UNIVERSITÀ DI PARMA

UNIVERSITA' DEGLI STUDI DI PARMA

DOTTORATO DI RICERCA IN

"*Scienze del Farmaco, delle Biomolecole e dei Prodotti per la Salute*"

CICLO XXXII

"Disclosing dynamics and structural features of Cysteine Synthase Complex"

Coordinatore:
Chiar.mo Prof. Marco Mor

Tutore:
Chiar.mo Prof. Stefano Bettati

Dottoranda: Brenda Rosa

Anni 2016/2019

TABLE OF CONTENTS

INTRODUCTION

1. Sulfur assimilation in bacteria

1.1 Cysteine pivotal role

Sulfur is one of the most important and abundant elements on Earth due to its chemical versatility. Inorganic sulfur is incorporated by bacteria into organic compounds with a very complex mechanism, that concludes with cysteine biosynthesis. Bacteria are capable to synthetize directly cysteine from sulfate through the reductive sulfate assimilation pathway (RSAP), that is well conserved in microbes. Cysteine gains a crucial role in this context, since this amino acid is the precursor of a number of key metabolites and sulfur compounds, that include reducing agents, cofactors, peptidic structures and membrane components [1–4]. It has been pointed out that cysteine biosynthesis is related to bacterial fitness and virulence, hence targeting this pathway could lead to reduction of pathogen infectivity, lower persistence of the microbe inside the host and decrease of antibiotic resistance [5,6].

Cysteine biosynthetic pathway is different between bacteria and mammals, since these latter cannot synthetize cysteine directly from inorganic sulfate. Indeed, mammals can only exploit a different pathway, that synthetizes cysteine from the amino acid methionine [7]. Considering that mammals lack RSAP enzymes, these proteins have been studied as potential drug targets for the development of new antibiotics or antibiotic enhancers against several pathogens.

Targeting enzymes involved in cysteine biosynthesis has been proposed for γ-proteobacteria (*Haemophilus influenzae* and *Salmonella enterica* serovar Typhimurium) [7–9], actinomycetales (*Mycobacterium tuberculosis*) [10,11] and protozoa (*Entamoeba histolytica*) [12].

1.2 Reductive sulfate assimilation pathway

RSAP in serovar Typhimurium and *Escherichia coli* follows the scheme illustrated in Figure 1 (adapted from [13]). To begin, sulfur has to be transported inside the cell as sulfate, through a permease. Subsequently, it encounters a series of oxido-reduction systems which allow it to reach the state of bisulfide. This process is connected to another branch, in which CoA-dependent acetylation of L-serine forms *O*-acetylserine (OAS). This reaction is catalysed by serine acetyltransferase (SAT or CysE). Finally, the replacement of β-acetoxy group of OAS with bisulfide concludes with the biosynthesis of the amino acid L-cysteine. The last step of cysteine biosynthesis is catalyzed by *O*-acetylserine sulfhydrylase (OASS), a dimeric pyridoxal 5'-phosphate (PLP)-dependent enzyme [4]. OASS has been identified in two isoforms: CysK (OASS A) and CysM (OASS B), respectively.

CysK was originally identified in *S.* Typhimurium [14,15] and is more abundant with respect to the other isoform. CysM instead can take part to another mechanism of sulfur incorporation, that was first observed in *Aspergillus nidulans* [16]. In this pathway, sulfur enters the cell as thiosulfate and then it reacts with *O*-acetylserine to form S-sulfocysteine. It has been hypothesised that the formation of L-cysteine from S-sulfocysteine is catalyzed by thioredoxin/glutaredoxin systems [17]. The capacity of CysM to accommodate different substrates is due to a larger active site and more structural flexibility associated to the catalysis in comparison with CysK [18].

Another critical difference between these two isoforms is the ability or not to associate with CysE, the prior enzyme in cysteine biosynthetic pathway. This bienzymatic complex is called cysteine synthase (hereafter referred to as CS complex) and was first characterized in *S.* Typhimurium [19].

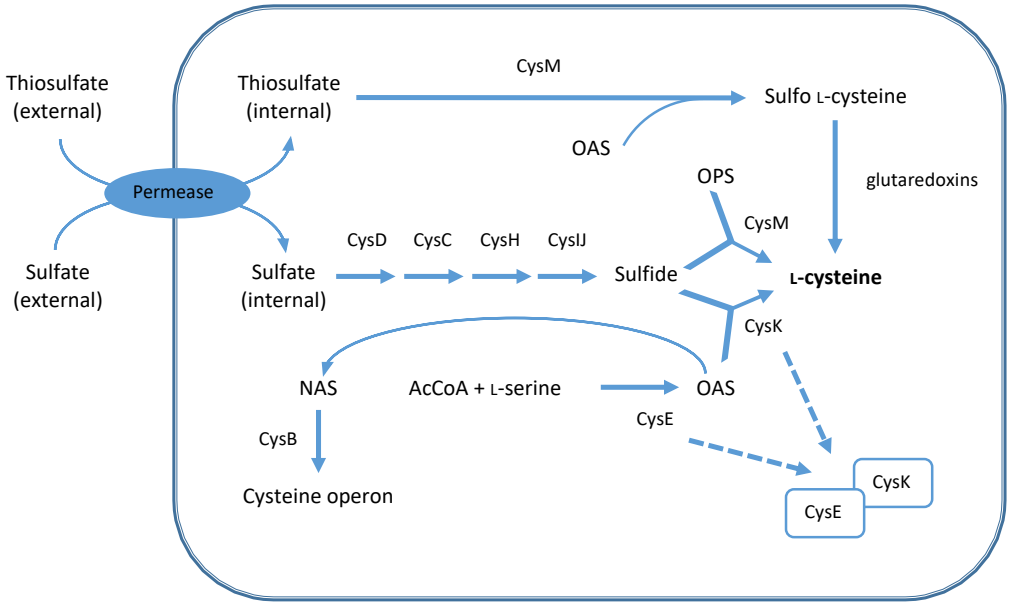


Figure 1. Sulfur assimilation pathway and cysteine biosynthesis in bacteria. Adapted from [13].

1.3 Regulation of cysteine biosynthesis in bacteria

As previously described, cysteine plays a key role in bacterial functions and metabolism. Moreover, high levels of this amino acid could lead to a potential toxicity into the cell. According to this, the biosynthesis of cysteine has to be strictly regulated at different levels, particularly: i) transcriptional control, ii) feed-back inhibition and iii) formation of protein complexes [20].

(i) Genes that encode proteins involved in cysteine biosynthesis are located on cysteine regulon, except for *cysE*. The transcription of the regulon is controlled by CysB, a LysR-type regulator. CysB is induced by *N*-acetylserine (NAS), formed by non-enzymatic conversion of OAS, whereas bisulfide and thiosulfate operate as anti-inducers [21–23].

(ii) Downstream, L-cysteine acts as a competitive feed-back inhibitor on CysE, leading to low levels of OAS. As a result, the sulfate reduction to sulfide is blocked [24,25]. Moreover, it has been demonstrated that consequently to cysteine binding, CysE undergoes a conformational rearrangement of the C-terminal region, that hampers acetyl-CoA entrance into the active site preventing serine acetylation [26].

(iii) The most peculiar mechanism of cysteine biosynthesis control relies on CS complex, that forms in prokaryotes and plants. Sulfur levels inside the cell and CS complex formation are connected. Indeed, bisulfide stabilizes CS complex, probably binding to an allosteric site of CysK, described in *S.* Typhimurium [27]. On the other side, OAS promotes dissociation of the complex, competing with CysE for CysK active site [19,28]. CysK is thus inhibited, as a result of CS complex formation, due to the engagement of its active site in binding with CysE [28].

2. Enzymes involved in the last steps of cysteine biosynthesis

2.1 Serine acetyltransferase

CysE is responsible for the second-last step of cysteine biosynthesis in bacteria and is encoded by *cysE* gene. The structure of hexameric CysE from *E. coli* in complex with cysteine has been described [29] and is illustrated in Figure 2. The crystal structure of CysE from *H. influenzae* in complex with cysteine and CoA has also been solved [26].

The enzyme's monomer comprises an amino-terminal α-helical domain and a carboxy-terminal left-handed β-helical domain, that typically belongs to *O*-acetyltransferase family. The quaternary assembly of the enzyme in the physiological state has been proposed to be a dimer of trimers, interfacing each other at the amino-terminal ends. This overall arrangement agrees with previous studies of crystal packing, as well as gel filtration and chemical cross-linking analysis [30].

O-acetylation of serine implicates the formation of a productive ternary complex of the enzyme with the substrate. The mechanism that has been proposed comes from the study of cysteine binding site, located in a crevice among each subunit of the trimer. Serine competes with cysteine to bind CysE active site and this explains the feed-back inhibition that controls cysteine biosynthesis in bacteria [24].

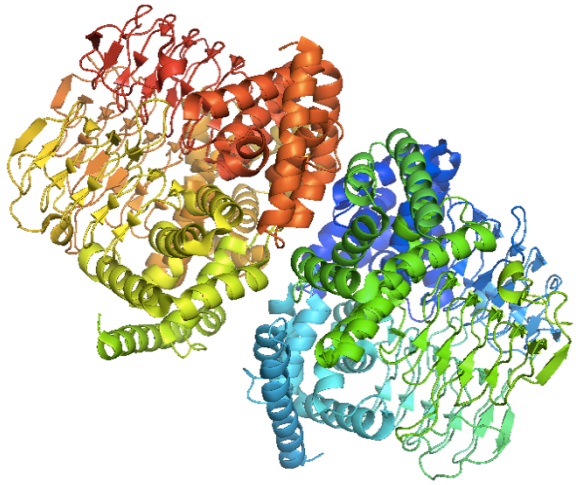


Figure 2. Three dimensional structure of CysE from *E. coli*. CysE dimer of trimers (PDB code: 1t3d).

2.2 *O*-acetylserine sulphydrylase

OASS is the last enzyme of cysteine biosynthetic pathway, that catalyses a beta-replacement reaction, leading to the formation of L-cysteine from OAS and sulfide.

OASS enzymes follow a ping-pong kinetic mechanism, illustrated in Figure 3 [15,31]. Initially, the PLP coenzyme is present in the internal aldimine form (IA), that binds OAS, releasing acetate as the first product. Afterwards, bisulfide binds to an intermediate form of the enzyme, the α-aminoacrylate (AA), and cysteine is produced. At the end, the form of IA is restored.

OASS is a homodimeric PLP-dependent enzyme of the fold type 2 family. Each subunit of the dimer contains a small and a large domain and the location of the active site is between them, in the protein core. In the active site a molecule of PLP cofactor is bound, forming an internal Schiff base with the ε-amino group of a lysine residue. This corresponds to Lys-41 in CysK of *S.* Typhimurium,

whose three dimensional structure has been solved [32]. Eight hydrogen bonds are responsible for anchoring the PLP cofactor to the protein pattern, two of which are formed with water molecules and the other six with residues located in the loop 176-180.

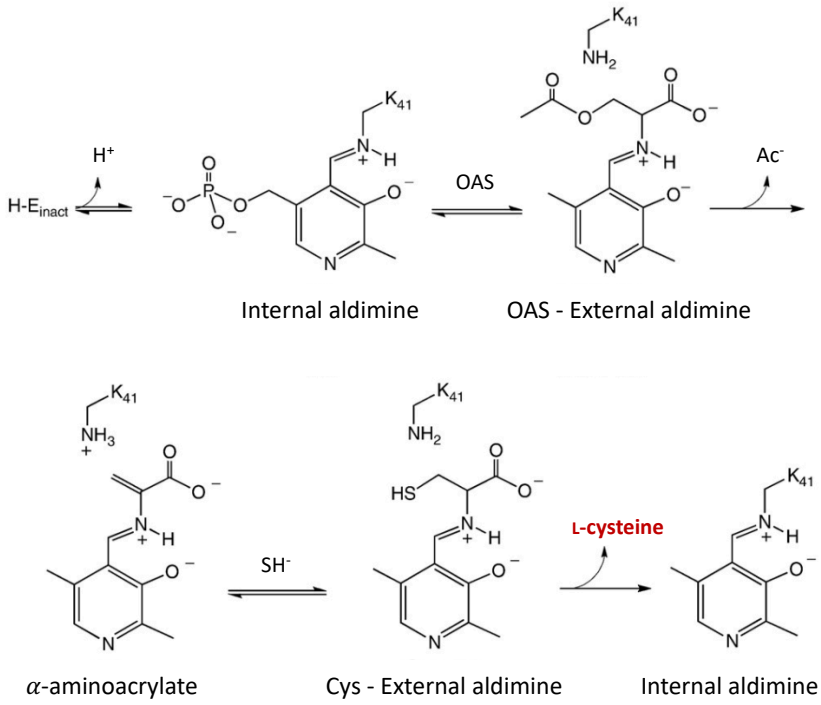


Figure 3. Reaction mechanism of CysK. Adjusted from [33].

It has been demonstrated that this open conformation (PDB code: 1oas) undergoes a large conformational change upon ligand binding [33], that involves mostly the substrate-binding loop (residues 67–71). The closure of the active site triggers another significant rearrangement in a sub-domain of the N-terminal domain (residues 87-131), that packs towards the C-terminal domain (residues 316-322). This latter, as a result, becomes more ordered in the closed conformation (PDB code: 1d6s, corresponding to K41A mutant of CysK in

complex with methionine bound to the active site). Figure 4 represents the superposition of the open and closed conformation.

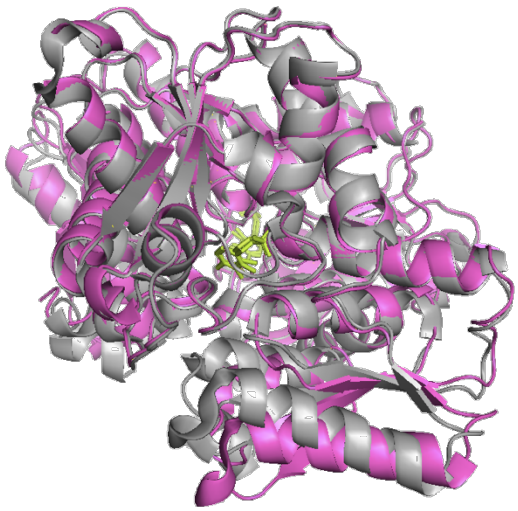


Figure 4. Alignment of the "open" and "closed" conformation of StCysK. The open state is coloured in magenta (PDB code: 1oas), whereas the closed state is depicted in grey (PDB code: 1d6s). The alignment has been made with Pymol software. The PLP in the active site is shown in yellow stick mode.

Another structure of CysK, in complex with chloride bound at an allosteric site [27], has been solved, that corresponds to an "inhibited" conformation. The latter is different from the open and closed arrangement and supports the role of the physiological inhibitor bisulfide.

The structure of CysM, the other isoform of OASS, has been solved for the enzyme from *E. coli* and *S.* Typhimurium [18,34]. CysM presents a similar structure and mechanism, compared to CysK. However, the presence of two polar residues (D281 and C280) allows CysM active site to be more hydrophilic. CysM also appears to be structurally more flexible.

A differential expression of CysM in comparison with CysK during anaerobic growth conditions has been hypothesized [35,36], but later findings are not in agreement with these observations [37].

2.3 Engagement of CysK in formation of complexes

CysK fulfils several biological functions, both acting alone or being involved in formation of protein-protein complexes. The organization of multienzymatic assemblies seems connected to metabolic processes as well as regulation of biochemical phenomena inside the cell, in a coordinated manner [38,39]. Nowadays, increasing research efforts are directed to investigate protein-protein interactions to better understand how could these be related to diseases development and thus, the possibility of targeting protein complexes association [40].

In this context, it has been pointed out that CysK acts as a promiscuous protein, exhibiting similar interaction patterns with different protein partners, besides CysE. These protein-protein interactions (PPIs) count for different functions, aside from its enzymatic role in cysteine biosynthesis [20]. CysK paralogs and interactors and the biological roles connected to these PPIs are reported in Table 1.

In Gram-positive bacteria, like *Bacillus subtilis*, for instance, it has been found that CysK associates with CymR, a transcriptional factor of the Rrf2-type family, that modulates the expression of genes involved in cysteine uptake and biosynthesis. Following the formation of this complex, CysK strengthens CymR and DNA binding leading to transcriptional repression of cysteine regulon [41].

Moreover, CysK has been related to a bacterial contact-dependent growth inhibition (CDI) system, that plays a role in Gram-negative inter-cellular competition. Two secretion proteins, namely CdiA and CdiB, coordinate CDI.

CdiA from uropathogenic *E. coli* strain 536 (UPEC536) releases its toxic domain CdiA-CT into target cells, but it's normally inactive. The association of CdiA-CT with CysK to form a stable binary complex (see chapter 4, page 24) triggers the activation of the toxin, that cleaves tRNA molecules [42].

The peculiar assembly with CysE, the previous enzyme in cysteine biosynthetic pathway, is particularly relevant to this PhD work and will be described in detail in chapter 3.

Recent studies identified in *Caenorhabditis elegans* CYSL-1, a CysK paralog, that interacts with the hydrolase EGL-9 and transduces signals to modulate the behavioral plasticity, in response to changes of O_2 level [43].

Table 1. CysK biological roles and its binding partners. Adapted from [20].

CysK interactor	Organism	Biological function
CymR	*B. subtilis* *S. aureus*	Transcription regulation (repression)
CdiA-CT $^{UPEC\ 536}$	*E. coli*	tRNase activation in CDI system
CysE	*E. coli* *S.* Typhimurium *H. influenzae* *A. thaliana*	Cysteine biosynthesis control in plants
EGL-9	*C. elegans*	Modulation of O_2-dependent behavioral plasticity

Although CysK binding to different partners is related to a specific biological function, all the moonlighting interactors could affect CysK activity in cysteine biosynthesis. Indeed, each partner binds to CysK active site, mimicking CysE C-terminal region, leading to CysK inhibition upon complex formation [20].

