine e rilascia il ferro) e centrifugata 5 minuti. 90 µL di surnatanti sono stati incubati con 72 µL di acido disolfonico di batofenantrolina a 1.7 mg/mL (con la funzione di chelare il ferro), 36 µL di ascorbato di sodio a 38 mg/mL (con la funzione di ridurre il ferro in forma ionica Fe^{2+}) e 27 µL di acetato d'ammonio (in grado di neutralizzare l'ambiente e rivelare il complesso rosso Fe-batofenantrolina). Dopo 30 minuti di incubazione, i campioni sono stati centrifugati 5 minuti a 10000 rpm e analizzati con spettroscopia UV-Vis, misurando la differenza di assorbanza tra 535 e 680 nm. La quantità di ferro contenuta è stata calcolata attraverso la retta di calibrazione.

1.2.8 Elettroforesi su gel

Elettroforesi su gel di poliacrilammide con agente denaturante laurilsolfato di sodio (SDS-PAGE)

L'elettroforesi su gel di poliacrilammide è una tecnica usata per separare le macromolecole biologiche, in base alla loro mobilità elettroforetica. La mobilità dipende dalla lunghezza, dalla conformazione e dalla carica delle molecole. Con l'aggiunta dell'agente denaturante SDS-PAGE, la struttura nativa della molecola viene distrutta e tutte le molecole si ricoprono della stessa carica. In questo modo la mobilità delle molecole dipende soltanto dalla dimensione e dal rapporto tra massa e carica. Quindi le molecole, lineari e cariche negativamente, si muovono verso l'anodo se soggette ad un campo elettrico, con una velocità che dipende solo dal loro peso molecolare. Sul gel sono stati caricati i campioni, assieme all'SDS-PAGE running buffer, e una miscela di proteine a peso molecolare noto, per identificare la proteina di interesse attraverso una curva di calibrazione.

Elettroforesi su gel di agarosio

L'elettroforesi su gel di agarosio è di solito utilizzata per separare una popolazione mista di macromolecole, di solito frammenti di DNA, in una matrice di agarosio 0.8%. Il DNA, a causa della

presenza di gruppi fosfato, è carico negativamente, perciò in presenza di un campo elettrico migra verso l'anodo. Le molecole di DNA vengono, quindi, separate in base alla loro dimensione, compiendo un percorso che è inversamente proporzionale al logaritmo del peso molecolare della molecola stessa. Dopo la separazione, i frammenti di DNA possono essere visualizzati attraverso luce UV, dopo aver colorato il gel con un apposito colorante.

1.2.9 Determinazione della concentrazione di una proteina

Per determinare la concentrazione di una proteina è stato utilizzato il metodo colorimetrico di Bradford, basato sul legame che si forma tra un colorante (Coomassie Brilliant Blue G-250) e la proteina in condizioni acide, che porta al viraggio del colorante dalla forma rossa a quella blu. All'interno del range lineare del saggio, esiste una relazione proporzionale tra il colorante e la proteina (maggiore è la quantità di proteina, maggiore è l'intensità del colore blu del colorante). Si misura l'assorbanza a 595 nm e la si compara con quella di una proteina standard (albumina di siero bovino). La concentrazione ottenuta con il saggio Bradford è stata confrontata con quella determinata attraverso il coefficiente di estinzione molare (ε_{280nm}).

1.2.10 Test di attività

Gascromatografia

La gascromatografia è stata usata per misurare elevate quantità di idrogeno, per esempio nella misura dell'attività specifica di MeH HydA e di MeHydA. Il gascromatografo è uno strumento usato per separare sostanze chimiche tra una fase stazionaria e una fase mobile gassosa. Ogni componente ha un tempo di ritenzione diverso, in base alla velocità con cui si muove lungo la colonna. All'uscita, un rivelatore è in grado di identificare elettronicamente ogni componente. I campioni per il test di attività sono stati preparati in fiale da 10 mL, sigillate con tappi di gomma in condizioni rigorosamente anaerobiche, in condizioni in cui HydA è in grado di catalizzare la produzione di idrogeno: HydA completamente ricostituita a 2-3 µM, ditionito di sodio a 100 mM come donatore di elettroni, metil viologeno a 10 mM come mediatore di elettroni e tampone fosfato di potassio a 100 mM a pH 6.8. Il campione così ottenuto è stato utilizzato per registrare un gascromatogramma attraverso un sistema GC (Shimadzu GC-2014). La quantità di idrogeno è stata determinata attraverso una curva di calibrazione. Il gascromatografo restituisce il valore dell'area di un picco da cui è possibile calcolare le moli di idrogeno prodotte nella fiala e l'attività specifica, secondo le equazioni 1.2 e 1.3 (dove n è il numero di moli):

$$n_{H2} \ nella \ fiala = \frac{area \ picco}{gradiente \ di \ calibrazione} * \frac{[volume \ totale \ fiala - volume \ liquido \ nella \ fiala]}{volume \ iniettato} \qquad (1.2)$$

$$attività \ specifica = \frac{n_{H2 \ nella \ fiala}}{peso \ molecolare * n_{catalizzatore} * tempo} \qquad (1.3)$$

Per attività specifica si intende la quantità di prodotto formata dall'enzima in un dato intervallo di tempo e in determinate condizioni rispetto alla massa totale di proteina (µmol di prodotto/min/mg di proteina).

Sensore di tipo Clark

L'elettrodo di Clark è stato usato per la misura di piccole quantità di idrogeno, nel caso dell'H_2 prodotto da HydF. È una tecnica amperometrica basata sulla misura della pressione parziale di idrogeno. L'idrogeno diffonde attraverso una membrana di silicone verso un anodo di platino, dove H_2 è ossidato. Il sensore è connesso ad un picoamperometro dove l'anodo è polarizzato rispetto al catodo interno in Ag/AgCl e che trasforma la corrente derivata dall'ossidazione dell'idrogeno in un segnale. Si può avere, quindi, una misura continua della quantità di idrogeno prodotto in funzione del tempo. Il sensore deve essere polarizzato a 1000 mV prima di effettuare le misure e occorre costruire una retta di calibrazione attraverso due punti corrispondenti a 0 e 100% di idrogeno puro a 20°C. I campioni sono stati preparati in condizioni anaerobiche e sigillati con silicone e olio di paraffina, unendo olo-

ΔTmeHydF 4 μM, metil viologeno 4 mM, ditionito di sodio 4 mM e il tampone fosfato di potassio a pH 6.0 contenente NaCl 100 mM. Il ditionito di sodio è stato aggiunto con una siringa subito prima dell'inizio della misura. Il test restituisce un diagramma che mostra l'andamento della micromolarità dell'idrogeno nel tempo. Da questo diagramma è possibile valutare l'attività come turnover number (TON) o come turnover frequency (TOF), secondo le equazioni 1.4 e 1.5.

$$TON = \frac{\Delta \mu M \ H_2}{\mu M \ enzima} \tag{1.4}$$

$$TOF = \frac{TON}{unità \ di \ tempo} \tag{1.5}$$

Il TON, che rappresenta il numero di volte che ogni enzima effettua il suo ciclo catalitico per secondo, e il TOF, che rappresenta il numero di cicli catalitici che si hanno ad ogni sito attivo per unità di tempo, possono essere correlati all'attività specifica, secondo la relazione 1.6:

$$attività \ specifica = \frac{TOF}{peso \ molecolare \ dell'enzima} \tag{1.6}$$

1.2.11 Diffusione dinamica della luce (DLS)

La diffusione dinamica della luce è una tecnica usata per misurare la dimensione delle particelle. Essa è basata sulla misura del moto browniano delle particelle, definito da una grandezza nota come coefficiente di diffusione traslazionale D e correlabile con la dimensione delle particelle attraverso l'equazione di Stokes-Einstein (1.7):

$$d(H) = \frac{kT}{3\pi\eta D} \tag{1.7}$$

($d(H)$ è il diametro idrodinamico, k è la costante di Boltzmann, T è la temperatura assoluta, η è la viscosità e D è il coefficiente di diffusione traslazionale). Il diametro idrodinamico rappresenta il diametro di una sfera che ha la stessa velocità di diffusione traslazionale della particella.

Le misure di DLS misurano, quindi, la velocità a cui fluttua l'intensità della luce diffusa; l'intensità del segnale è paragonata con se stessa a differenti intervalli di tempo attraverso un correlatore. I correlogrammi forniscono un gran numero di informazioni: il tempo a cui la correlazione inizia a decadere in modo significativo è un'indicazione della misura delle particelle del campione; la pendenza della retta indica il grado di polidispersità del campione (più è elevata la pendenza, più il campione è monodisperso). Dalla funzione di correlazione è possibile ottenere la dimensione delle particelle, attraverso il metodo cumulante che consente un fitting dei dati con un esponenziale singolo per ottenere la taglia media e l'indice di polidispersità, oppure attraverso un fitting con esponenziale multiplo per ottenere la distribuzione della dimensione delle particelle. Attraverso la teoria di Mie è possibile anche convertire la distribuzione della dimensione in una distribuzione dei volumi.

Lo strumento usato per le misure di DLS è un Zetasizer Nano (i cui elementi principali sono riportati in Figure 3.12). La proteina alla concentrazione desiderata è stata usata per preparare i campioni. Ogni campione è stato filtrato con filtro Millipore 0.1 μm per eliminare le impurezze e poi trasferito in una cuvetta in quarzo (ZEN2112).

1.2.12 Cristallizzazione delle proteine

La cristallizzazione è il processo che porta alla formazione di un cristallo, cioè un solido in cui gli atomi sono organizzati in strutture microscopiche ordinate. Una condizione necessaria per la cristallizzazione è la sovrasaturazione della soluzione di partenza. La cristallizzazione avviene in due fasi: la nucleazione è la nascita di nuclei a partire da una soluzione omogenea, mentre l'accrescimento è l'aumento delle dimensioni di un nucleo stabile già formato. Il diagramma di fase della cristallizzazione (Figure 3.13) mostra che, partendo da una soluzione sovrasatura, possono essere raggiunte tre diverse zone, a seconda del valore di diversi parametri. La zona di nucleazione si trova tra quella metastabile e quella di precipitazione, ma spesso è molto difficile ottenere cristalli in

maniera diretta, senza passare dalla zona di precipitazione o da quella metastabile (spesso la separazione di fase può essere un'indicazione della zona metastabile).

Il metodo usato per l'ottenimento di cristalli è la diffusione di vapore. Un serbatoio contenente i precipitanti assorbe acqua dalla goccia di cristallizzazione, portando la soluzione in condizioni di sovrasaturazione.

Il campione deve essere preparato in maniera molto accurata, evitando la presenza di impurezze che possono influenzare il processo di cristallizzazione. Alcuni campioni, a causa della loro sensibilità all'ossigeno, sono stati preparati in condizioni anaerobiche, in una glove box equipaggiata per la cristallizzazione delle proteine. Per le proteine non ricostituite è stato usato un altro macchinario (Mosquito di TTB Labtech) in condizioni aerobiche.

1.2.13 Spettroscopia UV-Visibile

La spettroscopia UV-Visibile è una tecnica analitica usata per una determinare quantitativamente differenti analiti. Le molecole contenenti elettroni di tipo π o non legati possono assorbire energia in forma ultravioletta o visibile (da 150 a 800 nm), causando un'eccitazione degli elettroni in orbitali molecolari a maggiore energia. La lunghezza d'onda di assorbimento dipende sia dalla natura chimica della molecola, sia dall'ambiente molecolare. L'assorbanza ottica delle proteine è misurata di solito tra 275 e 280 nm e a tali lunghezze d'onda l'assorbimento è dovuto principalmente a tre amminoacidi: triptofano, tirosina e cisteina. L'assorbanza è legata all'intensità della luce e dipende linearmente dalla concentrazione, secondo la legge di Lambert-Beer (1.8):

$$A = \epsilon * c * l \tag{1.8}$$

dove c è la concentrazione molare, l è il cammino ottico e ϵ è il coefficiente di assorbimento molare.

Tutte le misure sono state effettuate con Cary UV-Vis (Agilent) e una cuvetta in quarzo da 1 cm.

1.3 Risultati e discussione

1.3.1 Una forma troncata di HydF da *Thermosipho melanesiensis*

La prima parte del progetto è focalizzata sullo studio della maturasi TmeHydF, con l'obiettivo di ottenere la struttura cristallina della proteina completa del cluster [4Fe-4S] e del sub-cluster [2Fe]. Per ottenere questo obiettivo è stata progettata una forma troncata della proteina, eliminando il dominio GTP, fonte di disordine per il cristallo.

Per il design di tale proteina si è ricorso alle tecniche di biologia molecolare, che hanno permesso di eliminare i residui 1-169 di TmeHydF. Dopo alcuni tentativi falliti, la proteina troncata è stata espressa secondo un protocollo di induzione tardiva e aggiungendo una proteina chaperon (groES-groEL). Sulla proteina così ottenuta, chiamata ΔTmeHydF, sono stati effettuati dei test di espressione e solubilità (Figura 1.8, a sinistra) che hanno dimostrato che la proteina è ben espressa ed è solubile, ma ha perso la termostabilità che caratterizzava invece la forma wild-type, indicando che tale proprietà è collegata al dominio GTP. Per facilitare le operazioni di purificazione, è stato inserito all'interno della proteina un tag di sei istidine, in modo da poter utilizzare una cromatografia di affinità a ioni nickel come primo step di purificazione. Poiché il tag era già presente nel costrutto del plasmide, l'operazione è stata semplice e si è ricorso nuovamente alla tecnica della biologia molecolare. Esprimendo la proteina nelle stesse condizioni precedenti, si è ottenuto un prodotto solubile, come mostrato nei gel elettroforetici in Figura 1.8 in centro e a destra. E' stato, inoltre, osservato che è possibile ottenere una proteina solubile anche senza la presenza della proteina chaperon.

Il plasmide ottenuto per ΔTmeHydF con il tag di istidine è stato usato per l'espressione eterologa della proteina in *E. coli*, seguendo il protocollo precedentemente riportato e ottenendo una proteina solubile e ben espressa. Tale proteina è stata purificata con due step cromatografici, una cromatografia di affinità e una cromatografia di esclusione molecolare. Queste operazioni, condotte in aerobiosi, hanno

consentito di ottenere una proteina quasi completamente pura e omogenea, con una resa pari a 1.5 mg di proteina per grammo di cellule bagnate.
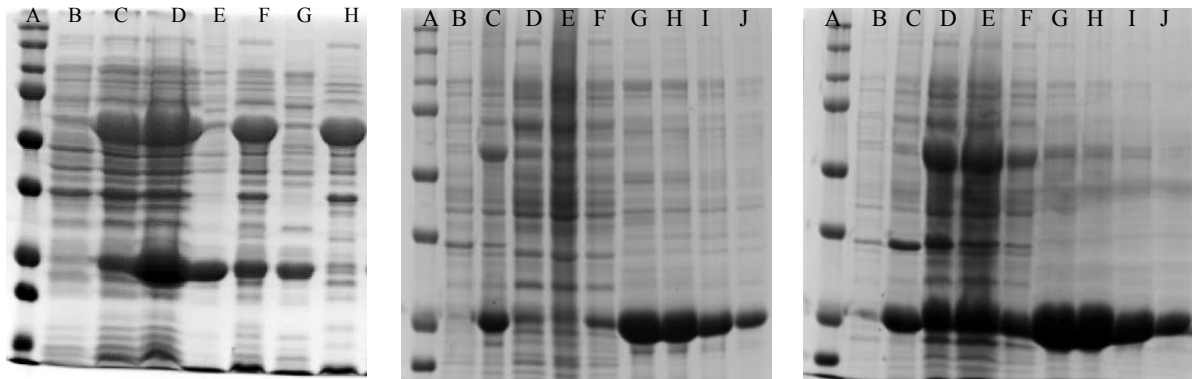

Figura 1.8 Elettroforesi su gel di poliacrilammide con SDS-PAGE. A sinistra, proteina senza 6his-tag: A) standard a pesi molecolari noti, B) Controllo: prima dell'induzione, C) espressione della proteina dopo un'ora dall'induzione, D) espressione della proteina dopo tutta la notte, E) precipitato dopo la lisi cellulare, F) surnatanti dopo lisi cellulare, G) precipitato dopo test di termostabilità (5 minuti a 50°C), H) surnatanti dopo test di termostabilità. In centro, proteina con his-tag espressa senza chaperon e a destra proteina con his-tag espressa con chaperon: (A) standard a pesi molecolari noti, (B) Controllo: proteina prima dell'induzione, (C) precipitato dopo lisi cellulare, (D) surnatanti dopo lisi cellulare, (E) flowthrough (F) espressione della proteina dopo 30 minuti dall'induzione (G) espressione della proteina dopo 100 minuti dall'induzione (H) espressione della proteina dopo 200 minuti dall'induzione (I) espressione della proteina dopo 350 minuti dall'induzione (J) espressione della proteina dopo 500 minuti dall'induzione.

Dopo aver valutato qualitativamente tramite spettroscopia UV-Vis che la quantità di ferro contenuta all'interno della proteina purificata aerobicamente è minima, risulta necessario ricostituire il cluster [4Fe-4S] in condizioni anaerobiche. La ricostituzione è stata effettuata incubando la proteina con solfato ferroso ammonico, L-cisteina, DTT e cisteina-desulfurasi per tutta la notte e seguendo la reazione attraverso spettroscopia UV-Vis. Si osserva la crescita di una banda di assorbimento a 400-420 nm che rappresenta la formazione del cluster [4Fe-4S] durante la reazione, come mostrato in Figura 1.9 a. In seguito, la proteina è stata purificata con cromatografia di esclusione molecolare per eliminare il ferro non specifico ed eventuali aggregati. La presenza del cluster è stata confermata qualitativamente attraverso spettroscopia UV-Vis (Figura 1.9 b, curva blu) che ha permesso di valutare il rapporto tra assorbanze a 400 e a 280 nm pari a 0.33 (come nel caso di altre ricostituzioni ben riuscite) e quantitativamente con il dosaggio del ferro, che ha fornito un valore di 3.3 nmol di ferro/nmol di proteina. Tale valore è distante dal valore teorico di 4 e può essere spiegato supponendo un errore del 10% nella misura della concentrazione della proteina.

Dopo la ricostituzione, la proteina contiene soltanto il cluster [4Fe-4S], ma manca ancora del sub-cluster binucleare. Incubando la proteina [4Fe-4S] con un eccesso molare del complesso adt in condizioni anaerobiche, è stato possibile incorporare il sub-cluster biferrico all'interno della proteina stessa. L'incorporazione è stata provata attraverso spettroscopia UV-Vis, osservando la presenza di una banda di assorbimento a 350 nm che indica la presenza del sub-cluster [2Fe] (Figura 1.9 b, curva rossa). Tale banda è dovuta al trasferimento di elettroni dai leganti CO e CN al ferro contenuto nel sub-cluster biferrico.
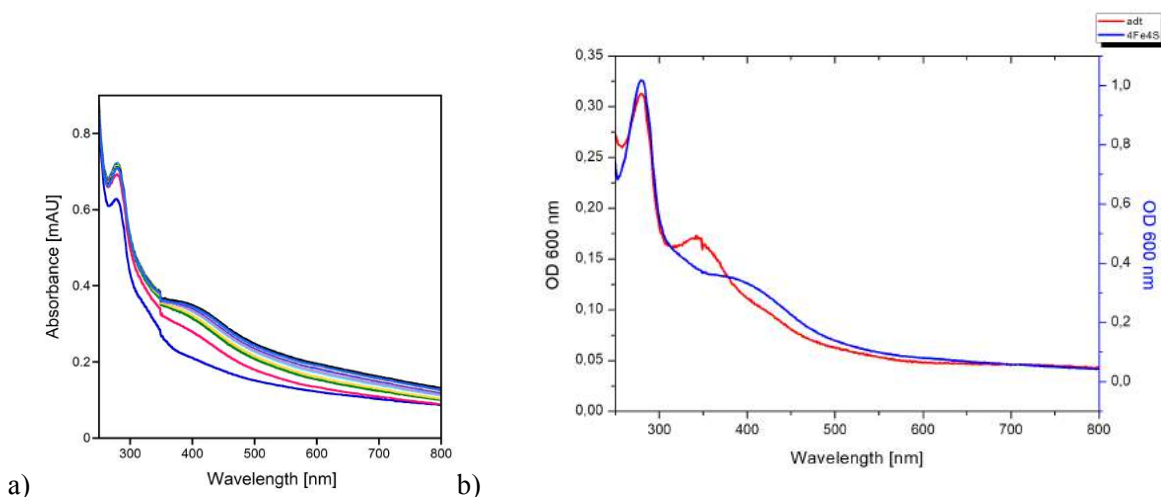
Figura 1.9 a) Spettro di assorbimento UV-Vis di ΔTmeHydF registrato ogni 20 minuti. Esso mostra la cinetica di ricostituzione del cluster [4Fe-4S]. **b)** Spettro di assorbimento UV-Vis di [4Fe-4S]-ΔTmeHydF (in blu) e di adt-ΔTmeHydF (in rosso).

Poichè TmeHydF ha la capacità di incorporare e trasferire il complesso adt a MeHydA e il dominio GTP favorisce la dissociazione di HydE con HydF[47] ma non sembra essere coinvolto nel trasferimento del subcluster [2Fe][51], è stato verificato se la forma troncata ΔTmeHydF conservi questa capacità. Adt-ΔTmeHydF è stata incubata con [4Fe-4S]-MeHydA per 30 minuti e, poi, tale soluzione di proteine è stata sottoposta ad un test di attività per monitorare la produzione di idrogeno. Attraverso gascromatografia, è stato ottenuto un valore di attività specifica di ~739 μmol H_2/min mg proteina, comparabile con quello ottenuto per MeHydA attivata direttamente col complesso adt. Questo risultato mostra che il complesso biferrico è stato correttamente trasferito da ΔTmeHydF a MeHydA e, quindi, che il dominio GTP non è necessario nell'interazione tra HydF e HydA.

La proteina è stata sottoposta ad un test di attività per monitorare la quantità di idrogeno prodotta nel tempo. Innanzitutto, [4Fe-4S]-ΔTmeHydF è stata incubata con il complesso pdt (è stato dimostrato che tale complesso fornisce attività maggiori rispetto al complesso adt per la proteina TmeHydF) in condizioni anaerobiche per un'ora. Successivamente, è stata sottoposta ad un test di attività con elettrodo di Clark ed è stato ottenuto un valore di TOF iniziale (calcolato nei primi 5 minuti) pari a 1 min^{-1}. Tale valore è 3 volte maggiore di quello ottenuto per TmeHydF attivata con il complesso pdt nelle medesime condizioni[45]. Poiché l'eliminazione del dominio GTP non coinvolge il sito attivo dell'enzima, non c'è ragione di aspettarsi un cambiamento apprezzabile dell'attività catalitica. Di conseguenza, tale risultato deve essere confermato con ulteriori esperimenti.

Attraverso la tecnica della diffusione dinamica della luce (DLS) è stato possibile osservare il comportamento della proteina in diverse condizioni. Gli aspetti principali che sono stati osservati sono i seguenti e sono riassunti in Figura 1.10:

- La proteina è stabile nel suo tampone (TRIS 50mM pH 8.0, NaCl 200mM, GlyOH 10%, DTT 5mM) ma subisce un fenomeno di aggregazione che diventa sempre più intenso al crescere della concentrazione della proteina. Infatti, il diametro del campione a 8 mg/mL è due volte maggiore di quello del campione a 2.4 mg/mL.

- Un aumento della concentrazione di proteina comporta un aumento del diametro delle particelle e, di conseguenza, una riduzione del coefficiente di diffusione, in accordo con la relazione lineare tra il coefficiente di diffusione e la concentrazione di proteina (Equation 4.1). Il fenomeno di aggregazione che si verifica al crescere della concentrazione è solo parzialmente reversibile (il diametro decresce fino a 8 nm per il campione a 1 mg/mL, mentre le dimensioni attese sono di 6 nm per il campione a 2.4 mg/mL).

- Poiché la proteina contiene molte cariche superficiali (come mostrato in Figure 4.19), la presenza di sale può influenzare la tendenza ad aggregare delle molecole proteiche. È stato

osservato che al crescere della concentrazione di NaCl nel tampone, diminuisce la dimensione della proteina, pertanto il sale stabilizza tali molecole, riducendone le interazioni intermolecolari.

- Il fenomeno di aggregazione dipende fortemente dal tampone in cui si trova la proteina. Il tampone che minimizza la tendenza all'aggregazione è Tris 50mM pH 8 con NaCl 200mM, GlyOH 10% e DTT 5mM. La pendenza della retta *d(h)*=f(*c*) (dove *d(h)* è il diametro idrodinamico e *c* è la concentrazione di proteina) è simile a quella osservata per TmeHydF nel tampone che favorisce la formazione dei cristalli.

a)

b)

c)

d)
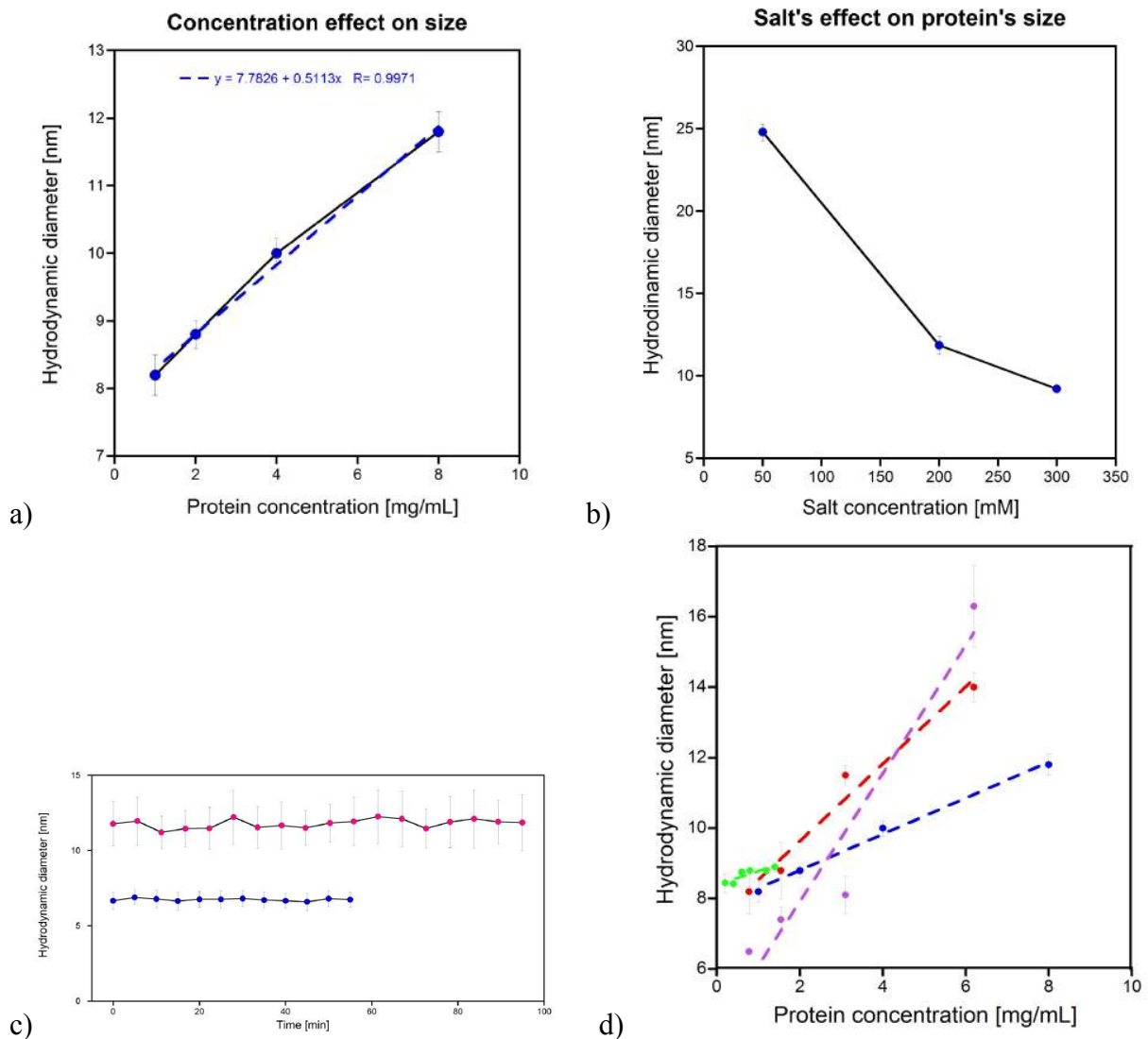
Figura 1.10 a) Variazione del diametro idrodinamico con la concentrazione della proteina nel buffer TRIS 50mM pH 8.0, NaCl 200mM, GlyOH 10%, DTT 5mM; **b)** variazione del diametro idrodinamico con la concentrazione di NaCl nel buffer TRIS 50mM pH 8.0, GlyOH 10%, DTT 5mM; **c)** variazione del diametro idrodinamico nel tempo a 8 mg/mL (curva rosa) e 2.4 mg/mL (curva blu); **d)** variazione del diametro idrodinamico con la concentrazione della proteina in diversi buffer: per ΔTmeHydF in TRIS 50mM pH 8, NaCl 200mM, GlyOH 10%, DTT 5mM (in blu), KP 25mM pH 7.5, NaCl 150mM, DTT 5mM (in rosso), HEPES 25mM pH 7.5, NaCl 200mM, GlyOH 30%, DTT 5mM (in violetto) e per TmeHydF in HEPES 25mM pH 7.5, NaCl 200mM, DTT 5mM (in verde).

Infine, sono stati effettuati diversi tentativi di cristallizzazione della proteina ΔTmeHydF, sia in condizioni anaerobiche (con la forma [4Fe-4S]-ΔTmeHydF) sia in condizioni aerobiche (con la forma apo). Diversi kit commerciali sono stati testati con la proteina a 7.5 e 16 mg/mL, ma è stato osservato un fenomeno di precipitazione reversibile a 16 mg/mL che indica che la proteina in quelle condizioni è

troppo concentrata e ha un'elevata tendenza ad aggregare. Nonostante questa osservazione, è stato possibile ottenere cristalli di [4Fe-4S]-ΔTmeHydF nel kit MbClass II in tre differenti condizioni. Tuttavia, i cristalli ottenuti sono incolori e questo è dovuto a due possibili ragioni: tutte queste condizioni sono a pH 5.6 e contengono citrato che può chelare il ferro del cluster; inoltre, i cristalli sono apparsi dopo lungo tempo, portando verosimilmente alla degradazione del cluster.

1.3.2 Una forma troncata di HydA da *Megasphaera elsdenii*

La seconda parte del progetto è focalizzata sullo studio della forma troncata MeH HydA, sulla quale diversi studi sono già stati effettuati. Durante il mio tirocinio, ho lavorato su alcune modifiche al protocollo di espressione e purificazione in modo da aumentare l'efficienza e la velocità di produzione.

Il protocollo di espressione già stabilito prevedeva l'utilizzo del plasmide pt7-7-MeH-HydA per l'espressione eterologa della proteina in *E. coli* e l'induzione dell'espressione proteica con IPTG in corrispondenza di una OD_{600} di 0.5. La proteina ottenuta con questo metodo risulta ben espressa e solubile, ma la crescita della biomassa osservata presenta una fase di latenza molto lunga. La proteina è, in seguito, stata purificata aerobicamente con due step cromatografici, una cromatografia di affinità (HiTrap chelating) e una cromatografia di esclusione molecolare (Superdex S200 26/60), ottenendo una proteina omogenea e pura. Oltre alla crescita lenta delle cellule, è stato riscontrato un altro problema, cioè un basso valore di resa finale (2.1 mg di proteine per grammo di cellule).

Pertanto, è stato necessario migliorare tale processo. La crescita cellulare lenta può essere dovuta a due motivi: è presente una mutazione del plasmide oppure la proteina è tossica per le cellule. Come prima verifica, è stato testato un nuovo plasmide, il pET22b, in cui è stato inserito il gene per MeH HydA e attraverso cui è stata effettuata la trasformazione delle cellule competenti. Seguendo lo stesso protocollo di coltura, è stata ottenuta una crescita delle cellule molto più veloce, come mostrato in Figura 1.11.
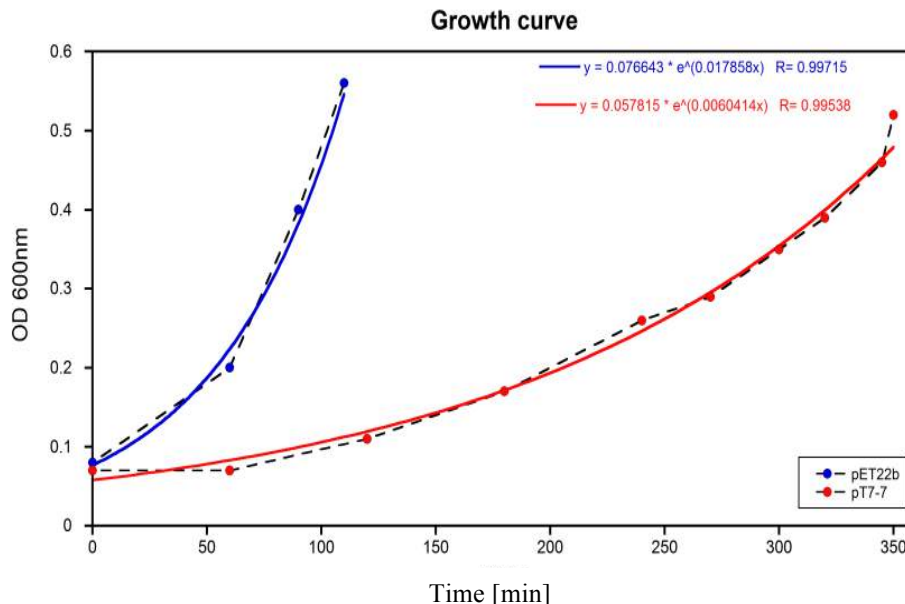
Figura 1.11 Crescita delle cellule TunerDE3 trasformate col plasmide pt7-7-MeH-HydA in rosso e di TunerBL21 DE3 trasformate con pET22b in blu.