3. Cysteine synthase complex

CS is a bienzymatic complex that has been studied since many years. It has been discovered and characterized for the first time in *S.* Typhimurium [19], but many reports confirm that CS assembly occurs also in other proteobacteria, such as *E. coli* [44–47] and *H. influenzae* [48–50]. A comprehensive study has been done also in plants, mostly on *Arabidopsis thaliana* [28,51–57].

Concerning the biological function of CS complex formation in bacteria, this association seems to be involved in protection of CysE towards cold inactivation and proteolysis [58,59], whereas the hypothesis of a substrate channeling role for the complex has been discarded [28]. Instead in plants, CS complex has been proposed to act as a metabolic sensor of sulfur supply in the cell [60].

3.1 Structural details of protein-protein interaction within the complex

The three dimensional structure of CS complex has not been solved yet, however an extensive characterization of CysK/CysE interaction has revealed interesting structural features. Biochemical studies have elucidated the important role of CysE C-terminal region, that is connected to CysE feedback inhibition by L-cysteine [45], but also to association with CysK. Indeed, the last 10 residues of CysE tether CysK active site to form the CS complex [44,61], as demonstrated by structural investigation of CysK crystals from *H. influenzae* complexed with CysE C-terminal decapeptide (illustrated in Figure 5, panel A) [50]. Both the isolated CysE C-terminal sequence and the full length protein interact with CysK active site, triggering a conformational change from an open to close state of CysK, likewise substrate binding [33,49]. Considering both cases, the affinity of the decapeptide has been found to be much weaker with respect to CysE [61].

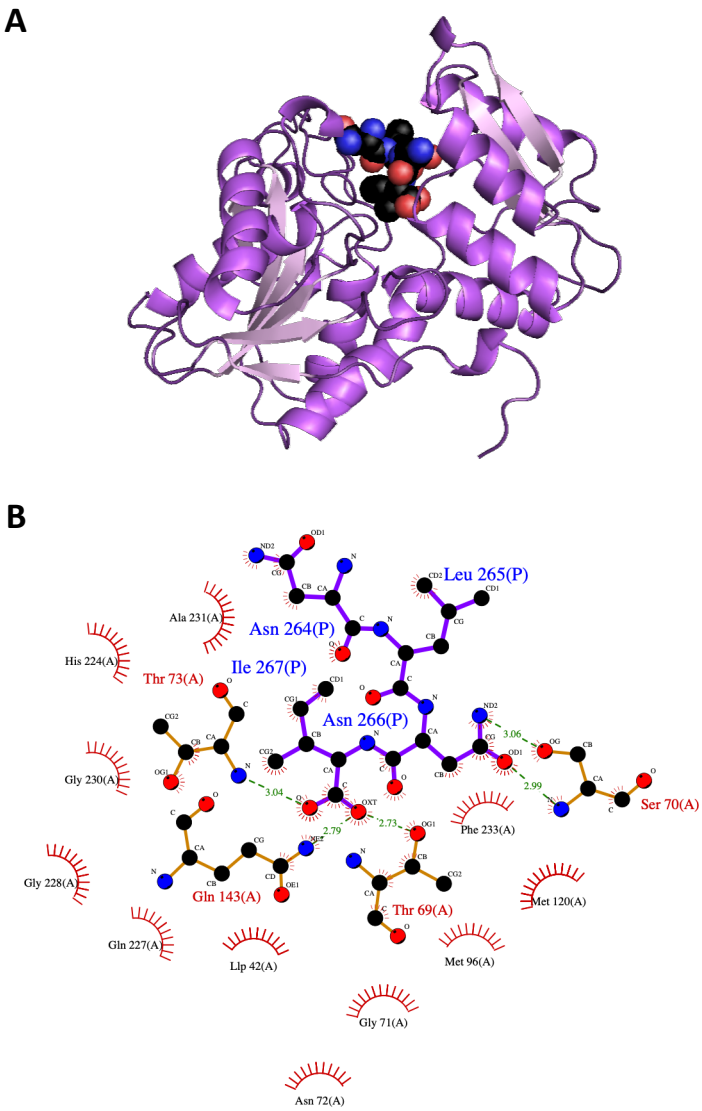


Figure 5. Last four residues of HiCysE C-terminal tail inserted into HiCysK active site. Panel A. Crystal structure (PDB code: 1y7l). CysK is represented in violet as a cartoon, the last 4 residues of CysE are shown as spheres. **Panel B.** LIGPLOT of interactions involving the last four residues of CysE (ligand Asn-Leu-Asn-Ile, depicted in sticks and balls) inside CysK active site. Made with PDBSum database (http://www.ebi.ac.uk/pdbsum/), with the entry 1y7l.

In particular, when OAS binds to CysK, its carboxylate forms a hydrogen bond with Thr69 in the active site and this interaction is thought to be responsible for the conformational transition. The carboxylate of CysE C-terminal Ile interacts with CysK similarly and could thus induce the closed form of the enzyme [33,48]. Furthermore, the insertion of CysE C-terminal part into CysK binding pocket inhibits competitively OAS binding to CysK.

Another interesting structural detail arises from the comparison of amino acids at CysE C-terminal region from different organisms. Indeed, a conserved isoleucine residue is highlighted. This amino acid is fundamental and anchors complex formation, so that CysE and CysK don't assemble in the complex following isoleucine removal from the C-terminal sequence of CysE [47]. These findings are confirmed by analysing the network of interactions established by the isoleucine residue of CysE decapeptide with HiCysK active site in the crystal structure [50]. Particularly, only the last four residues of the decapeptide display interpretable electron density in the structure.

A study based on docking of different pentapeptides, mimicking CysE C-terminal region (MNLNI), revealed that the carboxylate of Ile267 is involved in hydrogen bonds formation with residues Thr69, Asn72, Thr73 and Gln143 in HiCysK active site. Asn266 makes hydrogen bond contact with another residue in CysK binding pocket, that corresponds to Ser70 (Figure 5, panel B) [62].

In addition to this, further studies demonstrated that other two residues, an aspartic acid and a glutamic acid, are present in SAT C-terminal peptide from various sources, leading to the hypothesis that these residues are relevant to CS complex assembly [61].

Experiments directed to detect protein-protein interactions [63] have furthermore revealed other structural features near to CysK active site that are critical for CS complex association. Particularly, mutation of residues that belong to the

conserved loop β8A-β9A, located about 20 Å from the active site (Lys217, His221 and Lys222 in *A. thaliana* CysK), disturbs CS complex formation. Nonetheless, mutating these residues hasn't affected CysK activity.

3.2 Complex overall geometries explored, mechanism of formation and allosteric interactions

With respect to stoichiometry of the complex, it has been determined that two dimers of CysK are bound to one hexamer of CysE [49,51]. According to that, the quaternary geometry of CysK binding to CysE could respond to two different models, described in Figure 6. The first scenario, that is the most commonly accepted, illustrates only one CysK active site interacting with CysE C-terminal region (panel A). The other model, presents both CysK subunits involved in binding with the two trimers of CysE (panel B) [49].

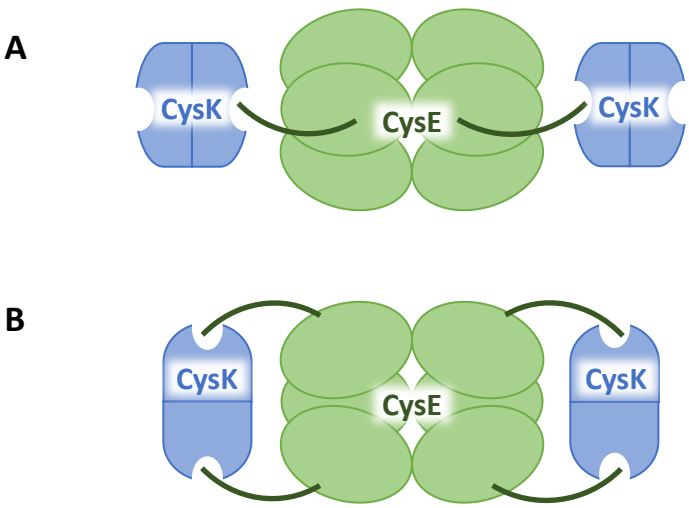


Figure 6. Models of quaternary geometry proposed for CS complex. Panel A. Scenario 1: one CysK (blue) active site engaged in binding with CysE (green) C-terminal peptide. **Panel B.** Scenario 2: both CysK (blue) active sites bind CysE (green) C-terminus. Adapted from [49].

Structural and functional studies have demonstrated that CysK is almost fully inhibited upon complex formation, though retaining about 10% of its activity, even when concentrations of CysE are saturating. These finding lead to hypothesize that an unoccupied CysK active site within the complex would be responsible for the residual activity, making more likely the case of the first scenario in Figure 6 [44,45]. In this model, CysE C-terminal tail insertion into CysK active site is connected to competitive inhibition with the substrate. Additionally, an allosteric communication seems to induce the closure of the other unoccupied CysK active site, accounting for the reduced catalytic activity.
The characterization of CS complex assembly mechanism has been deepened through equilibrium and pre-steady state kinetic experiments. It has been demonstrated that CysE association to CysK involves at least two steps and consequently three species of the complex, as shown in reaction 1:

$$\mathrm{CysK} + \mathrm{CysE} \underset{k_{-1}}{\overset{k_1}{\rightleftharpoons}} \mathrm{CysK{:}\,CysE} \underset{k_{-2}}{\overset{k_2}{\rightleftharpoons}} \mathrm{CysK{:}\,CysE'} \underset{k_{-3}}{\overset{k_3}{\rightleftharpoons}} \mathrm{CysK{:}\,CysE''} \quad (1)$$

The first step, corresponding to the docking of CysE C-terminal part, leads to the generation of an initial weak encounter complex. Subsequently, CysK undergoes a conformational transition [33] leading to the formation of a tight complex, that is the rate limiting step of CS assembly [48]. In this passage, the CysE structure is required to isomerize to an intermediate specie. The final configuration corresponds then to the fully isomerized form of the complex between CysE and CysK, that is responsible for CysK inactivation [48,64].
On the other side, there are evidences that CysK binding affects allosterically the communication between the two trimers of CysE [30] and alters its active site, leading to relief of substrate and feedback inhibition [45]. These effects could explain CysE activation within CS complex in bacteria [28,45].

3.3 Detection of allosteric effects

Many protein functions, including enzymes catalytic activity, oxygen transport and delivery, signal transduction and gene expression are allosterically controlled, as is the case of CS assembly [52,65,66]. Protein allostery occurs when regulation of the functional activity at the active site is operated through binding of effectors at a distal site of the same molecule. Concerning oligomeric complexes, allosteric interactions can be divided into homotropic, when the same ligand binds two or more equivalent binding sites of a multi-subunit complex (cooperative binding), or heterotropic, when different ligands bind to a distinct site and modulate the substrate binding affinity [65,67].

Both experimental and computational approaches have been employed to detect allosteric communications. X-ray crystallography provides structural details of the protein following perturbation, but it's limited by the lack of dynamical information, whereas transient conformations can be detected through nuclear magnetic resonance (NMR) spectroscopy [67]. Moreover, protein label based methods, e.g. hydrogen/deuterium exchange mass spectrometry (HDX-MS), and fluorescence resonance energy transfer (FRET) have been employed to track conformational exchange dynamics [68]. These experimental methods could be furthermore complemented by molecular dynamic simulation from computational approaches, to capture protein conformational motions [69].

3.4 Exploitability of CS complex

Despite a lot of literature concerning CS complex stoichiometry, mechanism of formation and some conserved structural features, many points have to be clarified yet. For instance, understanding the proper function of the complex in cysteine biosynthesis and the mechanism of reciprocal regulation of constituent enzymes. Moreover, its perspective exploitation as a pharmaceutical target makes

it an interesting subject to be characterized. Indeed, so far, different enzymes that belong to cysteine biosynthetic pathway have been studied for the development of novel antibiotics or antibiotic enhancers. For instance, inhibitors against both CysK and CysM isoforms have been investigated especially in *M. tuberculosis* and *E. histolytica*, due to their clinical relevance [70–73]. Also CysE, the previous enzyme in the biosynthetic pathway, has been targeted to design novel antiamoebic agents [12]. Furthermore, another enzyme of RSAP, that is the phosphoadenosine phosphosulphate reductase (CysH), has been considered as a potential therapeutic target against persistent *M. tuberculosis* [74].

To date, no molecules capable of inhibiting CS complex have been discovered, nor molecules able to interfere with complex formation, apart from OAS [19,75]. On the other side, targeting one of the constituent partners could lead to perturbation of the complex and perhaps affect not only the principal activity of the enzyme.

Elucidation of structural details and dynamics of protein-protein interaction could help understanding CS complex regulatory role and its exploitability as a pharmaceutical target.

4. CysK/CdiA-CT complex

4.1 Bacterial contact-dependent growth inhibition system

A significant role for CysK in contact-dependent growth inhibition (CDI) system in bacteria has been recently discovered and described [20,42]. CDI is a mechanism of bacterial competition, that blocks the growth of neighboring cells upon cell-to-cell contact. In uropathogenic *E. coli*, CDI takes place through two partner secretion proteins, CdiA and CdiB, depicted in Figure 7. Another protein, CdiI, provides immunity against the toxin released by CdiA, preventing autoinhibition.

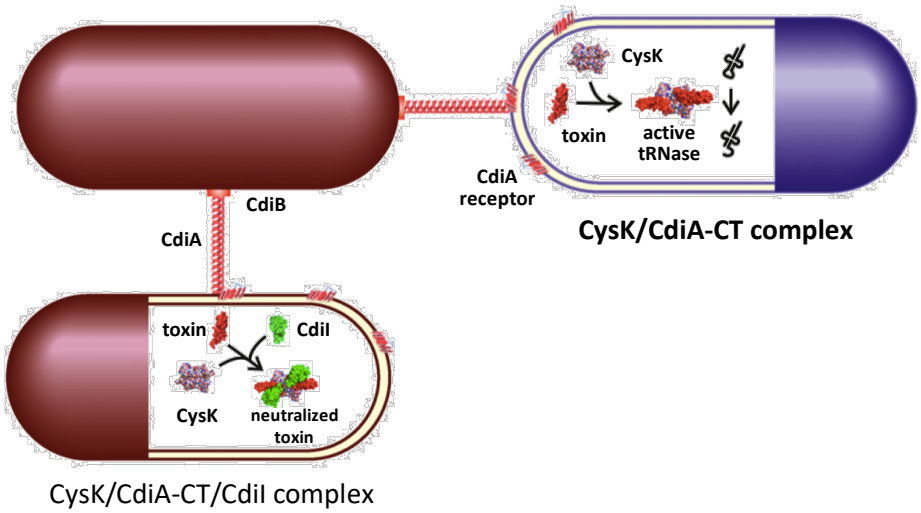


Figure 7. Scheme of CDI System in bacteria. In target cells, CysK forms a complex with CdiA-CT, activating the tRNase activity of the toxin (violet cell at top right, complex CysK/CdiA-CT). In isogenic cells, CdiI prevents autoinhibition, forming the ternary complex CysK/CdiA-CT/CdiI (dark red cell at bottom left). Modified from [20].

Particularly, CdiA are effector proteins that are exported by CdiB transport proteins onto CDI^+ inhibitors on cells surface. CdiA, after binding to specific receptors on neighboring cells, releases the protein C-terminal end (hereafter referred to as CdiA-CT), containing the toxic properties. In UPEC 536 CDI

system, the toxin enters the target cells and binds to CysK, that acts as a permissive factor activating the nuclease, whereas in sibling cells CdiI immunity protein inactivates the toxin complex [76–78]. CysK and CdiA-CT associate thus to form a stable complex and this interaction is connected to activation of the tRNase activity of the toxin. CdiA-CT binds both CysK and CdiI to form a ternary complex, whose biological role remains uncertain.

4.2 Features of CysK interaction with CdiA-CT

CdiA-CT in association with CysK appears to share the same interaction pattern of CysE within the CS complex. Indeed, CdiA C-terminus contains the sequence Gly-Tyr-Gly-Ile, that mimics *E. coli* CysE critical residues that anchor binding to CysK, that are Gly-Asp-Gly-Ile [20,42]. Moreover, OAS blocks CysK/CdiA-CT complex association, whereas L-serine doesn't affect the binding. These observations further support the fact that the molecular mimicry of CdiA-CT association with CysK active site is exploited as a conserved protein-protein interaction to promote the toxic activity [42].

The crystal structure of CdiA-CT^{EC536} in binary complex with CysK has been recently solved [79] and it is illustrated in Figure 8. Two toxin monomers are bound to each CysK dimer, without interacting among themselves. These structural studies have confirmed that CdiA-CT inserts its C-terminal tail into CysK binding pocket, making hydrogen bond contacts through Ile227 with conserved residues of CysK active site (Thr69, Asn72, Thr73 and Gln143). This network of interactions with CysK has been already observed for CysE association within CS complex [50,53,62].

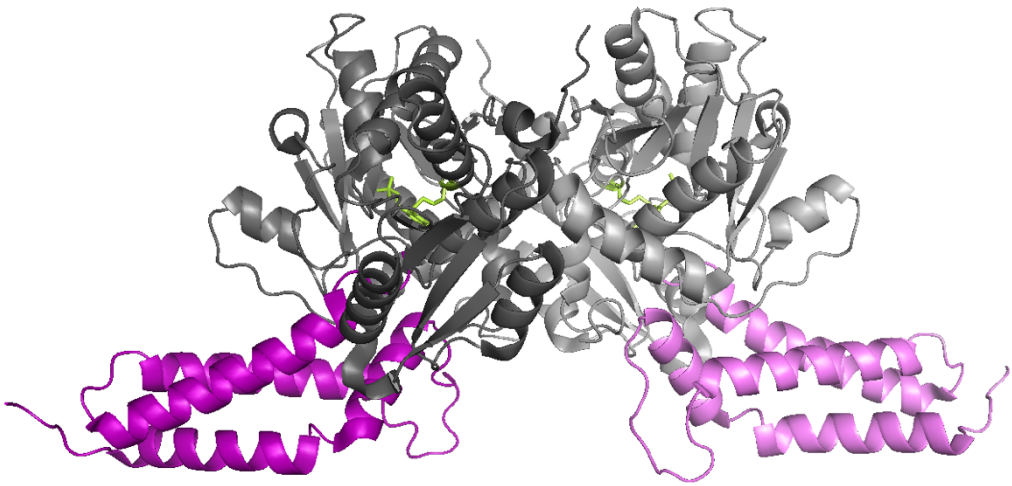


Figure 8. CysK/CdiA-CT complex crystal structure. CysK dimer is coloured in dark and light grey, the toxin is depicted in violet and purple (PDB code: 5j43). The PLP in the active sites is shown in yellow stick mode.