I test di espressione e solubilità hanno confermato che la proteina ottenuta è solubile e ben espressa. Il processo di purificazione è stato leggermente modificato: come cromatografia di affinità è stata utilizzata una HisTrap FF crude che ha maggiore capacità di legame e maggiore resistenza agli agenti riducenti, mentre come cromatografia di esclusione molecolare è stata usata la stessa colonna (Superdex S200 26/60). La resa finale con il nuovo plasmide è pari a 6.6 mg di proteina per grammo

di cellule. Di conseguenza, il cambiamento del plasmide ha portato a due risultati importanti: la riduzione del tempo di crescita cellulare e la triplicazione della resa.

Anche per questa proteina, è stato necessario ottenere la forma attiva effettuando una ricostituzione del cluster [4Fe-4S], seguendo il protocollo standard. L'effettiva ricostituzione del cluster è stata verificata qualitativamente attraverso uno spettro UV-Vis da cui è stato calcolato un valore del rapporto tra assorbanze 400/280 nm pari a 0.23, contenuto nel range riportato da esperimenti precedenti. Inoltre, la valutazione quantitativa attraverso il metodo di Fish ha fornito un valore di 3.6 nmol di ferro/nmol di proteina, molto vicino al valore teorico di 4.

Inoltre, in condizioni anaerobiche la proteina è stata attivata attraverso un'incubazione di 30 minuti con un eccesso molare del complesso adt. L'incorporazione è stata verificata attraverso gascromatografia, ottenendo un valore di attività specifica per [2Fe]-MeH HydA pari a 133+/-28 µmol H_2/min mg di enzima.

Al fine di studiare la struttura cristallina di MeH HydA, sono stati effettuati diversi tentativi di cristallizzazione. Fino ad ora, non è stato possibile ottenere cristalli, ma sono stati osservati alcuni trend che possono essere studiati più approfonditamente. La proteina, infatti, tende a precipitare a pH inferiori a 6.0, ma l'aggiunta di cloruro o nitrato d'ammonio può evitare la precipitazione anche a bassi pH. Inoltre, anioni e cationi possono indurre la precipitazione o stabilizzare la proteina, a seconda della loro concentrazione. L'attenzione è stata focalizzata soprattutto su due ioni: Zn^{2+} perché è stato osservato che può indurre la precipitazione della proteina; Cd^{2+} perché è noto come ottimo legante per i gruppi carbossilici ed è noto che la proteina contiene in superficie un elevato numero di carbossili, come confermato dal basso pI. In realtà, entrambi gli ioni hanno portato alla precipitazione di MeH HydA anche a basse concentrazioni.

1.3.3 HydA da *Megasphera elsdenii*

L'ultima parte del progetto tratta della proteina MeHydA. Essa è stata espressa eterologamente in *E. coli,* la sua espressione è stata indotta con IPTG in corrispondenza di una OD_{600} di 0.5 e la crescita delle cellule ha seguito un andamento esponenziale. La purificazione è stata condotta in due step cromatografici aerobici, uno di affinità e uno di esclusione molecolare, che hanno consentito di produrre una proteina pura e omogenea, con una resa pari a 2.1 mg di proteina per grammo di cellule bagnate. La ricostituzione del cluster [4Fe-4S] è stata condotta in anaerobiosi seguendo il protocollo standard per le ricostituzioni e l'incorporazione del sub-cluster [2Fe] è avvenuta incubando la proteina [4Fe-4S] con il complesso adt per un'ora in condizioni anaerobiche. L'incorporazione è stata verificata attraverso un test di attività con gascromatografia, che ha permesso di ottenere un'attività specifica di 659 µmol H_2/min mg proteina dopo un'ora e di 580 µmol H_2/min mg proteina dopo una notte.

Il laboratorio sta svolgendo degli esperimenti per capire come la proteina si comporta in presenza di ossigeno e durante il mio tirocinio ho preso parte ad uno di questi esperimenti. Il campione contenente la proteina attiva è stato esposto ad una quantità di ossigeno controllata e il suo comportamento è stato seguito attraverso spettroscopia UV-Vis e monitorando l'attività specifica nel tempo. È stato osservato che, in presenza di 10000 ppm di ossigeno, la proteina perde circa metà della sua attività dopo circa 3 ore (Figure 4.38).

1.4 Conclusioni

Durante la mia esperienza al *Laboratoire de Chimie des Processus Biologiques,* ho avuto l'opportunità di approfondire le mie conoscenze in biochimica, biologia molecolare e biologia proteica.

Da questo lavoro, è possibile affermare che il laboratorio è riuscito a ottenere una forma troncata stabile, solubile in acqua e pura di HydF da *Thermosipho melanesiensis*. Il principale obiettivo è ottenere una struttura cristallina completa di questa proteina, contenente il cluster [4Fe-4S] e il sub-cluster [2Fe]. Purtroppo, non è stato possibile ottenere cristalli contenenti i cluster, ma abbiamo individuato delle condizioni in cui la proteina cristallizza. Uno dei problemi che è stato necessario

affrontare durante la cristallizzazione è che la proteina a elevate concentrazioni ha mostrato un'elevata tendenza all'aggregazione. Abbiamo iniziato a studiare la stabilità della proteina in diverse condizioni, osservando che essa è influenzata da fattori come il tampone, il pH e la concentrazione di sale. Un approfondimento di tali studi può portare a individuare delle condizioni in cui la proteina è più stabile e questo potrebbe aiutare a cercare le condizioni ottimali di cristallizzazione. Un'alternativa può anche essere quella di soffermarsi su altre forme di questa proteina, dal momento che abbiamo dimostrato che la manipolazione di questa forma troncata non è molto semplice. Inoltre, in questo lavoro abbiamo fornito ulteriori prove del fatto che il dominio GTP non è coinvolto nell'interazione tra HydF e HydA. Un altro risultato interessante è che ΔTmeHydF mostra un'attività specifica maggiore di TmeHydF. Questo è inaspettato, poiché la differenza tra la forma troncata e quella wild-type risiede solo nel dominio GTP, ma non coinvolge il sito attivo. Pertanto, è necessario ripetere questi esperimenti per verificare il risultato ottenuto.

Un altro risultato notevole è lo sviluppo di un nuovo protocollo per la produzione di MeH HydA, che ha consentito di dimezzare il tempo di crescita delle cellule e di triplicare la resa finale. Tuttavia, rimane ancora molto lavoro da fare riguardo alla cristallizzazione della proteina. Non è stato, infatti, possibile ottenere cristalli di MeH HydA, ma sono state identificate delle buone condizioni che dovranno essere studiate meglio in futuro.

In conclusione, non c'è alcun dubbio nel dire che le idrogenasi [Fe-Fe] hanno un'incredibile attività nella riduzione di protoni e nell'ossidazione dell'idrogeno, ma esistono ancora diversi problemi che devono essere affrontati. In particolare, le idrogenasi sono molto sensibili all'ossigeno e devono essere manipolate in condizioni anaerobiche. Inoltre, sono difficili da produrre in forma attiva, anche se la maturazione sintetica ha parzialmente risolto questo problema. Siamo ancora molto lontani dalla realizzazione di un'idrogenasi con elevata attività, bassa sensitività all'ossigeno e possibili applicazioni in campo industriale, ma molti scienziati stanno lavorando per rendere questo possibile e diversi punti di partenza sono già esistenti.

I

2. Introduction

This work is part of a bigger project, which involves a basic research about hydrogenases in order to understand the reaction mechanism, the active site structure and how oxygen acts in their inactivation.

Hydrogenases are enzymes able to catalyse the reversible reaction of hydrogen production, through protons reduction. Every reaction involves energy transformations and hydrogenases use chemical energy. In particular, the use of chemical energy involves redox reactions with transfer of electrons and whole atoms of hydrogen. Thus, the reaction of protons reduction that gives hydrogen has to be coupled with a semi-reaction of oxidation, which provides electrons. The couple $2H^+/H_2$ has a reduction potential of -0.42 V and it is coupled with the oxidation of an electron donor, which has a more negative redox potential (in the case of this study methyl viologen was used as electron donor, which has a redox potential of -0.9 V).[1]

Hydrogenases are very sensitive to oxygen, in fact they are active only in anaerobic conditions. To catalyse the reaction, hydrogenases need electrons at a specific redox potential. Increasing the oxygen concentration leads to the increase of the redox potential and, in these conditions, hydrogenases are not active because oxygen has a more positive redox potential and it is a better acceptor of electrons and because it has a very high oxidative action which involves the degradation of the active site of the hydrogenase. In fact, in presence of very oxidative environments, oxygen is able to form very highly active species (ROS) that attack the active site of the hydrogenase, leading to its degradation. In the hydrogenase, the redox potential of clusters is influenced by the protein structure and it can be regulated by the surroundings. Thus, it is important to understand how the environment can influence the activity of the enzyme, in order to create new systems able to operate at different conditions. Furthermore, it is known that oxygen reacts with the H-cluster but the mechanism is still unknown and very debated. Thus, it is necessary to know how it reacts and which site of the H cluster is involved in the degradation.

The aim of these studies is to manage to use these enzymes in different biotechnological applications. We are interested in studying and using hydrogenases because they are the best catalysts for hydrogen production, with highest activity and low overpotential. Furthermore, they are easy to immobilize, thus they can have many applications in bioelettrodes. Experiments involving the immobilization of hydrogenases via adsorption on different materials are being carried out. For example, the challenge us to create devices for the biological hydrogen production and hydrogen evolution has been observed with hydrogenase absorbed on a pyrolytic graphite electrode.[2] The main application is as hydrogen photoelectrochemical production catalysts. In fact, so far most of the hydrogen is produced by steam reforming of natural gas but new renewable sources are needed. Water splitting is already used to produce hydrogen from natural sources but this process uses platinum as a catalyst, a very expensive and non-abundant element. Some examples of hydrogenases used as water splitting devices have already been reported. The main advantage of such a system is the use of a less expensive and largely available in nature metallic catalyst, like iron, taking inspiration from natural processes. For example, the hydrogenase was immobilized on a carbon-felt electrode and a porphyrin-sensitized TiO_2 nanoparticle was used as photoanode. A production rate comparable to a platinum electrode was obtained. The ultimate goal is to produce hydrogen and oxygen from water in a light-driven water splitting cell, miming the reaction that occurs in algae and cyanobacteria, coupling the an electrode for O_2 evolution containing the isolated PSII with an electrode for H_2 evolution containing hydrogenase and the isolated PSI. A big drawback of such a system is the very high oxygen sensitivity of this enzyme.[3]

Hydrogenases can also be used to prevent biocorrosion. In fact, these enzymes can induce the anaerobic microbial induced biocorrosion because they can act as chromate reductases, reducing chromate VI. Thus, they can also be applied in the decontamination of polluted environments.[4]

Hydrogenases can also be used as H_2 biosensors, such as the clay-PBV-hydrogenase electrode reported in the work of Quian et al.[5]. This biosensor was developed with the immobilized hydrogenase sandwiched between two layers of a montmorillonite clay and poly(butylviologen) (PBV) mixture on a glass carbon electrode. Such a system was able to transfer electrons by methyl viologen to the

electrode and to catalyse the oxidation of hydrogen to protons.

Hydrogenases can also have application in biofuel cells, exploiting their ability to catalyse the hydrogen oxidation. Fuel cells provide electrical energy from a chemical reaction producing only water as a by-product. They use only H_2 and O_2 and have higher efficiency than combustion engines. Biofuel cells use enzymes as electrocatalysts, in particular hydrogenases are used at the anode for the hydrogen oxidation. The advantage is that the enzyme is very specific for the substrate and that it is not necessary to separate cathode and anode with a membrane.[3]

Knowing how these enzymes act and their mechanism can lead to the rational design of synthetic mimics that can have large-scale applications. For example, some synthetic analogues of the [Fe-Fe] hydrogenase active site have been shown to be efficient catalysts for oxygen reduction in an aqueous medium.[6] The main issue still remains the low oxygen-tolerance. The development of catalysts that can operate under oxygen presence is the direction investigated by many researchers. Recently, it has been demonstrated that avoiding the inhibiting effects of oxygen with oxygen-tolerant catalysts can be more manageable than creating an oxygen-resistant catalyst. These catalysts are able to reduce oxygen and ROS without being damaged and in this way proton reduction is in competition with O_2 reduction and so hydrogen is produced at lower rate and yield. This problem can be faced designing a catalyst with higher kinetics for proton reduction than for oxygen reduction. The idea is to insert a sacrificial photosystem, which is able to reduce oxygen and to avoid ROS formation, maintaining an high activity into H_2 production. Some systems have been reported as possible applications for H_2 evolution, using redox active polymers containing viologen moieties that can immobilize the hydrogenase or 3D porous carbon electrodes with the enzyme.[7]

2.1 Hydrogen

The massive use of fossil fuels is starting to pose serious problems to global environment and climate change and this requires the development of new technologies for energy use and storage. The use of renewable energy sources is in expansion but the main problem is the development of storage systems that can store and deliver efficiently the energy when it is required.[8]

There are a variety of renewable sources that could be used for energy production. One promising alternative to fossil fuels is hydrogen. Hydrogen, in fact, is one of the molecules with the highest energy content (higher and lower heating values of hydrogen: 141.9 and 119.9 kJ/g)[9] and it is carbon-free. Moreover, its combustion produces energy and water according to the reaction: $2H_2 + O_2 \rightarrow 2H_2O$. It can also be safely transported and stored in many ways like compressed gas, cryogenic liquid and solid hydride. However, the use of hydrogen has also its drawbacks, which have to be faced and improved. Among these, molecular hydrogen is not widely available in nature, even if it is very abundant, but it is stored in water or hydrocarbons and it is necessary to efficiently extract it from these compounds.[10] Moreover, hydrogen is nowadays produced through very expensive and polluting methods.

Methods used for hydrogen production are mostly conventional methods, like hydrocarbon reforming and pyrolysis, or renewable processes, using biomass or water as sources (Figure 2.1). Among the hydrocarbon reforming methods, steam reforming, partial oxidation and autothermal reforming are high efficiency processes, widely used for hydrogen production, but they are based on biofuels and produce the greenhouse gas CO_2 as by-product of combustion. Nowadays, the attention is focused on the development of renewable sources based processes. A little quantity of the overall hydrogen is already produced with water electrolysis, which is able to produce very pure hydrogen, without pollutant by-products, but requires a big consumption of electricity. A very successful result would be using photo-catalysis to produce hydrogen from water splitting. Another way to produce hydrogen from renewable sources is biomass processes. A variety of processes exist and they are environmentally benign but the final yields are lower and depend on the raw material used.[8]
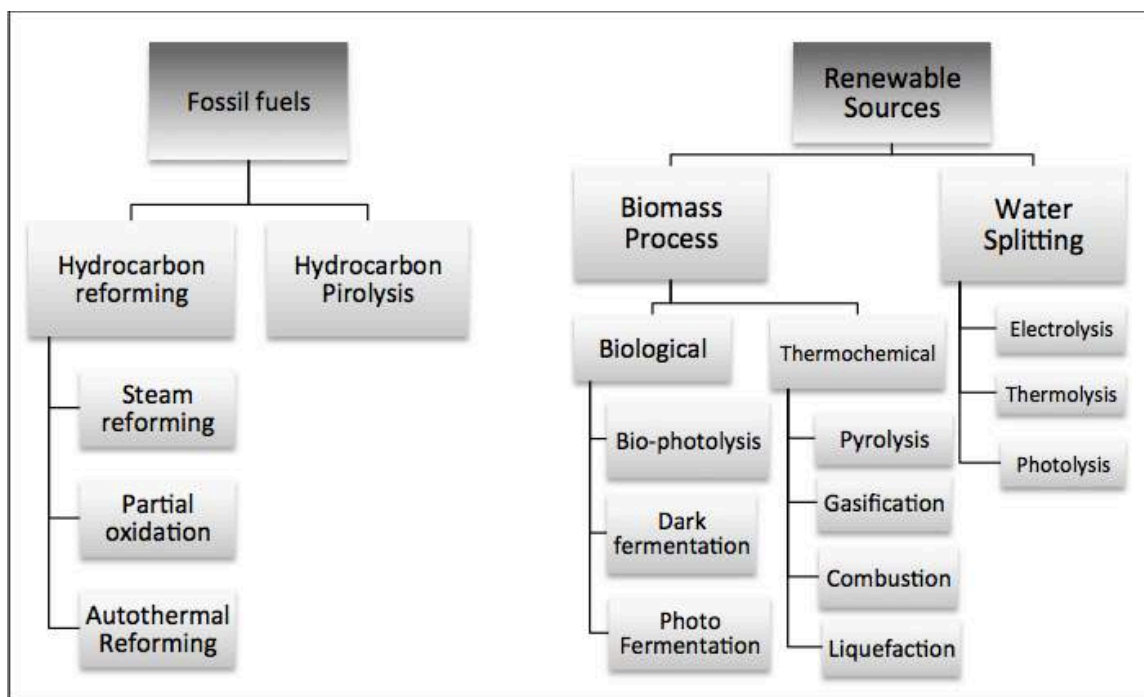
Figura 2.1 Hydrogen production methods

Always more technologies are developing using hydrogen as a source of energy, for example fuels cells. These have very promising potential applications because they are able to produce electricity and heat from H_2 and O_2, producing only water as a final product. The main problem in the application of this technology is the lack of molecular hydrogen in nature.

Thus, the necessity of new innovative methods to produce molecular hydrogen at high efficiency and lower cost caused a considerable increase in research in biological hydrogen production. This processes use mostly water or biomass as a source for hydrogen production by bacteria or algae.[8] These microorganisms use their hydrogenase and nitrogenase enzymes to catalyse the reversible reaction of proton reduction: $2H^+ + 2e^- \leftrightarrow H_2$.[11] Thus, our interest into hydrogenases comes from all these reasons.

2.2 Metalloproteins

Metalloprotein is a generic term for a protein containing a metal ion cofactor. It has been estimated that almost 30% of the proteins existing in nature are metalloproteins. The metal ion cofactor is usually coordinated by nitrogen, oxygen or sulphur atoms belonging to amino acids in the polypeptide chain, generally in its side-chains, or by a macrocyclic ligand incorporated into the protein (for example porphyrin, corrin and chlorine). The presence of the metal ion allows the protein to perform its functions that could not easily be performed without it. So, the metal ion is part of the active site and it has a crucial role.

Many efforts have been done to understand the structure and function of these proteins. They have many different functions in cells, such as transport (transferrin for iron), storage (ferritin for iron) or signal transduction. They are also involved in the catalysis of biological processes, such as photosynthesis, respiration, water oxidation, molecular oxygen reduction and nitrogen fixation. Some of them are enzymes and they are called metalloenzymes.[12]

Metalloenzymes are present in all enzyme families and catalyse a wide variety of reactions, so they cannot be associated to a particular group of enzymes, but each metal has been proved to have a specific role in the enzyme.[13] The metal ions that are commonly found in metalloenzymes are iron, zinc and copper, but some proteins also contain nickel, manganese, molybdenum, cobalt, vanadium and calcium. Table 2.1 reports the roles of the most common metal cofactors in enzymes. Some

examples are hemoproteins that contain an iron porphyrin as prosthetic group and hydrogenases that can contain iron-sulphur or nickel-iron clusters.

The properties of metalloproteins clearly depend on metal's nature but they can be strictly influenced by many other factors, such as ligands type and coordination number, second coordination sphere on the active site, accessibility of substrate to metal centre and three dimensional structure of the protein.

Table 2.1 Important functions of some metals in algae and bacteria.[13]

Metal	Specific roles in enzymes
K	Protein stabilization
Ca	Protein stabilization
Mg	Protein stabilization
Fe	Oxygen transport, storage and activation, electron transport, superoxide breakdown
Mn	Oxygen evolution, peroxide and superoxide breakdown
Zn	Protein stabilization, hydrolytic cleavage
Cu	Oxygen transport, storage and activation, electron transport, superoxide breakdown
Mo	Oxygen transfer, nitrogen activation
Co	Free radical reactions, nucleophile

Thus, knowing the tri-dimensional structures and the active site configuration is very important in the understanding of these proteins. Techniques for the investigation of chemical and structural properties are available. The most relevant are spectroscopic tools, such as electron paramagnetic resonance (EPR), electron-nuclear double resonance (ENDOR), electron spin resonance (ESR), resonance Raman or Mössbauer spectroscopy. Another powerful technique is X-ray protein crystallography which allows the resolution of the spatial arrangement of the atoms of the proteins.[14]

2.3 Iron-sulfur clusters

Among the metal clusters usually found in proteins, some of the most common are iron-sulfur clusters. Iron is one of the most abundant atoms on earth and it often acts as enzyme cofactor or prosthetic group. Iron can be bound to amino acidic chains directly or via other atoms, such as sulfur in the [Fe-S]-cluster.[15]

[Fe-S]-clusters often have an important role in proteins and were found to be very common in the oldest components of living matter, including Archea, bacteria, plants and animals. They are believed to have contributed to success of life on earth.[16] Studies on this cluster started in 1960s when they were found to be involved in oxidoreductive reactions in many plants and microorganisms. Later, in the 1970s it was shown that iron-sulfur complexes could be artificially synthesised as analogue complexes of the protein active site.[17]

Proteins with [Fe-S]-clusters can take part into a lot of reactions, but they are mainly involved in oxidation-reduction reactions and electron transfer. In fact, iron is able to switch very easily between the Fe^{3+} oxidized and the Fe^{2+} reduced forms, behaving as electron donor or acceptor in many biological reactions.[18] For example, iron is involved in cellular respiration in bacteria and mitochondria.[5] Another important function of this cluster is that, as they are able to reversibly bind Fe and S, they acts as Fe and S storage for the activation of enzymes or substrates, like in the enzyme aconitase, which has the function to catalyse the reversible transformation of citrate to isocitrate in the citric acid cycle.[19] Moreover, it has been shown that they are involved in DNA replication and repair. For example, DNA polymerase-ε requires the presence of a [Fe-S]-cluster to replicate DNA strands and repair damaged ones.[20]

Different kinds of [Fe-S]-cluster exist and they are generally coordinated to cysteines. The most common conformations are the rhombic arrangement [2Fe-2S], which contains 2 atoms of iron connected to four cysteines via two bridging sulfide ions, and the cubane form [4Fe-4S], that consists

in four iron atoms and four sulfur atoms coordinated to four sulfhydryl side chains of cysteines (Figure 2.2). Other clusters have been observed, like the [1Fe-0S], containing one single iron atom coordinated to four cysteines (such as in rubredoxin) and the [3Fe-3S] or more complex form like [7Fe-8S], [8Fe-7S] or [8Fe-8S] (observed in molybdenum-iron proteins).[12]

Studies on the [Fe-S]-cluster assembly process showed that this process occurs *in vivo* with three different machineries: the NIF system, observed the first time in *Azotobacter vinelandii*, for specific maturation of nitrogenase in azototrophic bacteria; the ISC assembly, observed in mitochondria of yeast and other eukaryotes, for the generation of [Fe-S] proteins under normal conditions; and the SUF system, observed in *Erwinia chrysanthemi* and *Escherichia coli* for the generation of [Fe-S] proteins under oxidative-stress conditions. During evolution, the latter two mechanisms were transferred from bacteria to eukaryotes, which were shown to contain [Fe-S]-clusters in mitochondria, cytosol and nucleus. In particular, mitochondria contain the ISC assembly machinery, while plastids host protein from SUF machinery. It was shown that the entire process was highly conserved from yeast to human. It always requires 1) a sulfur donor (normally a cysteine) that can release sulfur through a cysteine desulphurase, creating a persulphide as an intermediate; 2) a specific iron donor; 3) an electron transfer (normally performed by bacterial and mitochondrial components), in order to reduce S; 4) a scaffold protein, as platform for the biosynthesis of the [Fe-S] cluster. In fact they can bind [Fe-S]-clusters with labile binds, allowing an easy and stable transfer of the cluster to the target protein; 5) a cluster transfer protein which is needed for the correct cluster transfer to apo-proteins (Figure 2.3).[21]
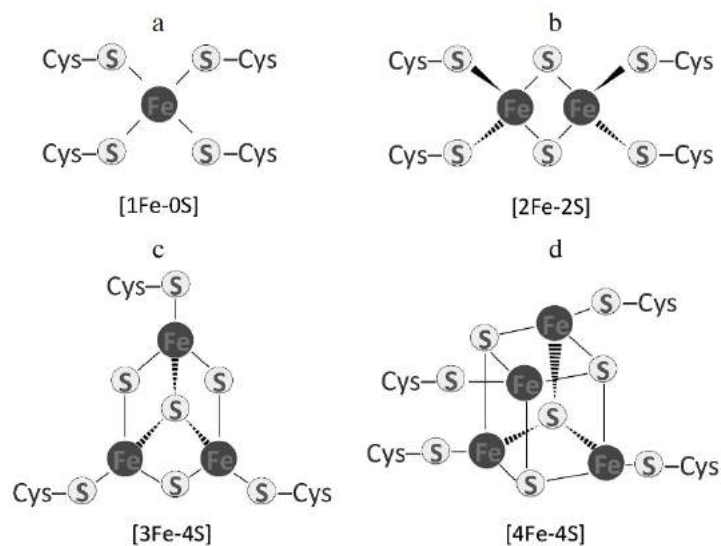

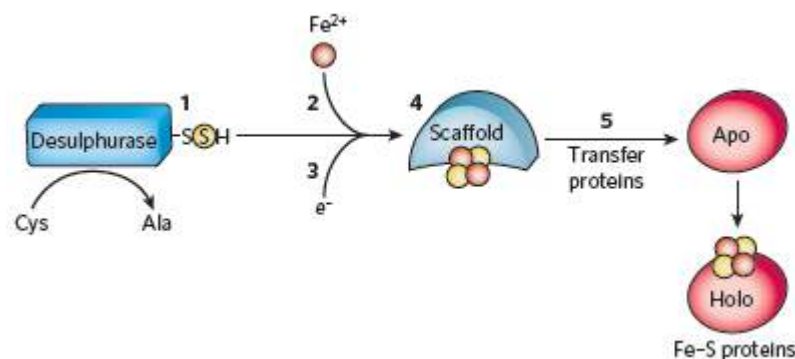**Figure 2.2** Iron-sulfur clusters commonly found in proteins.

Figure 2.3 Biosynthetic principles of [Fe-S]-protein biogenesis. The cysteine desulphurase releases sulphur from a Cys, converting it into an Ala. The sulphur obtained is combined with Fe^{2+} and electrons in order to assemble the [Fe-S]-cluster into a scaffold protein. The assembled cluster is then transferred to the apo-protein, converting it into an active holo-protein.[21]

2.4 Hydrogenases

Among iron-sulphur proteins, hydrogenases are a class of protein of special interest. Research on this class is significantly growing, since 2000, due to their inherent applicability in development of renewable energy technologies, based on hydrogen. New methods for producing hydrogen are needed from renewable sources and low cost processes.

Hydrogenases are metalloenzymes that can catalyse the reversible reaction of conversion of dihydrogen into protons and electrons at very high catalytic rates ($\sim$10 000s^{-1})[22]:

$$H_2 \ (\rightleftharpoons H^+ + H^-) \rightleftharpoons 2H^+ + 2e^- \qquad (2.1)$$

This reaction takes place in presence of an active site that contains a metal, which allows the pH of H$_2$ to dramatically decrease, leading to the heterolytic splitting of the molecule.

Hydrogenases can be classified in three classes: [Ni-Fe]-hydrogenases, [Fe-Fe]-hydrogenases and metal-free hydrogenases. The latter class contains one single iron atom and can activate hydrogen only in presence of another substrate and it is only found in methanogens. The most interesting classes are [Ni-Fe]- and [Fe-Fe]-hydrogenases.[3]

[Ni-Fe]-hydrogenases have been isolated as heterodimers, where the big subunit contains the bimetallic [Ni-Fe] center of the active site, deeply buried in the protein, and the little subunit contains some [Fe-S]-clusters, important for the electrons transfer and essential for the H$_2$ activation. The active site is thus made up of two atoms, one of iron and one of nickel, connected via two cysteines (Figure 2.4).[23] [Ni-Fe]-hydrogenases are involved in hydrogen consumption, while [Fe-Fe]-hydrogenases are normally involved in hydrogen production.

[Fe-Fe]-hydrogenases, normally known as HydA, have an active site made up of a [4Fe-4S]-cluster connected to a [2Fe]-subcluster via a thiolate of a cysteine. This complex is called H-cluster. The cysteine is the only bind between the [4Fe-4S]-cluster and the [2Fe]-subcluster. The two iron atoms of the di-iron subcluster are connected via an azapropanedithiolate bridge. The coordination sphere is filled by a CO and a CN group (Figure 2.4). HydAs differs greatly in size and number of accessory clusters, in fact, depending on the microorganism, HydA can have a single [4Fe-4S]-cluster (the one contained in the H cluster) or additional [4Fe-4S]-clusters that are involved in electron-transfer, from the surface to the active site of the protein.[24]
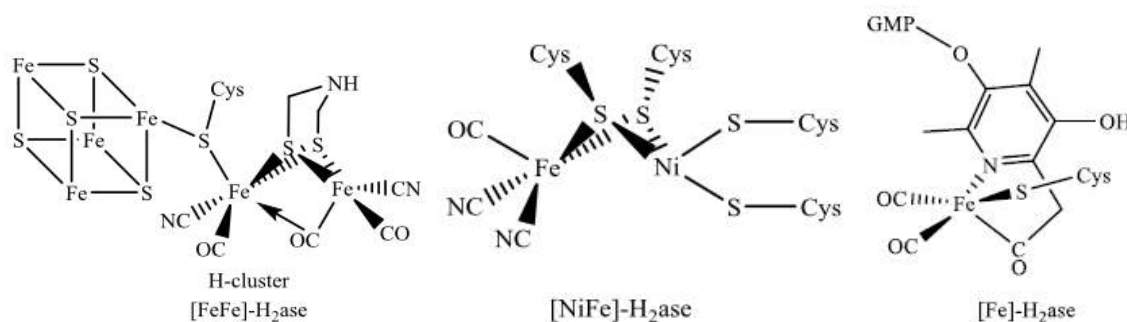

Figure 2.4 Active sites of the different classes of hydrogenases: H-cluster of [Fe-Fe]-hydrogenase (on the left), [Ni-Fe]-hydrogenase (at the centre) and [Fe]-only hydrogenase (on the right).

2.5 [Fe-Fe]-hydrogenases maturation

To become active, the [Fe-Fe]-hydrogenase polypeptide encoded by the *hydA* gene has to incorporate the [Fe-S]-cluster(s) and the di-iron subcluster. This is called maturation process, which is a very complex post-translational process and involves difficult reactions. In fact, it involves the synthesis of the CO and CN, the dithiolate ligand, the assembly of the diiron subcluster and its incorporation into the [4Fe-4S] protein and the assembly and transfer of the additional [Fe-S]-clusters. The whole process has to be tightly controlled because CN$^-$ and CO can be produced at toxic level, as well as hydrolytically sensitive dithiolate.

6

The pathway that occurs in nature for the hydrogenases maturation is being revealed by a variety of methods and is still partly unknown.

One of the first clues of the maturation process was the identification in the genome of *Chlamydomonas reinhardtii* of genes coding for the three accessory proteins, HydE, HydF and HydG, required for the maturation of an active hydrogenase. It was also shown that many organisms containing [Fe-Fe]-hydrogenases have in their genome genes coding for these three maturases.[25] Moreover, the first example of heterologous expression[25] and the first *in vitro* activation[26] of an active [Fe-Fe]-hydrogenase revealed that all the three proteins are necessary for the hydrogenase activation. This activation does not require the addition of cluster precursors, meaning that the maturation machinery is able to synthetize and transfer it to HydA.[26] However, the addition of iron, sulphide, SAM (S-adenosyl-L-methonine), cysteine and tyrosine was demonstrated to enhance the hydrogenase activity, suggesting that iron and sulphide reconstitute the iron-sulphur cluster and the SAM added is probably essential for the HydE and HydG proteins.[27] Moreover, it has been shown that the maturation process needs the [4Fe-4S] component of the H-cluster to be assembled before the [2Fe]-subcluster insertion into the protein. Thus, maturation requires the three maturases HydE, HydF and HydG for the synthesis of the [2Fe]-subcluster and its transfer and incorporation into the apo-hydrogenase.[28]

2.5.1 The Hyd protein machinery

Production of an active [Fe-Fe]-hydrogenase requires the simultaneous action of HydA with HydF, HydE and HydG that together compose the so-called HYD machinery.

Two of the additional proteins, HydE and HydG, belong to the radical SAM family. This family of enzymes generally catalyses difficult chemical reactions such as CH to C-S bond formation and glycyl radical formation.[29,30]

HydG contains two [4Fe-4S]-clusters: one is chelated by the typical radical SAM proteins motif $CysX_3CysX_2Cys$ in the N-terminal domain and the second is chelated by the $CysX_2CysX_{22}Cys$ motif in the C-terminal domain. Each cluster has its own role: the N-terminal cluster is required for SAM reductive cleavage and tyrosine cleavage to produce p-cresol, while the C-terminal cluster is required for the production of the diatomic groups CO and CN[31]. It has been shown that both the clusters are essential for activity.[32] Some experiments showed that HydG uses tyrosine as a substrate to synthetize CO and CN⁻. In fact, Pilet et al. in 2009 demonstrated that HydG was able to cleave tyrosine in p-cresol in a SAM-dependent reaction[33] and then, it was shown that HydG produces CN⁻ and CO using tyrosine as substrate.[34,35] Subsequently, Kuchernreuther et al. showed that on a cell-free system both the diatomic groups derive from tyrosine.[36] Furthermore, spectroscopic studied demonstrated that HydG provides both the diatomic ligands (CO and CN⁻) and the iron for the [2Fe]-cluster, necessary for the formation of the H-cluster.[37] In fact, Stopped-flow FTIR studies provided evidence for the formation on HydG of an iron-bound complex, containing CO⁻ and CN⁻, the $[Fe(CO_2^-)CN^-]$ synthon, which is considered to be the effective precursor of the H-cluster. Suess et al. showed that Cys binds the auxiliary iron-sulfur cluster of HydG and suggested that Cys serves as ligand platform where this synthon is built.[38]

HydE, as HydG, binds iron-sulfur clusters and catalyses the SAM reductive cleavage. Iron analysis, UV-Vis and EPR spectroscopy showed that it contains two [4Fe-4S]-clusters: one is bound to the radical SAM motif $CysX_3CysX_2Cys$ in the N-terminal domain and is very conserved in HydE; the second seems to be very labile and is presumed not to be involved in H-cluster biosynthesis.[39] X-ray structure of HydE from *Thermotoga maritima* showed that the second cluster is located on the molecular surface and it probably bounds variable amounts of iron and sulphide.[40] Pilet and al.[33] reported that HydE contains three anion-binding sites in a large cavity and one of them can bind SCN⁻ with high affinity. The substrate used by HydE is still unknown but some hypothesis can be made, based on early studies. Kuchernreuther et al.[41] reported that incubation of HydA from *C. reinhardtii* with SAM, cysteine, tyrosine, iron, sulphide and maturases increased 5 fold hydrogenase activity. Given the demonstration that tyrosine is the substrate of HydG[33-36] and SAM is required for radical SAM activity, cysteine reasonably seems to be a possible a substrate for HydE, but this is not been

demonstrated yet. Furthermore, Betz et al.[40] suggested that HydE substrate contains a thiol functional group. HydE is presumed to provide the dithiomethylamine bridge of the H-cluster, since it has been demonstrated that HydG provides the iron, CO and CN⁻ components of that cluster.

While the two proteins, HydE and HydG, are part of the radical SAM family, the third protein required for the hydrogenase activation, HydF, is a GTP-ase protein. Spectroscopic studies showed that this protein is able to bind a guanosintriphosphate and contains a [4Fe-4S]-cluster ligated by three cysteines and by a non-cysteinyl ligand, which has been subject of controversy and seems to vary depending on the local residues.[42] It was also observed the presence of an additional [2Fe]-subcluster, ligated by CO and CN⁻.[43] From size-exclusion chromatography, static light scattering data and the crystal structures[44,45], it was demonstrated that the protein is stable as a dimer where each monomer contains three domains, the GTP binding domain, the dimerization domain and the metal-cluster binding domain. Further information about HydF structure is discussed in Section 2.8.1.

Concerning the role of HydF in the HYD machinery, in 2008 McGlynn and al.[46] postulated that the GTP-ase domain is probably not involved in cluster transfer and that HydF acts as a scaffold or carrier protein in the H-cluster assembly. The hypothesis that HydF has a key role in the H-cluster assembly was encouraged by Czech et al.[43] and, later, it was proposed that a dimeric form of HydF interacts with HydE and HydG in a GTP-bound form and then the GTP is dissociated, in order to allow interactions with HydA.[44] HydF has a key role, being able to act as a carrier and scaffold protein. In fact, this indicates that it is able to interact (i) with HydE and HydG, in order to assemble the binuclear subcluster and (ii) with the apo-hydrogenase, when the [2Fe]-subcluster is transferred to the latter. Interactions between HydF and the other two maturases have been studied by Vallese et al.[47] who showed that (i) HydF can interact with both HydE and HydG, independently of its GTPase properties and (ii) HydE and HydG separately participate in creating structural modifications to HydF, allowing its interaction with the apo-hydrogenase.

The most likely mechanism is that the whole HYD machinery synthesizes and inserts a [2Fe]-subcluster and its ligands into the apo-HydA already containing the [4Fe-4S]-cluster. All the three maturases are essential: HydG carries out the synthesis of the [Fe(CO₂)CN⁻] synthon from tyrosine, HydE is supposed to provide the azadithiolate ligand from un unknown substrate and HydF interacts with the two maturases in order to assemble all the components and to form a diiron complex which is then transferred to apo-hydrogenase, resulting in the activation of the protein (Figure 2.5).
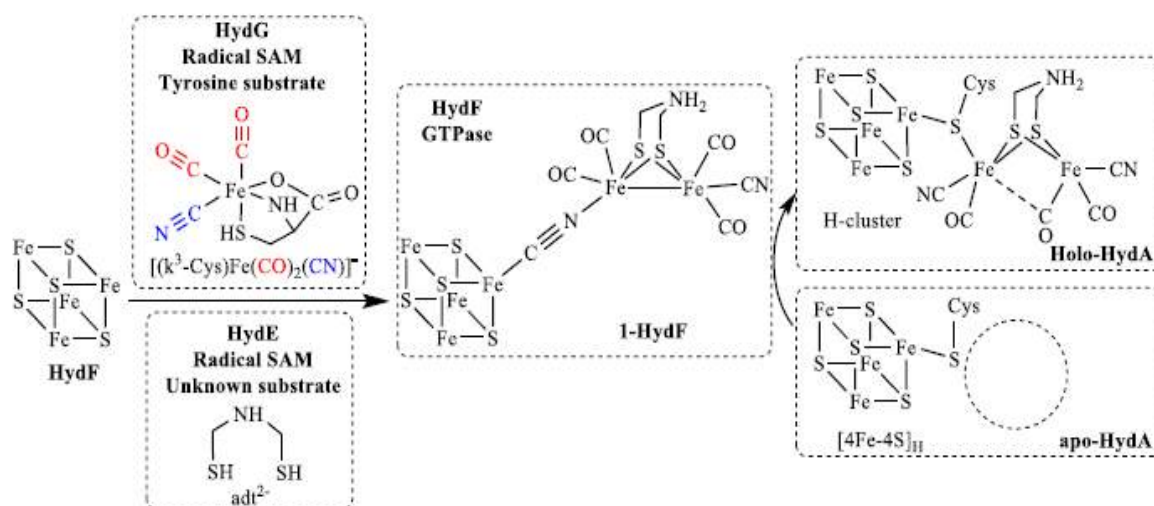

Figure 2.5 Schematic representation of apo-HydA maturation. HydF interacts with HydG and with HydE. HydG provides the [Fe(CO₂)CN⁻] synthon, using tyrosine as substrate, and HydE probably provides the azadithiolate ligand, from un unknown substrate. The three maturases interacts, HydF assemblees all the components creating a complex, 1-HydF, which contains the diiron subcluster. This complex is transferred to the apo-[Fe-Fe]-hydrogenase which is transformed into an active protein.[48]

Although the mechanism of hydrogenase maturation is being deeply studied, further investigations are required. In fact, many aspects still have to be defined, such as the role of the GTP-ase domain, HydE substrate and especially the order in which the maturases act. Broderick et al.[28] proposed two possible pathways for the [2Fe]-subcluster assembly, which particularly differ in the actions order: in the first one HydE and HydG separately interact with HydF which then transfer the diiron complex to HydA; in the alternate pathway HydG delivers iron bound cyanide and CO to HydF as cyano-carbonyl-Fe units, where azadithiolate addition connects the two mononuclear Fe species.

Another interesting remark is that in *C. reinhardtii* HydE and HydF have been observed to form a single peptide, named HydEF, suggesting that these two maturases are likely to form a complex during maturation process.[49]

2.6 Chemical maturation

As explained in the previous section, in nature hydrogenases are activated by a very complex mechanism, which involves three maturases and aims at the assembly of the H-cluster. Studies on the possibility to chemically activate hydrogenases are being carried out. Recently, it has been shown that hydrogenases can be chemically activated with a process that does not involve the co-expression of the maturases. In fact, chemically synthesized diiron complexes can be transferred, through the maturase HydF, to the [Fe-Fe]-hydrogenase, already containing the [4Fe-4S]-cluster. Even if these diiron complexes alone are not active, when inserted in the hydrogenase they showed a high activity, demonstrating the importance of the protein environment.[50]

Basing on the demonstration that anaerobically purified HydA from *C. reinhardtii* (*Cr*HydA1) could be activated by incubation with HydF from *Clostridium acetobutylicum*[43], in 2003, the first example of an artificial maturase system was reported. It was demonstrated that HydF from *T. maritima,* containing the [4Fe-4S]-cluster and expressed in *E. coli,* could incorporate synthetic diiron complexes **1**, **2** and **3** (Figure 2.6), leading to new hybrid proteins x-HydF, after a chemical reaction in anaerobic conditions. Spectroscopic characterization, such as FTIR, EPR and HYSCORE, provided evidence of this incorporation. Then, it was shown that only the hybrid **2**-HydF protein was able to activate HydA from *C. reinhardtii* containing only the [4Fe-4S]-cluster. So, this work demonstrated that complex **2** can be efficiently transferred from HydF to HydA and is able to activate the apo-hydrogenase, proving that the natural precursor is likely to be very similar to this complex.[51] Another step forward was made when it was shown that the synthetic di-iron complexes could be directly incorporated by HydAs.[52] This procedure allows obtaining an active hydrogenase with an activity comparable to the one measured for the wild type. Esselborn et al.[50] showed that the crystal structures of wild type active HydA from *Clostridium pasteurianum* (CpI) and the semisynthetic CpI matured with complex 2 (CpI^{ADT}) have identical structures.

These results demonstrated that the synthetic complex undergoes some transformations in order to be converted to the natural active substrate. Among these transformations, an isomerisation of the CN$^-$ ligand is needed, as well as a substitution of a CO by a cysteine ligand of the [4Fe-4S]-cluster of HydA.
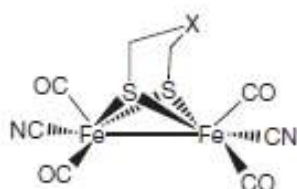

Figure 2.6 Structures of the synthetic mimics which were tried. X=CH$_2$ in complex **1**; X=NH in complex **2**; X=O in complex **3**.[51]

Thus, two possible patterns can be followed for the hydrogenase chemical maturation (Figure 2.7): activation of apo-HydA can be carried out directly with the diiron mimic or via HydF. Both the

mechanisms proved that only the complex containing the azadithiolatate bridge was able to obtain HydA with high enzymatic activities, underlining the central role of this nitrogen containing bridge in enzyme activity.

Recently, it has been shown the possibility to avoid not only the use of the HYD machinery but also the bypass of the strictly complete anaerobic purification procedure. In fact, it was shown that it was sufficient to heterologously express HydA from *Megasphaera elsdenii* in *E. coli*, aerobically purify the apo-protein and chemically activate it *in vitro*, in order to obtain an active hydrogenase. The chemical activation requires the anaerobic reconstitution of the [4Fe-4S]-clusters by treatment with iron and sulfur and the anaerobic incorporation of the [2Fe]-subcluster by reaction with the $[Fe_2(adt)(CO)_4(CN)_2]^{2-}$ complex. The effective incorporation of the complex inside MeHydA was proved by FTIR spectroscopy and activity assays.[53]

This synthetic maturation methodology was a very important discovery and helped significantly the identification of the nature of the central atom in the dithiolate bridge of the [2Fe]-subcluster. However, some severe drawbacks still exist, preventing the use of [Fe-Fe]-hydrogenases in biotechnological applications. Among these, the necessity to manipulate these proteins in strictly anaerobic conditions to avoid oxygen mediated inactivation is the biggest obstacle for their large-scale applications.
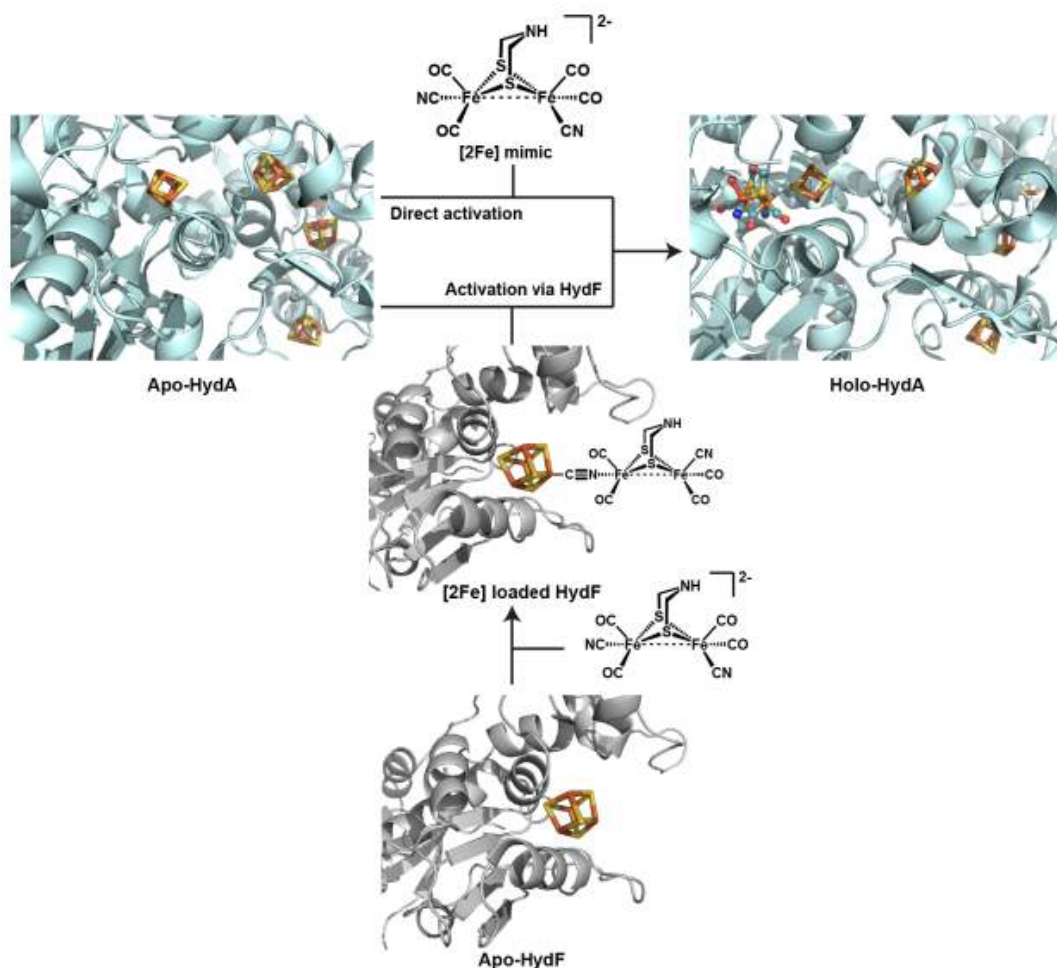

Figure 2.7 Possible patterns for apo-HydA activation. The first one consists in activation of apo-HydA via incubation with HydF. The latter is firstly incubated with a diiron complex, which is incorporated in the maturase. Then, [2Fe]-HydF is incubated with apo-HydA and the diiron subcluster is efficiently transferred to the hydrogenase, resulting in an active protein. The second process consists in direct activation of apo-HydA with diiron complex.[62]

2.7 Artificial hydrogenases

As already discussed before, hydrogenases have been widely studied for their potential technological applications, as they rely on non-noble, cheap and abundant metals. However, their use in large-scale applications is still limited by three factors: (i) they are difficult to produce in large quantities, (ii) they are extremely sensitive to oxygen, so they must be manipulated in strictly anaerobic conditions and (iii) it is very difficult to produce them in an active form. This latter has been partly faced with the chemical maturation.

In the last years, a variety of biomimetic and bioinspired metal complexes have been synthetized, allowing the reproduction of structures and functions of the natural metal sites through simple and low-molecular weight synthetic organometallic compounds. For example, many complexes have been synthetized as models for the active site of [NiFe]-hydrogenases. Among these, it was demonstrated that the heteronuclear Ni/Fe[54,55], Ni/Ru[56] and Ni/Mn[57] complexes showed catalytic activity for proton reduction.

However, none of these compounds showed an activity comparable to the one of hydrogenases, probably because of a lack of interactions with polypeptide chains. Thus, it was considered the idea to improve the performances of the mimics, incorporating them into the protein and designing an artificial hydrogenase.[58]

The concept of artificial enzymes has its origins in the work of Wilson and Whitesides[59] and is based on the idea of combining a biomolecular scaffold (a protein or an oligonucleotide) and a synthetic complex as active site. To date, a variety of biohybrid systems have been reported, offering the possibility to improve the catalytic performances.

The case of hydrogenases is very interesting, in particular for the possibility to optimize electrons and protons transfer. The concept of artificial hydrogenases implies the combination of a well-chosen and easily accessible host protein to a synthetic complex with proton reduction or hydrogen oxidation catalytic activity. The aim is to create a cheap, stable, water-soluble and active hybrid system. The protein would provide an environment of functional groups around the synthetic complex helping proton and electron transfers. For example, Roy et al.[60] managed to synthetize a bioinspired mimic binding a diiron hexacarbonyl cluster to an alanine-rich peptide via an artificial dithiol amino acid. The resulting biohybrid system was able to catalyse photo-induced production of hydrogen in water in acidic conditions and in presence of ruthenium as photosensitizer and ascorbate as sacrificial electron donor. Then, Sano et al.[61] showed that the hybrid system obtained treating apo-cytochrome c with a diiron complex worked as catalyst for hydrogen evolution in presence of ruthenium as photosensitizer and ascorbate as sacrificial reagent in aqueous media.

The fact that in [Fe-Fe]-hydrogenases the diiron cluster is attached to the protein only by a single cysteine allowed considering the direct maturation of an hydrogenase containing all the [4Fe-4S]-clusters but lacking the [2Fe]-subcluster (defined in the following apo-hydrogenase). In particular, HydF has been considered as the most suitable protein for a synthetic bioinspired complex, due to its ability to bind and transfer a diiron cluster. In this regard, it was shown that incubating HydF from *Thermotoga maritima* with azadithiol (adt-complex), propanedithiol (pdt-complex) and oxa-propandithiol (odt-complex) led to hybrid proteins similar to the target metalloenzyme. This result provided the first example of a rational structural design of an artificial hydrogenase. Furthermore, it was observed that only the pdt-hydrogenase showed a small activity in a standard chemical assay, using sodium dithionite and methyl viologen as source of electrons, as well as in a photochemical assay, using a metal−organic photosensitizer and a sacrificial electron donor under visible light irradiation. Although this activity was very low, this is a starting point for the development of more efficient artificial hydrogenases.[58]

Diiron mimics have also been evaluated in non-protein environments, for examples micelles dendrimers, polymers, oligo and polysaccharides and metal-organic frameworks.[62]

Unfortunately, to date these systems are considerably less active than natural enzymes but hope remains that a new, highly active and stable metalloenzyme could be generated, with potential for application in technological devices. This field in fact is still at its infancy and obtaining a working

artificial hydrogenase is very complex, as the reactions involve electrons and protons transfers and gas transfer. However, many studies have been and are still being carried out, demonstrating the possibility of creating an active artificial hydrogenase embedding an inefficient catalyst and increasing its efficiency due to the protein environment.

2.8 Aims of the project

2.8.1 HydF from *Thermosipho melanesiensis*

Many studies have been done on HydF structure. The first characterization of HydF was done on the protein from *Thermotoga maritima*[42], showing that it contains a [4Fe-4S]-cluster which binds GTP and catalyses its hydrolysis to GDP. Then, FTIR studies provided evidence of the presence of CO and CN⁻ ligands of the [2Fe] precursor and gel filtration studies of HydEFG showed that the proteins tend to assemble in a dimeric conformation.[63] Czech et al. showed that HydF from *Clostridium acetobutylicum* contains an iron cluster with similar structural features as the H-cluster of active HydA from *Chlamydomonas reinhardtii* (CrHydA1).[64] The first crystal structure was obtained from HydF from *Thermotoga neapolitana* (TnHydF)[44] and it showed that the protein is organised in three domains: (I) domain I corresponds to the GTP-binding domain, very similar to the other GTP-ases; (II) domain II is connected to domain I by a long stretch of amino acids and it is responsible for HydF dimerization; (III) domain III is the [4Fe-4S]-cluster binding domain and it is connected to this cluster by three highly conserved cysteine residues (Figure 2.8 a). The crystal contained one molecule per asymmetric unit and dimers were obtained by reconstruction of the crystal unit cell. However, the protein was crystallised as a dimer in the sample put to crystallize and a tetramer was formed during the course of the crystallization.

Another crystal structure of HydF was obtained from the [4Fe-4S]-reconstituted protein from *Thermosipho melanesiensis*.[45] The structure contained the three different domains already described for TnHydF (Figure 2.8 b). It was also shown that the iron-sulfur cluster is coordinated by three very highly conserved cysteines and by the carboxylate group of a glutamate. These residues are located in a positively charged cavity, which fits nicely the need to interact with negatively charged molecules like [4Fe-4S]$^{1-/2-}$ and the [2Fe]-subcluster. Many attempts were done in order to try to obtain crystals with the diiron subcluster, such as co-crystallisation and soaking, but it was not possible to obtain them so far. Interestingly, soaking [4Fe-4S]-TmeHydF crystals in a crystallization cocktail supplemented with the [2Fe] ligands led to the observation of a progressive break of the crystals upon ligand diffusion. Different hypothesis can be made to explain the phenomenon observed. First, it is possible that the diiron complex cannot easily access the cluster because of the presence of the Glu residue. Second, it has been shown that, in the crystal packing of the [4Fe-4S]-TmeHydF, the contact between dimers is in the area where the iron-sulfur cluster and the diiron subcluster are bound, leading to the disruption of the crystals, maybe due to the introduction of a negative charge of the [2Fe]-subcluster. Furthermore, it has been observed that GTP domains have significantly higher B factors with respect to the other two domains (Figure 2.9). As the B factors are related to the mean square isotropic displacement of the atoms, this implies that the GTP domain is not well ordered and is therefore a source of disorder in the crystal, leading to low-resolution crystals.[45]
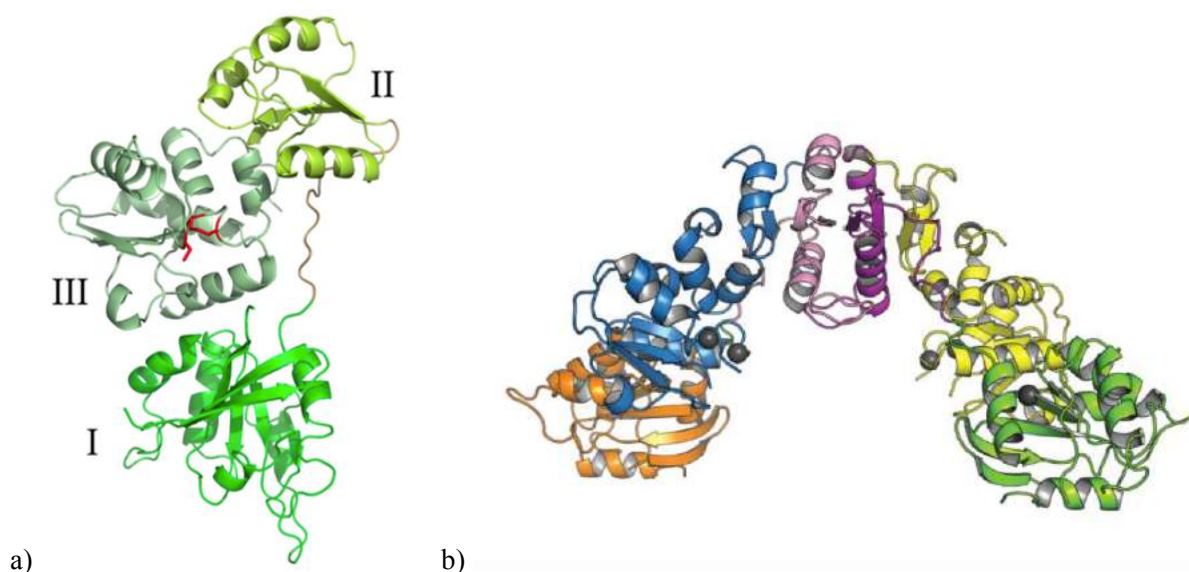
Figure 2.8 a) Representation of the structure of one subunit of the apo-TnHydF dimer. The three domains, labeled I, II, and III, are in different shades of green and the loop connecting domains I and II is in brown. The side chains of the three cysteines involved in subcluster binding are shown in red.[44] **b)** Representation of the [4Fe-4S]-HydF dimer from *Thermosipho melanesiensis*. GTP domains are in green and orange, dimerization domains in magenta and light pink and cluster binding domain in yellow and blue.[45]
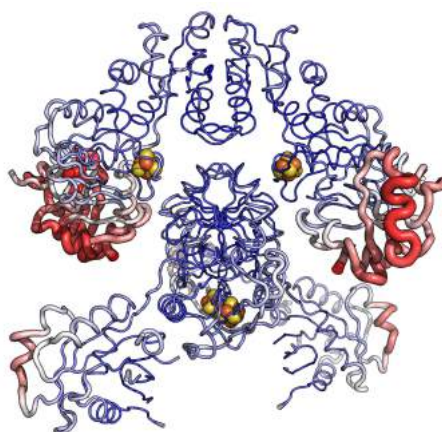
Figure 2.9 B factors of [4Fe-4S]-TmeHydF. B factors are coloured from blue to red for the range [40:170] Å.[45]

Thus, the first aim of my project was obtaining a crystal structure of HydF containing both the [4Fe-4S]-cluster and the [2Fe]-subcluster. In order to do that, it was necessary to design a truncated form of HydF from *Thermosipho melanesiensis,* cutting the GTP-ase domain from the wild type form. The idea was based on the fact that GTP domains have been observed to be not always good oriented and often disordered areas, causing a low-resolution diffraction of the crystals obtained and being a cause of trouble for the binding of the diiron subcluster.

2.8.2 HydA from *Megasphaera elsdenii*

Structures of HydA from three different microorganisms are already available.

The first structure was obtained from *Clostridium pasteurianum* (CpI) and it was composed of four distinct domains. The largest one is designated as active-site domain and it contains the active-site cluster, known as H-cluster. It is composed of six Fe atoms arranged as a [4Fe-4S]-cluster bridged to a [2Fe]-subcluster by a single cysteine. The [4Fe-4S] cluster is linked to the protein via four cysteines,

while the diiron subcluster is composed of two octahedrally liganded Fe atoms that contain five CO/CN ligands, three S ligands and one H_2O ligand. The other three domains contain the four accessory iron-sulfur clusters: three [4Fe-4S]-clusters and one [2Fe-2S]-cluster, arranged into three domains.[65]

Secondly, it was reported the structure of hydrogenase from *Desulfovibrio desulfuricans* (DdHydA), which presents an identical amino acid sequence with the hydrogenase from *Desulfovibrio vulgaris*. It reveals a conformation for the H-cluster similar to the one solved for CpI: in fact, it is composed of a ferrodoxin-type [4Fe-4S]-cluster, deeply buried into the protein, and a [2Fe]-subcluster, bridged by a cysteine. Each iron atoms of the binuclear cluster is coordinated to one CO, one CN^-, a monoatomic oxygen species and two bridging thiolates. The protein also contains three accessory iron-sulfur clusters, one of which is a [2Fe-2S]-cluster, as found in CpI.[66]

Finally, another structure was obtained from a truncated form of HydA from *Chlamydomonas reinhardtii* (CrHydA1), expressed in absence of the maturases HydE, HydF and HydG and therefore purified in its unmaturated form lacking the [2Fe]-subcluster. The structure presents many similarities with the structures of the other two proteins characterized. CrHydA1 it lacks any accessory cluster and therefore it has been appreciated for its simplicity that can be exploited for investigating maturation of the H-cluster.[67]

The three structures are reported in Figure 2.10.

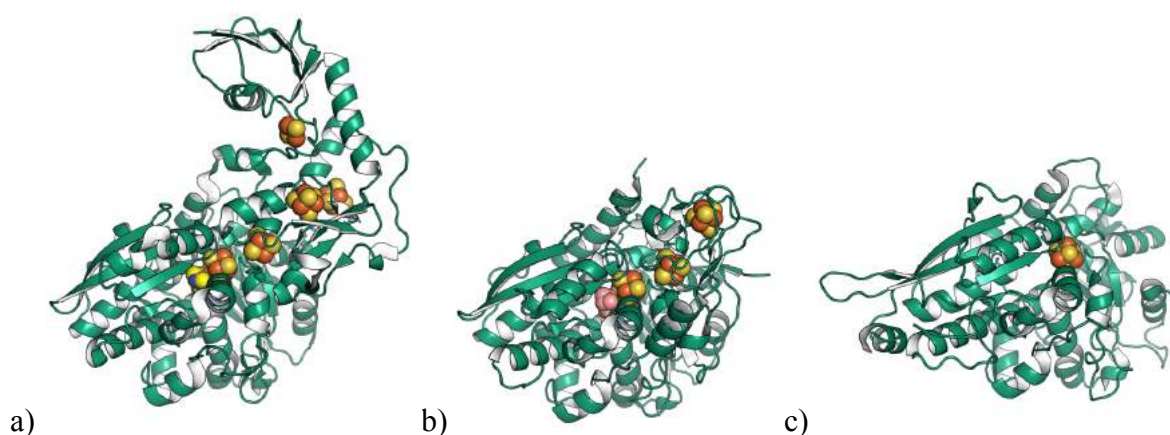a)   b)   c)

Figure 2.10 a) Structure of CpI. It is possible to distinguish the first domain containing the H-cluster, composed of a [4Fe-4S] cluster and a [2Fe] subcluster, and the three domains containing the four accessory iron-sulfur clusters. **b)** Structure of DdHydA. It is possible to distinguish the H-cluster, composed of a [4Fe-4S] cluster and a [2Fe] subcluster, and the three domains containing the two accessory [4Fe-4S] clusters. **c)** Structure of CrHydA. It contains only one [4Fe-4S] cluster and it lacks the [2Fe] subcluster and the accessory clusters. In each figure, iron atoms are represented in yellow and sulphur atoms are in orange.

Even if these structures are already available and provided important information about the behaviour of [Fe-Fe]-hydrogenases, another kind of HydA is being studied. This is the hydrogenase isolated from *Megasphaera elsdenii*, which showed very intriguing and difficulty explaining behaviours in terms of oxygen sensibility. In fact, the laboratory recently showed that it has higher resistance to oxygen, compared to the other HydAs. For this reason, we are interested in studying the structure of this protein in order to understand the cause of its higher oxygen resistance.

So, for the second aim of my project, I focused on the crystallisation of HydA from *M. elsdenii*. In particular, I worked on a truncated form, named MeH HydA, where all the accessory clusters have been eliminated, in order to obtain a simpler protein, in analogy to CrHydA1.

3. Methods and materials

3.1 Heterologous expressions of proteins

A heterologous expression is the expression of a gene in a host organism, which does not produce it naturally. The gene has to be inserted into its expression system through the recombinant DNA technology. The expression system can be a bacterium, a yeast or a plant cell but it has to have a property, known as competence. In molecular biology competence is the skill of a cell to uptake extracellular DNA from its environment. Cells can acquire this property with an artificial process. The laboratory already had competent cells from *E. coli* prepared with chemical method and these cells were used for all the transformations.

3.1.1   Transformation

The transformation is the operation that allows the competent cell to uptake the DNA molecule from the environment, through its plasmatic membrane. The DNA sequence that has to be amplified is contained in plasmids, small circular DNA molecules able to replicate independently. The schematic representation in Figure 3.1 shows an example of plasmid, in particular the pET22b, underlining its main components: it contains a sequence where the replication starts, called origin of replication (in blue), a gene coding for an enzyme giving the antibiotic resistance in order to select the cells that contain the plasmid (ampicillin resistance, in green), the gene of interest (in red) preceded by a t7 promoter (in cyan), necessary for T7 RNA polymerase binding and subsequent protein translation.
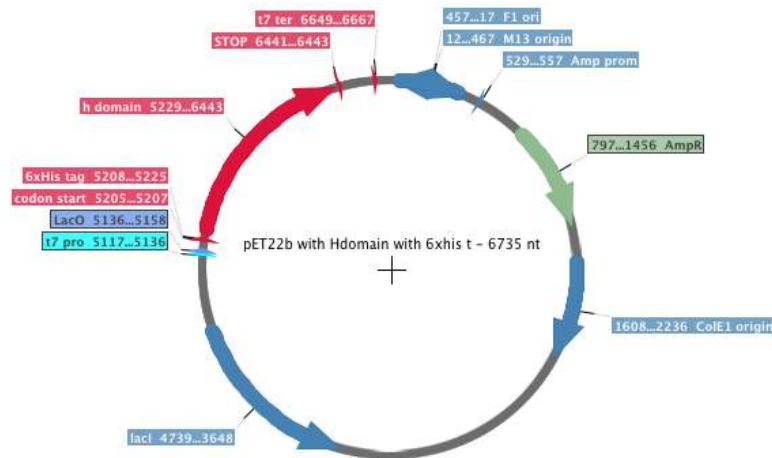

Figure 3.1 Schematic representation of pET22b plasmid for MeH HydA with 6 histidine tag.

MeHydA, MeH HydA and ΔTmeHydF without chaperones

The same protocol was used for MeHydA, MeH HydA and ΔTmeHydF without chaperones. Competent cells were transformed with the desired plasmid (respectively TunerDE3pLysS cells transformed with pT7-7-6HMeHydA, TunerBL21 DE3 cells transformed with pET22b and BL21DE3 cells transformed with pET22b plasmid): 1 µL of plasmidic DNA (corresponding to 50-100 ng) was added to 100 µL of cells and left 30 minutes on ice. The cells underwent a heat shock at 42°C for 45 seconds and were incubated 2 minutes on ice. After addition of 400 µL of warm Lysogeny broth (LB) at 37 °C, the cells were incubated at 37°C for 50 minutes. LB was prepared in the laboratory and it is a complex medium containing tryptone as source of peptides, yeast extracts (containing vitamins) and sodium chloride as sodium ions for transport and osmotic balance. Then, the cells were centrifuged at 2000 g for 5 minutes and part of the supernatants was discarded, in order to concentrate the cells. After re-suspension, 20% and 80% of the transformation product were loaded on two agar plates containing

Ampicillin 100 µg/mL and incubated overnight at 37°C. In these conditions, only cells containing the plasmid can grow, thanks to the antibiotic resistance provided by beta-lactamase.

ΔTmeHydF with chaperones

For the expression of ΔTmeHydF with chaperones, BL21DE3 cells were used as competent cells. 50 ng of pGRO7plasmid and 50 ng of ΔTmeHydF plasmid were added to 100 µL of cells. Then, the protocol was the same as before. The agar plates contained Ampicillin and Chloramphenicol at 50 and 34 µg/mL respectively.

3.1.2   Preculture

Three colonies were taken from the plates, put in 130 mL of growth broth containing antibiotic and left at 37°C overnight with an agitation of 200 rpm. Table 3.1 shows the medium used for each protein.

Table 3.1 Kind of antibiotic contained and medium used for the preculture of the different proteins produced.