Functional data [80] have demonstrated that both complexes associate through a two steps binding mechanism where the formation of an encounter complex is followed by a slow isomerization process.

The crystal structure shows that CysK in complex with CdiA-CT adopts an open conformation, as well as the unliganded state of the protein [81], suggesting that different from CysE the toxin doesn't induce a conformational change of CysK active site [33].

According to this observation, the rate-limiting conformational change in CysK/CdiA-CT complex formation could be ascribed to the toxin active site rearrangement towards its catalytically active form.

CdiA-CT binding to CysK displays similar affinity compared to CysE, nonetheless the two CysK binding partners are not in competition with each other

[80]. Moreover, a ternary complex including the three proteins together (CysE/CysK/CdiA-CT) has not been isolated in *E. coli*.

Thus, the model that has been proposed for CysK/CdiA-CT and CS complex formation is illustrated in Figure 9: CysK adopts an open conformation when not complexed with other proteins or substrates and maintains this configuration when CdiA-CT binds to its active sites. CysK undergoes instead a conformational change to a closed state following CysE binding, as explained in chapter 3.1 [80].

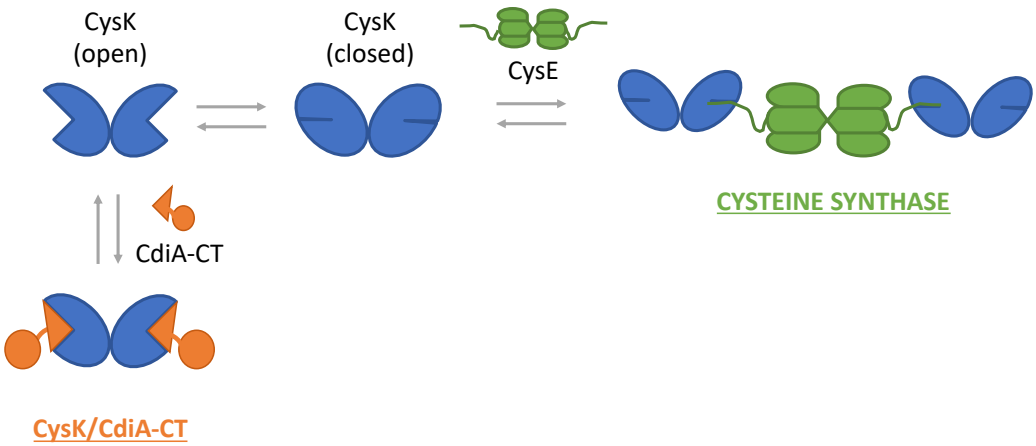


Figure 9. Formation of CS complex (indicated as cysteine synthase) and CysK/CdiA-CT complex. Adapted from [74].

As described in this chapter, CdiA-CT shares different common features with CysE concerning the interaction with CysK active site. Moreover, CdiA-CT is the only partner that forms a complex with CysK for which it has been possible to obtain a crystal structure until now, and therefore it could be used to better understand CysK association with CysE and help to gain insights into CS complex structure.

AIM OF THE THESIS

The focus of this PhD work was the characterization of the bienzymatic complex formed by CysK and CysE, the proteins that catalyse the last two steps of cysteine biosynthesis in bacteria and plants. This assembly, called Cysteine Synthase (CS), has been the object of an extensive investigation over the years. In particular, enzymes involved in cysteine biosynthetic pathway, like CysK and its isoform CysM, but also CysE and CysH have already been proposed as targets of antibiotic therapies, due to cysteine connection to many functions relevant to bacterial fitness and pathogenicity. In this context, CS assembly could be itself exploited as a pharmaceutical target or, on the other side, its association/dissociation could also affect the activity of the constituent enzymes. Challenging missing information concerning the structure and the regulatory role of bacterial CS complex could thus be useful for the development of novel antibiotics or antibiotic enhancers.

This PhD work applied a combination of different approaches with the aim to further characterize different aspects of CS structure and dynamics, in particular:

- exploring protein-protein interaction hotspots of CS assembly;
- answering questions concerning CS complex quaternary geometry in solution, to clarify the binding mode of the two dimers of CysK with the hexamer of CysE;
- investigating the respective changes in dynamics affecting the constituent enzymes, as a consequence of their interaction.

As a first step, protein painting technique helped to characterize protein-protein interaction (PPI) hotspots between CysK and CysE within the CS complex, after the validation of the assay on a known binary complex, composed by CysK and a bacterial tRNase (CdiA-CT), involved in CDI system. Subsequently, these results were used as constraints to create a model for the three dimensional organization of CS assembly, based on the envelope obtained by SAXS measurements. Finally, investigation on how CS formation induces changes in

dynamic properties of constituent enzymes, both CysK and CysE, was performed by using Hydrogen/Deuterium Exchange coupled to Mass spectrometry (HDX-MS).

CYSTEINE SYNTHASE PROTEIN-PROTEIN INTERACTION HOTSPOTS REVEALED BY PROTEIN PAINTING

1. Introduction

Protein-protein interactions (PPIs) and assemblies are crucial for many biological processes and signaling pathways and due to their functional and regulatory role, could potentially be exploited as drug targets. Despite their relevance, PPIs are difficult to characterize, moreover many multienzymatic complexes lack crystal structure information, as is the case of CysK interacting with CysE. With the aim to identify hotspots of PPIs in the CS complex, a protein painting assay was employed. This recently developed technique exploits small molecular dyes to cover the solvent-exposed surface of proteins and protein complexes in solution. In the native protein conformation, dyes cannot gain access to buried areas, such as interprotein interfaces. Upon denaturation, previously buried sequences, unprotected by dyes, become accessible to solvent and proteolytic enzymes. The following tryptic digestion workflow allows identification of peptides, containing only residues that in the native protein conformation are part of protein-protein interfaces or buried regions. This assay was first validated on CysK/CdiA-CT complex, that shares the same pattern of interactions adopted by CysE in complex with CysK and for which the three dimensional structure is available. Afterwards, the same workflow was applied to CysK assembly with CysE, disclosing interesting features of this PPI.

Main results of this work are reported in the paper "Combination of SAXS and Protein Painting Discloses the Three-Dimensional Organization of the Bacterial Cysteine Synthase Complex, a Potential Target for Enhancers of Antibiotic Action", that was published in *International Journal of Molecular Sciences* [13].

2. Materials and methods

2.1 Materials

All chemicals and reagents were purchased from Sigma - Aldrich (St. Louis, MO, USA) and were of the highest grade commercially available, unless otherwise stated. Materials were used as received.

2.2 Proteins expression and purification

CysK and CysE from *E. coli*, cloned in pET21P and pSH21p vectors, respectively, were over-expressed in the bacterial strains BL21(DE3) and BL21(DE3) Tuner^{TM}. Cells were grown in Luria Bertani medium at 37 °C and induced with 1 mM isopropyl β-D-1-thiogalactopyranoside (IPTG).

CysK and CysE were purified and the tag was removed from CysE following a procedure described in details in a previous work [45], with minor modifications. CysK concentration was calculated based on the absorbance of the coenzyme pyridoxal 5'-phosphate (PLP), using an extinction coefficient at 412 nm of 9370 $M^{-1} \cdot cm^{-1}$, calculated by the alkali denaturation method [82]. CysE concentration was determined using an extinction coefficient at 280 nm of 26,900 $M^{-1} \cdot cm^{-1}$.

Protein purity was assessed by SDS-PAGE and was estimated to be higher than 95%.

CdiA-CT was expressed in *E. coli* BL21(DE3) Tuner^{TM} and was purified as CdiA-CT:CdiI-His6 complex, as previously described in [77]. CdiA-CT and CdiI were separated by metal-affinity chromatography in 8 M urea. CdiA-CT was refolded by dialysis in 20 mM sodium phosphate (pH 7.0), 85 mM sodium chloride, 10 mM 2-mercaptoethanol (2-MCE), 2 mM EDTA. Its native structure was evaluated by circular dichroism spectroscopy. The protein was further purified using size-exclusion chromatography (SEC) on a fast protein liquid

chromatography (FPLC) column packed with Ultrogel AcA44 resin (exclusion limit 200 kDa, operating range 17–175 kDa, column volume 63 ml and void volume 20.4 ml, Pall Corporation, Port Washington, NY, USA), run at 0.2 ml/min in 20 mM sodium phosphate, 85 mM sodium chloride, 10 mM 2-MCE, 2 mM EDTA, pH 7.0.

CdiA-CT concentration was calculated using an extinction coefficient at 280 nm of 13,300 $M^{-1} \cdot cm^{-1}$.

2.3 Enzyme activity assays

The specific activity of CysK, determined with the discontinuous assay of Gaitonde [83] in the presence of 0.6 mM NaHS, 10 mM OAS, and 3 nM enzyme (on a monomer basis), was 280 U/mg.

The specific activity of CysE, determined at 20 °C by measuring the disappearance of acetyl coenzyme A (AcCoA) signal at 232 nm in the presence of 1 mM L-Ser, 0.25 mM AcCoA, and 7 nM enzyme (monomer), was 83 U/mg, in agreement with previously-reported kinetic data [45].

2.4 Spectroscopy

Absorption measurements were carried out at 20.0 ± 0.5 °C using a Varian (Palo Alto, CA, USA) CARY400 spectrophotometer. All spectra were corrected for buffer contribution.

Fluorescence emission spectra were collected using a FluoroMax-3 fluorometer (HORIBA Jobin Yvon, Kyoto, Japan) at 20 ± 0.5 °C, equilibrating samples for 5 min prior to spectra acquisition. All spectra were corrected for buffer contribution, and the slit width was set to optimize the signal-to-noise ratio.

CysE/CdiA-CT stoichiometric binding to CysK was monitored by measuring PLP fluorescence emission at 500 nm following excitation at 412 nm [45,48,49,80]. For equilibrium binding experiments, the dependence of emission intensity on ligand concentration was determined using the binding isotherm equation:

$$I = I_0 \cdot \frac{I_{max} \cdot [L]}{K_d + [L]} \quad (1)$$

or a quadratic equation that describes tight binding:

$$I = I_0 + I_{max} \cdot \frac{([P] + [L] + K_d) - \sqrt{([P] + [L] + K_d)^2 - 4 \cdot [P] \cdot [L]}}{2} \quad (2)$$

where I is the fluorescence intensity at 500 nm in the presence of CysE or CdiA-CT, I_0 is an horizontal offset, I_{max} is the maximum change in fluorescence at saturating [L], [L] is the total ligand concentration, [P] is the total protein concentration and K_d is the dissociation constant for the complex.

2.5 Size Exclusion Chromatography

The oligomeric state of CysE and CysE/CysK complex in native conditions was determined on an analytical HPLC-SEC Superdex 200 increase 3.2/300 column (GE-Healthcare, Chicago, IL, US) in PBS in the presence of 1 mM tris(2-carboxyethyl)phosphine (TCEP). A calibration curve was obtained by running five commercial standards for size exclusion chromatography (blue dextran, ferritin 440 kDa, conalbumin 75 kDa, ovalbumin 43 kDa, and carbonic anhydrase 29 kDa, GE-Healthcare) and the home-made standard glyceraldehyde 3-phosphate dehydrogenase (GAPDH, 144.2 kDa).

2.6 Protein painting for Mass Spectrometry

All proteins solutions, either alone (CysK) or in complex with binding partners (CysK/CdiA-CT or CysK/CysE), were incubated for 15 min with a mixture containing the following molecular dyes at 5 mM concentration each, according to a published protocol [84]: disodium 1-amino-9,10-dioxo-4-[3-(2-sulfonatooxyethylsulfonyl)anilino]anthracene-2-sulfonate (RBB), disodium 4-amino-3-[[4-[4-[(1-amino-4-sulfonatonaphthalen-2-yl)diazenyl]phenyl]phenyl] diazenyl]naphthalene-1-sulfonate (CR), sodium 4-(4-(benzyl-et-amino)-ph-azo)-2,5-di-cl-benzenesulfonate (AO50).

In addition to this, another set of molecular paints was tested, containing: 8-Anilino-1-naphthalenesulfonic acid ammonium salt (ANSA, K&K Laboratories, New York, MA, USA), 4-(3,6-dimethyl-1,3-benzothiazol-3-ium-2-yl)-N,N-dimethylaniline chloride (Thioflavine T), 2- amino -5- [(4-amino-3-sulfophenyl)(4-imino-3-sulfo-2,5-cyclohexadien-1-ylidene)methyl] -3- methyl-benzenesulfonic acid, sodium salt (Acid Fuchsin), 4',5'-dibromo-3',6'-dihydroxy-2',7'-dinitro-spiro[isobenzofuran-1(3H),9'[9H]xanthen]-3-one (Eosin B, Merck, Darmstadt, Germany). Dye solutions were prepared dissolving each molecular paint in PBS.

Proteins alone and the complexes were incubated with the different sets of molecular dyes for 15 min at room temperature. To allow proper complexes formation, CysK/CdiA-CT and CS complex were pre-incubated 30 min on ice, before mixing each complex with the molecular paints. Proteins were incubated in a total volume of 50 µl, at the following concentrations:

- 5.13 µM CysK alone;
- 3.33 µM CysK and 5 µM CdiA-CT for the complex formed by CysK and the toxin;
- 5 µM CysK and 7.5 µM CysE for CS complex.

Protein concentrations within the complexes were chosen according to binding stoichiometries. Each sample was prepared in triplicate.

Molecular paints excess was removed by acetone precipitation: a four-fold sample volume of cold acetone was added to the sample, mixed and incubated for 1 hour at -80 °C. Then, samples were centrifuged for 15 min at 16,000×g at 4 °C and the supernatant discarded. The precipitates were resuspended in 50 µl PBS and samples were denatured with urea (final concentration 0.5 M), reduced in 13 mM 1,4-dithiothreitol (DTT) for 15 min at room temperature and alkylated with 16 mM iodoacetamide (15 min at room temperature, in the dark). Finally, digestions were performed with trypsin for 2 hours at 37 °C, at the following protease/protein ratio: 1:4.4 w/w for CysK alone and CysK in CS complex and 1:3.1 w/w for CysK in complex with CdiA-CT. To stop trypsin reactions, trifluoroacetic acid (TFA) at 0.1% final concentration was added to the samples. Tryptic peptides were desalted, using 10 µl Pierce® C18 Tips (Thermo Scientific, Waltham, MA, USA), for the following mass spectrometric (MS) analysis.

2.7 Mass Spectrometric analysis

Mass spectrometric analyses were carried out using a 4800 Plus MALDI TOF/TOF™ spectrometer (Ab Sciex, Framingham, MA, USA) in positive ion reflectron mode combining 400 shots in the mass range 500–3600 Da. MALDI spots were prepared using the dried droplet method. Briefly, 1 µL of sample was mixed with 1 µL of 10 mg/ml α-cyano-4-hydroxycinnamic acid (HCCA) in 75% v/v acetonitrile and 2.5% v/v TFA, and 0.5 µL of the solution was spotted onto the plate. Each sample was analysed in triplicate. External calibration was performed using the following mixture of standard peptides: bradykinin fragment 1–7 (m/z 757.3997), angiotensin II (human) (m/z 1046.5423), P14R (m/z 1533.8582), ACTH fragment 18–39 (human) (m/z 2465.1989), and insulin chain

B oxidized (m/z 3494.6513). Calibrations were accepted at the following conditions: 15 ppm mass tolerance and max outlier error, 4 minimum peaks to match, 5 as minimal signal-to-noise ratio (S/N). Resulting MS spectra were submitted to an in-house created database search using the Mascot search engine. Methionine oxidation was selected as variable modification and carbamidomethylation of cysteine as fixed modification, two missed cleavages were tolerated, and peptide mass tolerance was set at 100 ppm. To ensure the quality of database searching results, the peak assignment for every peptide was manually checked, considering S/N = 3 as the limit of detection (LOD).

3. Results and discussion

Formation of protein complexes and the respective important contact points, defined as hotspots of the interaction, are fundamental in many biological processes and respond to structural and regulatory functions. Furthermore, protein-protein interactions (PPIs) could be exploited as drug targets, though their identification and characterization remain a challenging goal [40]. Indeed, protein-protein complexes often lack crystal structures and yet, these structures could not perfectly describe physiological protein-protein assembly conditions [85].

In this context, a protein painting technique has been recently developed [84,86], with the aim to spot protein-protein interface residues through binding of a set of small molecular dyes or "paints".

The workflow is schematically illustrated in Figure 10.

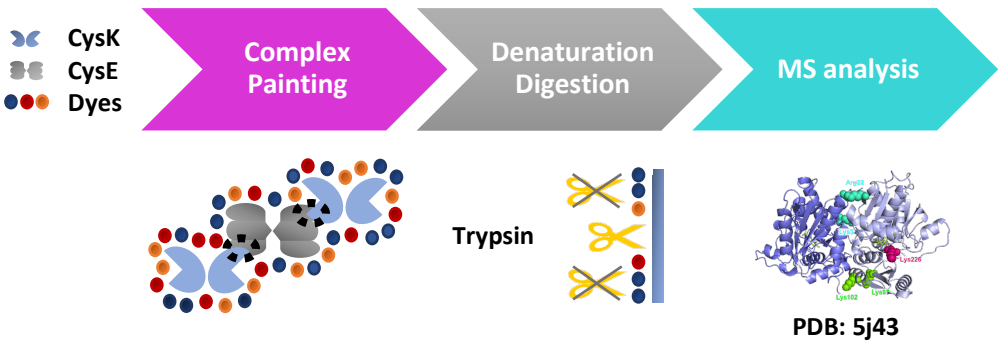


Figure 10. Protein painting workflow scheme.