PROTEIN	MEDIUM	ANTIBIOTIC CONTAINED
MeH HydA	TB	Ampicillin at 100 µg/mL
MeHydA	LB	Ampicillin at 100 µg/mL
ΔTmeHydF without chaperones	LB	Ampicillin at 100 µg/mL
ΔTmeHydF with chaperones	LB	Ampicillin at 50 µg/mL and Chloramphenicol at 34 µg/mL

Terrific Broth (TB) has been prepared in the laboratory and contains tryptone, yeast extracts, glycerol and potassium phosphate.

3.1.3   Culture

MeHydA and MeH HydA

After overnight preculture, a volume of preculture was added to 1 L of LB medium with the appropriate antibiotic such as to reach an initial value of OD at 600 nm between 0.05 and 0.08. The cells were grown at 37°C with continuous shaking at 200 rpm in the INFORS incubator. The OD at 600 nm was followed over time until it reached the value of 0.5-0.6. Then, protein expression was induced with Isopropyl β-D-1-thiogalactopyranoside (IPTG) at 0.5 mM. IPTG is a lactose metabolite able to trigger the transcription of the *lac* operon and it is therefore used to induce protein expression when the gene is under *lac* operon control. Cells were incubated for 5 hours at 20°C, then harvested by centrifugation at 6000 g at 4°C for 15 minutes and stored at -20°C.

Normally, we obtained ~2 g of wet cells per litre of culture.

ΔTmeHydF

After overnight preculture, a volume of preculture was added to 1 L of LB medium with the appropriate antibiotic such as to reach an initial value of OD at 600 nm between 0.05 and 0.08. The cells were grown at 37°C at 200 rpm in the INFORS incubator. When the OD at 600 nm the value of ~1.3, the chaperones expression was induced with arabinose at 2 mg/mL. When the OD reached the value of 1.8, the cultures were left at 4°C until the temperature in the incubator reached 16°C (20-30 minutes), then IPTG 0.1 µM was added and cells were grown overnight at 16°C. Cells were harvested by centrifugation at 6000 g for 15 minutes at 4°C and stored at -20°C.

Normally we obtained ~6 g of wet cells per litre.

3.1.4   Aerobic purification

MeHydA and MeH HydA

Cells were re-suspended with 4 mL/g wet cells of lysis buffer (50 mM Tris pH 8, 300 mM NaCl, 0.5% v/v Triton, 1 mM DTT and 5 mM β-met) and were sonicated on ice for 6 minutes in total (with cycles of 10 seconds ON, 40 seconds OFF) with a maximum amplitude of 40%. Sonication allows the cell lysis by cavitation, caused by high frequency sound waves emitted by the sonicator. Benzonase and magnesium were added to the cells to degrade nucleotidic polymers. The lysis product was centrifuged at 190000 g at 4°C for 45 minutes with an Optima XPN-80 Ultracentrifuge (Beckman Coulter) and the insoluble fraction was discarded. This step allowed the separation of cellular debris. The first purification step was an affinity chromatography, the HisTrap crude FF column (its chromatogram is reported in Figure 3.2): this column contains Ni^{2+} ions that can selectively retain the six histidine tag of the protein, while proteins which do not have this tag pass through the column. The protein is then eluted with an imidazole gradient. The supernatants were loaded on a HisTrap column equilibrated with Tris 50 mM pH 8 and 300 mM NaCl. After elution, fractions containing the protein of interest were pulled, stabilised with 5 mM DTT and concentrated by ultrafiltration to ~5-8 mL. GlyOH 20% v/v was added to stabilise the sample which was flash frozen in liquid nitrogen and stored at -80°C. The second purification step was a size-exclusion chromatography with a Superdex S200 26/600: this column allows the separation of the proteins based on their hydrodynamic size, thus small proteins will have a higher elution volume. The protein was thawed, centrifuged 15 minutes at 10000 g at 4°C and then loaded on the column equilibrated in 50 mM Tris-HCl pH 8.0, 300 mM NaCl, 10 % glycerol and 5mM DTT. Fractions containing the protein of interest were pulled and concentrated by ultrafiltration to 100 μM. The concentration was verified with Bradford assay, using bovine serum albumin as standard (see section 3.8). The sample was divided in aliquots of 1 mL, frozen in liquid nitrogen and stored at -80°C.

SDS-PAGE gel (see section 3.7) was done after each elution, in order to identify the fractions that contained the protein of interest and the protein purity.

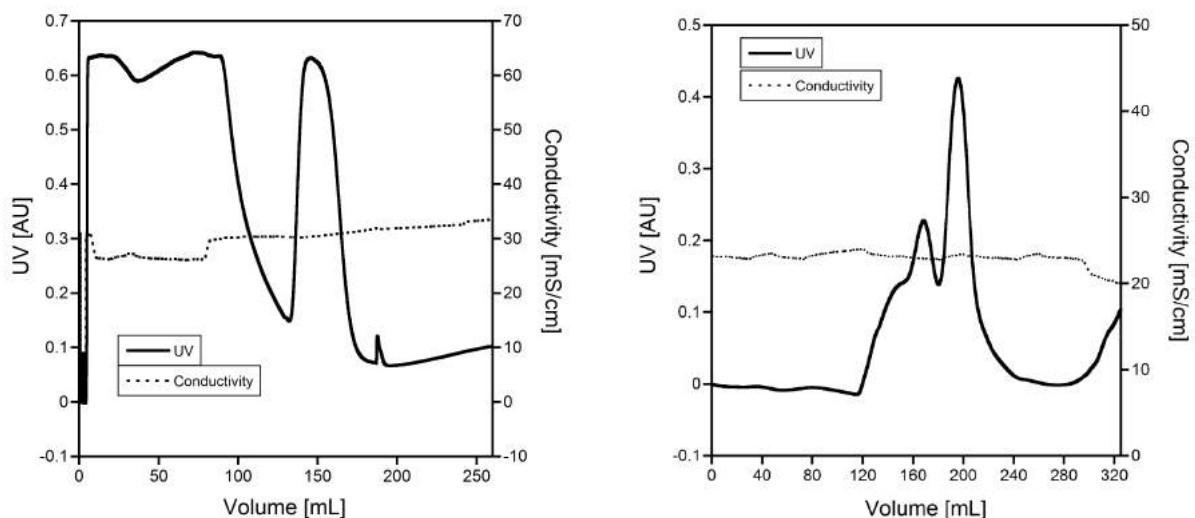The final yield was about 6.6 mg of protein/g wet cells.

Figure 3.2 Examples of chromatograms obtained during aerobic purification of MeH HydA.

ΔTmeHydF

The protocol is very similar to the one followed for HydA. Cells were resuspended in lysis buffer (50 mM Tris pH 8, 350 mM NaCl, 5% Glycerol, 5 mM imidazole, 1 mM DTT and 5 mM β-met) and

sonicated on ice. Lysozyme, benzonase and magnesium were added to help the lysis. The lysis product was centrifuged at 190000 g at 4°C for 45 minutes with Optima XPN-80 Ultracentrifuge (Beckman Coulter) and the insoluble fraction was discarded. Supernatants were loaded on a HisTrap column equilibrated with 50 mM Tris pH 8, 350 mM NaCl, 5% Glycerol, 5 mM imidazole and elution was performed with linear imidazole gradient from 10 to 500 mM. ΔTmeHydF containing fractions were pulled, stabilised with 5 mM DTT and concentrated by ultrafiltration to ~14 mL. GlyOH 20% v/v was added to stabilise the sample which was frozen in liquid nitrogen and stored at -80°C. The protein was thawed, centrifuged 10 minutes at 10000 g at 4°C and then loaded on Superdex S200 26/600, equilibrated in 50 mM Tris pH 8.0, 300 mM NaCl, 5 % glycerol and 5mM DTT. Fractions containing ΔTmeHydF were pulled and concentrated by ultrafiltration to 100 µM. The concentration was verified with Bradford assay, using bovine serum albumin as standard. The sample was divided in aliquots, frozen in liquid nitrogen and stored at -80°C.

The final yield was about 1.53 mg of protein/g wet cells.

3.1.5   Anaerobic reconstitution and purification

The same protocol was followed for all the proteins. [4Fe-4S]-cluster reconstitution was conducted in strictly anaerobic conditions in a M. Brown glovebox with less than 0.5 ppm of oxygen. An aliquot of protein at 100 µM was treated with 10 mM of DTT for 10 minutes, in order have all the cysteines reduced. Then, a 6 molar excess of Fe(II) from Mohr's salt was added as a source of iron and a 6 molar excess of L-cysteine as a source of sulphur. In addition, a catalytic amount of cysteine desulfurase CsdA was added to generate sulphide from L-cysteine. The reaction was incubated overnight and followed with UV-visible spectroscopy.

The purification of the reconstituted protein was conducted in anaerobic conditions and by size-exclusion chromatography. To remove precipitated protein and iron/sulphur polymers, the protein was centrifuged 20 minutes at 12.5 krpm then concentrated to 500 µL by ultrafiltration with a Vivaspin with 10 kDa cut-off. The sample was then injected on a Superdex S200 10/300 equilibrated with 50 mM TRIS pH 8.00, 300 mM NaCl, 5% glycerol, 5 mM DTT. The peak was collected and a UV-Vis spectrum was recorded to evaluate the Abs 400/280 ratio. An example of elution profile is reported in Figure 3.3.
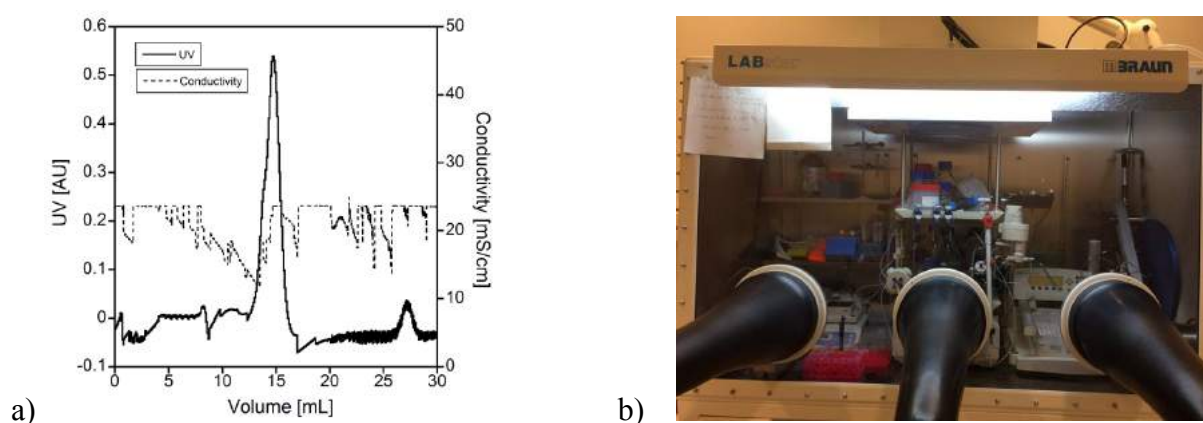

Figure 3.3 a) Example of elution profile of Superdex S200 10/300 for ΔTmeHydF. The peak at 15 mL represents the fractions containing the protein of interest. **b)** M. Brown glovebox with less than 0.5 ppm of oxygen.

3.2 Chemical maturation and [2Fe]-subcluster transfer from HydF to HydA

The [4Fe-4S] protein underwent a chemical maturation in order to incorporate the [2Fe]-subcluster. As [2Fe] precursor I used $(Et_4N)_2[Fe_2(adt)(CO)_4(CN)_2]$ (adt-complex) or $(Et_4N)_2[Fe_2(pdt)(CO)_4(CN)_2]$ (pdt-complex), already prepared in the laboratory. The protein was washed with a DTT free buffer, 100 mM potassium phosphate at pH 6.8, in order to avoid that the reducing agent chelate metals. The

protein was incubated with a 10 molar excess of the adt- (or pdt-) complex in strictly anaerobic conditions. After one hour, excess chemical was removed with a desalting column (NAP-10) and the protein was concentrated to 0.5-1 µM. The concentrated protein was frozen at -80°C or directly used for experiments. The incorporation of the diiron subcluster was assessed by comparing the UV-Vis spectra of [4Fe-4S]-hydrogenase before and after treatment with adt (or pdt). Incorporation of the mimic was previously shown by iron quantitation.

The reconstituted [2Fe]-ΔTmeHydF was used for activity test with Clark electrode or to test the ability of the protein to transfer the diiron subcluster to HydA. In order to do this latter, a 10 molar excess of [2Fe]-ΔTmeHydF was incubated with [4Fe-4S]-MeHydA and activity was tested after 15 minutes in order to verify the correct transfer of the subcluster.

3.3 Digestion with enzymes to cut a protein

Trypsin is an enzyme belonging to the serine proteases group, able to digest proteins, cutting the peptidic bond between amino acids. They are non-specific enzymes but trypsin cuts preferentially at the carboxyl side of the amino acids lysine or arginine.

This enzyme was used as an attempt to cut the GTPase domain from TmeHydF.

The digesting conditions were 1.5 mg/mL of protein in 50 µL of total volume. The protease was added in different quantities in order to test different trypsin to TmeHydF ratios, from 7,8125 to 1000. The digestion was conducted at 37°C for 30 minutes. The proteases action was verified with two SDS-PAGE gels: one was done to observe the effect of using different trypsin/protein ratios and the other followed the reaction in time, at one trypsin/protein ratio.

3.4 Molecular biology

3.4.1 Primers design and PCR

The molecular biology techniques were used in order to cut the GTPase domain from the wild type HydF from *Thermosipho melanesiensis*. First of all, it was necessary to remove the part of the gene coding for the GTPase domain. To do that, we designed the PCR primers to obtain the new gene. The primers were ordered from Eurofins and the sequences are reported in Table 3.2.

Table 3.2 Nucleotidic sequences of the designed primers.

Name	Sequence FW	Sequence REV
ΔTmeHydF	5' ATGGAAATTCCGTATCTGTCG 3'	5' ATGTATATCTCCTTCTTAAAGTTAAAC 3'
ΔTmeHydF HisTag	5' CTCGAGCACCACCACCAC 3'	5' GCGCATCAGCGGTTTAATC 3'

Once received the primers, the gene was cut with a Polymerase Chain Reaction (PCR). This technique consists in amplifying a region of the DNA molecule. Primers are necessary because they furnish the two extremities complementary to 3' needed to start the replication of DNA. It is necessary to mix the two oligonucleotidic primers and the DNA that has to be amplified with DNA polymerase and a sufficient quantity of nucleotides.

Primers and DNA were diluted at ~10 µM and concentration was verified with NanoDrop. Each PCR tube was prepared mixing:

- 12.5 µL of Q5 high-fidelity Master Mix (providing polymerase and nucleotides)
- 1 µL of forward primer
- 1 µL of reverse primer

- 1 µL of DNA
- 9.5 µL of water

for a total volume of 25 µL. The samples underwent a PCR in T100$_{TM}$ Thermal cycler by Biorad. The PCR is based on repeated cycles of DNA denaturation, primers hybridation and DNA synthesis by polymerase. Thus, the PCR program was made up of different steps, reported in Table 3.3.

The gene presence was verified by agarose gel electrophoresis, based on the separation of DNA molecules according to their length, applying an electric field. Migration is done at 100 V in Tris/Acetic acid/EDTA and DNA is revealed with SYBR Safe.

Table 3.3 Thermocycling conditions for PCR.

STEP	TEMPERATURE	TIME	CYCLES
Initial Denaturation	98°C	30 seconds	1
Denaturation	98°C	10 seconds	
Annealing	50–60°C	10 seconds	30 cycles
Extension	72°C	20 seconds/kb	
Final Extension	72°C	2 minutes	1
Hold	4°C		

3.4.2 Ligation of PCR product

The PCR product consists of a linear double DNA strand and it is necessary to recreate the circular form of the plasmid, in order to insert it in a cell with higher efficiency. So, the PCR product is incubated with kinase, ligase and the restriction enzyme DpnI. Ligase is able to join together DNA fragments, kinase can perform the phosphorylation, where the substrate gains a phosphate group and the high-energy ATP molecule donates a phosphate group (energy is produced and ligase uses this energy to settle the sugar-phosphate structure of DNA) and DpnI is a restriction endonuclease that allows the template removal. Actions of the three enzymes are shown in Figure 3.4. 1 µL of PCR product was digested 5 minutes at room temperature with 5 µL of KLD buffer and 1 µL of KLD Mix that contains the three enzymes.
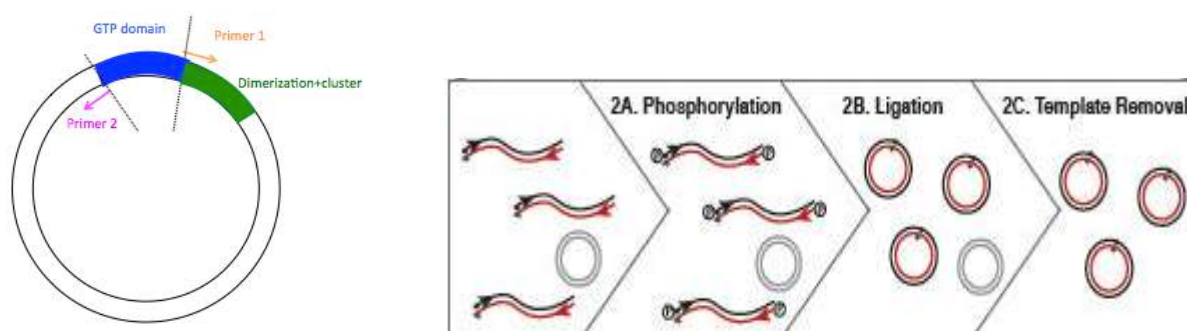
Figure 3.4 On the left, schematic representation of the TmeHydF gene that we want to cut: the two primers were designed in order to amplify only the part of the gene in which we are interested and not the GTPase domain part. On the right, schematic representation of the digestion of PCR product with KLD Mix, showing the three actions of ligase, kinase and DpnI.

3.4.3 Transformation

Then, the gene has to be incorporated into an expression system. We used *Escherichia coli*, a gram-negative bacterium. Commercial competent cells of *E. coli* were transformed with 5 µL of the digestion product, following the standard protocol for transformations.

3.4.4 PCR on colony

After one night incubation at 37°C, different colonies were obtained and it was necessary to verify if they contained the modified gene. So, one colony was added to a mix of

- 1 µL of forward primer corresponding to the beginning of the sequence of the new gene
- 1 µL of reverse primer
- 10 µL of DreamTaq Green PCR Master Mix that contains the DNA polymerase DreamTaq, the green buffer DreamTaq, $MgCl_2$ and deoxynucleosides triphosphate (dNTPs).
- 8 µL of water

and the operation was repeated for other colonies. The PCR program's steps are reported in Table 3.4. The PCR product was checked through agarose gel electrophoresis and some of the colonies that contained the right size DNA strand were grown in 10 mL of LB containing Ampicillin 50 µg/mL.

Table 3.4 Thermocycling conditions for the PCR

STEP	TEMPERATURE	TIME	CYCLES
Initial Denaturation	95°C	3 minutes	1
Denaturation	95°C	1 minute	
Annealing	53°C	1 minute	25 cycles
Extension	68°C	1 minute	
Final Extension	68°C	5 minutes	1
Hold	8°C		

The plasmid of interest was then purified with QIAprep Spin Miniprep Kit by Qiagen, concentrated to 70 ng/µL (the concentration was measured with NanoDrop) and sent for sequencing to verify the presence of the modified gene.

3.5 Expression and solubility test

Tests were done on 150 mL of preculture diluted in 100 mL of LB containing 100 µg/mL of Ampicillin so that the initial OD at 600 nm was 0.1. Growth and induction conditions varied during each test and are all reported in Section 4. After protein expression's induction with IPTG, aliquots of 500 µL were taken each hour for 4 hours and centrifuged at 6000 g for 5 minutes at 4°C. The pellet was dissolved in 50 µL of SDS-PAGE loading buffer and heated at 96°C for 5 minutes. These samples were used to test the expression. After 4 hours, two aliquots of 2 mL were taken and centrifuged at 6000 g for 5 minutes. The pellet was dissolved in 600 µL of lysis buffer (4.25 mL buffer A, 500 µL bug buster, 250 µL lysozyme at 10 mg/mL, where buffer A contains 30 mM Tris pH 8, 200 mM NaCl) with 1.5 mM of $MgCl_2$ and 0.1 µL of benzonase. After 15 minutes of incubation at room temperature, the samples were centrifuged at 18000 g for 15 minutes at 4°C. The first sample was used to test the protein solubility: 20 µL of SDS-PAGE loading buffer were added to 100 µL of supernatants, while the pellet was dissolved in 600 µL of lysis buffer, sonicated 5 minutes in sonication bath and then 100 µL were added to 20 µL of SDS-PAGE loading buffer. The second sample was used to test the thermo stability of the protein: all supernatants were incubated at 50°C and aliquots of 50 µL were taken every 5 minutes for 20 minutes and centrifuged at 18000 g for 5 minutes. 10 µL of SDS-PAGE loading buffer were added to supernatants, while the pellet was re-suspended in 50 µL of lysis buffer and then added to 10 µL of SDS-PAGE loading buffer. All samples were heated at 96°C for 5 minutes and then charged on SDS-PAGE gel (8 µL).

3.6 Iron quantitation with Fish method

To determine the quantity of iron contained in the protein, Fish method was used. Before starting the quantification, a calibration curve has to be made. A commercial standard solution of iron at 1000 mg/mL was diluted 36 times to obtain a 0.5 mM solution. This solution was then diluted in water in order to have a range from 0 to 10 nmol. The test was done with different protein quantities, normally between 0.5 and 4 nmol, depending on the expected amount of iron. Normally, the quantification was done on two different concentrations and the measure was repeated three times for each concentration. The protein was incubated 15 minutes at room temperature with 45 µL of a 1 M solution of perchloric acid which denatures the protein and releases iron. After 5 minutes of centrifugation at 10000 rpm, 90 µL of supernatants were incubated with 72 µL of bathophenantroline disulfonic acid at 1.7 mg/mL that has the function to chelate the iron, 36 µL of sodium ascorbate at 38 mg/mL that is able to reduce iron in Fe^{2+} form and 27 µL of ammonium acetate (saturated solution diluted 1/3) which neutralizes the environment and reveal the complex Fe-bathophenantroline (a red complex). After incubation of 30 minutes, the samples are centrifuged 5 minutes at 10000 rpm and analysed at UV-Vis spectrum. The absorption difference Abs 535-680 was measured and the quantity of iron was calculated using the calibration curve in Figure 3.5.
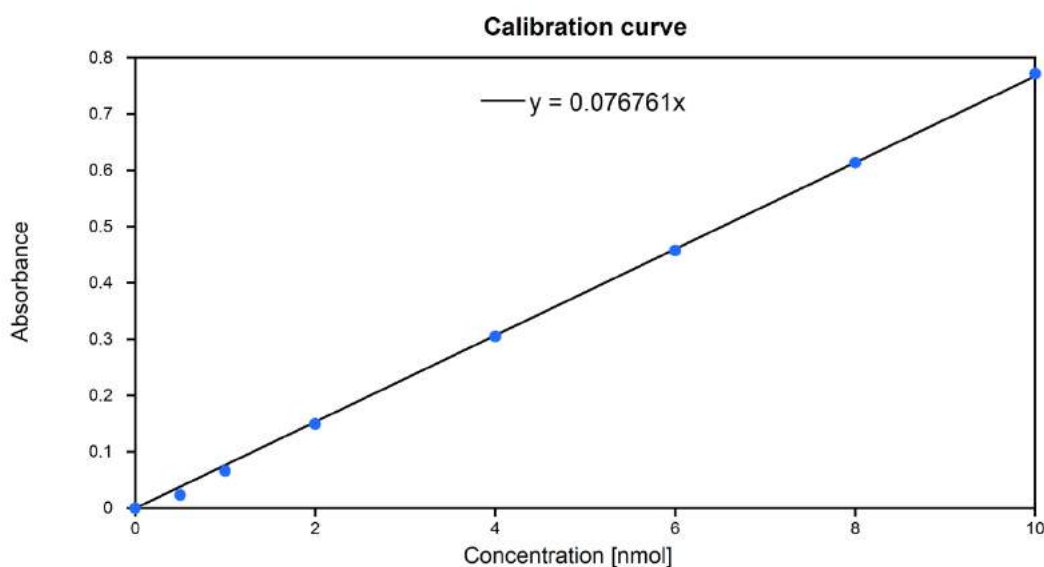
Figure 3.5 Calibration curve for iron quantitation in a range of Fe between 0 and 10 nmol. The line slope is used as coefficient to determine the iron quantity in the protein. (R^2=0.9994)

3.7 Gel electrophoresis

3.7.1 Sodium dodecyl sulphate polyacrylamide gel electrophoresis (SDS-PAGE)

Polyacrylamide gel electrophoresis is a technique used to separate biological macromolecules based on their electrophoretic mobility. Mobility depends on length, conformation and charge of the molecule. With SDS-PAGE electrophoresis, a denaturant (sodium dodecyl sulphate) is added to destroy the native structure of the molecule and turn it into a molecule whose mobility is only function of length and mass to charge ratio. When the detergent binds the protein, it transforms the molecule into a negatively charged linear chain that can move toward the anode if subjected to an electric field. The mobility of the molecule depends only on its molecular weight.

The gel is made up of two parts:

- The higher part is the *stacking gel* that allows concentration of the loaded sample.

- The lower part is called *resolving gel* and it allows the separation of molecules based on their molecular weight.

The components of the two layers are reported in Table 3.5. After polymerisation of the acrylamide, samples were loaded on gel with 20 μL of SDS-PAGE running buffer. The running buffer is composed of Tris 25 mM, Glycine 200 mM and SDS 0.1%.

A molecular marker containing protein standards is also loaded on the gel to estimate the molecular weight of all the bands on the gel.

Table 3.5 Polyacrylamide gel recipes for SDS-PAGE electrophoresis.

Resolving gel 10%		Stacking gel 4%	
Tris pH 8.9	4.44 mL	Tris pH 6.8	0.84 mL
Acrylamide	2.5 mL	Acrylamide	0.5 mL
SDS 20%	50 μL	SDS 20%	50 μL
Milli-Q water	2.85 mL	Milli-Q water	3.51 mL
APS	100μL	APS	50μL
Temed	10 μL	Temed	10 μL

3.7.2 Agarose gel electrophoresis

Agarose gel electrophoresis is another method of gel electrophoresis used to separate a mixed population of macromolecules, such as DNA, RNA or proteins in a matrix of agarose. This is an easy technique particularly suitable for separating DNA fragments of size range common found in laboratories.

The matrix is made of natural polysaccharide polymers extracted from seaweed. During gelation, agarose polymers associate non-covalently and form a network of bundles whose pore sizes determine a gel's molecular sieving properties.

The phosphate backbone of the DNA molecule is negatively charged, therefore, when an electric field is applied, DNA fragments migrate to the positive charged anode. DNA molecules are separated by size within an agarose gel in a pattern such that the distance travelled is inversely proportional to the logarithm of its molecular weight. The distance between DNA bands of different lengths is influenced by the percentage of agarose in the gel, with higher percentages requiring longer run times. We used gel with 0.8% agarose. The rate of migration of a DNA molecule through a gel is influenced by size of DNA molecule, agarose concentration, DNA conformation, voltage applied, type of agarose and electrophoresis buffer. After separation, the DNA molecules can be visualized under UV light after staining with an appropriate dye.

3.8 Determination of protein concentration

The concentration of a protein was determined with the Bradford protein assay. The Bradford assay is a colorimetric protein assay, based on the binding of the Coomassie Brilliant Blue G-250 dye to the protein. In fact, under acidic pH, the red form of the dye is converted into its blue form, binding to the protein being assayed. Within the linear range of the assay, there is a proportional bind between the dye and the proteins, so that the more the protein is present, the more the dye binds and becomes blue. The absorbance at 595 nm is measured and compared to a protein standard. In my case, the standard protein was Bovine Serum Albumin (BSA).

The concentration calculated with Bradford assay was always compared to the one determined using the molar extinction coefficient (ε_{280nm}), as shown in Figure 3.6.
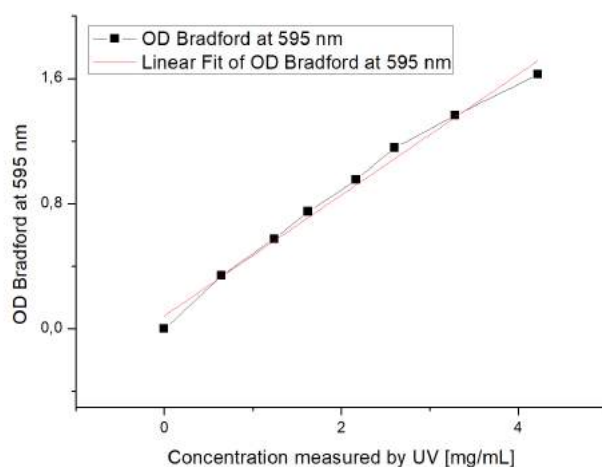

Figure 3.6 Comparison between concentrations measured with UV and with Bradford assay for MeH HydA.

3.9 Activity test: in vitro enzymatic assays

Two methods were used to quantify the specific activity of the enzyme, based on the expected amount of hydrogen.

3.9.1 Gas chromatography

To detect high quantities of hydrogen gas chromatography was used. In particular, I used it to calculate the specific activity of MeH HydA and MeHydA. Gas chromatograph is an instrument used for the separation of chemicals in a complex sample between a stationary phase and a gas mobile phase, depending on their physical properties and on their affinity with the stationary phase. In this way, each component exits at a different retention time, depending on its rate through the column. At the end of the column, a detector is able to electronically identify each component.

The samples for the activity test were prepared in 10 mL vials sealed with rubber stoppers in strictly anaerobic conditions. The sample contained 2-3 μM of reconstituted HydA, 100 mM of sodium dithionite as electron donor, 10 mM of methyl viologen as electron mediator and 100 mM of potassium phosphate buffer at pH 6.8. In this way, HydA is able to catalyse the production of hydrogen. The sample was immediately used for the activity test: a GC system (Shimadzu GC-2014 with thermal conductivity detector and Quadrex column) recorded a gas chromatogram like the one reported in Figure 3.7. The amount of hydrogen was determined through a calibration curve (Figure 3.8). The gas chromatogram returns a value of peak area. From the area of the peak we are able to calculate the moles of hydrogen produced in the vials and, from there, to calculate the specific activity (Equations 3.1 and 3.2).

$$n_{H2} \ in \ the \ vial = \frac{pick \ area}{calibration \ gradient} * \frac{[total \ volume \ of \ vial - volume \ of \ liquid \ in \ vial]}{injected \ volume} \qquad (3.1)$$

$$specific \ activity = \frac{n_{H2 \ in \ the \ vial}}{molecular \ weight * n_{catalyst} * time} \qquad (3.2)$$

where n stands for number of moles.

The specific activity represents the quantity of product formed by an enzyme in a given amount of time and under given conditions per mass of total protein. It is the activity of an enzyme per milligrams of total protein and it is expressed as (μmol of product)*min^{-1}*(mg of protein)$^{-1}$.
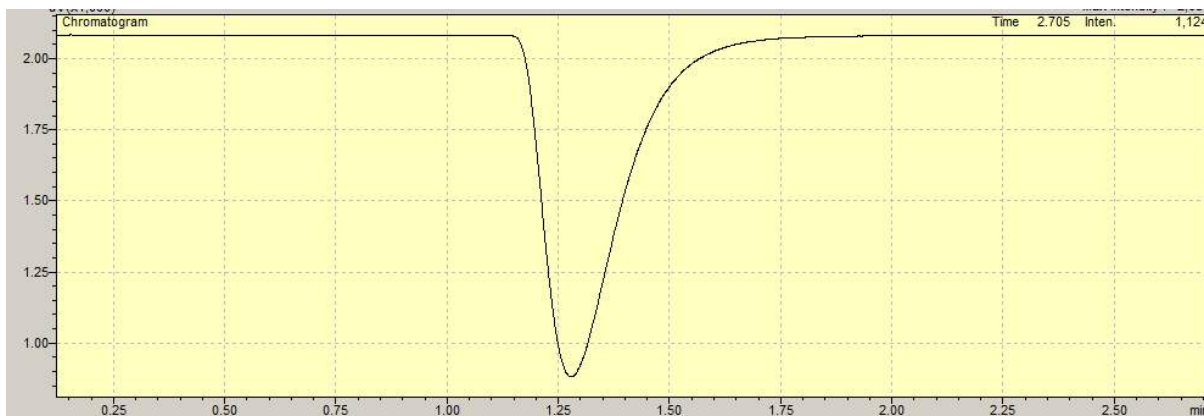
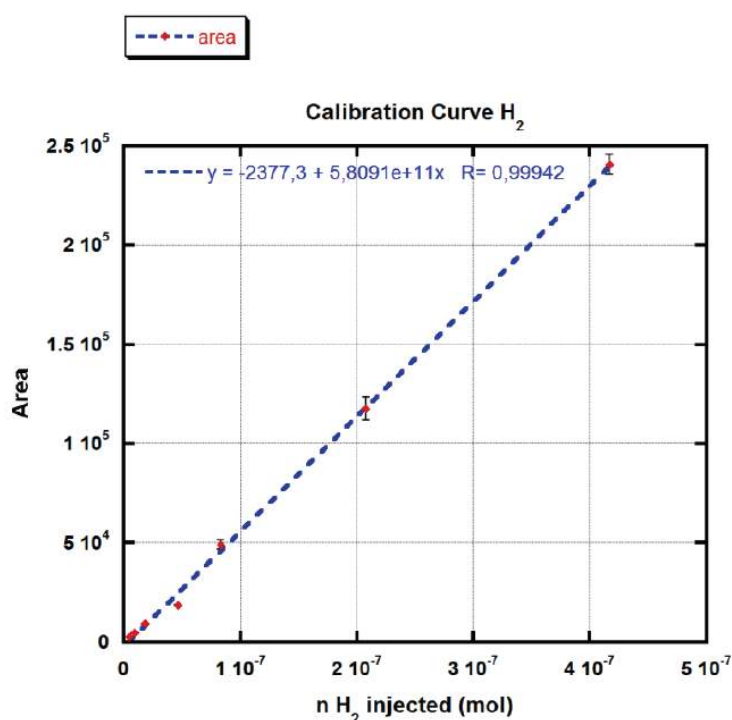**Figure 3.7** Example of a GC chromatogram. The area of the peak is proportional to the amount of hydrogen detected.