A protein or a preformed protein complex is pulsed, under native conditions, with molecular dyes that bind non-covalently to solvent-exposed surfaces, without gaining access to protein-protein contact regions. The paints remain bound during the following steps of protein treatment: denaturation, reduction, alkylation, prior

to digestion with trypsin, a protease that preferentially cleaves at Arg and Lys residues. Thus, bound paints prevent tryptic digestion of residues exposed to solvent in the native protein conformation, allowing mass spectrometric (MS) detection and subsequent identification of peptides containing Arg or Lys located only at protein-protein interfaces or buried areas, selectively.

Solvent accessible surfaces are identified in unpainted samples and become coated by molecular dyes in painted proteins, hampering peptides located in these areas to be identified by MS. Comparing a protein alone and a protein complex both painted, residues that are engaged in PPIs upon complex formation, become inaccessible to molecular dyes, suggesting their involvement in PPI hotspots or areas that are buried following protein-protein assembly [84].

Protein painting was applied to protein-protein complexes formed by CysK with different binding partners. First, the technique was tested on the assembly of CysK with the toxin CdiA-CT, for which the crystal structure has been recently solved [79]. Moreover, CysK/CdiA-CT association shares the same structural features adopted by CysE within CS complex. Once validated, the same protein painting workflow was applied to CysK/CysE complex, whose three dimensional structure has not been determined yet.

3.1 Structural analysis of CysK/CdiA-CT complex from *E. coli*

Analysis of protein-protein interfaces when CysK associates with the toxin CdiA-CT was made with the database PDBSum (http://www.ebi.ac.uk/pdbsum/ [87]). Residues within complex interaction surface are 30 for CysK and 21 for CdiA-CT, whereas the interface area extends for 1229 $Å^2$ for CysK and 1281 $Å^2$ for the toxin, respectively. Moreover, interactions at binding interface include 14

hydrogen bonds and 116 non-bonded contacts. Overall results are reported in Figure 11.

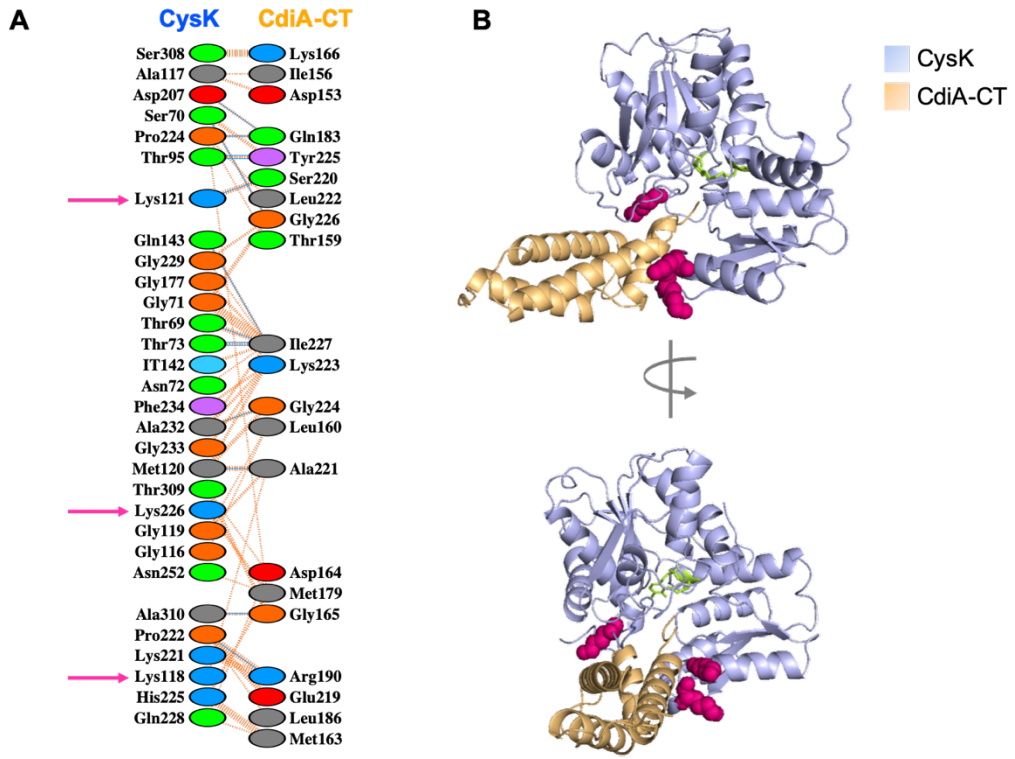


Figure 11. Structural analysis of CysK/CdiA-CT complex from *E. coli*. Adapted from [13]. **Panel A.** Residue interactions across CysK and CdiA-CT interface within the complex. The calculation was made with PDBSum database (http://www.ebi.ac.uk/pdbsum/), with the entry 5j43. CysK interface is represented on the left side, whereas CdiA-CT interface is represented on the right side. Residues are coloured by type: Positive (H,K,R); negative (D,E); neutral (S,T,N,Q); aliphatic (A,V,L,I,M); aromatic (F,Y,W); Pro and Gly (P,G). Hydrogen bonds are illustrated in blue and non-bonded contacts are indicated with an orange dashed line. Lys residues on CysK interface are pointed out by a magenta arrow. **Panel B.** CysK/CdiA-CT complex (PDB code: 5j43). One monomer of CysK (light-blue) interacting with one monomer of CdiA-CT (light-orange) are depicted. Lys118, Lys121 and Lys226 (theoretical tryptic cleavage sites at CysK/CdiA-CT interface) are shown in magenta, spheres mode. Pyridoxal 5'-phosphate (PLP) is shown in yellow stick mode.

Concerning CysK partner of the complex, 3 residues that could be cleaved by trypsin make contact with the toxin interface (indicated by a magenta arrow in Figure 11): Lys118, Lys121 and Lys226.

3.2 Validation of protein painting on CysK/CdiA-CT complex

First, the protein painting assay was tested on CysK assembly with the toxin CdiA-CT, to validate the workflow [42,79]. Indeed, the characterization of hotspots between CysK and CdiA-CT interaction surfaces could be confirmed by the analysis of the known crystal structure (see chapter 3.1).

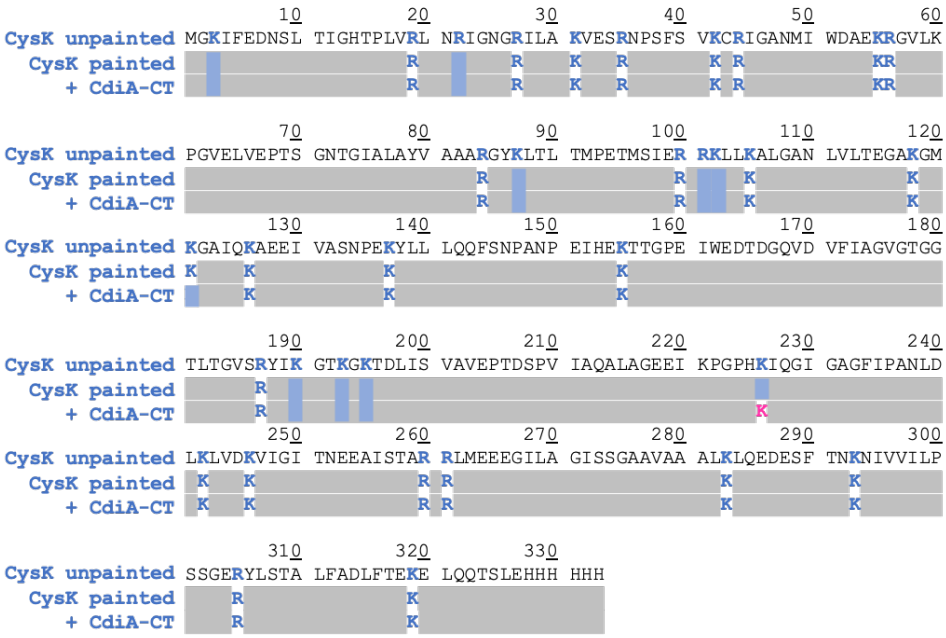
Application of the tryptic digestion workflow coupled to MS detection on unpainted CysK, allowed to identify all 34 available trypsin cleavage sites on CysK sequence (first line in Scheme 1 and 2).

CysK alone and CysK/CdiA-CT complex were pulsed with a set of 3 paint molecules, composed by: RBB, AO50 and CR (see chapter 2.6, "Protein painting for Mass Spectrometry", for chemical names and details). These dyes were tested by Luchini and co-workers on carbonic anhydrase and were selected out of a larger set of candidate molecules, due to their high association rate and low dissociation rate [84].

The results from the experiment on CysK partner of the complex are illustrated on Scheme 1: comparing the identified cleavage sites for the protein alone or complexed with the toxin (CdiA-CT), two residues behave differently.

Particularly, in CysK alone Lys226 is covered by dyes, thus accessible to the solvent, and the corresponding peptide is not identified in the painted protein. On the contrary, this residue is cleaved by trypsin in CysK/CdiA-CT complex (raw data shown in Figure 12A), proving it as a PPI hotspot of the assembly, as confirmed by the structural analysis (chapter 3.1).

Other three residues, Lys118, Lys121 and Lys221, were predicted to be part of the interaction surface observing the crystal structure (Figure 11). Lys221 couldn't be taken into account, since a proline residue at its C-terminus blocks the cleavage by trypsin.



Scheme 1. Trypsin cleavage sites on CysK sequence. Trypsin cleavage sites are shown, identified on CysK alone both unpainted (first line, residues cleaved are in bold blue, white background), and painted (second line, residues colour code is identical to line 1) and on CysK complexed with CdiA-CT (third line, see residues colour code in line 1, except for K226, highlighted in bold magenta on a white background). Dyes are presented as blue rectangles. Scheme modified from [13].

Considering the other two residues, while Lys118 is cleaved in CysK alone and after its association with the toxin, Lys121 is cleaved in CysK alone, but becomes covered by dyes after complex formation. Indeed, a subdomain of CysK that contains Lys121 residue moves to a different conformation, following substrate binding (see chapter 2.2 of the Introduction, page 13) [33]. This rearrangement could be responsible for Lys121 exposure to the solvent, hence to molecular paints, only following CysK interaction with CdiA-CT.

After this first series of experiments, another set of four molecular paints, composed by 8-Anilino-1-naphthalenesulfonic acid ammonium salt (ANSA), Thioflavine T, Acid Fuchsin, and Eosin B (see chapter 2.6 of "Materials and

methods") was tested on CysK. The exploitation of other dyes, belonging to different chemical classes, was explored with the aim to increase the number of "coloured" sites of CysK partner of the complex, hence possibly gaining more information about other protein-protein interface residues. Unfortunately, the sequence coverage of tryptic residues wasn't increased accordingly. Testing other chemical dyes, to cover more residues within protein assembly interfaces, and exploiting other proteases, to increase the resolution, through overlapping coverage in hotspot regions, are possible future perspectives of this project.

The results provided from the protein painting set-up that was used in this PhD work are in agreement with CysK/CdiA-CT crystal structure [79] information.

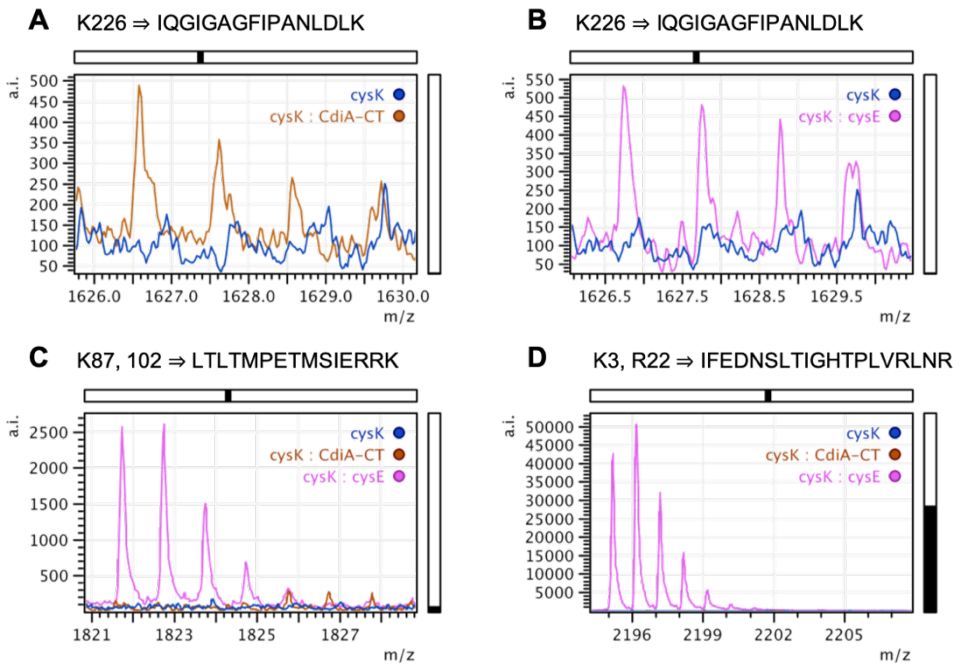


Figure 12. Mass spectra (raw data) of peptides differentially identified in CysK alone or in the complex with either CdiA-CT toxin or CysE. Panel A. Peptide 227-242 from CysK in the complex with toxin and with an observed [M+H]+ of 1626.61 m/z. **Panel B.** Peptide 227-242 from CysK in the CS complex with an observed [M+H]+ of 1626.75 m/z. **Panel C.** Peptide 88-102 from CysK in the CS complex with an observed [M+H]+ of 1821.78 m/z. **Panel D.** Peptide 4-22 from CysK in the CS complex with an observed [M+H]+ of 2195.22 m/z. Figure adjusted from [13].

3.3 Application of protein painting technique to CS complex

Since protein painting allowed to identify residues at the protein-protein interface between CysK and CdiA-CT consistent with the available crystallographic information, its application to CS complex could be useful, considering that the three dimensional structure of the assembly is unavailable. Moreover, CysK association with both protein partners (the toxin and CysE) is similar, with respect to the mechanism of complex formation, encompassing the insertion of CysE or CdiA-CT C-terminal tail into CysK active site [20,42,80].

To study PPI hotspots of CS complex, first the association of CysK with CysE was proved by size-exclusion chromatography (SEC) and fluorescence spectroscopy assays, as illustrated in Figure 13.

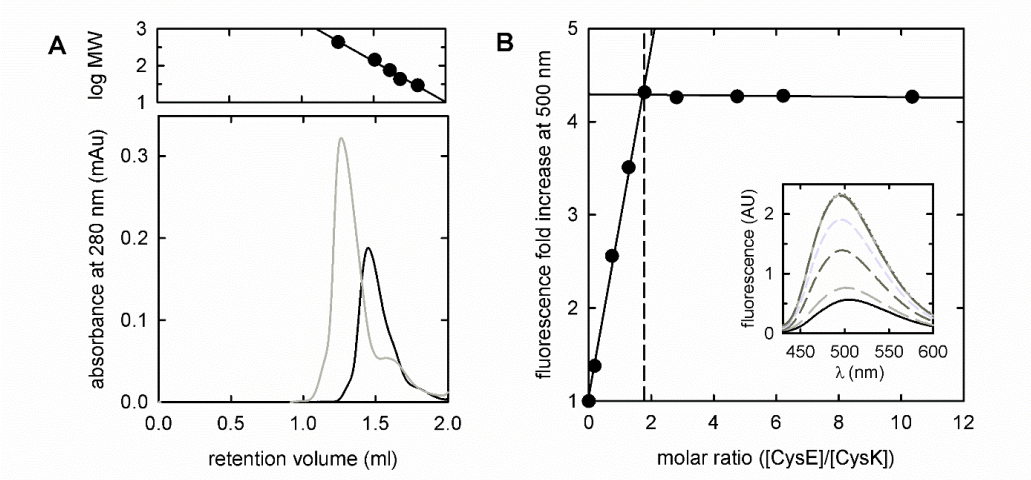
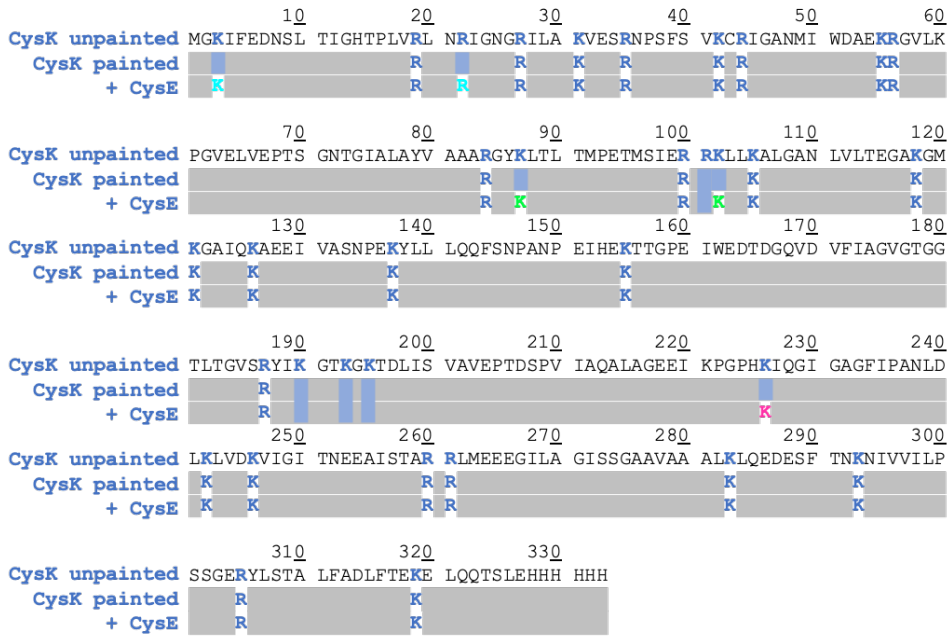


Figure 13. Panel A. Size exclusion chromatography. The black chromatogram indicates 39 μM CysE and the grey chromatogram a mixture of 39 μM CysE and 26 μM CysK (1.5 molar ratio). CysE elutes at 1.44 ml, that coincides with a 181 kDa molecular mass, whereas CysK elutes at 1.61 ml, corresponding to a mass of 77 kDa. **Panel B. Binding stoichiometry calculation for CS complex by fluorescence emission spectroscopy.** The stoichiometric ratio was found to be 1.7. Inset: fluorescence emission spectra upon excitation at 412 nm of CysK cofactor, PLP, in the presence of increasing concentrations of CysE. Adapted from [13].

Indeed, fluorescence emission spectroscopy can be used to study CysE association with CysK, exploiting the fluorescence properties of pyridoxal 5'-phosphate (PLP) cofactor, which are responsive to microenvironmental changes near the active site (Inset of panel B, Figure 13) [49,80]. Stoichiometric titration of CysK with CysE (Figure 13, panel B), results in a stoichiometric ratio of 1.7, indicating that two dimers of CysK associate with a hexamer of CysE, as proposed by previous studies [19,49,51] and confirmed by SEC results (Figure 13, panel A). Concerning SEC analysis, a 1.5 molar excess of CysE, with a predicted molecular mass of 174 kDa for the hexameric form, over CysK, with predicted molecular mass of 71 kDa for the homodimer, should form a CS complex, eluting at a theoretical molecular weight of 314 kDa. This is not in agreement with the MW of 470 kDa, calculated by SEC analysis, that is substantially larger. However, this result is in line with previous observations [45], and this discrepancy could be ascribed to the elongated shape of CS complex, as already predicted by molecular docking [51,52] and binding experiments [48,49,64].

As previously introduced, protein painting assay with the optimized set-up was extended to CS complex and the results are depicted in Scheme 2 (tryptic cleavage sites on CysK), Figure 14 (differential residues on CysK structure) and Figure 12 (MS spectra, panel B-D). Making a comparison between painted CysK alone and in association with CysE, 5 residues that were covered by molecular dyes, thus exposed to the solvent, became buried in the complex, and the respective peptides could be identified by the following MS analysis upon complex formation.

Among these, Lys226 (highlighted in magenta in Scheme 2, Figure 14 and raw data illustrated in Figure 12B) was again spotted as a differently covered residue, assessing the similar interaction pattern between CysK and its two 2 partners, as previously reported in literature [20,42], confirming the initial hypothesis of this work.



Scheme 2. Trypsin cleavage sites on CysK sequence. Trypsin cleavage sites are shown, identified on CysK alone both unpainted (first line, residues cleaved are in bold blue, white background), and painted (second line, residues colour code is identical to line 1) and on CysK within CS complex (third line, residues identified also on CysK alone are colour coded as in line 1 and 2, the 5 differential residues are coloured in cyan, green and magenta, based on their location on CysK structure, as in Figure 14). Dyes are presented as blue rectangles. The scheme is adapted from [13].

Moreover, this residue is located on a conserved loop, β8A-β9A, nearby CysK active site, that contains crucial residues for complex formation [63]. Indeed, it has been demonstrated that mutation of K217, H221 and K222 in *A. thaliana* CysK disrupts its assembly with CysE. These residues correspond respectively to K221, H225 and K226 in *E. coli*, that were analysed in this work.

Other two residues, Lys87 and Lys102, were cleaved by trypsin only following CysK association with CysE (see MS spectrum in Figure 12C, tryptic cleavage site and residues on CysK structure coloured in green in both Scheme 2 and Figure 14). These residues are not located in proximity to CysK active site on the

crystal structure (see Figure 14), however they belong to a subdomain of the N-terminal domain, involved in a large conformational change that follows substrate binding, as it has been described by Burkhard and co-workers for *S.* Typhimurium CysK [33] (see chapter 2.2 of the Introduction, page 13). This rearrangement allows the protein to adopt a closed conformation, bringing the region that encompasses residues 87-131 in CysK sequence, closer to the active site. Based on these results, CysE binding seems to stabilize the closed state of CysK. On the other side, these two residues were exposed to solvent and thus covered by dyes in both CysK alone and CysK interacting with CdiA-CT. Interestingly, these findings support a model for the formation of CysK/CdiA-CT and CS complexes, that has been hypothesized by our group, based on previous functional data [80] (see Figure 9, page 27). Finally, Lys3 and Arg22 (identified peptide in Figure 12D, residues coloured in cyan in Scheme 2 and Figure 14), despite presenting the same behaviour as the other 3 residues, i.e. identification of peptides only following CysK association to CysE, are not involved as PPI hotspots of complex formation. Indeed, these residues are located nearby the interface between the two monomers of CysK. These results are consistent with a compaction/stabilization of the dimeric interface, suggesting an allosteric communication between the two monomers of CysK, that follows CysE binding. These considerations are corroborated by functional studies, where occupation of one CysK active site by the C-terminus of CysE seems to induce the closure of the other unoccupied active site of CysK dimer [44,45].

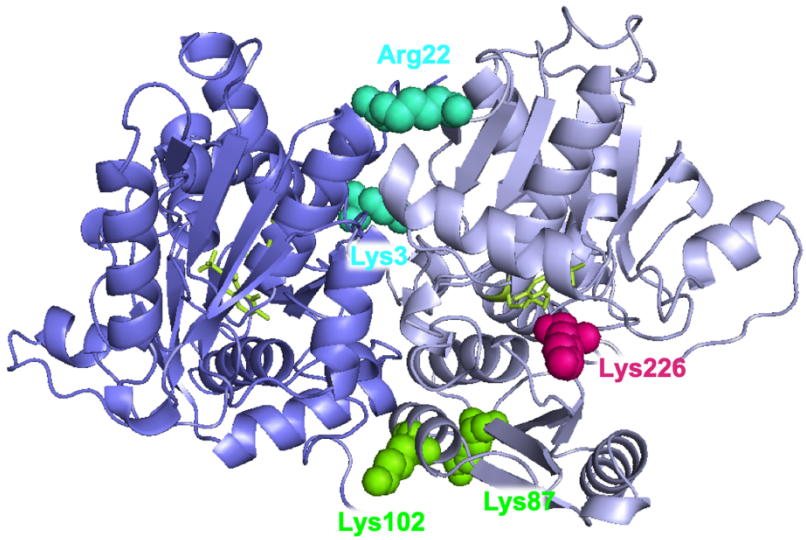


Figure 14. Trypsin cleavage sites on CS complex. Residues cleaved by trypsin following CS complex formation are presented in sphere mode on one monomer (light blue) of CysK three dimensional structure (PDB code: 5j43). The colour code is different, according to residue position. Lys226, coloured in magenta, is located at the active site entrance. Lys102 and Lys87, coloured in green, are part of the moveable domain. Lys3 and Arg22, presented in cyan, are part of the interdimeric interface of CysK. CysK monomers are illustrated in blue and light blue. The toxin monomers were removed. The PLP is shown in yellow stick mode. Modified from [13].

A MODEL FOR THE SOLUTION STRUCTURE OF CYSTEINE SYNTHASE COMPLEX

1. Introduction

Concerning the stoichiometry of CS complex [49,51], it has been demonstrated that one hexamer of CysE binds two dimers of CysK. The quaternary geometry of the assembly could hence correspond to two possible scenarios: both CysK active sites interacting with CysE, or only one active site engaging CysE C-terminus in binding, exposing the other monomer to the solvent (see Figure 6, page 20).

In order to create a low resolution model for the solution structure of CS complex, Small Angle X-ray Scattering (SAXS) data were collected.

SAXS probes the structure and interactions of biomolecules and their complexes in solution, at 1-2 nm resolution. This is achieved recording the X-rays scattering behaviour or intensity as a function of the scattering angle [88].

Parameters derived from SAXS analysis, through the Guinier formula, include the radius of gyration (R_g) and the forward scattering intensity $I(0)$, that is related also to the molecular weight (MW). From the scattering profile, the distribution function $p(r)$ of intramolecular atomic distances can be obtained, where D_{max} is the maximum intramolecular distance [89].

Beyond the determination of key parameters, algorithms are used to reconstruct low resolution 3-D electron density maps of the molecule from the 1-D scattering profiles and to obtain structural information. This procedure, where usually bead models are used to fit SAXS data, leads to the ab initio determination of the molecular shape [90]. To prove the fitting robustness between different bead model solutions, the comparison and averaging of the results from different reconstruction runs is normally done [91].

Eventually, quaternary structure determination of molecular complexes, as is the case of CS assembly discussed in this PhD work, can be simplified if high-resolution structures of the single components or prior data obtained from different techniques (like NMR [92], footprinting and mass spectrometry [93] or

fluorescence) are available and could be integrated. Specific tools are moreover used to perform global rigid body modelling of macromolecular complexes against SAXS data [94].

Concerning CS complex, SAXS results were obtained in close connection with protein painting approach, using spotted PPI hotspots of CS complex to model CysK and CysE structures into the SAXS envelope. SAXS analysis was performed by our collaborators at Elettra Synchrotron in Trieste. Main results of this work are reported in the paper "Combination of SAXS and Protein Painting Discloses the Three-Dimensional Organization of the Bacterial Cysteine Synthase Complex, a Potential Target for Enhancers of Antibiotic Action", that was published in *International Journal of Molecular Sciences* [13].

2. Materials and methods

Proteins expression and purification, enzymes activity assays, spectroscopy and size exclusion chromatography are described in the chapter “Cysteine synthase protein-protein interaction hotspots revealed by protein painting”, part 2.2-2.5.

2.1 Materials

Unless otherwise stated, reagents were purchased from Sigma Aldrich (St. Louis, MO, USA) and were of the highest grade commercially available. Materials were used as received.

2.2 Sample preparation for SAXS measurements

CysK, CysE, and CS complex were prepared in a stoichiometric ratio. Afterwards, proteins were re-purified and buffer exchanged in 20 mM sodium phosphate, 85 mM NaCl, 2 mM EDTA, 10 mM 2-mercaptoethanol (2-MCE), pH 7.5, using a Superdex 200 10/300 GL (GE-Healthcare, Chicago, IL, US). After the chromatographic run, proteins were dialyzed against their storage buffer. Storage buffer was also used to record the SAX-scattering baseline. Proteins were concentrated by ultrafiltration and their final concentration was determined by Bradford assay. SAXS measurements were performed using freshly diluted protein samples at different concentrations: 1.5, 1.0, and 0.5 mg/ml. Data sets of samples that showed indication of aggregation were merged with the corresponding data collected at lower concentration.

2.3 Small Angle X-ray Scattering

SAXS data of all samples were collected at the Austrian beamline at the Elettra Synchrotron (Trieste, Italy) using a Pilatus3 1M detector system at a sample-

detector distance of 1.232 m and at a wavelength λ = 0.154 nm. Measurements were carried out at 10 °C in capillaries of 1.5 mm outer diameter/0.01 mm wall thickness made from borosilicate (Hilgenberg, Maisfeld, Germany), enclosed within a custom-made thermostatic compartment connected to an external circulation bath and a thermal probe for temperature control. Raw data were radially averaged and calibrated to absolute units (cm^{-1}) by using a freshly prepared 5 mg/ml BSA solution in 50 mM Hepes, pH 7.5. The scattering curves have been normalized to the primary beam intensity, corrected for sample transmission, and normalized to absolute scattering units, using IGOR Pro (Wavemetrics, Lake Oswego, OR, USA). Each set of scattering patterns was carefully checked and the average after a positive control over radiation damage was performed. Radiation damage was not observed on samples presented in this study. GIFT [95] was used to test for residual constant background. The pair distance distribution function (PDDF) was calculated with GNOM [89], which was also used to determine the radius of gyration and maximum dimension of all protein structures.

Bead ab initio modelling was conducted using DAMMIF [96], from the ATSAS package. For each run 10 ab initio models were generated and subsequently analysed and averaged using DAMCLUST and DAMAVER [91] from ATSAS package.

The CysK and CysE crystallographic model structures were superimposed to the ab initio bead models using SUPCOMB [97]. The multi-subunit CS complex was manually fitted in the de novo bead envelope using one hexameric CysE model (PDB code: 1t3d) and two CysK dimers (PDB code: 1d6s) in SASpy [98]. The manual pre-alignment was the starting point for the automatic rigid body refinement with SASREF [94], where a number of putative contact points were imposed as specified in Results and Discussion (chapter 3), based on experimental data retrieved by the protein painting approach (see chapter

“Cysteine synthase protein-protein interaction hotspots revealed by protein painting”).

3. Results and discussion

Since CS complex crystal structure is not available, SAXS measurements were employed to better describe the quaternary organization that CysK adopts in assembly with CysE. Indeed, two possible scenarios could take place for CS complex, according to the determined stoichiometry [49,51]. In the first model, both CysK active sites are bound to distinct CysE C-terminal tails whereas, in the other scenario, only one CysK monomer is engaged in complex formation, exposing the other active site to the solvent (see Figure 6, page 20).

3.1 Validation of CysK and CysE structures with SAXS

As a first step, the structures of both CysK and CysE alone were analysed by SAXS. Small-angle X-ray scattering data for CysK, CysE and CS complex are depicted in Figure 15.

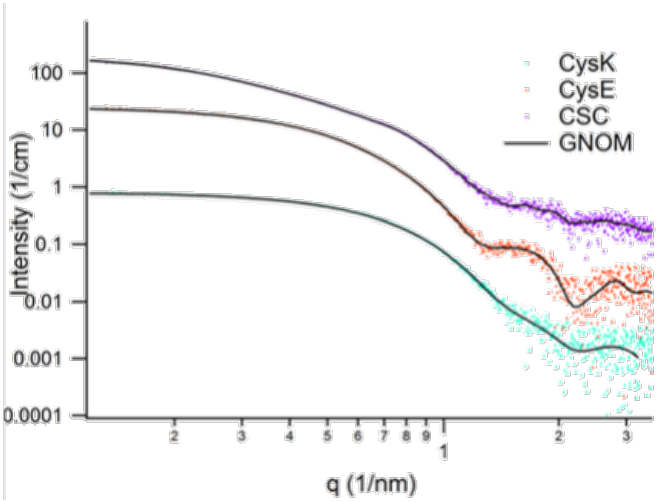


Figure 15. SAXS data for CysK, CysE and CS complex. Merged solutions scattering data of CysK (cyan dots), CysE (orange dots) and CS complex (indicated as CSC, purple dots), rescaled to 1 mg/ml and vertically shifted to improve the representation with the GNOM software fit (black lines). Adapted from [13].

The pair distance distribution function (PDDF) data, related to SAXS measurements of CysK, CysE and CS complex, are instead illustrated in Figure 16.

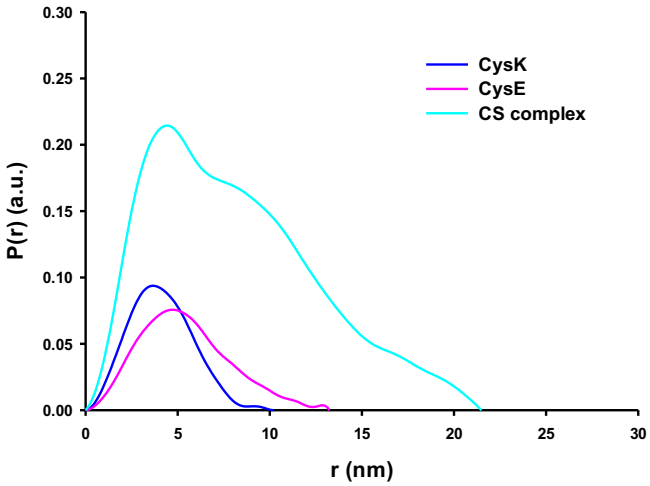


Figure 16. Pair distance distribution function (PDDF) data of GNOM. CysK (blue), CysE (pink) and CS complex (cyan). The data were rescaled for sake of clarity. Modified from [13].

Table 2 reports the comparison between the derived structural data (R_g, that is the Guinier radius, and D_{max}, the maximum dimension) and the data obtained for the crystallographic models using the program CRYSOL [99].

DAMMIF modelling results for CysK and CysE are illustrated in Figure 17 and 18. CysK ab initio DAMMIF model was calculated by averaging and filtering. SAXS bead model fits CysK dimeric structure, in both the open conformation (PDB code: 1oas, illustrated in Figure 17) and the closed conformation (PDB code: 1d6s, not shown) [100].

Table 2. Summary of the derived SAXS data. Adapted from [13].

Parameter	CysK	CysE	CS complex
GNOM			
Max. dimension D_{max} (nm)	8.5	13.0	22.0
Guinier radius R_g (nm)	2.623	3.856	6.606
q-range (nm^{-1})	0.1-3.5	0.1-3.5	0.1-3.5
CRYSOL			
PDB file	1oas	Hexamer built from 1t3d	
Guinier radius R_g (nm)	2.572	3.361	
Molecular weight (kDa)	67.57	171.6	
Max. dimension D_{max} (nm)	8.675	11.0	
DAMMIF			
MW estimate for proteins (kDa)	58.6	155	300
Phase radius of gyration (nm)	2.62	3.86	6.61
Maximum phase diameter (nm)	10.500	14.800	24.300

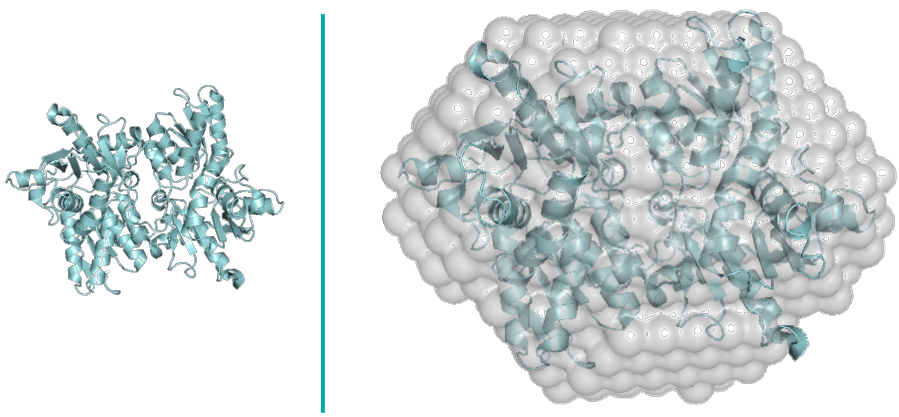


Figure 17. Crystal structure and DAMMIF modelling result for CysK. The ab initio DAMMIF model (filtered) of CysK in white spheres was overlaid to the crystallographic structure (PDB code: 1oas, presented in cyan shades, ribbon mode). Adjusted from [13].