Figure 3.8 Calibration curve.

3.9.2 Clark type hydrogen sensor

Clark electrode was used to detect little amounts of hydrogen, for example H_2 produced by HydF proteins. This is an amperometric technique based on the measure of hydrogen partial pressure and it allows the detection of low hydrogen quantities, in order to avoid big consumptions of protein. The working principle of the sensor is based on the diffusion of hydrogen through a silicone membrane to a platinum anode where the H_2 is oxidized. The sensor is connected to a high quality picoammeter where the anode is polarized against the internal Ag/AgCl cathode. The picoammeter transforms the current derived from the oxidation into a sensor signal. With this method is possible to have a continuous measure of produced hydrogen as a function of time. The sensor has to be polarised at 1000 mV before starting the measures and the H_2 concentration is quantified with a calibration curve made with two points at 0 and 100% of pure hydrogen at 20°C.

Tests were performed in a miniaturized glass chamber and samples were prepared in anaerobic conditions and sealed with silicon and paraffin oil. Samples contained 4 µM (or 1.35 µM) of holo-ΔTmeHydF, 4 mM of methyl viologen as electron mediator and 4 mM of sodium dithionite as electron donor in 100 mM of potassium phosphate buffer at pH 6.0, containing 100mM NaCl. Sodium dithionite was added with a syringe immediately before the start of the measure. The sample is mixed with a magnetic stirrer in order to promote the diffusion of the produced hydrogen in the solution.

The test returns a diagram, which shows the variation of hydrogen micro molarity during time. From this, it is possible to evaluate the activity as turnover number (TON) and turnover frequency (TOF).

$$TON = \frac{\Delta \mu M \, H_2}{\mu M \, enzyme} \tag{3.3}$$

$$TOF = \frac{TON}{time \, unit} \tag{3.4}$$

The turnover number represents the number of times each enzyme molecule carries out its catalytic cycle per second. The turnover frequency is a measure of the number of catalytic cycles occurring at the catalytic centre per unit time. It is possible to correlate these values to the specific activity, known the enzyme weight:

$$specific \, activity = \frac{TOF}{enzyme \, molecular \, weight} \tag{3.5}$$

3.10   Dynamic light scattering

Dynamic light scattering (DLS) is a technique used to measure the size of particles. It is based on the measure of Brownian motion of particles, which is defined by a property known as translational diffusion coefficient D and which can be related to size of particles. The larger the particle, the slower the Brownian motion will be. The size of the particle is related to the diffusion coefficient by the Stokes-Einstein equation:

$$d(H) = \frac{kT}{3\pi\eta D} \tag{3.6}$$

where $d(H)$ stands for hydrodynamic diameter that represents how a particle diffuses within a fluid, k is the Boltzmann's constant, T is the absolute temperature, η is the viscosity and D is the translational diffusion coefficient. The diffusion coefficient is affected by particle's size but also by many other factors, such as surface structure, concentration and ionic strength of the medium. Ions in the medium, in fact, can change the Debye length (K^{-1}), a property that represents the thickness of the electric double layer, influencing the diameter value. All the particles are approximated to spherical particles, thus the hydrodynamic diameter is the diameter of a sphere that has the same translational diffusion speed as the particle.

The light scattering study is based on two theories:

- Rayleigh theory asserts that the intensity of light scattered I is proportional to the diameter to the sixth power d and inversely proportional to the laser wavelength to the fourth power λ.

$$I = I_0 * \frac{(1+cos^2\theta)}{2R^2} \left(\frac{2\pi}{\lambda}\right)^4 \left(\frac{n^2-1}{n^2+2}\right)^2 \left(\frac{d}{2}\right)^6 \tag{3.7}$$

  This means that big particles will scatter much light than little ones and so the measure can be very affected by contribution of larger particles.

- Mie theory is the theory that gives an explication of the maxima and minima in the plot of intensity with angle and is used for the conversion of intensity distribution in volume distribution.

So, DLS measures the Brownian motion and, in order to do that, it measures the rate at which the intensity of the scattered light fluctuates. The intensity of a signal is compared with itself at different time intervals through a correlator. Large particles produce slow changes in the signal and so the correlation will persist for a long time. Correlograms (like the one reported in Figure 3.9) give a lot of information: the time at which the correlation starts to significantly decay is an indication of the

sample size; the steep of the line indicates the polydispersity degree of the sample (the steeper the line, the more monodiperse the sample). In order to obtain the size from the correlation function, two methods can be used:

- Cumulants analysis is based on fitting a single exponential to the correlation function, obtaining mean size and polydispersity index.

- The other method is based on the fitting of a multiple exponential to the correlation factor, obtaining the distribution of particle sizes (intensity size distribution).

The intensity size distribution can be converted into a volume distribution, using the Mie theory, and that will change the value of the diameter, giving more importance to other peaks' influence (Figure 3.10).
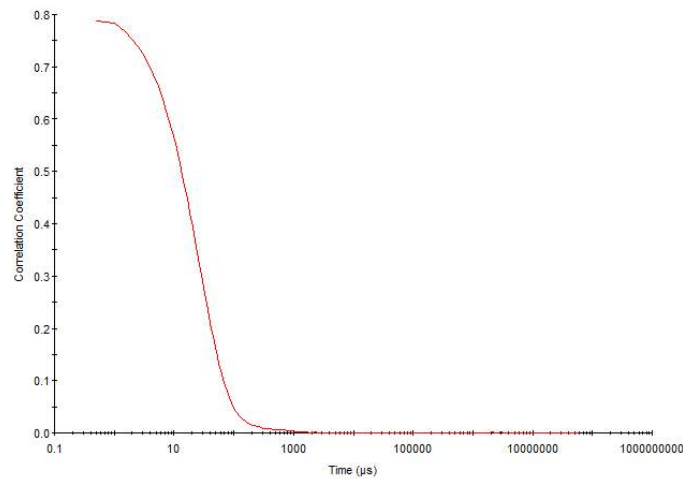
Figure 3.9 Typical correlogram from a sample containing protein molecules.
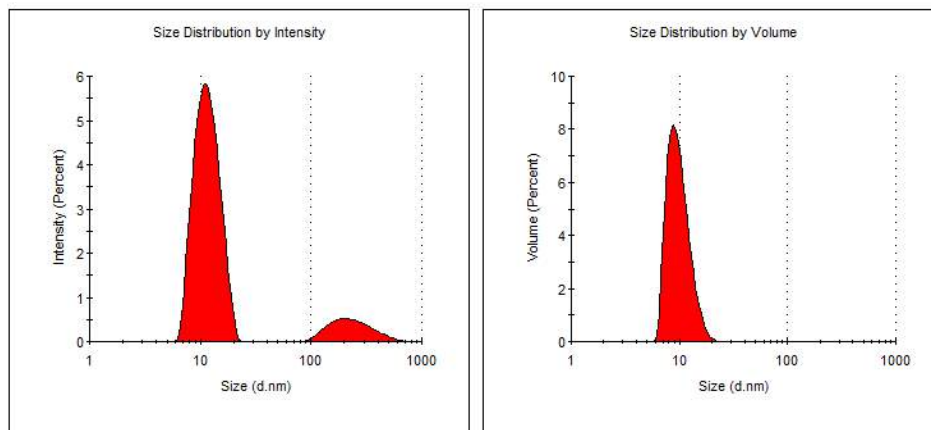
Figure 3.10 Typical size distribution based on intensity (on the left) and based on volume (on the right).

DLS measures were done with the Zetasizer Nano composed of six main components. The laser is the light source that illuminates the sample, placed in a cell. Scattered light is measured by a detector, an attenuator is needed to reduce the intensity of scatter and avoid the saturation of the detector. The scattering intensity signal is processed by a correlator, connected to a computer (Figure 3.11).

27
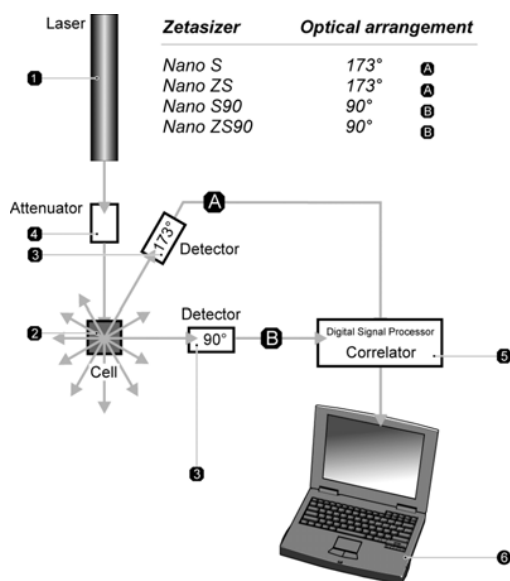
Figure 3.11 Schematic representation of optical configurations of Zetasizer Nano series for dynamic light scattering measurements.

For every test, samples were prepared with the protein at the desired concentration and filtered with Millipore 0.1 µM for 3 minutes at 12000 g in order to eliminate all the particles bigger than 100 µL. Low-volume quartz cuvette (ZEN2112) from Malvern Instruments was used for the measures with a volume of sample of 20 or 40 µL.

3.11   Protein crystallisation

Crystallisation is the process that leads to the formation of a crystal that is a solid where atoms are highly organised in ordered microscopic structures. Protein crystallisation is quite difficult because of the fragility of the molecules and the big quantities of solvent that they contain. One necessary condition to get crystals is supersaturation of the starting solution, but many factors can influence the process: temperature, air pressure, pH, ions concentration and precipitant and protein concentrations.

Crystallisation occurs in two steps:

• Nucleation is the appearance from a homogeneous solution of solid clusters, called nuclei. They are microscopic and not stable, but they can become stable if they reach a critical size.

• Crystal growth is the increase of the size of stable nuclei already formed.

The aim is to find the correct conditions to be in the nucleation zone. Observing the crystallisation phase diagram reported in Figure 3.12, it appears that, starting from a supersaturated solution, three zones can be reached, depending on the value of different parameters. The nucleation zone is between the metastable and the precipitation zone, but generally it is very hard to get crystals directly without passing through precipitation or metastable zones. Phase separation can be an indication of a metastable solution and this could be a promising starting point to reach the nucleation zone, changing chemical and physical parameters.
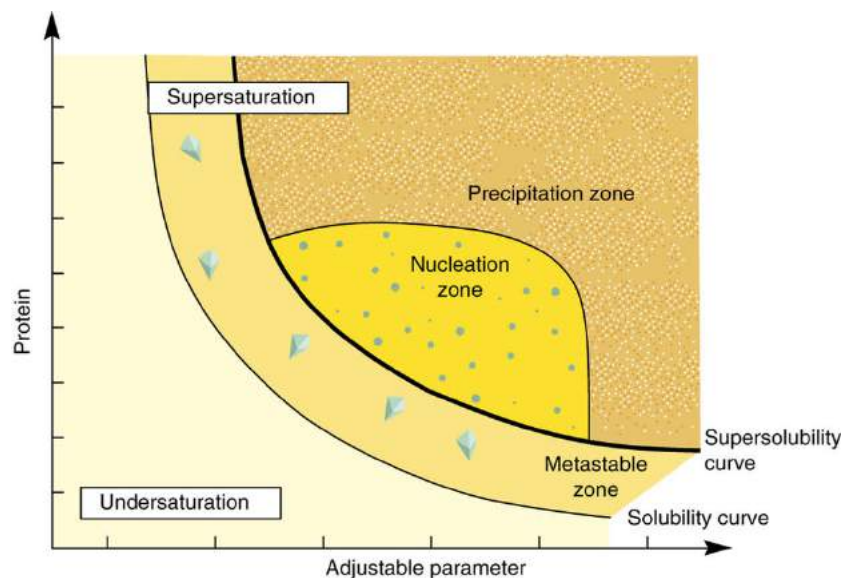
Figure 3.12 Crystallisation phase diagram. Undersaturated and supersaturated solutions are separated by the solubility curve. Between the two zones, there is a metastable zone, bounded by the solubility curve and the supersolubility curve. Above the supersolubility curve, it is possible to distinguish two zones: the nucleation zone and the precipitation zone.[68]

Different methods can be used to obtain crystals, but the most employed one is vapour diffusion. It is based on the presence of a reservoir of precipitants in vapour equilibrium with the crystallisation drop. Vapour diffusion can be performed in two ways. The first one is hanging-drop vapour diffusion in which a drop of protein solution is placed on a cover slide and mixed with a certain amount of precipitant solution from a reservoir. The system is closed with a cover slide and water vapour diffuses from the hanging drop to the reservoir, increasing proportionally both protein concentration and precipitant concentration, until vapour tension equilibrium is reached. The second way is sitting-drop vapour diffusion and in this case the drop rests on a shelf but the followed principle is the same as before (Figure 3.13).
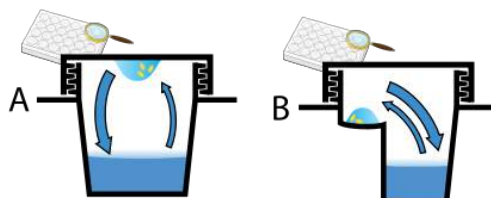
Figure 3.13 Schematic representation of the two kinds of vapour diffusion crystallisation. A) hanging-drop B) sitting-drop.

The preparation of the sample had to be very accurate because impurities can highly influence the crystallisation process. Some of the samples, because of their oxygen sensibility, were prepared in anaerobic conditions, in a glove box equipped for protein crystallisation with a robot OryxNano from Douglas instruments. For non-reconstituted proteins, samples were prepared in aerobic conditions using the Mosquito robot from TTP Labtech.

Different commercial kits were used to screen a set of mostly non-redundant conditions. To further investigate, some screens were prepared with a Dragonfly (TTP Labtech). All the commercial screens used are reported in tables 3.6 and 3.7.

Table 3.6 Commercial kits tried in order to crystallize MeH HydA.

apo-MeH HydA		[Fe-S]-MeH HydA	
Commercial kit	**Protein concentration**	**Commercial kit**	**Protein concentration**
PACT Premier	13, 10 and 5 mg/mL	PACT premier	16 and 12 mg/mL
PEGs II	13, 10 and 5 mg/mL	JCSG+	16, 12 and 8 mg/mL
MIDAS	13, 10 and 5 mg/mL	Mb Class II	20 mg/mL
Mb Class II	13, 10 and 5 mg/mL	PEGs II	17 and 14 mg/mL
JCSG+	13, 10 and 5 mg/mL		
Solubility screen	13 mg/mL		
Wizard Cryo	12.5 and 8 mg/mL		
SALT Rx	12.5 and 8 mg/mL		
MPDs Suite	12.5 and 8 mg/mL		
Classic II Suite	12.5 and 8 mg/mL		

Table 3.7 Commercial kits tried in order to crystallize ΔTmeHydF.

apo-ΔTmeHydF		[Fe-S]-ΔTmeHydF	
Commercial kit	**Protein (at 4.3 mg/mL)/reservoir ratio**	**Commercial kit**	**Protein concentration**
PACT Premier	1/2, 1/1, 2/1	PACT premier	8 mg/mL
PEG Rx	1/2, 1/1, 2/1	JCSG+	16, and 8 mg/mL
Solubility screen	1/2, 1/1, 2/1	Mb Class II	15 and 7.5 mg/mL
Mb Class II	1/2, 1/1, 2/1	PEGs II	15 and 7.5 mg/mL
JCSG+	1/2, 1/1, 2/1	Solubility Screen	7.5 mg/mL

3.11.1 Sample preparation

apo-proteins

Samples of apo-MeH HydA were prepared in aerobic conditions. The buffer was changed using a NAP-10 into 10 mM Tris pH 8.0, 150 mM NaCl and 1 mM DTT. The protein was then concentrated to 13 mg/mL using a Vivaspin with a 10kDa cutoff and measuring the concentration with Bradford protein assay. Apo-ΔTmeHydF in buffer Tris 50 mM pH 8.0, NaCl 300 mM, GlyOH 5% v/v and 5 mM DTT was concentrated at ~4.5 mg/mL.

[Fe-S]-proteins

Samples of [Fe-S]-MeH HydA were prepared in anaerobic conditions. The protein was reconstituted, working in a M Braun glove box with less than 0.5 ppm of oxygen and concentrated with a Vivaspin with a 10kDa cut-off to the desired concentration, measured with Bradford protein assay. The buffer was Tris 10 mM pH 8.0, NaCl 300 mM, GlyOH 10% v/v and 5 mM DTT for HydA and Tris 50 mM pH 8.0, NaCl 300 mM, GlyOH 5% v/v and 5 mM DTT for HydF.

3.11.2 Plates preparation

Aerobic conditions

The reservoirs of 96-wells-plates were manually filled with commercial screening kits or, in the case of manual kits, using the Dragonfly. Then, the Mosquito was used to mix normally protein and reservoir solution (100 nL ranges) and the plate was covered to avoid evaporation.

Anaerobic conditions

The plates are filled with reservoir solutions in aerobic conditions and then transferred in a glove box with less than 0.5 ppm of oxygen. Inside the glove box, a machine is used to mix reservoir solutions and protein. The plate has to be suddenly covered to avoid evaporation.

All the plates were stored at 18°C and observed with microscope.

Figure 3.14 reports examples of instrumentation used for crystallography.

Figure 3.14 Some of the instrumentations used for crystallography assays.

3.12 UV-Vis spectroscopy

UV-Vis spectroscopy is an analytic technique used for quantitative determination of different analytes. Molecules containing electrons of π or non-bounding type are able to absorb energy in form of ultra violet or visible light (150 to 800 nm). The absorption of this energy causes an excitation of the electrons, which are promoted to higher energy molecular orbitals. The more easily the electron is excited, the longer is the wavelength of light it can absorb. However, the wavelengths of absorption of a molecule are not only dependent on the chemical nature of the molecule, but also on the molecular environment. For this reason, UV-Vis spectroscopy is a good technique for the evaluation of ligand-binding reactions, enzymes catalysis and protein conformation. This kind of spectroscopy is normally performed with liquid solutions, but it is possible to study gas and solids too. It is a non-destructive technique, very sensitive and it requires small amounts of protein.

The optical absorption of proteins is usually measured between 275 and 280 nm. At this wavelength, in fact, absorption is mainly due to three amino acids: tryptophan, tyrosine and cysteine, with their molar absorption coefficients decreasing in that order.

The absorbance is related to intensity of light and it depends linearly on concentration, according to the Lambert-Beer equation (Equation 3.8), which is a fast way to calculate protein concentration through this technique.

$$A = \epsilon * c * l \tag{3.8}$$

where c is the molar concentration, l is the path length in cm and ϵ is the molar absorption coefficient. The value of ϵ changes from one protein to another one and it can be evaluated by a variety of experimental techniques or adding up the contributions of the aromatic amino acids of the protein.

All the measures were registered with a Cary UV-Vis provided by Agilent and a quartz cuvette of 1 cm length was used.

4. Results and discussion

4.1 A truncated form of HydF from Thermosipho melanesiensis

The first part of the project was focused on the study of the maturase HydF. As largely explained in the introduction, crystallographic structures of HydF are already available, from different microorganisms, including one with a [4Fe-4S]-cluster (see Figure 2.8). However, no structure has been obtained so far with the [2Fe]-subcluster. It has been observed that GTP domains are not very well ordered and are therefore a source of disorder in the crystal (see Figure 2.9). Thus, the idea was to remove the GTPase domain of the protein and try to obtain crystals with more diffracting power.

4.1.1 Strategy 1: digestion with trypsin

In order to cut the GTPase domain of the wild type protein, different strategies were tried. The first one was based on trypsin, an enzyme which is able to cut the proteins and cuts at the C terminus of a lysine or arginine. The idea was to perform limited proteolysis on the protein in correspondence and run a SDS-PAGE to see if individual and soluble domains were released corresponding to the GTPase and dimerization/cluster binding domains (trypsin was expected to cut after lysine in position 167). The digestion was conducted at 37°C for 30 minutes to determine the best trypsin/protein ratio. From the electrophoresis in Figure 4.1 a, a band at around 30 kDa was observed for different protein concentrations and, looking at the intensity profile with the trypsin ratio, we concluded the 1/125 ratio was a good starting point. So, the same experiment was done again with a nmol trypsin/nmol TmeHydF ratio of 1/125 and the digestion was followed for two hours. In the experiment followed in time, no intense band was observed at both 26 (dimerization and cluster binding domains) and 17 kDa (GTPase domain), indicating no expected domain was released upon proteolysis with trypsin (Figure 4.1 b). Though we could have identified the bands at ~30 and ~25 kDa by mass spectrometry, we decided to use another strategy.
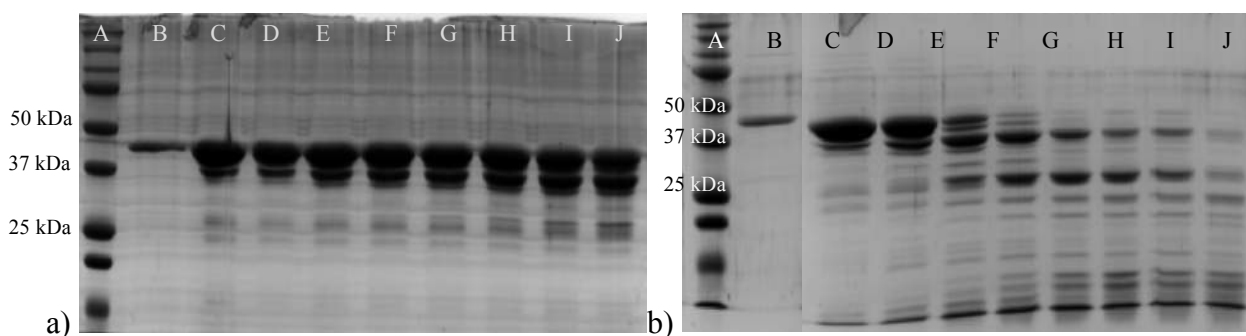
Figure 4.1 a) SDS-PAGE gel showing the digestion of TmeHydF with trypsin at different nmol trypsin/nmol TmeHydF ratios: A) molecular weight ladder, B) control: TmeHydF before digestion, C) ratio 1/7.813, D) 1/15.625, E) 1/31.25, F) 1/62.5, G) 1/125, H) 1/250, I) 1/500, J) 1/1000. **b)** SDS-PAGE gel showing the digestion of TmeHydF at nmol trypsin/nmol TmeHydF ratio of 1/125 at different times: A) molecular weight ladder, B) control: TmeHydF before digestion, C) after 5 minutes, D) after 10 min, E) after 20 min, F) after 30 min, G) after 45 min, H) after 60 min, I) after 90 min, J) after 120 min.

4.1.2 Strategy 2: molecular biology

The second strategy was the site directed mutagenesis on TmeHydF in order to cut the GTPase domain. First of all, it was necessary to design the primers, which were ordered from Eurofins. The primers were designed to amplify all the plasmid except the nucleotides coding for the residues 1-169 of TmeHydF. Then, a polymerase chain reaction was done with a gradient from 50 to 60°C, in order to

get the new construct, and the dimension of the PCR product was verified after electrophoresis on agarose gel (Figure 4.2). A band at ~6000 bp was observed, consistent with the expected size of the new plasmid construct (6054 bp).
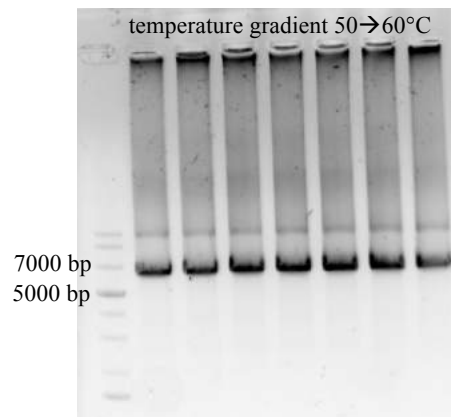

Figure 4.2 Agarose electrophoresis to verify the size of the new plasmid.

Kinase, ligase and DpnI were added to the PCR product and incubated for 5 minutes at room temperature to digest initial plasmid DNA and circularize the PCR product before transformation of DH5α competent cells. After overnight growth on LP-agar plate, some colonies were selected to test the presence and the size of the gene in the plasmid construct by PCR on colony using T7 and T7term primers. Some PCR products show a band at ~900 bp consistent with the expected size of 860 bp of the ΔTmeHydF gene. The other colonies have a PCR product of higher size below 1500 bp which corresponds to the initial plasmid with expected PCR product size of 1374 bp (Figure 4.3). This indicates that the DpnI digestion was not long enough to degrade the initial DNA template.

Figure 4.3 Agarose electrophoresis gel to verify the presence of the gene: column A contains the molecular weight ladder, the other columns show the gene contained in different selected colonies.

The gene was extracted from colonies 7 to 9 and the samples were sent for DNA sequencing to verify the correct sequence of the gene.

The next step was to verify the solubility and the expression of the recombinant protein. As shown on the SDS-PAGE gel in Figure 4.4, the protein had a very high expression level after 3 hours from induction with 0.5 mM IPTG at 37°C but all the protein was in the insoluble fraction after cells lysis. This indicates that the protein was not produced in soluble form in these conditions.

In order to have a soluble protein, another attempt was done using a lower induction temperature and a lower concentration of IPTG (0.1 mM at 20°C), exploiting the fact that lower temperatures and inducer concentration can slow the synthesis speed and hopefully help producing the protein in soluble form. As shown in Figure 4.5, the protein resulted to be well expressed after 4 hours but still insoluble, as only a negligible fraction of the protein is in the supernatants after cells lysis.
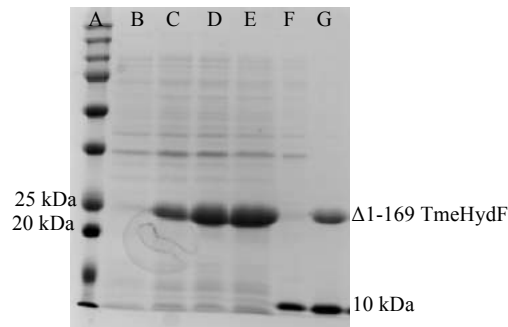
Figure 4.4 SDS-PAGE gel showing the expression and solubility test results. A) molecular weight ladder, B) Control: protein before induction, C) protein expression after 1h from induction, D) protein expression after 2h from induction, E) protein expression after 3h from induction, F) supernatants after cell lysis, G) precipitate after cell lysis. The band at 25.7 kDa represents ΔTmeHydF, the band at around 10 kDa represents the lysozyme that was added in the lysis buffer.
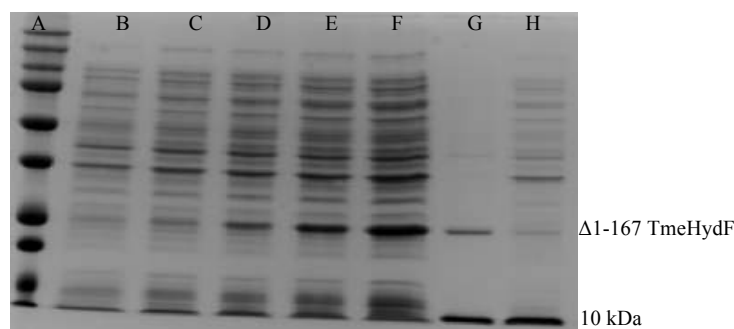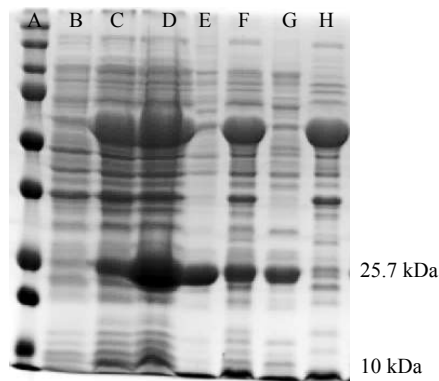

Figure 4.5 SDS-PAGE gel showing the expression and solubility test results. A) molecular weight ladder, B) Control: protein before induction, C) protein expression after 1h from induction, D) protein expression after 2h from induction, E) protein expression after 3h from induction, F) protein expression after 4h from induction, G) precipitate after cell lysis, H) supernatants after cell lysis. The band at 25.7 kDa represents ΔTmeHydF, the band at around 10 kDa represents the lysozyme that was added in the lysis buffer.

Failing to obtain a soluble form of ΔTmeHydF with this strategy, the other idea was using a late lag-phase induction protocol and adding to the cells a plasmid containing the chaperones groES-groEL in order to help the protein folding. The cells were co-transformed with the same quantity of plasmids for the chaperones and for the protein. Cells grew at 37°C, when the culture reached an OD at 600 nm around 1.3, arabinose was added to induce the chaperones expression. When OD at 600 nm reached 1.8, the temperature was decreased at 16°C and the protein expression was induced with IPTG at 0.1 mM. The culture was left overnight at 16°C. A thick band at 27.5 kDa was observed (Figure 4.6) after all night incubation, indicating that the protein is very well expressed. After cell lysis, the protein is fairly distributed between precipitate and supernatants, fact that allows the conclusion that this strategy led to the production of a soluble protein. A remarkable aspect is that the protein is not thermostable: after cells lysis the supernatants were separated from the pellet and were incubated at 50°C for a period of time. Then, the sample was centrifuged, the supernatants were separated from the pellet and the pellet was resuspended in lysis buffer. It was observed that already after 5 minutes at 50°C all the protein precipitated in the pellet and no protein was found in the supernatants. Therefore, while the wild type TmeHydF is thermostable, the deletion of the GTPase domain induced the loss of the protein thermostability.

Figure 4.6 SDS-PAGE gel showing the expression and solubility test results. A) molecular weight ladder, B) Control: before induction, C) protein expression after 1h from induction, D) protein expression after all the night, E) precipitate after cell lysis, F) supernatants after cell lysis, G) precipitate after thermostability test (5 minutes at 50°C), H) supernatants after thermostability test (5 minutes at 50°C). The band at 25.7 kDa represents ΔTmeHydF, the band at around 10 kDa represents the lysozyme that was added in the lysis buffer and the band at ~60 kDa represents the groEL.

So, this strategy worked for the production of a soluble protein, but the purification to homogeneity of that protein could be time-consuming due to the need of developing a protocol from scratch that involves a lot of trial and error. The purification procedure is also complicated by the fact that ΔTmeHydF is not thermostable anymore, a property that is extremely convenient to remove a lot of contaminants from *E. coli*.

This is why we decided to insert a six-histidine tag at the end of the polypeptidic chain to use nickel affinity chromatography as a purification step. This is an easy and quick process as a six-histidine tag is already present in the plasmid construct; only the deletion of a stop codon at the end of our gene is needed.

First of all, the primers were designed to amplify all the plasmid except the nucleotides coding for the residues corresponding to the stop codon of ΔTmeHydF. Then, a polymerase chain reaction was done with a gradient from 58 to 68°C. The dimension of the PCR product was verified after electrophoresis on agarose gel (Figure 4.7). After 5 minutes of incubation of the PCR product with kinase, ligase and DpnI at room temperature, a transformation of competent DH5α cells was done. After overnight growth on LP-agar plate, some colonies were selected to test the presence and the size of the gene in the plasmid construct by PCR on colony using T7 and T7term primers. Some PCR products show a band at ~900 bp consistent with the expected size of 860 bp of the ΔTmeHydF gene.

The plasmid was produced and purified from colonies 2, 4 and 8 and the samples were sent for DNA sequencing. Only clone 8 had the correct sequence.

In Figure 4.8 are shown the solubility and expression test results for the protein produced with and without chaperones.
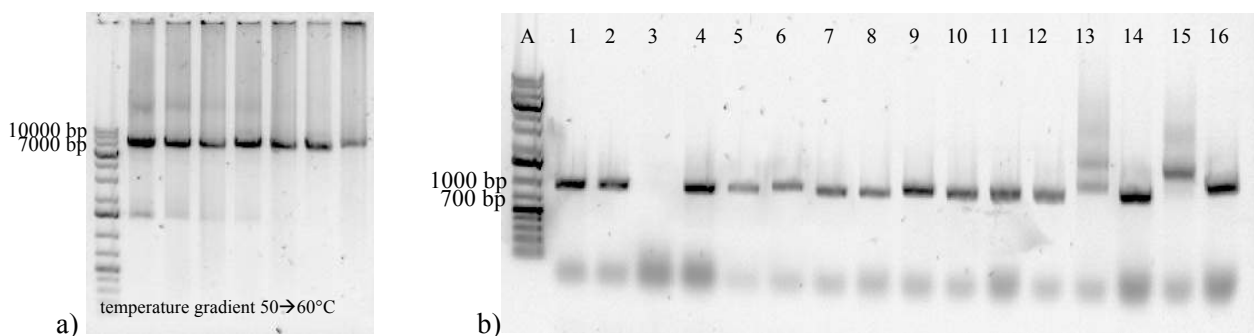
Figure 4.7 a) Agarose electrophoresis to verify the PCR product for the stop codon deletion; b) Agarose electrophoresis gel to verify the presence of the gene: column A contains the molecular weight ladder, the other columns show the gene contained in different selected colonies.
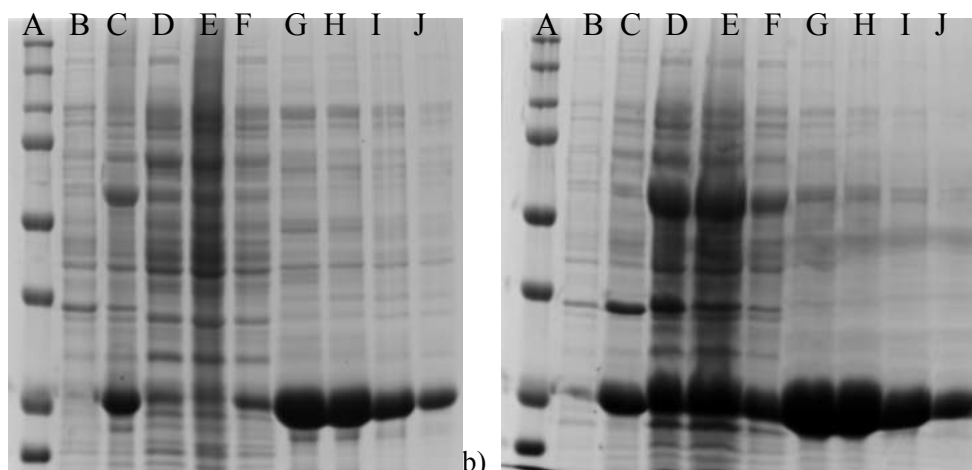
36

Figure 4.8 SDS-PAGE gel of (A) molecular weight ladder, (B) Control: protein before induction, (C) precipitate after cell lysis, (D) supernatants after cell lysis, (E) flowthrough (F) protein expression after 30 minutes from induction (G) protein expression after 100 min from induction (H) protein expression after 200 min from induction (I) protein expression after 350 (J) protein expression after 500 min from induction from induction for (a) ΔTmeHydF expressed without chaperones and (b) ΔTmeHydF expressed with chaperones.