CysE ab initio DAMMIF model well overlaps the hexameric form, obtained from the three dimensional structure (PDB code: 1t3d) [26,29], as depicted in Figure 18. Furthermore, CysE crystal structure doesn't show significant electron density corresponding to the 11 C-terminal residues. On the other side, SAXS model was collected from a protein construct where the C-terminal peptide was expressed. Indeed, observing SAXS bead model, an empty space at both CysE C-terminal regions is noticeable and suitable for the missing residues to be accommodated.

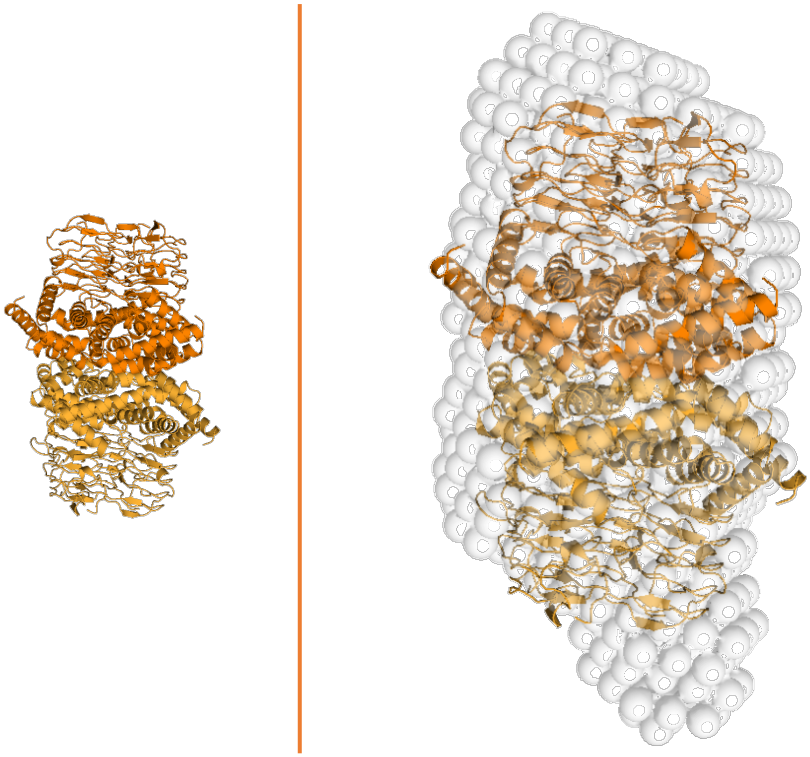


Figure 18. Crystal structure and DAMMIF modelling result for CysE. CysE ab initio DAMMIF model (filtered) in white spheres is overlaid with the hexameric crystal structure (PDB code: 1t3d, trimers are shown in ribbon mode, orange shades). Adjusted from [13].

CysK and CysE solution scattering behaviours are both in agreement with the respective solved three dimensional structures, validating them in order to apply SAXS measurement to the CS complex.

3.2 SAXS analysis of CS complex

Gel filtration and dialysis of the CS complex sample were performed, prior to collect SAXS measurements. The ab initio DAMMIF model together with the derived structural data (R_g = 6.606 ± 0.003 and D_{max} = 22.0 nm), obtained after reduction and rescaling to 1 mg/ml, present an elongated S-shape envelope. This model fits better with the scenario illustrated in Fig. 6A (see page 20), where only one CysK subunit is engaged in binding with CysE hexamer. A simulation of CS assembly with both CysK active sites bound to each trimer of CysE (Fig. 6B, page 20) wasn't in agreement with the experimental data reported in Figure 16, since it yielded a smaller value of both R_g and D_{max} (18.5 nm). Moreover, this model involves a high surface of CysK interacting with CysE within complex formation, that is not supported by CS PPI hotspots number and position revealed by protein painting approach (see part 3.3 of the chapter "Cysteine synthase protein-protein interaction hotspots revealed by protein painting"). It has furthermore been demonstrated by functional studies [44,45] that CysK retains about 10 % of its activity upon complex formation, even when concentrations of CysE are saturating.

As a starting point, a preliminary CS model created from one hexamer of CysE (PDB code: 1t3d) and two dimers of CysK (PDB code: 1oas) was manually overlaid to the ab initio bead envelope, by the use of SASpy.

Afterwards, protein painting interaction hotspots, in particular Lys226 on CysK and Arg242 on CysE sequences, were used as constrains to model the quaternary CS assembly inside the SAXS envelope. Arg242 on CysE was chosen since it

belongs to a conserved part of the sequence (PARIV), that is solvent exposed and thus is reasonably supposed to interact with CysK.

Automatic rigid body refinement with SASFER in the q-range between 0.15 and 3.9 nm^{-1} was performed. Three different constrained distances between the selected residues were tested: 1.0, 1.5 and 2.0 nm.

Three independent solutions for each distance were ran. A satisfactory fit was obtained for all the three independent solutions (chi^2 values of 1.22, 1.23 and 1.24) for the distance set to 1.5 nm, whereas 1.0 nm (chi^2 value around 1.99) and 2.0 nm distance (constrain too loose, chi^2 value obtained of 0.85, 1.81 and 6.84) didn't yield a good fit. Figure 19 illustrates the results setting the constrain distance to 1.5 nm: the three models nicely fit the ab initio model and well overlay the SAXS pattern and the q-space data regularized with GNOM (data not shown).

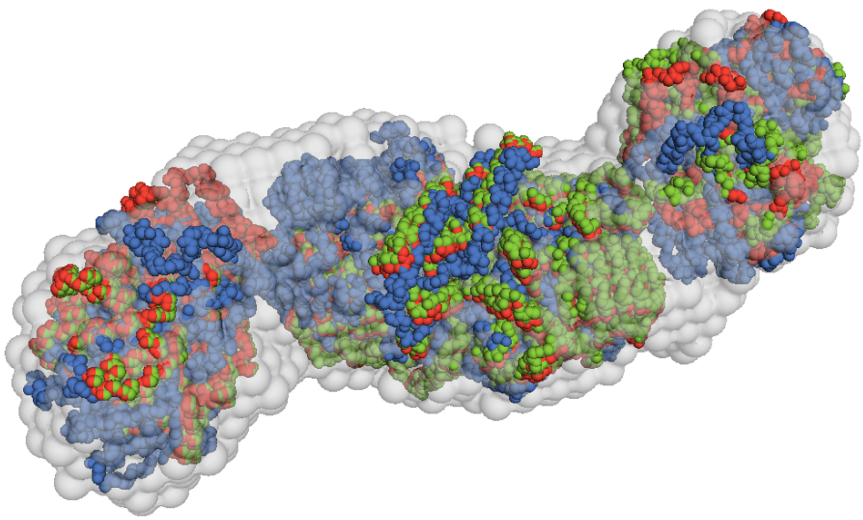


Figure 19. Superposition of ab initio model of CS assembly with 3 independent rigid-body refined models performed with SASREF. The constrained distance was imposed to 1.5 nm. Real-space ab initio model (grey surface, transparency set to 60%), overlaid with the three models of the different solutions (the models are presented in backbone traces, coloured in red, blue and green respectively). Adapted from [13].

These experimental evidences, together with the previous observations we made (see page 67), lead to the following considerations:

- the exclusion of the model with both CysK subunits interacting with CysE trimers;
- the hypothesis that the binding of one subunit of CysK to CysE C-terminal tail is transmitted through CysK dimeric interface to the other subunit, that undergoes an allosterically-driven inhibition for 80% of its activity. The existence of an allosteric communication between the two monomers of CysK, following CysE binding, is moreover corroborated by protein painting data (see part 3.3 of the chapter "Cysteine synthase protein-protein interaction hotspots revealed by protein painting");
- the S-shape of the complex also suggests that the two binding surfaces between CysE and CysK are correlated. Indeed, engagement of one CysK subunit to one CysE trimer seems to be allosterically communicated to the other trimer within the hexamer, hence dictating the orientation of the binding of the second subunit of CysK;
- the molecular weight of CS assembly derived from SAXS experiments (300 kDa, reported in Table 2) well agrees with the theoretical one [101].

Focusing on CysK active site interacting with C-terminal part of CysE, a cleft that is still accessible to the solvent can be noticed in the built model (Figure 20). This effect could be ascribed to the lack of 11 residues in the C-terminal region of CysE crystal structure [29] (see chapter 3.1, page 66).

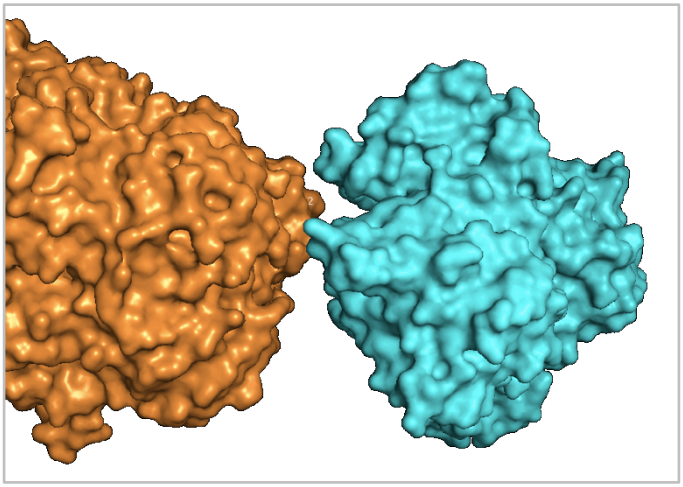


Figure 20. Close-up of CysK active site. CysK (cyan, surface mode), interacting with CysE (coloured in orange, surface mode) within the CS complex model obtained, that is shown in Figure 19. To be noticed the presence of a solvent exposed cleft at CysK active site entrance. Adjusted from [13].

A comparison with the three dimensional structure of CysK complexed with the toxin CdiA-CT (PDB code: 5j43) [79] could suggest the potential orientation of this C-terminal peptide (INHTFEYGDGI). Indeed, when CysK structure assembled with the toxin was overlapped on CysK structure in the model of the CS complex, the C-terminal residues of CdiA-CT (KIESALKGYGI) well fit into the solvent accessible pocket, previously unoccupied, as depicted by Figure 21. Together these observations agree and further confirm the model of the complex obtained from SAXS measurements.

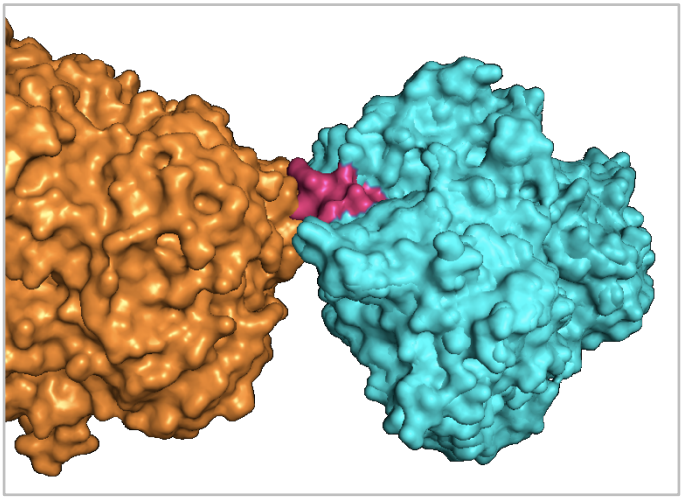


Figure 21. Close-up of CysK active site with overlapped CdiA-CT C-terminal peptide. CysK (cyan, surface mode), interacting with CysE (coloured in orange, surface mode) within the CS complex model obtained, that is shown in Figure 19. Superposition of this model with the complex of CysK and CdiA-CT (PDB code: 5j43) helped to visualize the position of the C-terminal peptide of the toxin. The cavity, present in Figure 20, is occupied by the toxin C-terminal peptide (shown in pink, surface mode). Adjusted from [13].

PROBING DYNAMIC PROPERTIES OF CYSTEINE SYNTHASE COMPLEX BY HYDROGEN-DEUTERIUM EXCHANGE MASS SPECTROMETRY

1. Introduction

As already mentioned, an extensive biophysical investigation of CysK interaction with CysE has been done over the years. However, combining functional data with information about proteins dynamics during and upon the binding process could deepen the understanding on the role and function of CS assembly.

Hydrogen/deuterium exchange mass spectrometry (HDX-MS) provides a valuable approach to investigate protein structural aspects and dynamics in solution [102]. This technique has indeed a wide range of applications, ranging from protein folding/unfolding studies [103], dynamics of complex molecular assemblies [104] and binding experiments, including protein-protein interactions [105]. Investigation on binding of protein:polypeptide, protein:small molecule, protein:lipid or protein:nucleic acid are moreover reported in literature [102,106].

By this technique, the hydrogen to deuterium exchange of backbone amides is detected by mass spectrometry. Amide HDX is sensitive to the presence and stability of local hydrogen bonds and solvent accessibility, hence is related to protein structure and conformational dynamics [107–109]. Moreover, HDX doesn't involve alteration of protein covalent structure in order to label it and it's thus a non-perturbing technique. Traditionally, HDX method has been coupled with NMR analysis [110], but nowadays electrospray ionization (ESI)-MS is the most widely used acquisition method.

Two kind of approaches are normally used to perform HDX studies, namely continuous and pulsed labeling [111,112]. Continuous labeling strategy monitors the deuterium incorporation of a native protein as a function of the exposure time. In pulsed HDX experiments, a perturbing agent induces a conformational change of a population of protein molecules that are then exposed to deuterium pulse for a short period of time, allowing the detection of folding intermediates. Following the labeling reaction, a global exchange information can be obtained, with mass

analysis at intact protein level. An increase of spatial resolution provides instead HDX behaviour of short fragments of the peptide backbone [113].

The continuous labeling HDX-MS workflow [109] was applied to study our system in this PhD work. Briefly, the protein of interest is diluted in a deuterated buffer and incubated for different time periods. The deuterium incorporation is then stopped by reducing pH to 2.5 and temperature to 0 °C, that corresponds to the quenching step. At this stage, the back-exchange (BE) of deuterium with hydrogen has to be controlled, to avoid label losses. Minimization of BE is achieved with short analysis time and maintaining the temperature at 0 °C. Quenched samples undergo proteolytic digestion with acidic proteases, usually pepsin. The resulting peptides are desalted and separated by reversed-phase liquid chromatography, then analysed by high-resolution MS. The following relative HDX level of individual segments of the protein at different labeling time points provides localized exchange information [114].

In this work, HDX-MS was applied to detect regions of CysK and CysE that undergo dynamic movements and changes in solvent accessibility upon CS complex formation.

This work was done during a secondment of 6 months at Copenhagen University, Department of Pharmacy, Protein Analysis Group, under Prof. Kasper Dyrberg Rand supervision.

2. Materials and methods

Proteins expression and purification, enzymes activity assays, spectroscopy and size exclusion chromatography are described in the chapter “Cysteine synthase protein-protein interaction hotspots revealed by protein painting”, part 2.2-2.5.

2.1 Materials

All reagents were purchased from Sigma Aldrich (MO, USA) in analytical grade except the following: immobilized pepsin beads (Thermo Scientific, MA, USA), acetonitrile (Biosolve, Dieuze, France).

2.2 Sequence coverage optimization

First, a list of peptides was created for both CysK and CysE proteins.

Four different quenching buffers were tested to assess the best sequence coverage, numbers of peptides identified and redundancy for each protein.

Four replicates of 6 μM CysK and CysE in 5 μl of equilibration solution (PBS, pH 7.45) were prepared. The samples were diluted 1:9 with equilibration solution and quenched 1:1 into ice-cold buffer.

Four different quenching solutions were tested:

- 300 mM phosphate buffer, pH 2.3;
- 300 mM phosphate buffer, pH 2.3, 6 M guanidinium chloride;
- 300 mM phosphate buffer, pH 2.3, 6 M urea;
- 2 M glycine buffer, pH 2.3.

Quenched samples were immediately frozen at –80 °C and stored until measurements by liquid chromatography mass spectrometry (LC-MS analysis).

2.3 Hydrogen/Deuterium Exchange (HDX)

Prior to HDX labeling, CysK and CysE proteins were diluted to 5 μM in 5 μl equilibration solution (PBS, pH 7.45). For CS complex formation, proteins were diluted to 5 μM CysK and 7.5 μM CysE in 5 μl total volume of equilibration buffer and incubated 30 min at room temperature to allow proper complex formation. The relative concentration of the proteins in the complex was chosen according to binding stoichiometry (1:1.5).

The HDX reactions were initiated by diluting the samples 1:9 with 99% deuterated PBS buffer at 25 °C. HDX reactions were incubated for specific time periods (0.03, 1.67, 16.67, 60 and 720 min). For the shorter deuterium labeling time (0.03 min), two people were required for sample preparation.

Shorter time points (0.03, 1.67 min) were prepared in triplicate, whereas the other time periods were performed in single replicate. Non deuterated samples were prepared in triplicate for CysK and 1 replicate for CysE and CS complex.

Maximum deuterated controls were prepared for the CS complex in triplicate. 5 μM CysK and 7.5 μM CysE in 5 μl total volume of equilibration buffer (PBS) were diluted 1:9 with the deuterated buffer (20 mM Tris, pH 7.4, 6 M guanidinium chloride) and incubated at 37 °C for 24 hours.

After incubation with deuterium, each sample was quenched by 1:1 dilution into ice-cold quench buffer (300 mM phosphate, pH 2.3). Quenched samples were immediately frozen at –80 °C and stored until measurements by liquid chromatography mass spectrometry (LC-MS analysis).