4.1.3   Expression and aerobic purification

The plasmid obtained for ΔTmeHydF with the histidine tag was used for the heterologous expression of the protein in *E. coli*, using the BL21DE3 strain. The cells grew at 37°C (as shown in Figure 4.9), arabinose was added when OD at 600 nm was around 1.8 and the temperature was decreased at 16°C. When 16°C were reached, 0.1 mM IPTG was added to induce protein expression.
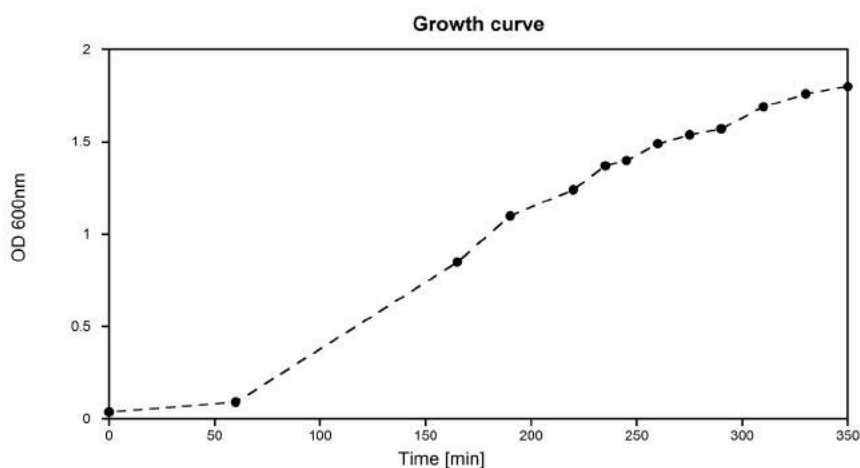
Figure 4.9 Growth curve of BL21DE3 cells transformed with pET-22b plasmid.

This strategy resulted in the production of a soluble and well expressed recombinant protein, as shown in Figure 4.10. From the gel, it appears that no chaperone is observed due to the fact that arabinose was added at OD 1.8 and temperature of 16°C, causing a poor expression of the chaperones. That implies that it is not necessary to obtain a soluble protein and only the late lag phase induction is relevant.
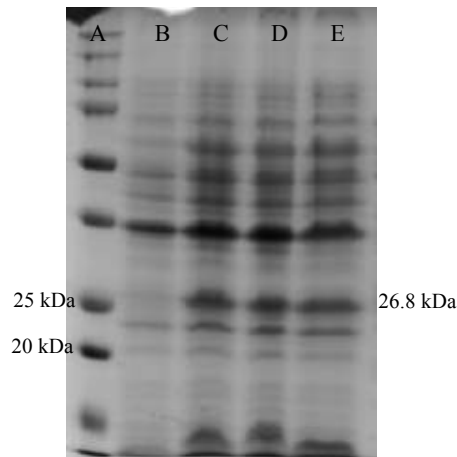
Figure 4.10 SDS-PAGE gel of (A) molecular weight ladder, (B) control before IPTG induction, (C) ΔTmeHydF expression after IPTG induction, (D) ΔTmeHydF in the pellet after cells' lysis, (E) ΔTmeHydF in the surnatants after cells' lysis.

The protein was then purified aerobically with two chromatographic steps: an affinity chromatography (HisTrap FF crude) that can separate the protein of interest from all the proteins without the six-histidine tag and a size-exclusion chromatography (Superdex S200 26/600) that can separate proteins based on their hydrodynamic size. In fact, the bigger is the size, the lower is the retention volume. For the protein, the retention volume is around 200 mL (Figure 4.11).

After these two steps an almost completely pure and homogeneous protein was obtained, as shown in the SDS-PAGE gel in Figure 4.11 c, with a final yield of 1.5 mg of proteins per gram of wet cells. As all the purification steps were done in aerobiosis, the protein was in its apo-form with a little content of Fe, as confirmed by the small shoulder at ~400 nm, observed by UV-visible (Figure 4.12).
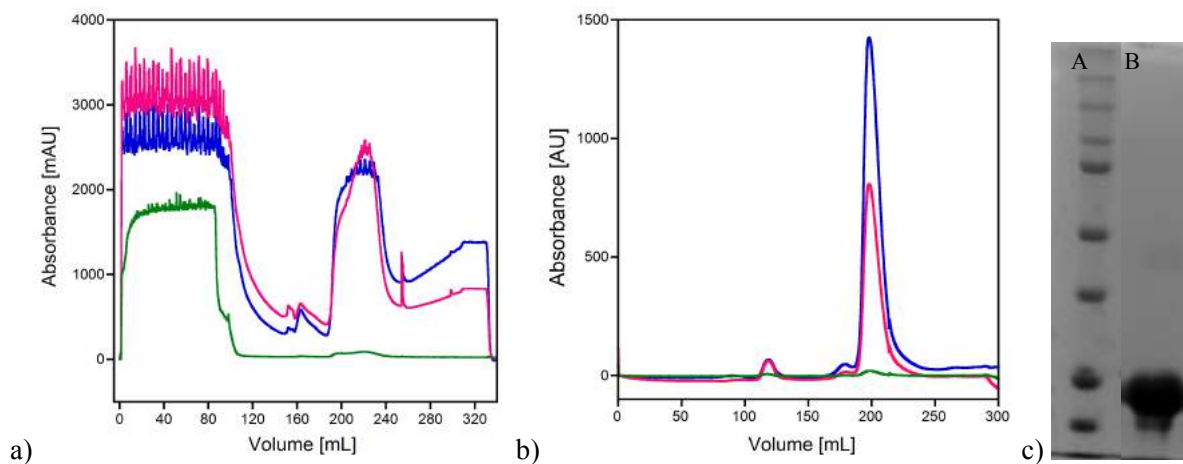
Figure 4.11 a) Chromatogram of HisTrap FF crude, showing the absorbance at 280 (blue), 260 (pink) and 400 nm (green); b) Chromatogram of Superdex S200 26/600, showing the absorbance at 280 (blue), 260 (pink) and 400 nm (green); c) SDS-PAGE gel of (A) molecular weight ladder and (B) ΔTmeHydF after purification.
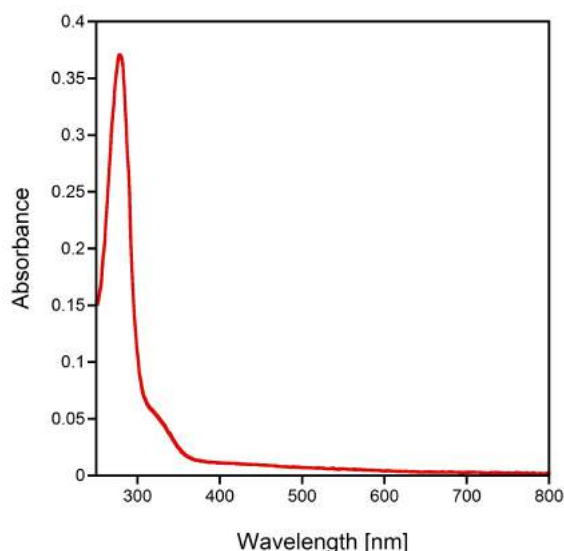
Figure 4.12 UV-Vis spectrum of purified apo-ΔTmeHydF.

4.1.4 [4Fe-4S] cluster reconstitution of apo form of truncated TmeHydF

As the purified protein was in its apo form, it was necessary to reconstitute the [4Fe-4S]-cluster. This incorporation was carried out using the same protocol used for HydA proteins. 100 µM protein was incubated in anaerobic conditions (<0.5 ppm of oxygen) with a molar excess of 6 of iron ammonium sulphate and L-cysteine, in presence of DTT as a reducing agent, in order to have all the cysteines reduced, and the cysteine desulfurase CsdA. The reaction was monitored overnight by UV-Visible absorption. An absorbance band at 400-420 nm grows and absorbance ratio between 400 and 280 nm increases until it reaches a constant value after the first 250 min (see Figure 4.13).

a)   b)
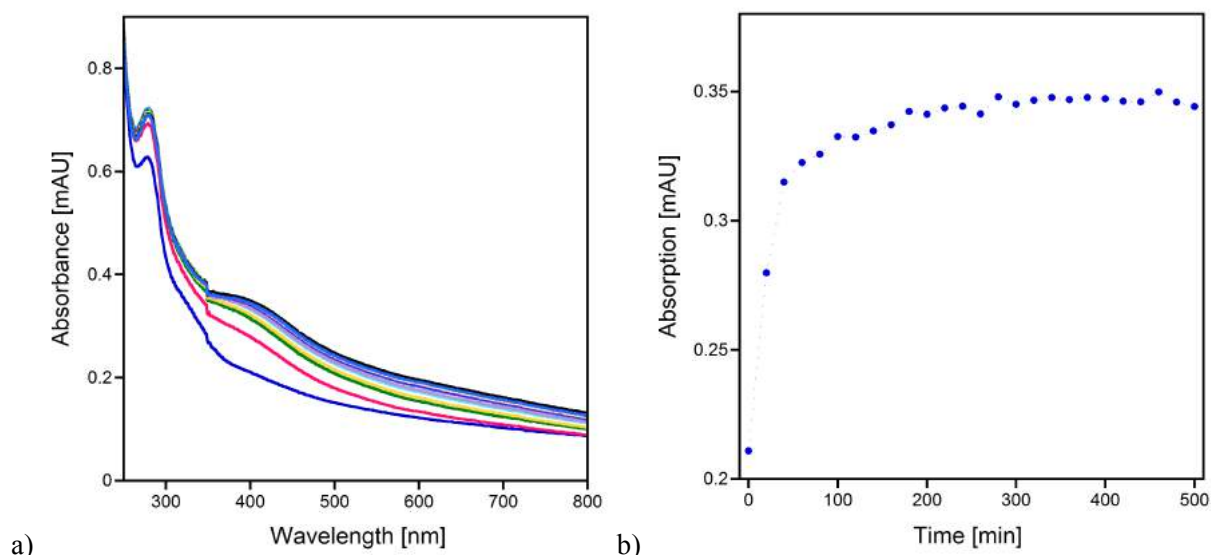
Figure 4.13 (a) UV-visible absorption spectra of ΔTmeHydF recorded every 20 minutes, showing the [4Fe4S] cluster reconstitution' kinetics. (b) Absorbance ratio 400/280 nm during time.

The protein was purified by size exclusion chromatography to eliminate all the non-specific iron as well as the possible aggregates that were formed during the reaction. The elution profile is shown in Figure 4.14 a. After the purification, the presence of the [4Fe-4S]-cluster was verified qualitatively with an UV-visible absorption spectrum and quantitatively with iron quantification using the Fish method. From the UV-Vis absorption spectrum (shown in Figure 4.14 b) the value of the absorbance

ratio 400/280 nm was calculated and resulted to be 0.33. From reconstitutions of other aliquots resulted that the absorbance ratio 400/280 nm is always around 0.33-0.34. The iron quantification gave a value of the ratio between nmol of iron and nmol of protein around 3.3, which is not very near to the theoretical value of 4. Probably, this low value can be explained considering a 10% error on the concentration evaluation.

An important issue that has to be solved is the yield, which is less than 50%. The loss is mainly due to the concentration step, probably because the protein binds the concentrator's membrane.

a)   b)
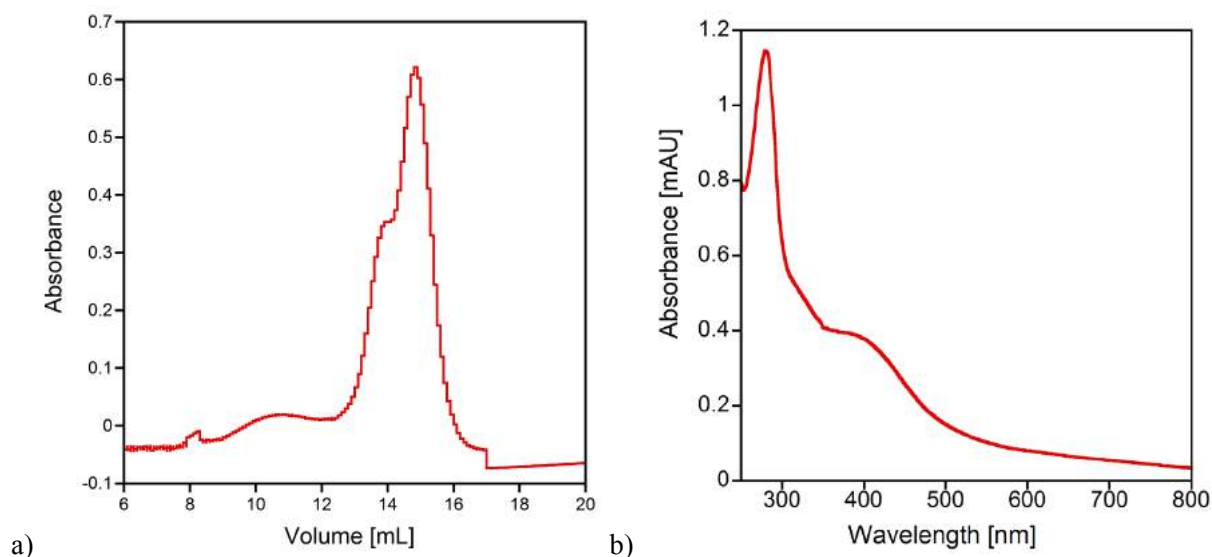
Figure 4.14 (a) elution profile of Superdex S200 10/300 (b) UV-Vis spectrum of [4Fe-4S]-ΔTmeHydF.

4.1.5   Chemical maturation

After the reconstitution, the protein contains only the [4Fe-4S]-cluster, but the binuclear subcluster is still missing, as explained in the introduction. It has been shown that treating HydF proteins with the synthetic complex results in a hybrid containing the [2Fe]-subcluster and able to activate HydA. This means that the natural complex synthesized by HydG and HydE is similar to the synthetic one.

After washing the protein with potassium phosphate at pH 6.8, [4Fe-4S]-ΔTmeHydF was reacted with a 10 molar excess of the $[Fe_2(adt)(CO)_4(CN)_2]^{2-}$ complex, in anaerobic conditions and for one hour, and then purified with a desalting column to remove all the mimic in excess. The incorporation of the [2Fe]-complex was proved by the UV-Vis spectrum shown in Figure 4.15 where a band appears at ~350 nm, indicating the presence of the diiron complex. This band is originated by the electron transfer from CO and CN ligands to the iron contained in the binuclear cluster.

As a maturase, the wild type form of TmeHydF has the ability to incorporate the adt-complex and to transfer it to MeHydA. It is known that GTP favours dissociation of HydE with HydF[47] and that GTP is not necessary for the [2Fe]-subcluster loaded HydF to activate HydA[51]. However, it has not been shown that the GTPase domain is necessary for HydF to interact with HydA. To this end, we checked if the truncated form of HydF from *Thermosipho melanesiensis* was still able to transfer the [2Fe]-complex to MeHydA. To do that, a 10 fold excess of adt-ΔTmeHydF was incubated with [4Fe-4S]-MeHydA for 30 minutes and then used to measure the hydrogen evolution. Once added methyl viologen and sodium dithionite to the protein, the sample was analysed on gas chromatography in order to quantify the amount of H_2 produced by the reaction. We reached values of the specific activity ~739 μmol H_2/min mg protein, comparable to the ones of MeHydA directly activated with adt-complex. As it was previously shown in the laboratory that the mimic does not dissociate from TmeHydF in this period of time, this indicates that the diiron complex was correctly transferred from ΔTmeHydF to MeHydA.
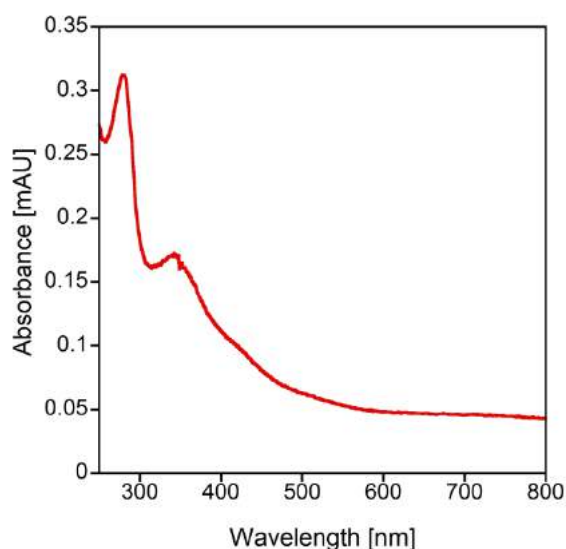
40

Figure 4.15 UV-Vis spectrum of adt-ΔTmeHydF.

4.1.6 Activity of ΔTmeHydF

Eliminating the GTPase domain of the TmeHydF led to a new hydrogenase which could potentially be an artificial hydrogenase. Thus, we measured the activity of the ΔTmeHydF, monitoring the quantity of hydrogen produced in time.

After washing the protein with potassium phosphate at pH 6.8 100 mM, [4Fe-4S]-ΔTmeHydF was reacted with a 10 molar excess of $[Fe_2(pdt)(CO)_4(CN)_2]^{2-}$ (pdt-complex), in anaerobic conditions for one hour, and then purified with a desalting column to remove all the mimic in excess. The pdt-complex was used (instead of adt-complex) because the laboratory already proved that this complex leads to higher activity levels for TmeHydF. To measure the activity of this protein the Clark's electrode was used, allowing the detection of little quantities of hydrogen. Tests were performed adding to the protein at a known concentration a ten fold excess of methyl viologen and sodium dithionite. The electrode measured the potential and the quantity of hydrogen produced was derived through a calibration curve. Following the hydrogen evolution in time, the activity of the protein was calculated and it resulted to have an initial TOF (calculated in the first 5 minutes) of 1 min^{-1}. This value is 3 times bigger than the activity obtained with TmeHydF activated with pdt-complex (Figure 4.16).[45]
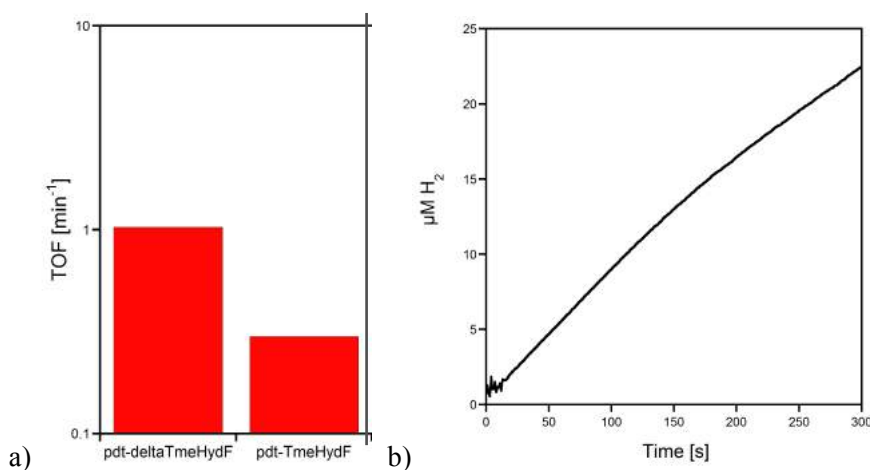
Figure 4.16 a) Activity comparison of different proteins activated with pdt-complex. **b)** Hydrogen evolution as a function of time for pdt-ΔTmeHydF.

These results have to be confirmed with further experiments, as the operation of eliminating the GTPase domain is not involving the active site of the protein, so there is no reason to expect a higher activity of the truncated protein.

4.1.7 Characterization by Dynamic Light Scattering (DLS)

Using Dynamic Light Scattering, we checked the stability of the purified protein in buffer TRIS 50 mM pH 8.0, NaCl 200 mM, GlyOH 10%, DTT 5 mM at two different concentrations. As reported in Figure 4.17 the value of the hydrodynamic diameter is quite constant, indicating that the sample is stable in the buffer over this period of time. It is clear though that there is an aggregation phenomenon which increases with increasing protein concentration. Indeed, the diameter of the sample at 8 mg/mL is almost twice bigger (11.8 nm) than the one at 2.4 mg/mL (6.5 nm, value calculated from HydroPro with x-ray structure with GTPase domain deleted).
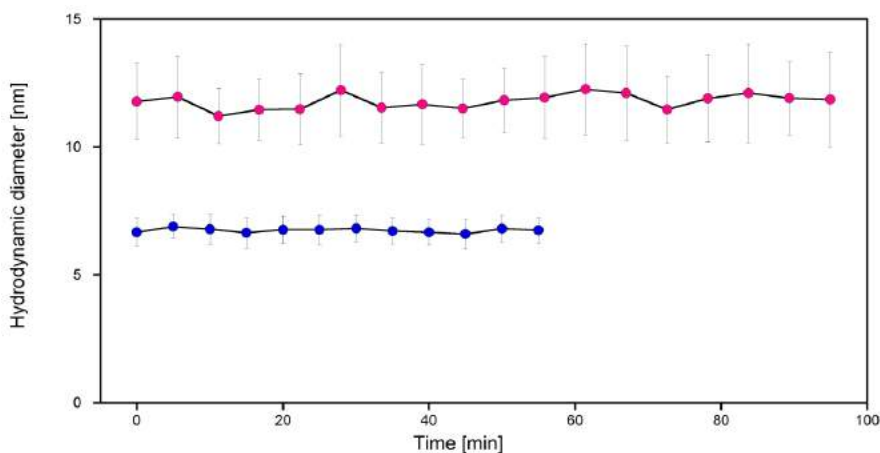
Figure 4.17 Hydrodynamic diameter evolution in time for ΔTmeHydF at 8 mg/mL (pink curve) and 2.4 mg/mL (blue curve).

In order to better understand the effect of protein concentration on molecules size, other measures were done with DLS. The value of the hydrodynamic diameter was measured for protein at different concentrations, starting from 8 mg/mL and diluting the sample 1:2 until 1mg/mL. It was observed from Figure 4.18 a that increasing the protein concentration increases the molecule's diameter. The protein forms aggregates when concentrated and this aggregation is only partially a reversible event, because the diameter decreases until the value of 8 nm at 1 mg/mL, while the expected dimensions are about 6 nm at 2.4 mg/mL (this concentration was obtained by dilution of a protein sample at 8 mg/mL). On the other hand, increasing the protein concentration decreases the diffusion coefficient (Figure 4.18 b). This is in accord to what is expected, because a more concentrated protein corresponds to bigger oligomers and the bigger the oligomer is, the lower is the diffusion coefficient. It is possible to conclude that there is a linear relation between the diffusion coefficient and the concentration[69], according to what is expected:

$$D_m = D_0(1 + k_D * C) = 42.037 * (1 - 0.0376 * C) \tag{4.1}$$

where D_m is the measured diffusion coefficient, D_0 is the diffusion coefficient at zero concentration, C is the sample concentration and k_D is defined as the difference between a thermodynamic component k_T and a hydrodynamic component k_H.
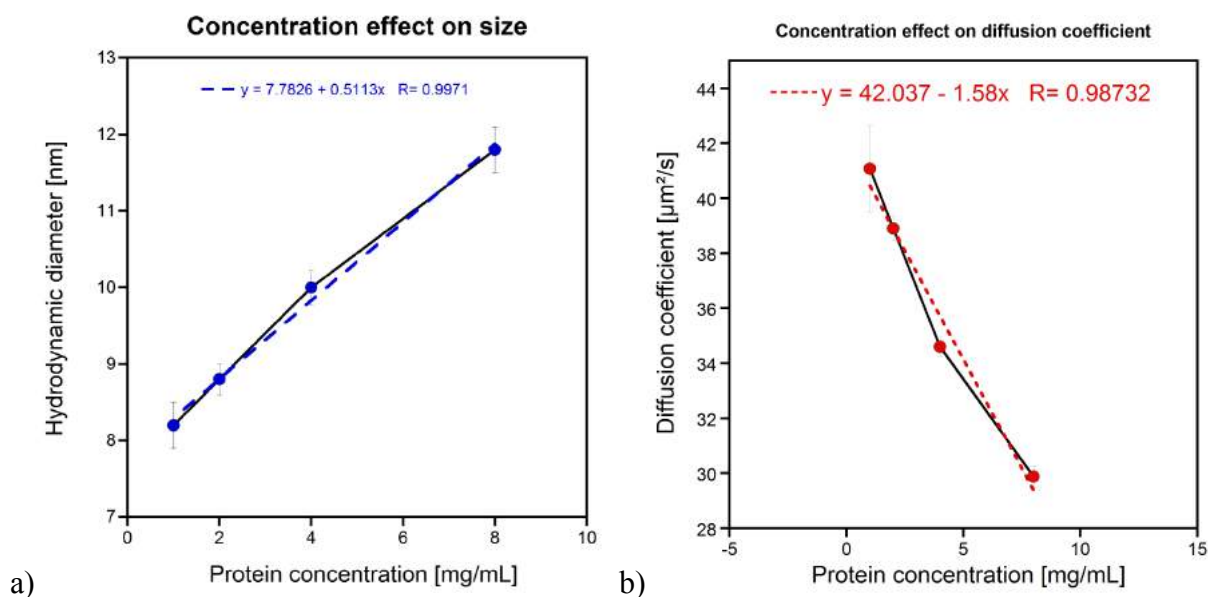
Figure 4.18 a) Evolution of hydrodynamic diameter with the protein concentration in buffer TRIS 50mM pH 8.0, NaCl 200mM, GlyOH 10%, DTT 5mM; b) Evolution of diffusion coefficient with the protein concentration in buffer TRIS 50mM pH 8.0, NaCl 200mM, GlyOH 10%, DTT 5mM.

The results obtained are due to the protein and solvent interaction. We want to find conditions that can limit the aggregations between the molecules, in order to have a light aggregation effect. Salt molecules, for example, can influence the behaviour of the protein in terms of aggregations. It is important to consider that the protein has a lot of charges, so adding ions from salt molecules can partially neutralize these charges and minimize molecules interactions (Figure 4.19).
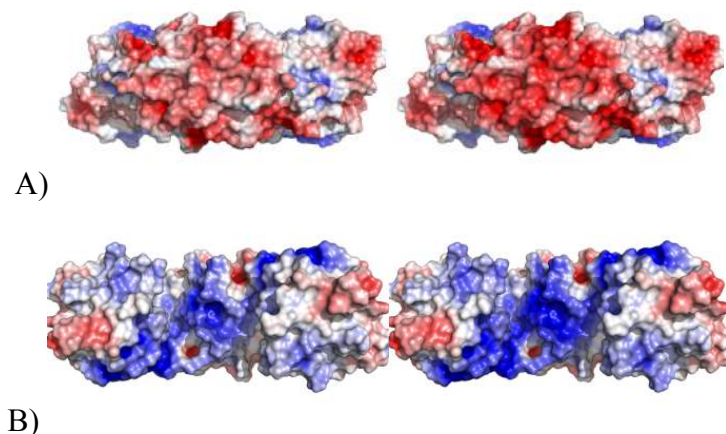
Figure 4.19 NaCl effect on protein's charges. A) Negative charged size of the protein with 500 mM NaCl (left figure) and 50 mM NaCl (right figure). B) Positive charged size of the protein with 200 mM NaCl (left figure) and 50 mM NaCl (right figure). Red represents negative charges, blue positive charges and white neutral atoms. The figure shows that adding salt reduces the surface charges (as the colour is less intense in the molecule with higher concentration of salt).

Some measures on DLS were done with the protein at 8 mg/mL in buffer TRIS 50 mM pH 8.0, GlyOH 10%, DTT 5 mM containing different concentrations of NaCl. Observing Figure 4.20, it appears that increasing the quantity of NaCl in the buffer reduces the size of the protein. So, for this protein, salt is necessary in order to stabilize the molecules and reduce intermolecular interactions, due to the bipolar properties of this protein.
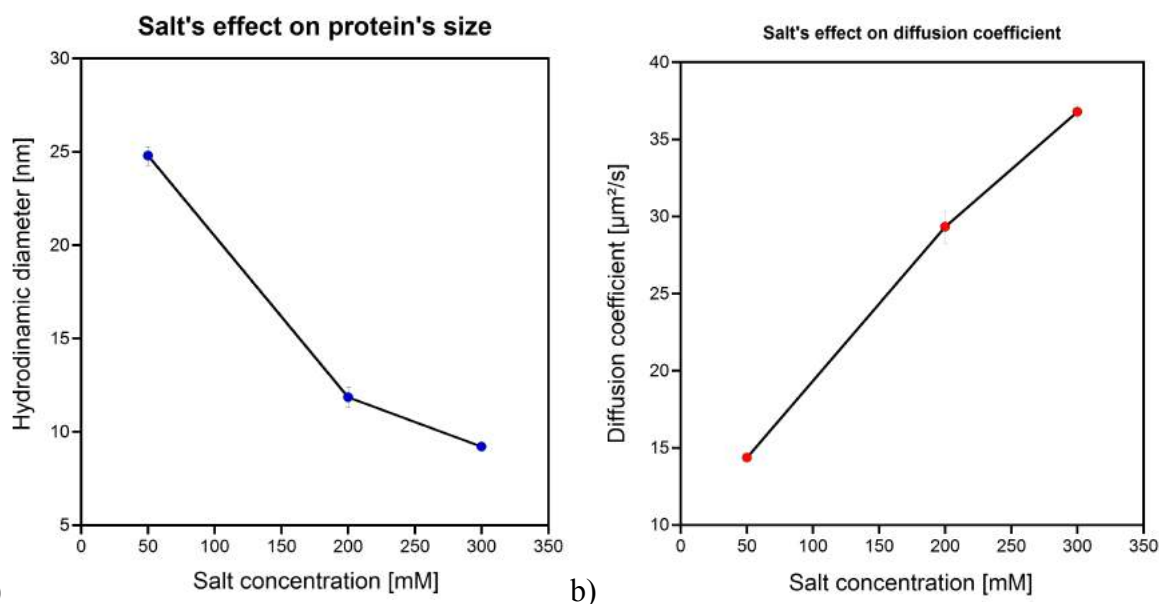
Figure 4.20 a) Evolution of hydrodynamic diameter with concentration of NaCl in buffer TRIS 50mM pH 8.0, GlyOH 10%, DTT 5mM; b) Evolution of diffusion coefficient with concentration of NaCl in buffer TRIS 50mM pH 8.0, GlyOH 10%, DTT 5mM.

Finally, to try to find a better buffer for crystallization conditions screening, we used DLS to study the effect of different buffers on the sample size.

From this experiment, it is clear that self-association is very dependent on the buffer, as revealed by the very different slopes of the linear regressions of the points. It is clear that buffers KP 25 mM pH 7.5 NaCl 150 mM DTT 5 mM and HEPES 25 mM pH 7.5 NaCl 200 mM GlyOH 30% DTT 5 mM strongly favour self-association. The "best" buffer, among the three analysed, is Tris 50 mM pH 8 with NaCl 200 mM, GlyOH 10% and DTT 5 mM. The slope of the plot $d(H)$=f(c) is similar to that observed for the sample of TmeHydF in the buffer that favours the formation of crystals. The results of the experiment realized with Hepes and 30% glycerol (in violet in Figure 4.21) are intriguing, given that the value at zero concentration should be close for all the experiments as it should be the true $d(H)$. It appears that the quality of the data for this experiment is lower as the sample had more contaminants than for the other experiments.
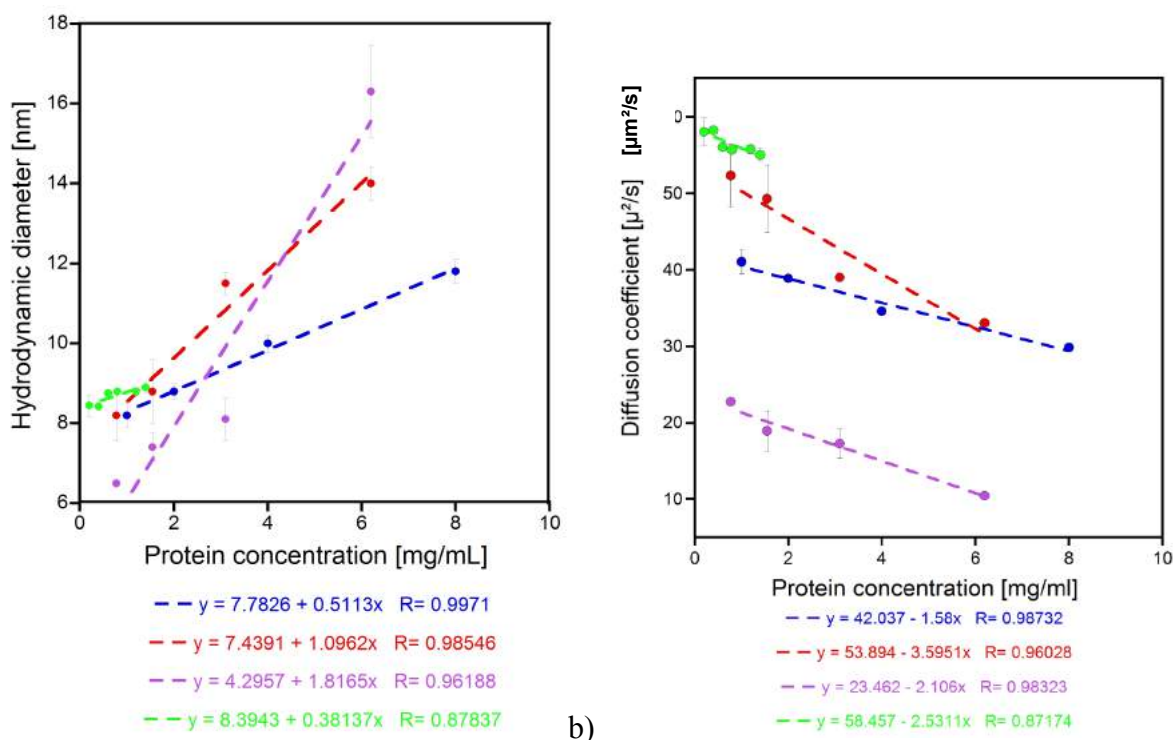
Figure 4.21 a) Evolution of hydrodynamic diameter with protein concentration and **b)** evolution of diffusion coefficient with protein concentration for ΔTmeHydF in buffer TRIS 50mM pH 8 NaCl 200mM GlyOH 10% DTT 5mM (in blue), KP 25mM pH 7.5 NaCl 150mM DTT 5mM (in red), HEPES 25mM pH 7.5 NaCl 200mM GlyOH 30% DTT 5mM (in violet) and for wild type TmeHydF in HEPES 25mM pH 7.5 NaCl 200mM DTT 5mM (in green, published by Caserta et al.[45]).

4.1.8   Crystallography

One of the long-term aims of the project is trying to get crystals of the truncated form of HydF from *Thermosipho melanesiensis* with the diiron subcluster.

We started testing some commercial screens in aerobic conditions that allow an easier and faster work. In this case, we observed that the protein easy precipitated in many conditions, but we could not manage to identify trends or effects of different factors.

We also tried to observe if the protein was able to crystallize in the same conditions as the full-length form does (0.1 M HEPES pH 7.5 and 4% w/v PEG 8000). Thus, we studied the behaviour of the protein in HEPES at pH 7.0 and at different concentrations of PEG 8000. After 4 weeks we obtained that many drops are still clear and only a little part precipitated, but we could not obtain crystals.

Other attempts were done in anaerobic conditions with different commercial kits at concentration between 7.5 and 16 mg/mL, but we observed a reversible precipitation of the protein at 16 mg/mL that was slightly whitish, indicating that the protein was too concentrated in these conditions and had a big tendency to aggregate. Despite this remark, we obtained promising results in three different conditions of the commercial kit MbClass II:

- Crystals with the shape of small stacked plates were obtained in 0.1 M Li_2SO_4, 0.1 M tri-sodium citrate pH 5.6, 12% PEG 4000 (Figure 4.22 a);

- spherulites were obtained at 7.5 and 16 mg/mL in 0.1 M NaCl, 0.1 M tri-sodium citrate pH 5.6, 30% PEG 400 (Figure 4.22 b);

- spherulites were obtained only at 16 mg/mL in 0.1 M Li_2SO_4, 0.1 M tri-sodium citrate pH 5.6, 30% PEG 400.

Crystals and spherulites appeared in 3 weeks. Crystals obtained in the condition PEG 4000 look transparent and are not brownish. This suggests that the [4Fe-4S]-cluster was lost during crystallisation. This might be due to the cluster degradation over time but one must consider that these conditions are at pH 5.6 and contain citrate that can chelate the iron of the cluster which is solvent accessible.
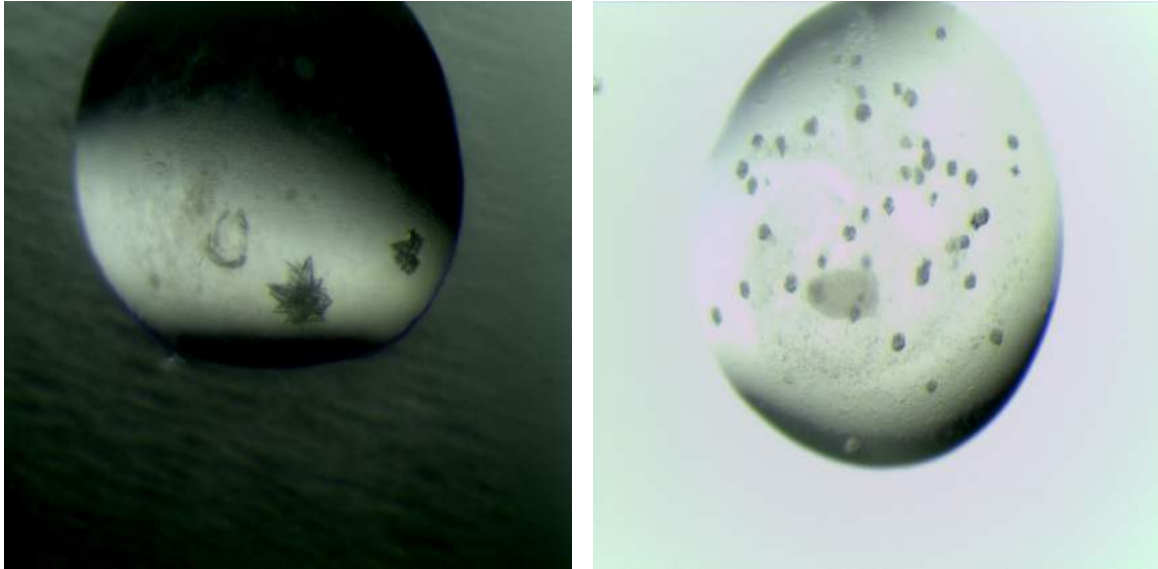

Figure 4.22 a) crystals obtained from [4Fe-4S]-ΔTmeHydF at 7.5 mg/mL in 0.1 M Li$_2$SO$_4$, 0.1 M tri-sodium citrate pH 5.6, 12% PEG 4000; **b)** spherulites obtained from [4Fe-4S]-ΔTmeHydF at 7.5 mg/mL in 0.1 M NaCl, 0.1 M tri-sodium citrate pH 5.6, 30% PEG 400. Images at optical microscope with 153x.

4.2 A truncated form of HydA from Megasphera elsdenii

The second part of the project was focused on the study of a truncated form of HydA from *Megasphaera elsdenii*. In the laboratory, studies were already been done on this protein and some protocols had already been established. During my internship, I studied some modifications in order to increase efficiency and speed of the protein's production.

4.2.1   Expression and aerobic purification with pt7-7-MeH-HydA plasmid

The pt7-7-MeH-HydA plasmid was used for the heterologous expression of the protein in *E. coli*, using TunerDE3 cells and following the protocol already developed. The protein expression was induced with IPTG when the OD at 600 nm was around 0.5. The cell growth followed an exponential trend and it was very slow, compared to other proteins growth's protocols, as shown in Figure 4.23. We observe a quite long lag phase at the beginning, when bacteria have to settle in the new environment.

This strategy resulted in the production of a soluble and well expressed recombinant protein, as shown in Figure 4.24.

The protein was then purified aerobically with two chromatographic steps: an affinity chromatography (HiTrap chelating), that can separate the protein of interest from all the proteins without the six-histidine tag, and a size-exclusion chromatography (Superdex S200 26/60), that can separate proteins based on their mass. The elution profiles of the two columns are reported in Figure 4.25. After these two steps a pure and homogeneous protein was obtained, as shown in the SDS-PAGE gel in Figure 4.25 c, with a final yield of 2.1 mg of proteins per gram of wet cells. All the purification steps were done in aerobiosis, so finally the protein was in its apo-form with a little content of Fe.
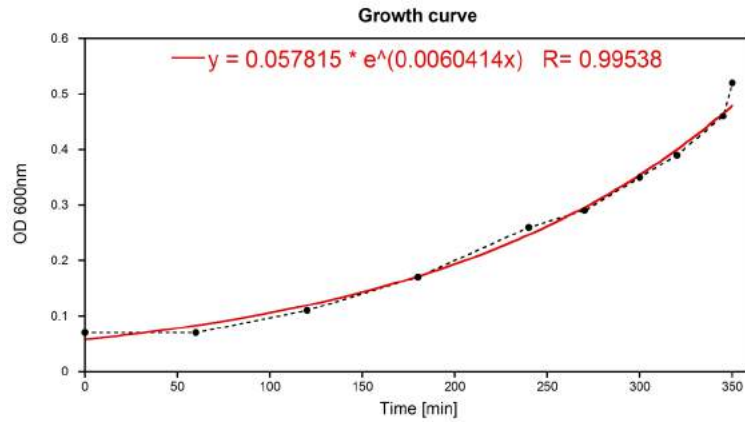
Figure 4.23 Growth curve of TunerDE3 cells transformed with pt7-7-MeH-HydA plasmid.
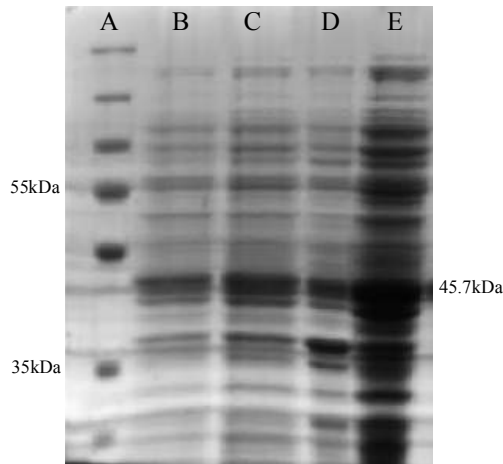

Figure 4.24 SDS-PAGE gel of (A) molecular weight ladder, (B) before IPTG induction, (C) cells 5h after IPTG induction, (D) insoluble fraction after cells lysis, (E) soluble fraction after cells lysis.
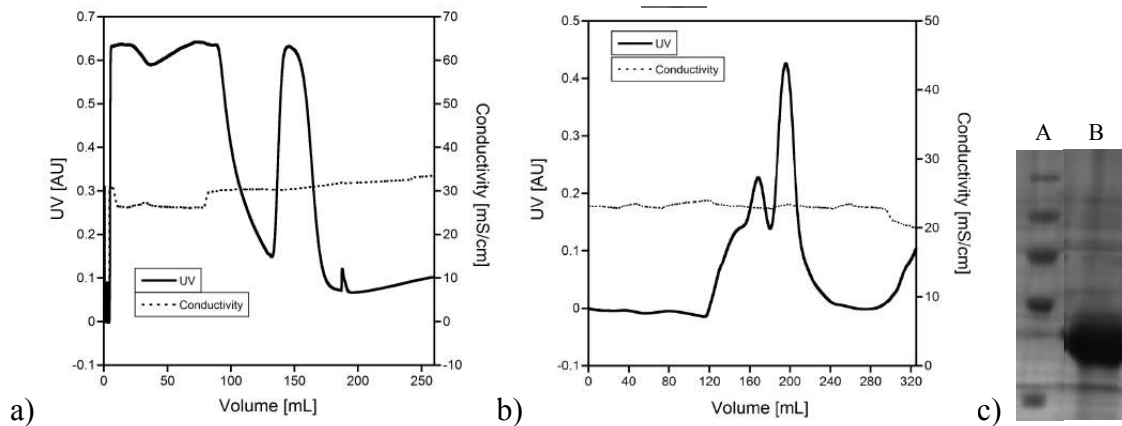

Figure 4.25 a) Elution profile of HiTrap column: the first peak is due to the proteins which do not contain the 6 histidine tag (se they are not able to bind the column), while the protein of interest is identified by the second peak, corresponding to an elution volume ~150 mL. b) Elution profile of Superdex column. The protein of interest corresponds to the second peak, at ~200 mL. c) SDS-PAGE gel of (A) molecular weight ladder and (B) MeH HydA after purification.

4.2.2 Expression and aerobic purification with pET22b plasmid

As showed before, the growth of the cells with pt7-7-MeH-HydA plasmid was really slow and that could be due to two reasons: 1) there is a mutation in the plasmid leading to an unidentified issue; 2) the protein is toxic for the cells.

First of all, we tested a new plasmid, the pET22b one. MeH HydA gene was inserted in this plasmid. Competent TunerBL21 DE3 cells were transformed with this plasmid. The same protocol used for the pt7-7-MeH-HydA was followed. As shown in Figure 4.26, the growth of the cells was much faster and the time needed to reach an OD at 600 nm of 0.5 was 6 times shorter.
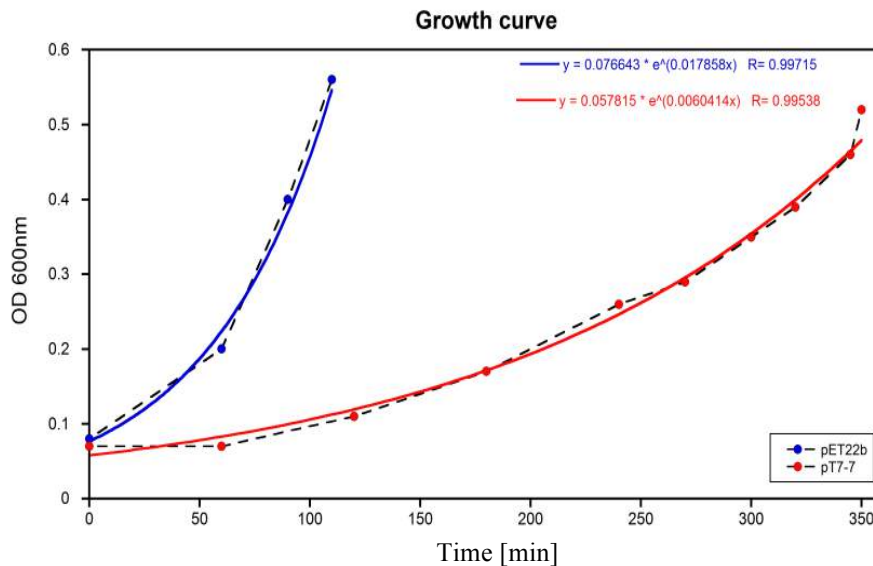
Figure 4.26 Growth curve of TunerDE3 cells transformed with pt7-7-MeH-HydA plasmid in red and TunerBL21 DE3 cells transformed with pET22b plasmid in blue.

It was then necessary to verify the expression and the solubility of the protein obtained. As shown on the SDS-PAGE gels in Figure 4.27, the protein was soluble and had a good level of expression after 5 hours induction with IPTG.

a) b)
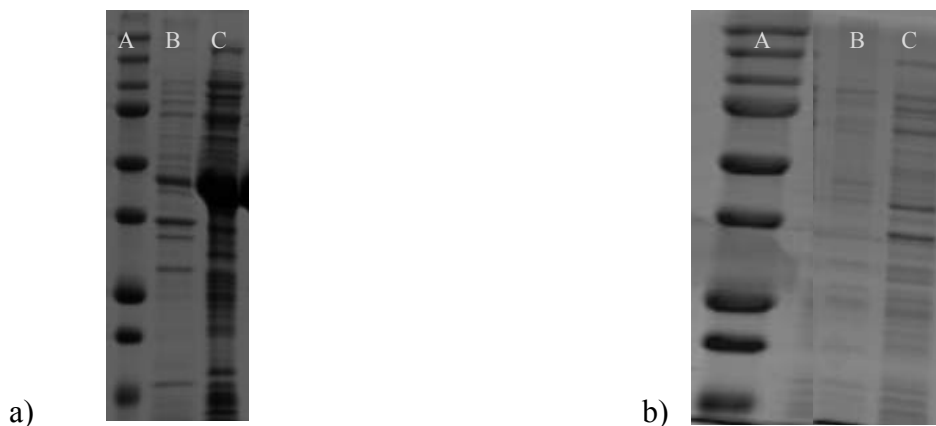
Figure 4.27 a) SDS-PGA gel showing MeH HydA expression. (A) molecular weight ladder, (B) MeH HydA before induction, (C) MeH HydA 5h after induction. **b)** SDS-PAGE gel showing the solubility of MeH HydA. (A) molecular weight ladder, (B) MeH HydA in the pellet after cells' lysis, (C) MeH HydA in the surnatants after cells' lysis.

For the purification of the protein, the protocol was slightly changed compared to the pT7-7 MeH HydA. The affinity chromatography step was done with HisTrap FF crude which normally has higher binding capacity and higher resistance to reducing agents than the HiTrap column. For the size-

48

exclusion chromatography the same column was used (Superdex S200 26/60). The elution profiles of the two columns are reported in Figure 4.28. After the two purification steps, it was possible to obtain a very pure protein (Figure 4.28 c) with a final yield of 6.6 mg of protein per gram of wet cells. So, the plasmid change allowed the achievement of two important improvements: the reduction of growth time (changing from >11h to <7h) and the triplication of the yield. It indicates that the protein is not toxic for the cells and that there was a problem with the pt7-7-MeH-HydA plasmid. As the growth time was a major consideration, one could propose to do an overnight expression at low temperature, but the laboratory has shown that an overnight expression protocol generates insoluble MeH HydA.

a)   b)   c)
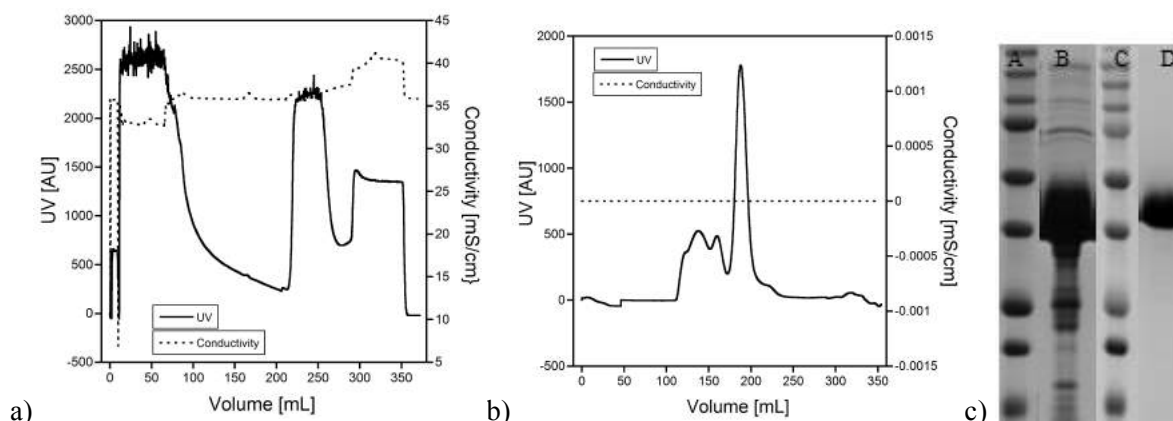
Figure 4.28 Elution profiles of (a) HisTrap column and (b) Superdex S200 26/60. c) SDS-PAGE gel of (A) and (C) molecular weight ladder, (B) MeH HydA after purification with affinity chromatography and (D) MeH HydA after purification with size-exclusion chromatography.

4.2.3   [4Fe-4S] cluster reconstitution of apo form of MeH HydA

In order to obtain the active form of the protein, it was first necessary to reconstitute the [4Fe-4S]-cluster. This incorporation was carried out using a standard protocol. The protein was incubated in anaerobic conditions (<0.5 ppm of oxygen) at 100μM with a 6 molar excess of iron ammonium sulphate and L-cysteine, in presence of DTT as a reducing agent in order to have all the cysteines in reduced form and the cysteine desulfurase CsdA. The reaction was followed overnight by UV-Visible absorption spectroscopy (see Figure 4.29).

Non-specific iron and protein aggregates were removed by size exclusion chromatography. The elution profile is shown in Figure 4.30 a. After the purification, the presence of the [4Fe-4S]-cluster was verified qualitatively with an UV-Visible spectrum and quantitatively with iron quantification. From the UV-Vis absorption spectrum (shown in Figure 4.30 b) the value of the absorbance ratio 400/280 nm was calculated and resulted to be 0.23, which is in the range between 0.21 and 0.24, as reported by previous experiments with the same protocol. The iron quantification gave a value of the nmol of iron to nmol of protein ratio around 3.6, which is not very far from the theoretical value of 4 and is a regular value obtained in the laboratory. Thus, the reconstitution can be considered complete.

In other reconstitutions we obtained an absorbance ratio 400/280 nm around 0.18, even if the iron quantification confirmed the insertion of the [4Fe-4S]-cluster into the protein. A contamination of the glove box by methyl viologen can explain this behaviour, as this complex can reduce the $[4Fe-4S]^{2-}$ to $[4Fe-4S]^{+}$ which has a significantly lower absorption coefficient at these wavelength.
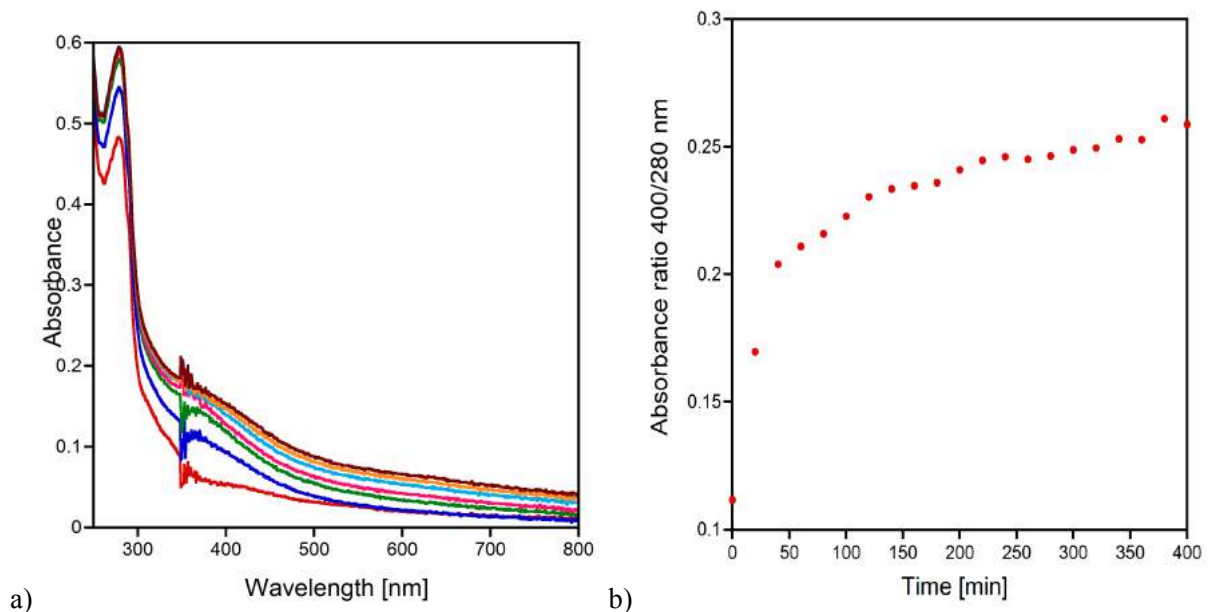
Figure 4.29 (a) UV-visible absorption spectra of MeH HydA recorded every 20 minutes, showing the [4Fe4S] cluster reconstitution' kinetics. **(b)** Absorbance ratio 400/280 nm during time.
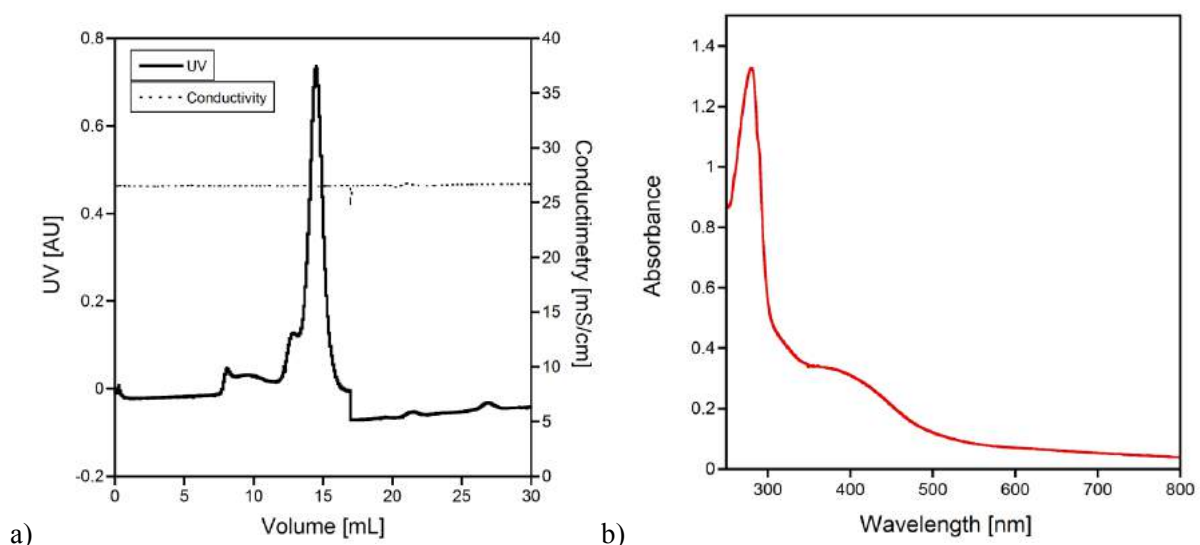

Figure 4.30 (a) elution profile of Superdex S200 10/300 that shows the presence of some oligomers. (b) UV-Vis spectrum of [4Fe-4S]-MeH HydA.

As previously observed for HydF, an important issue that has to be solved is the low yield which is less than 50%. The loss is mainly during the concentration steps, probably because the protein adsorbs on the concentrator membrane.

4.2.4   Direct activation of MeH HydA with adt-complex

The protein with the [4Fe-4S]-cluster is not completely active. It is necessary to reconstitute the whole active site, recreating the diiron subcluster. As already done before, incubating the protein with a 10 fold excess of the $[Fe_2(adt)(CO)_4(CN)_2]^{2-}$ complex for 30 minutes in strictly anaerobic conditions, allows the binuclear subcluster to be incorporated into the enzyme which is activated. Measuring the activity of the [2Fe]-MeH HydA with Gas Chromatography appears that the enzyme was activated after 30 minutes with a specific activity of 133+/-28 µmol H_2/min mg of enzyme.

4.2.5 Dynamic Light Scattering

The quality of the protein was verified through measure with Dynamic Light Scattering which can measure the diffusion speed of particles and derive a particle size from the Stokes-Einstein equation. The sample appeared to be stable and contaminated with an aggregate at ~100 nm. The species with average hydrodynamic diameter of 7.4 nm is monodisperse with %Pd of 18.9%. As indicated by the single peak in size distribution in %mass, the aggregate at 100 nm represents less than 0.1% of the sample (see Figure 4.31). This indicates that the sample is of good quality allowing the setup of crystallization trays.
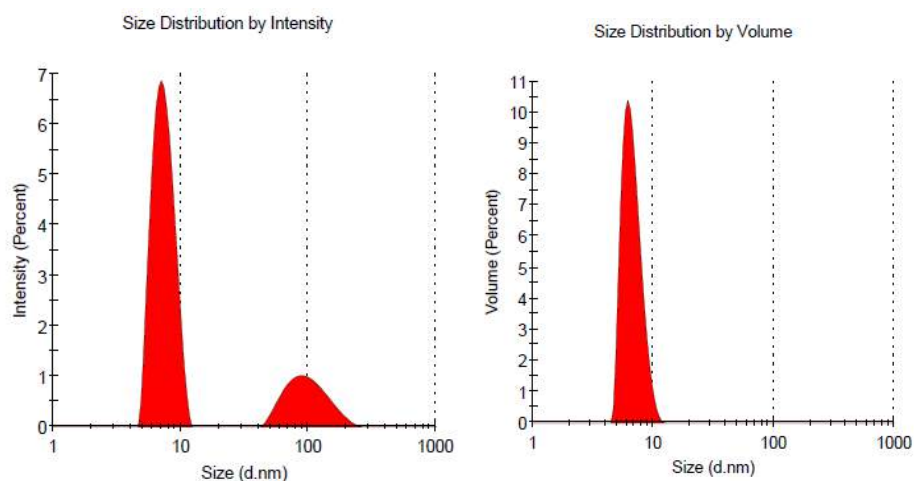
Figure 4.31 Size distribution of a solution of MeH HydA at XX mg/mL in TRIS 50 mM pH 8.0, NaCl 150 mM and DTT 1 mM particles, based on diffusion intensity (on the left) and based on %volume, also called %mass, (on the right).

4.2.6 Crystallography

As already explained in the Introduction section, studying the structure of HydA from *Megasphaera elsdenii* can be very useful to understand the properties of this protein. Crystallography can help to study the structure, because crystals can be analysed with X-rays, providing a diffraction pattern which can be studied to discern the protein tertiary structure.

We tried to crystallize the protein in its apo-form in aerobic conditions and in its [Fe-S]-reconstituted form in anaerobic conditions, at different concentrations ranging from 5 to 20 mg/mL. Different commercial kits were tried to screen different conditions and to find the most favourable ones. All the details about conditions tried and concentrations are reported in the Methods and Materials section.

We could not manage to get crystals, but we obtained some potentially interesting conditions that can be deeper investigated in future. For example, we obtained spherulites in 0.1 M Tris pH 8, 60% v/v Polypropylene glycol 400 (Figure 4.32 a) and phase separation with a phase containing a lot of protein in 0.2 M Lithium chloride, 0.1 M MES pH 6, 20% w/v PEG 6000 (Figure 4.32 b).
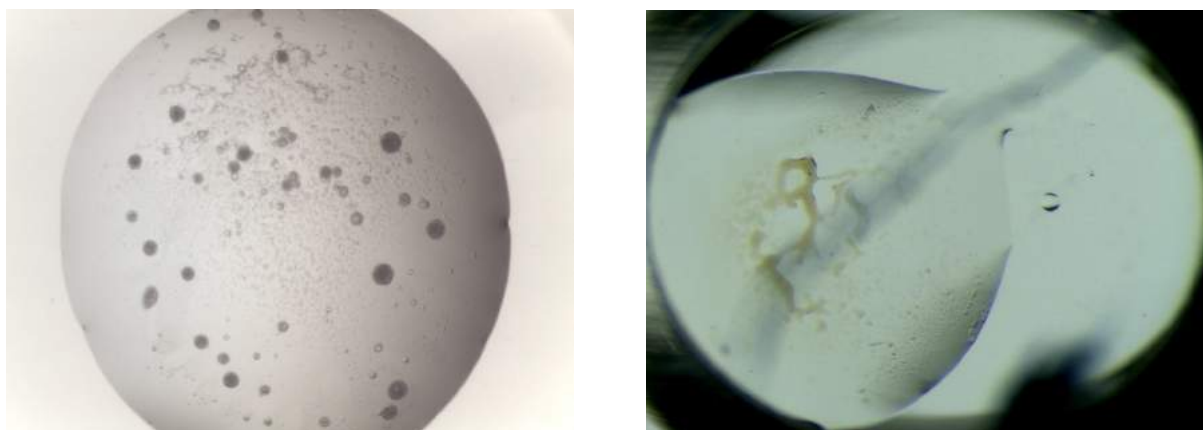
Figure 4.32 a) spherulites obtained from apo-MeH HydA at 13 mg/m in 0.1 M Tris pH 8, 60% v/v Polypropilene glycol 400; **b)** phase separation obtained from [Fe-S]-MeH HydA at 12 mg/mL in 0.2 M Lithium chloride, 0.1 M MES pH 6, 20% w/v PEG 6000. Images at optical microscope with 153x.

From all the conditions we tested, it is possible to define trends with pH, cations and anions present in the drop. The protein, in fact, precipitates at pH under 6.0 but adding ammonium chloride or nitrate can avoid precipitation even at low pH. Anions and cations can induce precipitation or stabilise the protein, depending also on their concentration: too high concentrations lead to precipitation.

PACT screen, as a systematic screen that tests different parameters, was very useful to understand how the protein behaves in different conditions. In fact, it tests effects of pH, cations and anions. We observed that the protein precipitated at pH under 6.0 and that Zn induced precipitation, favouring intermolecular contacts.

Basing on these observations, we focused our attention on two particular ions: Zn^{2+} and Cd^{2+}. The former was analysed because from PACT and other commercial kits we noticed that Zn^{2+} could induce the precipitation of the protein and we wanted to understand if less concentrations had other effects. Cd^{2+} was tried because it is known to be a good carboxylate ligand[70] and we know that the protein contain a lot of carboxylates on its surface, as the low pI confirms. An example where crystals were obtained exclusively in presence of cadmium is shown in Figure 4.33.
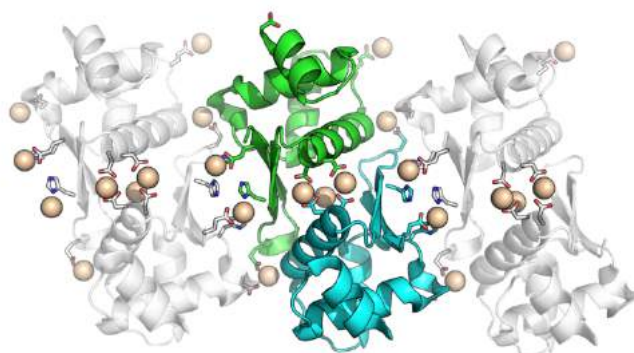

Figure 4.33 Example of a successful crystallization only in presence of Cd^{2+} **(**pdb 2fu24): the E. coli ferric uptake regulator N-terminal domain that could be crystallized.

Both with Zn^{2+} and with Cd^{2+}, we observed that the ions led to the precipitation of MeH HydA even at low concentrations.

4.3 HydA from Megasphaera elsdenii

The last part of the project deals with the full-length form of HydA from *Megasphera elsdenii*. This protein was expressed, purified aerobically and reconstituted anaerobically. Then it was used to test

the skill of ΔTmeHydF to transfer the binuclear subcluster to apo-hydrogenase, in the experiment already discussed above.