2.4 Liquid Chromatography and Mass Spectrometry

Frozen quenched samples were quickly thawed in a table top centrifuge and injected into a refrigerated (0 °C) UPLC system (nanoAcquity UPLC and HDX manager, Waters, Milford, MA).

The samples were subjected to online pepsin digestion at 20 °C through a home-packed column (2 cm in length, 2 mm i.d.) containing pepsin immobilized agarose beads (Thermo Scientific, 125 μM).

The resulting peptides were trapped on a trap column (VanGuard BEH C18 precolumn, 1.7 μm, 2.1 mm x 5 mm, Waters) and desalted for 3 min at a flow rate of 200 μl/min with solvent A (0.23% formic acid in MQ water, pH 2.5).

Subsequently, the peptides were separated over an analytical column (Acquity UPLC BEH C18, 1.7 μm, 1.0 x 100 mm, Waters) at a flow rate of 40 μl/min, with a 9 min linear gradient rising from 8 to 40% solvent B (0.23% formic acid in acetonitrile).

Following the chromatographic separation, the peptides were analysed using a hybrid ESI-Q-TOF mass spectrometer (Synapt G2-Si, Waters).

The MS was set in positive ionization mode and the peptides were further separated by ion mobility to enhance peak capacity. Glu-fibrinopeptide B was used to lock mass correction of all spectra.

For peptide identification, non-deuterated samples were injected and subjected to 2 different chromatographic separation conditions: a short gradient, identical to deuterated samples, and a long gradient (35 min linear gradient rising from 5 to 30% solvent B, at a flow rate of 40 μl/min). The chromatographic separation was performed using the same set-up as the deuterated samples.

MS/MS analyses were performed using a combination of data independent (MS^E) and data dependent (DDA) acquisition mode.

2.5 HDX-MS data analysis

MS/MS data were processed with ProteinLynx Global Server (PLGS) version 3.0 (Waters, Milford, MA) for peptide identification. Peptides identified by DDA had to have PLGS Ladder Score above 1.0, below 15 ppm mass error for the

precursor ion. Manual inspection of the fragment spectrum for each selected peptide was also done. The same parameters were applied to peptides, added from the DDA analysis with a long chromatographic gradient, that were added in some regions of the proteins sequence, to increase sequence coverage.

DynamX version 3.0 (Waters, Milford, MA) was used to filter peptides identified by data independent acquisition (DIA), based on the following parameters: minimum 2 product ions, minimum 0.2 product ions per amino acid, maximum 10 ppm mass error on precursor ion. Further, the peptides were filtered if identified in a minimum of three out of four consecutive MS^E runs.

Analysis of deuterium uptake of all identified peptides was manually verified.

Back-exchange (BE) for individual peptides was calculated using the average of the maximum deuterated controls, according to the following equation:

$$BE\ (\%) = \left(1 - \frac{m_{max} - m_{0\%}}{N\ D_{frac}}\right) \times 100\% \quad (1)$$

where m_{max} is the theoretical maximum deuterium uptake of the peptide (N-terminal residues and proline residues are excluded from the calculation), $m_{0\%}$ corresponds to the mass of the non-deuterated peptide, N is the number of amide hydrogens of the peptide and D_{frac} is the fraction of deuterium in the HDX labeling buffer. BE was calculated to validate HDX-MS system and to support the analysis of CysK and CysE peptides.

The plotted deuterium uptake profiles were not corrected for BE, but y axis was normalized to show the maximum deuterated control at 80 %.

To determine significant differences in HDX comparing two states, the standard deviation of time points performed at least in triplicate was used to calculate a confidence interval (CI). The standard deviation of individual peptides (n = 3)

was averaged and a pool standard deviation (SD_{pool}) was calculated, using the equation:

$$SD_{pool} = \sqrt{SD_A^2 + SD_B^2} \tag{2}$$

where SD_A and SD_B are the states being compared (CysK alone and CysK within the CS complex, or CysE alone and CysE within the CS complex).

The SD_{pool} was then used to calculate the 99% confidence interval (CI), using the equation:

$$CI = \pm \frac{(t_{n-1} \cdot SD_{pool})}{\sqrt{n}} \tag{3}$$

where t_{n-1} is the table value for the two-tail 99% CI with two degrees of freedom ($t_{99\%,\ n=3} = 9.925$) and n is the number of replicates ($n = 3$). In the comparative HDX analysis of different states, a peptide was considered to have a significant difference in HDX only if it showed a significant difference in deuterium incorporation, above the calculated CI.

3. Results and discussion

HDX-MS experiments were performed in order to explore CysK and CysE dynamic properties, following their assembly to form CS.

Proteolysis with pepsin allowed the identification of 79 CysK peptides, covering 98.5% of the protein sequence (Figure 22), and 72 peptides covering 97.8% of CysE sequence (Figure 23), that were used to measure HDX in local regions of the two enzymes.

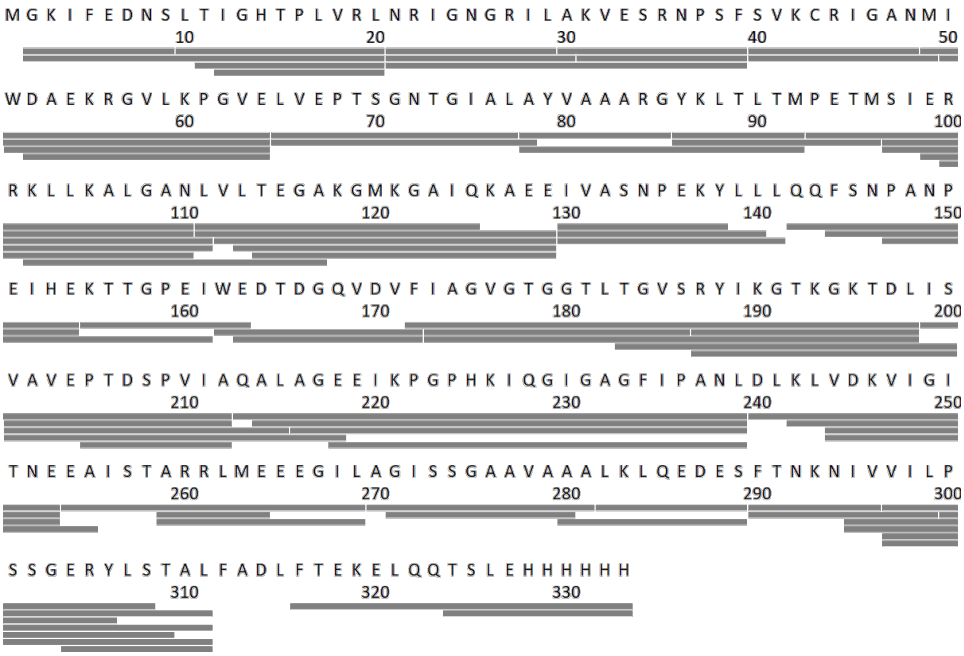


Figure 22. Sequence coverage for HDX-MS analysis of CysK. Digestion of the protein with immobilized pepsin yielded a total of 79 peptides. The peptides are depicted as grey bars and are aligned with the protein sequence. The coverage corresponds to 98.5%.

Since HDX-MS measurements were performed for the first time on this system, composed by the two proteins (CysK and CysE) alone and the CS complex, a continuous labeling approach was chosen. The experiment included deuterium

incorporation time points spanning from seconds to hours, respectively: 0.03, 1.67, 16.67, 60 and 720 min.

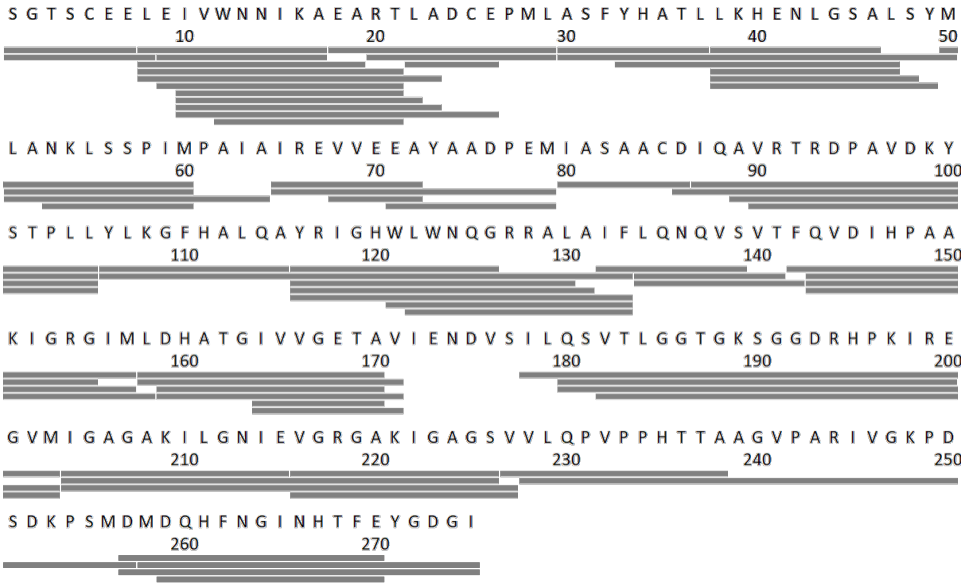


Figure 23. Sequence coverage for HDX-MS analysis of CysE. Online pepsin proteolysis identified a total of 72 peptides, covering 97.8 % of CysE sequence. The peptides, depicted as grey bars, are aligned with the protein sequence.

Although manually difficult to perform, the shorter time period (0.03 sec) was chosen since CysK and CysE bind with high affinity to assemble in CS complex (K_d about 5 nM for *E. coli*) [[80]]. Moreover, this labeling time was the start point of a comparison between a traditional approach and a fast HDX technique, performed on a microfluidic chip [upgraded from [115]]. These data are part of an on-going project and will not be reported in this PhD work.

Local HDX as a function of time was studied. Two different states, corresponding to CysK and CysE alone, were examined and compared with their states in the CS complex.

3.1 CysK dynamics upon CS complex formation

To investigate CysK dynamics upon CS complex formation, HDX-MS experiments of CysK alone and of the protein interacting with CysE were performed. Based on these results, multiple regions of CysK showed significant changes in HDX upon binding to CysE. Figure 24 and Figure 25 illustrate the overall comparison of the HDX between the two states of CysK protein, engaged or not in CS assembly. In Figure 24, the average relative deuterium uptake of CS complex is subtracted from the average value of CysK state for each peptide and time point. Differences in HDX (ΔHDX) in Da are plotted, according to individual identified CysK peptides.

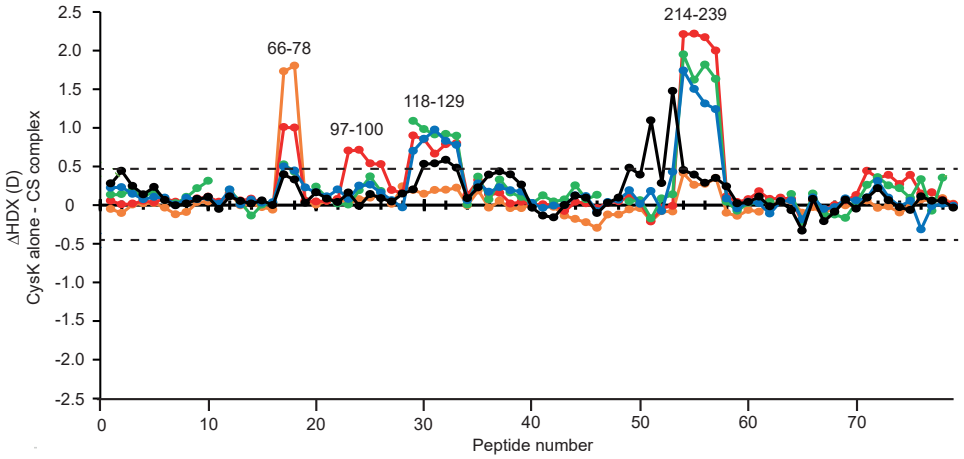


Figure 24. CysK average deuterium uptake difference plot. Difference plot of the average deuterium uptake (ΔHDX) between two states: CysK alone and CysK assembled with CysE (CS complex) for the 79 identified peptides at the five measured time points (orange–0.03 min; red–1.67 min; green–16.67 min; cyan–60 min; black–720 min). The peptides are arranged along the x-axis according their position in CysK sequence, from N-terminal to C-terminal. Positive and negative values along y-axis indicate reduced or increased HDX, respectively, following CysK binding to CysE. Values at 0.03 and 1.67 min correspond to the mean of three replicates, whereas values at 16.67, 60 and 720 min are single measurements. The dotted line, plotted at ± 0.46 D, indicates the threshold value for significant differences in HDX (99% confidence interval, calculated from the pooled standard deviations for all time points performed in replicates, at both states). Significant regions of CysK sequence, discussed in the main text, are indicated above the chart.

A 99% confidence interval (CI) was calculated as a threshold for significance of changes in CysK dynamics, corresponding to 0.46 Da (dotted line in Figure 24). Significant differences in HDX between the two states of CysK compared are also mapped onto CysK crystal structure (Figure 25).

According to this, decreased HDX was observed for the region encompassing residues 66-78 (peptides 65-77/78), located in proximity of the active site. Region 97-100 (comprising peptides 93/97-110, 97/99-111, but not 86-96 and 100-110) and residues 118-129 (peptides 111-125/129, 112-129, 113-129, 114-129, but not 102-117), belonging to a subdomain of the N-terminal domain, also showed reduced HDX following CS complex formation. An effect of protection from HDX upon binding to CysE occurred starting from the end of helix 10 (H10, residues 214-216), encompassing the β-loop up to residue 239 (peptides 213/214/216/218-239). Regions of the protein that undergo significant changes in dynamics upon binding with CysE, are depicted with different colours in Figure 25 and are indicated above the chart showing differences in deuterium uptake for the identified peptides, between the two states compared (see Figure 24). Together, these results revealed that several parts of CysK became protected from HDX upon complex formation.

The region including residues 66-78 (indicated in Figure 24 and coloured in pink in Figure 25) showed this behaviour. Representative mass spectra for CysK peptide 65-77, covering this region, are shown in Figure 26. Residues in this part of the sequence underwent EX2 exchange kinetics [116], as demonstrated by the single isotopic distribution that shifts at higher mass-to-charge (m/z) values, following protein deuteration at different time points. This type of exchange was observed for all peptides with significant differences in HDX, analysed in this work.

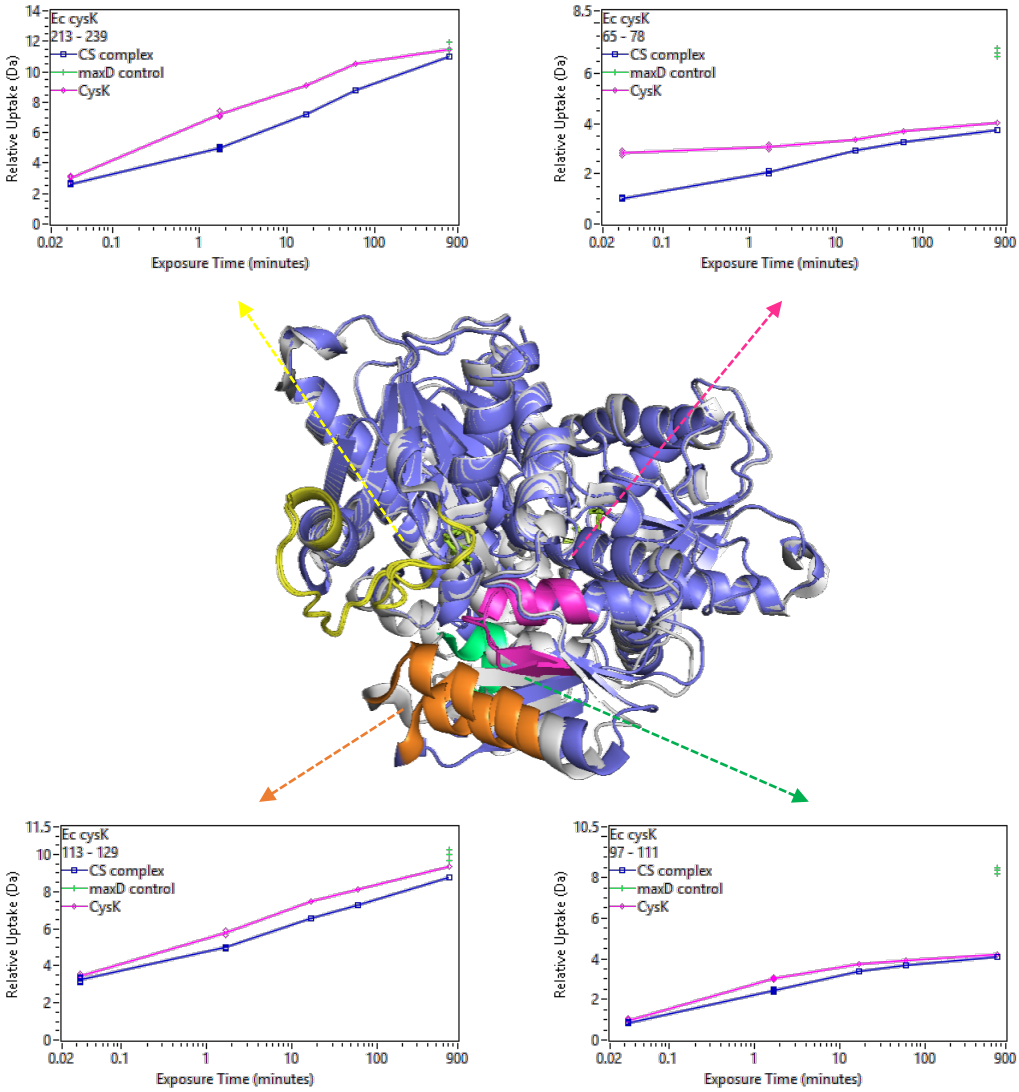


Figure 25. Conformational dynamics of CysK alone and following CS complex formation. Regions that present significant decrease of HDX following CysK association with CysE (described in the main text) are coloured on CysK structure (pink–residues 66-78, green–residues 97-100, orange–residues 118-129, yellow–residues 214-239). CysK crystal structures are aligned and correspond to two different states adopted from the protein (see below in the main text): the "open" conformation (PDB code: 1oas, coloured in light blue) and the "closed" conformation (PDB code: 1d6s, coloured in light grey). For each region, showing significant changes in CysK dynamics upon complex formation, are reported the relative deuterium uptake plots of representative peptides, as a function of the measured time points (0.03, 1.67, 16.67, 60 and 720 min). CysK alone is plotted in magenta, CysK in CS complex is plotted in blue and the maximum deuterated controls are plotted as green stars at top right (3 replicates). The protein name and the peptide sequence start and end are indicated in the plot (top, left).