4.3.1   Expression and aerobic purification

The pT7-7-6HMeHydA plasmid was used for the heterologous expression of the protein in *E. coli*, using TunerDE3pLysS cells and following the protocol already developed. The protein expression was induced with IPTG when the OD at 600 nm was around 0.5. The cell growth followed an exponential trend (as reported in Figure 4.34) and OD of 0.6 was reached after 3 hours.
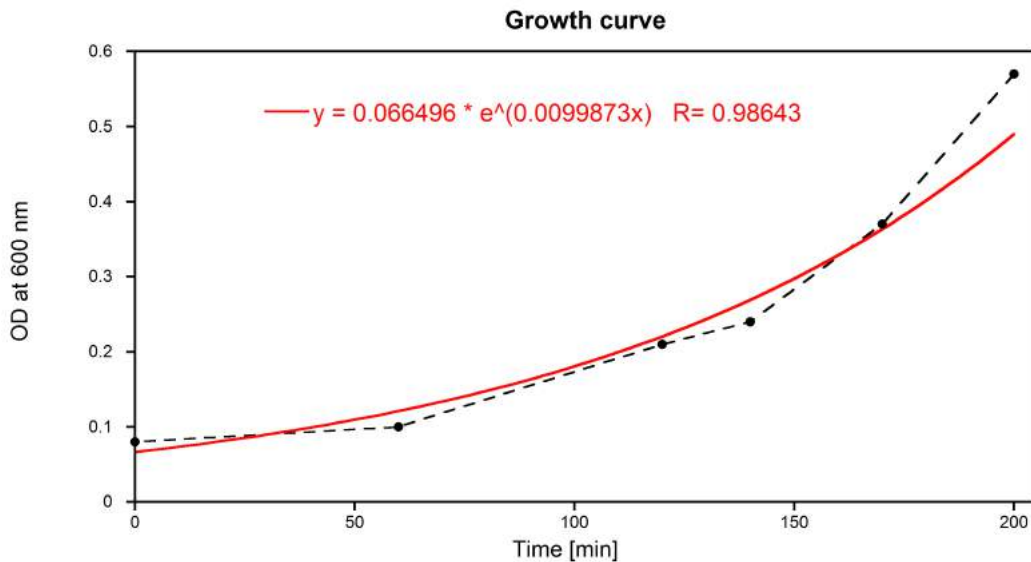
Figure 4.34 Growth curve of TunerDE3pLysS cells transformed with pt7-7-6MeHydA plasmid.

This strategy resulted in the production of a soluble and well expressed recombinant protein, as shown in Figure 4.35.

The protein was then purified aerobically with two chromatographic steps: an affinity chromatography (HisTrap column) and a size-exclusion chromatography (Superdex S200 26/600). After these two steps a quite pure and homogeneous protein was obtained, as shown in the SDS-PAGE gel in Figure 4.36, with a final yield of 2.1 mg of proteins per gram of wet cells. All the purification steps were done in aerobiosis, so finally the protein was in its apo-form with a little content of Fe.
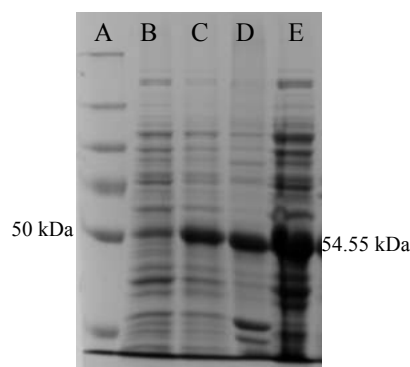

Figure 4.35 SDS-PAGE gel of (A) molecular weight ladder, (B) before IPTG induction, (C) cells 5h after IPTG induction, (D) insoluble fraction after cells lysis, (E) soluble fraction after cells lysis.
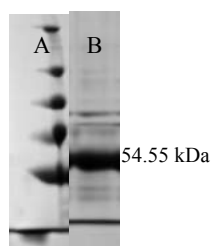
Figura 4.36 SDS_PAGE gel of (A) molecular weight ladder and (B) MeH HydA after purification

4.3.2 [4Fe-4S] cluster reconstitution of apo form of MeH HydA

In order to obtain the active form of the protein, first of all it was necessary to reconstitute the [4Fe-4S]-cluster. This incorporation was carried out using a standard protocol. The protein was incubated in anaerobic conditions (<0.5 ppm of oxygen) with a 17 molar excess of iron ammonium sulphate and L-cysteine, in presence of DTT and the cysteine desulfurase CsdA. The protein was then purified with a size-exclusion chromatography to eliminate all the non-specific iron and the possible aggregates that were formed during the reaction. The elution profile is shown in Figure 4.37 a. After the purification, the presence of the [4Fe-4S]-cluster was verified qualitatively with an UV-Visible spectrum (Figure 4.37 b).
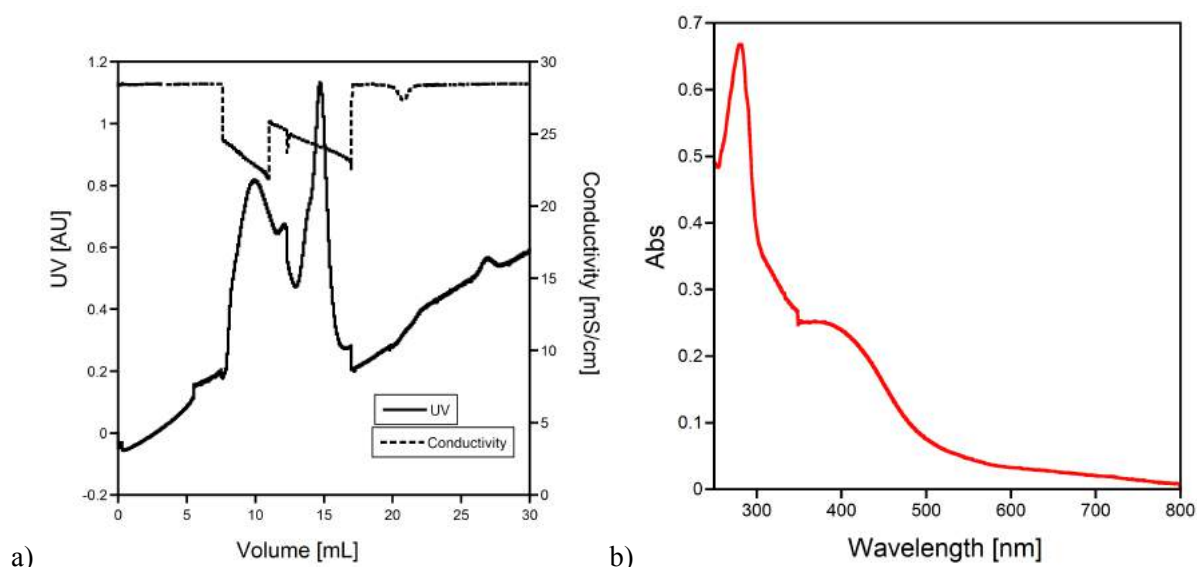

a)   b)

Figure 4.36 (a) elution profile of Superdex S200 10/300 (b) UV-Vis spectrum of [4Fe-4S]-MeHydA.

The protein was then incubated with 10 fold excess of $(Et_4N)_2[Fe_2(adt)(CO)_4(CN)_2]$, called in the following adt-complex, in order to reconstitute the [2Fe]-subcluster. After one hour of incubation, the incorporation of the diatomic complex was verified through an activity test with gas chromatography that resulted in a specific activity of the protein of 659 µmol H_2/min mg protein. The same test was repeated after one night and the measured activity was around 580 µmol H_2/min mg protein.

4.3.3 Activity test with oxygen

To test protein resistance to oxygen, it is necessary to study how the active protein behaves in presence of oxygen. This kind of experiments is being carried on in the laboratory and, during my internship, I participated to one of these tests.

The sample containing the active protein was exposed to a controlled quantity of oxygen and its behaviour was followed by a UV-Vis spectrum and by monitoring the specific activity in time. With a quantity of oxygen of 10000 ppm, the protein lost almost the half of its activity after nearby 3 hours.

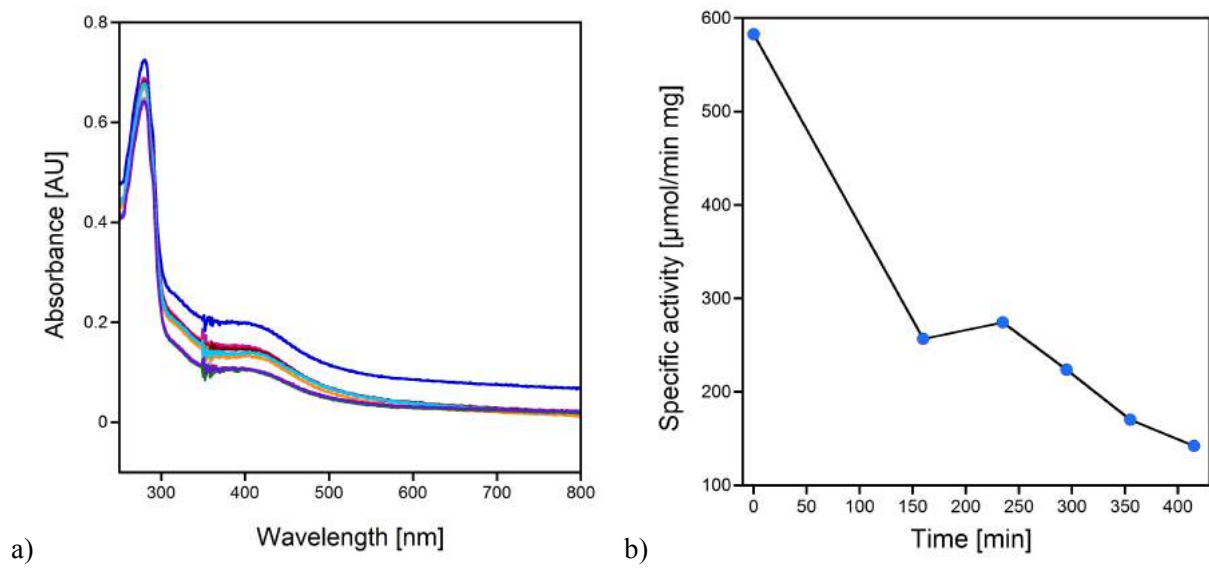The UV-Vis spectrum and the results of activity tests with gas chromatography are reported in Figure 4.38.

Figure 4.37 a) UV-Vis spectrum showing the degradation of the protein in time in presence of 10000ppm of O_2. b) Lost of specific activity during time.

5. Conclusions

During my experience at the *Laboratoire de Chimie des Processus Biologiques*, I had the opportunity to work in a very interesting and priority project and to improve my knowledge in many fields, such as protein biology, biochemistry and molecular biology.

From this work and from the analysis carried out, it can be claimed that the laboratory managed to obtain a stable, water-soluble and pure form of a truncated HydF from *Thermosipho melanesiensis*. We used molecular biology to generate the truncated HydF gene, we expressed the protein in *E. coli* and we developed a protocol for its purification. The main aim with this protein is to solve its crystal structure, containing both the diiron sub-cluster and the [4Fe-4S]-cluster. Unfortunately, we did not manage to obtain crystals containing clusters so far, but we have some crystallization hits that can be further investigated. One of the main problems of crystallisation attempts that we dealt with was the fact that the protein in its buffer was not stable at high concentration and showed a big tendency for aggregation. We already started to study how the protein stability is influenced by many factors, such as buffer, pH and salt concentration, using Dynamic Light Scattering. We saw that the presence of salt can help stabilising the protein, neutralising the surface charges and reducing intermolecular interactions. Focusing on this kind of studies could allow finding the conditions in which the protein is more stable and this could maybe help to find the best crystallization conditions. Since this protein crystallizes in acidic condition and more specifically in presence of citrate, which might chelate the iron of the [4Fe-4S]-cluster, another alternative could be to design other mutants of the full-length protein, in order to obtain a more easily crystallizing protein.

Furthermore, in this work new evidences were provided about the fact that the GTPase domain is not necessary for transferring the [2Fe]-subcluster to HydA. Indeed, we demonstrated that ΔTmeHydF is able to interact and activate both MeH HydA and MeHydA.

It was also shown that this protein presents a higher activity compared to the full-length form. This was unexpected, because the difference between the truncated form and the wild type resides only in the GTPase domain, but there were not changes in terms of active site. Thus, these experiments have to be repeated.

Another remarkable result obtained during my internship was the development of a new protocol for MeH HydA production. In the laboratory, a protocol for MeH HydA expression and purification was already established, but it required a lot of time and yield was not very high. Changing the plasmid used for cells transformation and the first chromatographic column, we managed to half the time of cells growth and to the triplicate the yield.

A lot of work still remains to be done concerning the crystallization of this protein. As already said before, no structures of this protein are available to date but obtaining a crystal structure could be very useful in order to understand the behaviour of this protein and results that are still unexplained so far. We focused on the truncated form of the protein, supposing it was easier to manipulate and to crystallize. We did not manage to obtain crystals but we obtained some interesting starting conditions for optimisation, which will be further investigated.

A complete knowledge of the structures of these proteins, with or without clusters, can help to understand the interaction between proteins and the effect of oxygen on protein deactivation, which are two of the main problems that still remain to be solved.

In conclusion, there is no doubt in saying that [Fe-Fe]-hydrogenases have an incredible activity towards proton reduction and hydrogen oxidation, but drawbacks still have to be faced. The main issue with these hydrogenases is their high sensitivity to oxygen, which degrade the active site in a still incompletely known way. This fact complicates manipulations because it is necessary to work in strictly anaerobic conditions, using a glove box. Moreover, they are difficult to produce in an active form, even if the synthetic maturation partially solved this issue, avoiding the necessity to produce the three maturases.

I would like to conclude saying that though many groups are working in this competitive field, there is still a long way to engineer a hydrogenase with high activity, low sensitivity to oxygen and with possible application in the industry.

6. Abbreviations

6-his tag	Six-histidine tag
Abs	Absorbance [-] or [mAU]
adt	Azadithiolate
Ala	Alanine
c	Concentration [mg/mL]
C. reinhardtii	*Chlamydomonas reinhardtii*
CpI	HydA from *Clostridium pasteurianum*
CpI^{ADT}	HydA from *Clostridium pasteurianum* maturated with adt-complex
CrHydA1	HydA from *Chlamydomonas reinhardtii*
CsdA	Cysteine desulphurase
Cys	Cysteine
D	Diffusion coefficient [$\mu m^2/s$]
d	Diameter [nm]
d(H)	Hydrodynamic diameter [nm]
DdHydA	HydA from *Desulfovibrio desulfuricans*
DLS	Dynamic light scattering
DNA	Deoxyrinonucleic acid
dNTP	Deoxynucleoside triphosphate
DTT	Dithiothreitol
E. coli	Escherichia coli
ENDOR	Electron-nuclear double resonance
EPR	Electron paramagnetic resonance
ESR	Electron spin resonance
FTIR	Fourier-transform infrared spectroscopy
GC	Gas Chromatography
GDP	Guanosine diphosphate
Glu	Glutamate
GTP	Guanosine-5'-triphosphate
His	Histidine
HYSCORE	Hyperfine sublevel correlation spectroscopy
I	Intensity of light scattered
IPTG	Isopropyl β-D-1-thiogalactopyranoside
k	Boltzmann's constant $=1{,}38 \times 10^{-23}$ J/K
l	Path length [cm]

LB	Lysogenic Broth
M. elsdenii	*Megasphaera elsdenii*
MeH HydA	Truncated form of HydA from *Megasphaera elsdenii*
MeHydA	HydA from *Megasphaera elsdenii*
n	Refractive index [-]
OD_{600}	Optical density at 600 nm [-]
odt	oxa-propandithiol
PCR	Polymerase Chain Reaction
PSI	Photosystem I
PSII	Photosystem II
pdt	propanedithiol
R	Distance to the particle [m]
SAM	S-adenosyl-L-methonine
SDS-PAGE	*Sodium Dodecyl Sulphate - PolyAcrylamide Gel Electrophoresis*
T	Temperature [K]
T. maritima	*Thermotoga maritima*
TB	Terrific Broth
TmeHydF	HydF from *Thermosipho melanesiensis*
TmHydF	HydF from *Thermotoga maritima*
TnHydF	HydF from *Thermotoga neapolitana*
TOF	Turnover frequency [min^{-1}]
TON	Turnover number
UV	Ultraviolet
Vis	Visible
ΔTmeHydF	Truncated form of HydF from *Thermosipho melanesiensis*
ε	Molar absorption coefficient [$M^{-1} cm^{-1}$]
$ε_{280nm}$	Molar extinction coefficient at 280 nm [$M^{-1} cm^{-1}$]
η	Viscosity [Pa s]
$θ$	Scattering angle [rad]
λ	Wavelength [nm]
μM	Micromolarity

7. References

[1] Ruggeri B., Tommasi T., Sanfilippo S., 2015, BioH2 & BioCH4 Through Anaerobic Digestion. *Green Energy and Technlogy, Springer-Verlag London.*

[2] Lojou E., Bianco P., 2004, Electrocatalytic Reactions at Hydrogenase-Modified Electrodes and Their Applications to Biosensors: From the Isolated Enzymes to the Whole Cells. *Electroanalysis 16 (13-14).*

[3] Lubitz W., Ogata H., Rüdiger O., Reijerse E., 2014, Hydrogenases. *Chemical reviews* 114 (8), 4081-4148.

[4] Mertens R. and Liese A., 2004, Biotechnological applications of hydrogenases. *Current opinion in Biotechnology 15 (4), 343-348.*

[5] Quian D. J., Nakamura C., Wenk S. O., Ishikawa H., Zorin N., Miyake J., 2002, A hydrogen biosensor made of clay, polybutylviologen and hydrogenase sandwiched on a glass carbon electrode. Biosensors and Bioelectronics 17 (9), 789-796.

[6] Dey S., Rana A., Crouthers D., Mondal B., Kumar Das P., Darensbourg M. Y., Dey A., 2014, Electrocatalytic O_2 Reduction by Fe-Fe hydrogenase active site models. Journal of the American Chemical Society 136 (25), 8847-8850.

[7] Wakerley D. W. and Reisner E., 2015, Oxygen-tolerant proton reduction catalysis: much O_2 about nothing?. *Energy environmental science* 8, 2283-2295.

[8] Nikolaidis P., Poullikkas A., 2017, A comparative overview of hydrogen production processes. *Renewable and Sustainable Energy Review* 67, 597-611.

[9] Dincer I., Acar C., 2015, Review and Evaluation of hydrogen productions methods for better sustainability. *International Journal of Hydrogen Energy* 40, 11094-11111.

[10] Alternative fuels data center. Available on: https://www.afdc.energy.gov/fuels/hydrogen_basics.html.

[11] Hallenbeck P.C., Benemann J.R., 2002, Biological hydrogen production; fundamentals and limiting processes. *International Journal of Hydrogen Energy* 27, 1185-1193.

[12] Fontecille-Campus J. C. and Nicolet Y., 2014, *Metalloproteins: methods and protocols*, Methods in molecular biology, vol. 1122, Springer Science and Business Media New York.

[13] Barton L. L., Goulhen F., Bruschi M., Woodards N. A., Plunkett R. M., Rietmeijer F. J. M., 2007, The bacterial metallome: composition and stability with specific reference to the anaerobic bacterium *Desulfovibrio desulfuricans. BioMetals*, 20, 291–302.

[14] Hoppert M., 2011, Metalloenzymes, in *Encyclopedia of Geobiology* (Reitner J., Thiel V.)*, part of the Encyclopedia of Earth Sciences Series*, 558-563.

[15] Stiban J., Minyoung S., Kaguni L., 2016, Iron-sulfur clusters in mitochondrial metabolism: Multifaceted roles of a simple cofactor. *Biochemistry* (Moscow), 81(10), 1066-1080.

[16] Rouault T. A., 2012, Biogenesis of iron-sulfur clusters in mammalian cells: new insights and relevance to human disease. *Disease Models & Mechanisms* 5(2), 155-164.

[17] Beinert H., Holm R.H., Münck E., 1997, Iron-sulfur clusters: nature's modular, multipurpose structures. *Science* (New York, N. Y.) 277(5326), 653-659.

[18] Bridwell-Rabb J., Fox N. G., Tsai C. L., Winn E., Barondeau D. P., 2014, Human frataxin activates Fe-S cluster biosynthesis by facilitating sulfur transfer chemistry. *Biochemistry 53, 4904-4913*.

[19] Kiley P. J., Beinert H., 2003, The role of Fe-S proteins in sensing and regulation in bacteria. *Current opinion in Microbiology* 6(2), 181-185.

[20] Jain R., Vanamee E. S., Dzikovski B. G., Buku A., Johnson R. E., Prakash L., Prakash S., Aggarwal A. K., 2014, An iron-sulfur cluster in the polymerase domain of yeast DNA Polymerase ε. *Journal of Molecular Biology* 426(2), 301-308.

[21] Lill R., 2009, Function and biogenesis of iron-sulphur proteins, *Nature* 460 (7257), 831-838.

[22] Madden C., Vaughn M. D., Diez-Pérez I., Brown K. A., King P. W., Gust D., Moore A. L., Moore T. A., 2012, Catalytic Turnover of [Fe-Fe]-hydrogenase Based on Single-Molecule Imaging. *Journal of American Chemical Society* 134(3), 1577-1582.

[23] Vignais P. M., Billoud B., Meyer J., 2001, Classification and phylogeny of hydrogenases. *FEMS Microbiology Reviews* 25, 455-501.

[24] Vignais P. M., Billoud B., 2007, Occurrence, Classification and Biological Function of Hydrogenases: An Overview. *Chemical Reviews* 107, 4206-4272.

[25] Posewitz M. C., King P. W., Smolinski S. L., Zhang L., Seibert M., Ghirardi M., 2004, Discovery of Two Novel Radical S-Adenosylmethionine Proteins Required for the Assembly of an Active [Fe] Hydrogenase. *Journal of Biological Chemistry* 279(24), 25711-25720.

61

[26] McGlynn S. E., Ruebush S. S., Naumov A., Nagy, L. E., Dubini A., King P. W., Broderick J. B:, Posewitz M. C., Peters J. W., 2007, In vitro activation of [Fe-Fe] hydrogenase: new insights into hydrogenase maturation. *Journal of Biological Chemistry* 12(4), 443-447.

[27] Kuchenreuther J. M., Stapleton J. A., Swartz J. R., Delprato A. M., 2009, Tyrosine, Cysteine and S-Adenosyl Methionine Stimulate In Vitro [FeFe] Hydrogenase Activation ([FeFe] Hydrogenase Maturation). *PLoS ONE* 4(10), e7565.

[28] Broderick J. B., Byer A. S., Duschene K. S., Duffus B. R., Betz J. N., Shepard E. M., Peters J. W., 2014, H-cluster assembly during maturation of the [Fe-Fe]-hydrogenase. *Biological Inorganic Chemistry* 19 (6), 747-757.

[29] Fontecave M., Ollagnier-de-Choudens S., Mulliez E., 2003, Biological radical sulfur insertion reactions. *Chemical Reviews* 103, 2149–2166.

[30] Frey P.A., Hegeman A.D., Ruzicka, F.J., 2008, The radical SAM superfamily. *Critical Reviews in Biochemistry and Molecular Biology* 43, 63–88.

[31] Driesener R.C., Duffus B.R., Shepard E. M., Bruzas I. R., Duschene K. S., Coleman N. J. R., Marrison, A. P. G., Salvadori E., Kay C. W. M., Peters J. W., Broderick J. B., Roach P. L., 2013, Biochemical and kinetic characterization of radical S -Adenosyl- L -Methionine enzyme HydG. *Biochemistry*, 52(48), 8696-8707.

[32] King P. W., Posewitz M. C., Ghirardi M. L., Seibert M., 2006, Functional studies of [Fe-Fe] Hydrogenase Maturation in an *Escherichia coli* Biosynthetic System. *Journal of Bacteriology* 188(6), 2163-2172.

[33] Pilet E., Nicolet Y., Mathevon C., Douki T., Fontecilla-Camps J. C., Fontecave M., 2009, The role of the maturase HydG in [FeFe]-hydrogenase active site synthesis and assembly. *FEBS Letters* 583, 506-511.

[34] Driesener R. C., Challand M. R., McGlynn S. E., Shepard E. M., Boyd E. S., Broderick J. B., Peters J. W., Roach P. L., 2010, [FeFe]-Hydrogenase Cyanide Ligands Derived From S-Adenosylmethionine-Dependent Cleavage of Tyrosine. *Angewandte Chemie International Edition* 49, 1687-1690.

[35] Shepard E. M., Duffus B. R., McGlynn S. E., Challand M. R., Swanson K. D., Roach P. L., Peters J.W., Broderick J.B., 2010, [FeFe]-Hydrogenase Maturation: HydG-Catalyzed Synthesis of Carbon Monoxide, *Journal of the American Chemical Society* 132, 9247-9249.

[36] Kuchenreuther J. M., George S. J., Grady-Smith C. S., Cramer S. P., Swartz J. R., 2011, Cell-free H-cluster Synthesis and [FeFe] Hydrogenase Activation: All Five CO and CN_2 Ligands Derive from Tyrosine. *PLoS ONE* 6(5).

[37] Kuchenreuther J. M., Myers W. K., Suess D. L. M., Stich T. A., Pelmenschikov V., Shiigi S. A., Cramer S. P., Swartz J. R., Britt R. D., George S. J., 2014, The HydG Enzyme Generates an $Fe(CO)_2(CN)$ Synthon in Assembly of the FeFe Hydrogenase H-Cluster. *Science* 343, 424-427.

[38] Suess D. L. M., Bürstel I., De La Paz L., Kuchenreuther J. M., Pham C.C., Cramer S. P., Swartz J. R., Britt R. D., 2015, Cysteine as a ligand platform in the biosynthesis of the FeFe hydrogenase H cluster. *Proceedings of the National Academy of Sciences* 112, 11455-11460.

[39] Betz J. N., Boswell N. W., Fugate C. J., Holliday G. L., Akiva E., Scott A. G., Babbitt P. C., Peters J. W., Shepard E. M., Broderick J. B., 2015, [FeFe]-Hydrogenase Maturation: Insights into the Role HydE Plays in Dithiomethylamine Biosynthesis. *Biochemistry* 54, 1807−1818.

[40] Nicolet Y., Rubach J.K., Posewitz M.C., Amara P., Mathevon C., Atta M., Fontecave M., Fontecilla-Camps J.C., 2008, X-ray structure of the [FeFe]-hydrogenase maturase HydE from Thermotoga maritima. *Journal of Biological Chemistry* 283, 18861–18872.

[41] Kuchenreuther J. M., Stapleton J. A., Swartz J. R., 2009, Tyrosine, Cysteine, and S-Adenosyl Methionine Stimulate In Vitro [FeFe] Hydrogenase Activation. *PLoS ONE* 4(10).

[42] Brazzolotto X., Rubach J. K., Gaillard J., Gambarelli S., Atta M., Fontecave M., 2006, The [Fe-Fe]-Hydrogenase Maturation Protein HydF from *Thermotoga maritima* Is a GTPase with an Iron-Sulfur Cluster. Journal of Biological Chemistry 281(2), 769-774.

[43] Czech I., Silakov A., Lubitz W., Happe T., 2010, The [FeFe]-hydrogenase maturase HydF from Clostridium acetobutylicum contains a CO and CN^- ligated iron cofactor. *FEBS Letters* 584, 638-642.

[44] Cendron L., Berto P., D'Adamo S., Vallese F., Govoni G., Posewitz M. C., Giacometti G. M., Costantini P., Zanotti G., 2011, Crystal Structure of HydF Scaffold Protein Provides Insights into [FeFe]-Hydrogenase Maturation. *The Journal of Biological Chemistry* 286(51), 43944-43950.

[45] Caserta G., Pecqueur L., Adamska-Venkatesh A., Papini C., Roy S., Artero V., Atta M., Reijerse E., Lubitz W., Fontecave M., 2017, Structural and Functional Characterization of the hydrogenase-maturation HydF protein, *Nature Chemical Biology* 13(7), 779-784.

[46] McGlynn S. E., Shepard E. M., Winslow M. A., Naumov A. V., Duschene K. S., Posewitz M. C., Broderick W. E., Broderick J. B., Peters, J. W., 2008, HydF as a scaffold protein in [FeFe] hydrogenase H-cluster biosynthesis. *FEBS Letters* 582, 2183–2187.

[47] Vallese F., Berto P., Ruzzene M., Cendron L., Sarno S., De Rosa E., Giacometti G. M., Costantini P., 2012, Biochemical Analysis of the Interactions between the Proteins Involved in the [FeFe]-Hydrogenase Maturation Process. *The Journal of Biological Chemistry* 287(43), 36544-36555.

[48] Johnson M. K. and Scott R. A., 2017, *Metalloprotein Active Site Assembly,* John Wiley & Sons

[49] King P. W., Posewitz M. C., Ghirardi M. L., Seibert M., 2006, Functional studies of [Fe-Fe] Hydrogenase Maturation in an *Escherichia coli* Biosynthetic System. *Journal of Bacteriology* 188(6), 2163-2172.

[50] Esselborn J., Muraki N., Klein K., Engelbrecht V., Metzler-Nolte N., Apfel U. P., Hofmann E., Kurisu G., Happe T., 2016, A Structural view of synthetic cofactor integration into [Fe-Fe]-hydrogenases. *Chemical Science* 7, 959-968.

[51] Berggren G., Adamska A., Lambertz C., Simmons T. R., Esselborn J., Atta M., Gambarelli S., Mouesca J. M., Reijerse E., Lubitz W., Happe T., Artero V., Fontecave M., 2013, Biomimetic assembly and activation of [Fe-Fe]-hydrogenases. *Nature* 499, 66-70.

[52] Esselborn J., Lambertz C., Adamska-Venkates A., Simmons T., Berggren G., Noth J., Siebel J., Hemschemeier A., Artero V., Reijerse E., Fontecave M., Lubitz W., Happe T., 2013, Spontaneous activation of [FeFe]-hydrogenases by an inorganic [2Fe] active site mimic. *Natural Chemical Biology* 9(10), 607-609.

[53] Caserta G., Adamska-Venkatesh A., Pecqueur L., Atta M., Artero V., Roy S., Reijerse E., Lubitz W., Fontecave M., 2016, Chemical assembly of multiple metal cofactors: The heterologously expressed multidomain [FeFe]-hydrogenase from Megasphaera elsdenii. *Biochimica et Biophysica Acta* 1857(11), 1734-1740.

[54] Canaguier S., Field M., Oudart Y., Pecaut J., Fontecave M., Artero V., 2010, A structural and functional mimic of the active site of NiFe hydrogenases. *Chemical Communications* 46, 5876−5878.

[55] Barton B. E., Whaley C. M., Rauchfuss T. B., Gray D. L., 2009, Nickel-Iron Dithiolato Hydrides Relevant to the [NiFe]-Hydrogenase Active Site. *Journal of the American Chemical Society* 131, 6942−6943.

[56] Canaguier S., Fourmond V., Perotto C. U., Fize J., Pecaut J., Fontecave M., Field M. J., Artero V., 2013, Catalytic hydrogen production by a Ni-Ru mimic of NiFe hydrogenases involves a proton- coupled electron transfer step. *Chemical Communications* 49, 5004−5006.

[57] Fourmond V., Canaguier S., Golly B., Field M. J., Fontecave M., Artero V., 2011, A nickel-manganese catalyst as a biomimic of the active site of NiFe hydrogenases: a combined electrocatalytical and DFT mechanistic study. *Energy and Environonmental Science* 4, 2417−2427.

[58] Artero V., Berggren G., Atta M., Caserta G., Roy S., Pecqueur L., Fontecave M., 2015, From Enzyme Maturation to Synthetic Chemistry: The Case of Hydrogenases. *ACS Publications* 48, 2380-2387.

[59] Wilson M. E., Whitesides G. M., 1978, Conversion of a protein to a homogeneous asymmetric hydrogenation catalyst by site-specific modification with a diphosphinerhodium(I) moiety. *Journal of the American Chemical Society* 100(1), 306-307.

[60] Roy A., Madden C., Ghirlanda, G., 2012, Photo-induced hydrogen production in a helical peptide incorporating a [FeFe] hydrogenase active site mimic. *Chemical Communications* 48, 9816–9818.

[61] Sano Y., Onoda A., Hayashi T., 2011, A hydrogenase model system based on the sequence of cytochrome c: photochemical hydrogen evolution in aqueous media. *Chemical Communications* 47, 8229–8231.

[62] Caserta G., Roy S., Atta M., Artero V., Fontecave M., 2015, Artificial hydrogenases: biohybrid and supramolecular systems for catalytic hydrogen production or uptake. *Current Opinion in Chemical Biology* 25, 36-47.

[63] Shepard E. M., McGlynn S. E., Bueling A. L., Grady-Smith C. S., George S. J., Winslow M. A., Cramer S. P., Peters J. W., Broderick J. B., 2010, Synthesis of the 2Fe subcluster of the [FeFe]-hydrogenase H cluster on the HydF scaffold. *Proceedings of the National Academy of Sciences of the United States of America* 107(23), 10448-10453.

[64] Czech I., Stripp S., Sanganas O., Leidel N., Happe T., Haumann M., 2011, The [FeFe]-hydrogenase maturation protein HydF contains a H-cluster like[4Fe4S]–2Fe site. *FEBS Letters* 585(1), 225-230.

[65] Peters J. W., Lanzilotta W. N., Lemon B. J., Seefeldt L. C., 1998, X-ray Crystal Structure of the Fe-Only Hydrogenase (CpI) from Clostridium pasteurianum to 1.8 Angstrom Resolution. *Science* 282(5395), 1853-1858.

[66] Nicolet Y., Piras C., Legrand P., Hatchikian C. E., Fontecilla-Camps J. C., 1999, Desulfovibrio desulfuricans iron hydrogenase: the structure shows unusual coordination to an active site Fe binuclear center. *Structure* 7(1), 13-23.

[67] Mulder D. W., Boyd E. S., Sarma R., Lange R. K., Endrizzi J. A., Broderick J. B., Peters J. W., 2010, Stepwise [FeFe]-hydrogenase H-cluster assembly revealed in the structure of HydAD^{EFG}. *Nature* 465(7295), 248-251.

[68] Khurshid S., Saridakis E., Govada L., Chayen N. E., 2014, Porous nucleating agents for protein crystallization. *Nature Protocols* 9, 1621-1633.

[69] Narayanan J., Liu X. Y., 2003, Protein Interactions in Undersaturated and Supersaturated Solutions: A Study Using Light and X-Ray Scattering. *Biophysical Journal* 84, 523-532.

[70] Trakhanov S., Kreimer D. I., Parkin S., Ames G. F., Rupp B., 1998, Cadmium-induced crystallisation of proteins: II. Crystallization of the Salmonella typhimurium histidine-binding protein in complex with L-histidine, L-arginine or L-lysine. *Protein Science* 7(3), 600-604.

8. Acknowledgements