Peptide 65-77

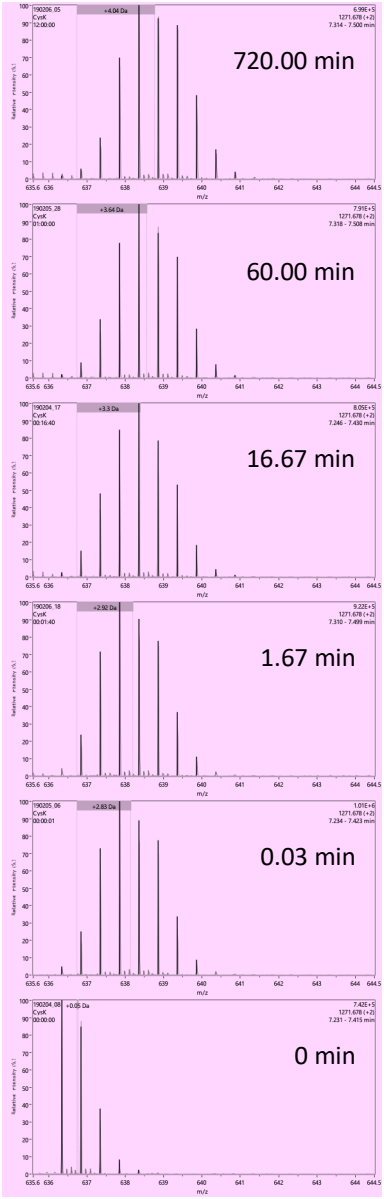
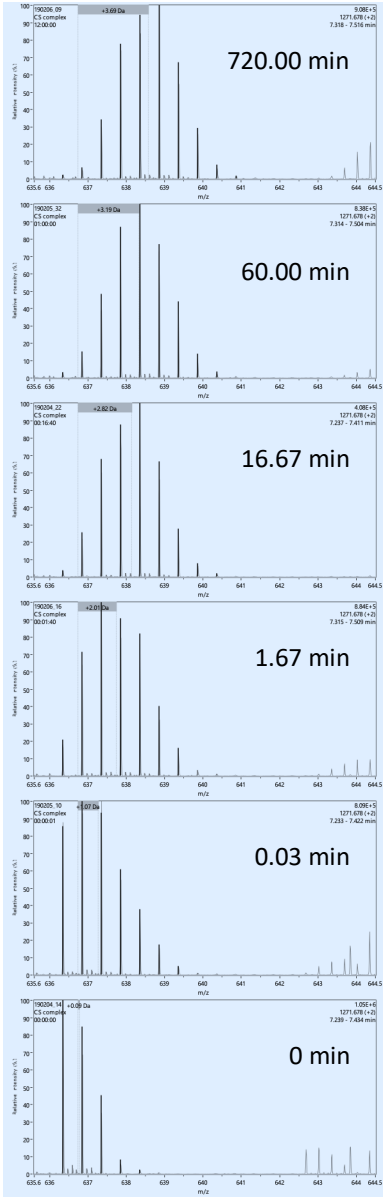
CysK alone	CS complex
	

Figure 26. Mass spectra of CysK peptide 65-77. See previous page. Representative mass spectra for CysK peptide 65-77 (charge state +2) in both states of the protein: alone (light pink panel) and assembled with CysE (CS complex, light blue panel). The isotopic distributions at 0 min time point correspond to the non-deuterated samples. Afterwards, the isotopic envelopes migrate to higher m/z values as a function of time (0.03 to 720 min), upon deuterium incorporation.

The relative deuterium uptake of CysK alone and within CS complex for a representative peptide of this region (65-78), plotted as a function of the labeling time (0.03-720 min), is shown in Figure 25 (top, right).

Concerning residues 66-78, moreover, a great difference in deuterium uptake could be already observed at the earliest measured time point (2 seconds). This finding is in very good agreement with previous experiments of site-directed mutagenesis and functional studies [[50,62]], showing that CysE C-terminal tail, inserting into CysK active site, makes several contacts with the surrounding residues, including Ser70, Asn72 and Thr73 in *H. influenzae* CysK (that correspond to the same residues in *E. coli* CysK, analysed in this work). This region was thus confirmed to be part of the binding interface with CysE within CS complex.

Observing the crystal structure of HiCysK in complex with CysE C-terminal tetrapeptide (PDB code: 1y7l, see Figure 5 at page 18), the residue Gln143 has been found to be involved in hydrogen bond formation with CysE C-terminal Ile267. Indeed, a stabilization of that region over the time could be observed, regarding peptide 142-154 and neighboring residues (see the region between peptides 35 and 40 in Figure 24). However, we decided to use the 99% CI as a threshold, to be more selective with respect to significance of variations in CysK dynamics. Therefore, these changes in HDX weren't considered to be significant according to the analysis made in this work.

Decreased HDX in the CS complex was observed also for residues 97-100 and 118-129. Relative deuterium uptake plots for representative peptides belonging to these regions are depicted in Figure 25 (bottom, right and bottom, left).

For the first part of the sequence (97-100), a significant difference could be noticed at the time point 1.67 min, whereas for the latter residues (118-129), the effect on protein dynamics started at 1.67 min and extended to the longer time points.

These two groups of residues are located on a peculiar region of the protein. Indeed, it has been demonstrated [33] that residues 87-131 of *S.* Typhimurium CysK belong to a flexible domain, that undergoes a large conformational movement upon ligand binding. The same rearrangement, that allows CysK to adopt a closed state, has been hypothesized by our group to take place also following interaction with CysE [80]. In the closed form, this region is brought into proximity of CysK active site, where CysE inserts its C-terminal peptide, and is thus stabilized, as it could be observed from HDX data.

Segmental movements in CysK upon transition between the open and closed state can be better observed in Figure 27, where regions showing significant differences in HDX are highlighted. Residues belonging to the moveable domain, described in this part, are coloured in green and orange.

Intriguingly, the present results correlate with our findings from the protein painting assay (see chapter "Cysteine synthase protein-protein interaction hotspots revealed by protein painting"), where two residues belonging to this subdomain became buried upon CS complex assembly [13]. These observations further corroborate the hypothesis that a stabilization of the closed conformation is operated by CysE within CS complex.

Residues 214-239 (see Figure 24 and the region coloured in yellow in Figure 25) were also stabilized upon CysK interaction with CysE. The decrease of HDX could be observed at all the time points, but was considered significant for 1.67,

16.67 and 60 min. The relative deuterium uptake of a representative peptide of this region is plotted as a function of labeling time (0.03-720 min) in Figure 25 (top, left).

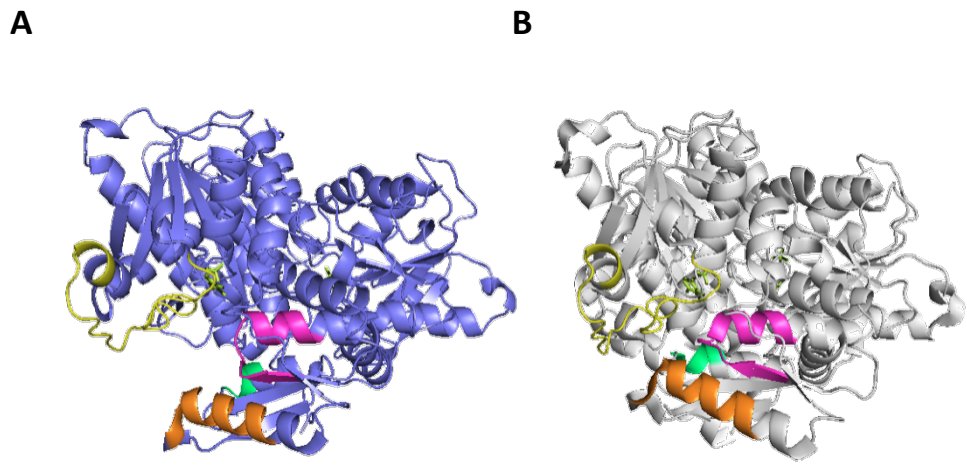


Figure 27. Changes in dynamics in CysK, upon CS complex formation. Regions showing significant changes of CysK dynamics (i.e. HDX) upon CS complex formation are highlighted on CysK crystal structures, represented both in the open conformation (PDB code: 1oas, coloured in light blue, **Panel A.**) and in the closed conformation (PDB code: 1d6s, coloured in light grey, **Panel B.**). The colour code for regions with significant dynamics is the same adopted in figure 25. Residues that are part of the flexible domain are coloured in green (97-100) and orange (118-129).

In particular, these residues (214-239) belong to a conserved part of CysK sequence, the loop β8A-β9A, whose location is adjacent to CysK active site. The relevance of this region was already discussed in this PhD thesis, since applying protein painting approach to the CS assembly and to the complex formed by CysK and the toxin (CdiA-CT), a Lys residue (K226), belonging to this region, became buried, i.e. protected from the solvent, in both complexes arranged by CysK, with respect to the protein alone [13]. Furthermore, previous findings [63] indicated that 3 residues located in this part of CysK sequence are crucial for its interaction with CysE: mutation of K217, H221 and K222 in *A. thaliana*

(corresponding to K221, H225 and K226 in *E. coli*), indeed, disrupts CS complex formation.

An effect of structural stabilization upon CysK assembly with CysE could be noticed also in the C-terminal region, starting from residue 300 (around peptide number 70 in Figure 24). This effect could be attributed to the movement toward the closed state of the protein. Indeed, it has been demonstrated that residues belonging to the C-terminal domain of CysK become more ordered in the closed structure [33]. However, although peptides in this region show differences in HDX that are borderline with respect to the CI, these changes were not considered above the threshold of significance that was chosen for this analysis.

3.2 CysE dynamics following binding to CysK

Investigation on the effects of CS assembly formation on dynamics and local structural changes of the protein was applied also to CysE partner of the complex. CysK interaction with CysE impacted different regions of the latter protein, in comparison with the unbound state. Overall results from the continuous labeling experiment that was conducted are illustrated in Figures 28 and 29, on a difference plot, comparing the two states of the protein studied, and on the crystal structure of CysE (regions showing significant differences in HDX are depicted on both trimers of the hexamer), respectively.

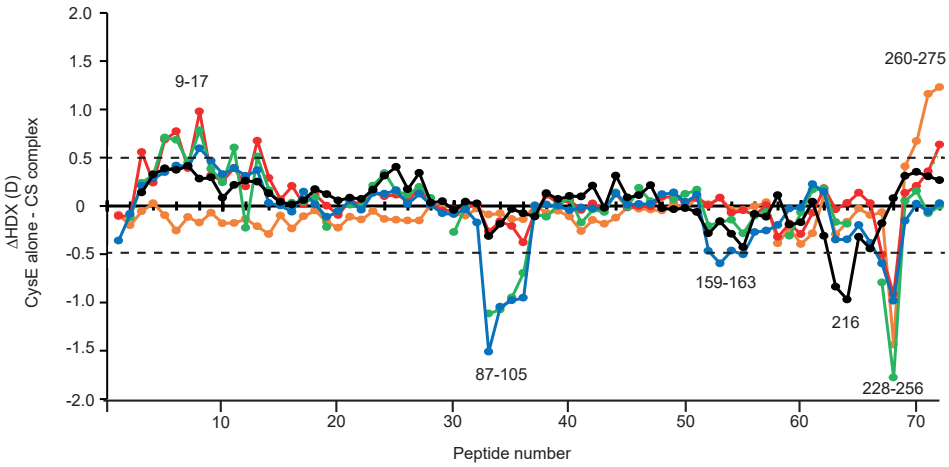


Figure 28. CysE average deuterium uptake difference plot. Difference plot of the average deuterium uptake (ΔHDX) between two states: CysE alone and CysE assembled with CysK (CS complex) for the 72 identified peptides at the five measured time points (orange–0.03 min; red–1.67 min; green–16.67 min; cyan–60 min; black–720 min). The peptides are arranged along the x-axis according their position in CysE sequence, from N-terminal to C-terminal. Positive and negative values along y-axis indicate reduced or increased HDX, respectively, following CysE binding to CysK. Values at 0.03 and 1.67 min correspond to the mean of three replicates, whereas values at 16.67, 60 and 720 min are single measurements. The dotted line, plotted at ± 0.49 D, indicates the threshold value for significant differences in HDX (99% confidence interval, calculated from the pooled standard deviations for all time points performed in replicates, at both states). Significant regions of CysE sequence, discussed in the main text, are indicated in the chart.

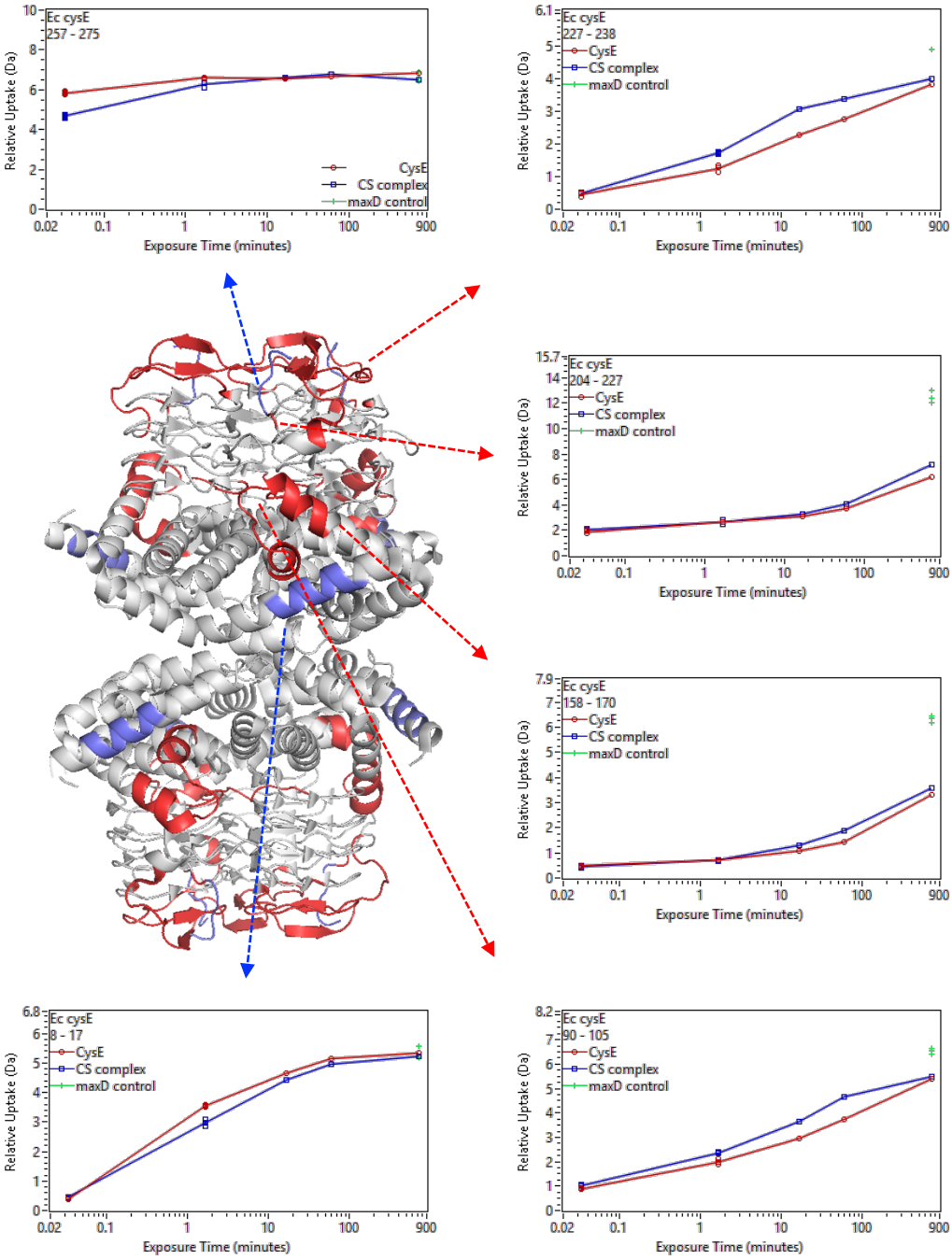

Ec cysE
257 - 275
Relative Uptake (Da)
Exposure Time (minutes)
CysE
CS complex
maxD control
Ec cysE
227 - 238
Ec cysE
204 - 227
Ec cysE
158 - 170
Ec cysE
8 - 17
Ec cysE
90 - 105

Figure 29. Conformational dynamics of CysE alone and following CS complex formation. Significant differences in HDX upon CS assembly are mapped onto CysE crystal structure (PDB code: 1t3d). Regions coloured in blue showed a decrease of HDX (structural stabilization), whereas regions depicted in red underwent an increase of HDX (structural destabilization). Regions coloured in light grey indicate insignificant difference in HDX between the two states compared, or regions that were not covered by peptides. For each region, showing significant changes in CysE dynamics upon complex formation, are reported the relative deuterium uptake plots of representative peptides, as a function of the measured time points (0.03, 1.67, 16.67, 60 and 720 min). CysE alone is shown in red, CysE in CS complex is plotted in blue and the maximum deuterated controls are depicted as green stars at top right. The protein name and the peptide sequence start and end are indicated in the plot (top, left).

In order to establish which changes in HDX between CysE alone and the protein engaged in CS complex had to be considered significant, a 99% confidence interval (CI) was calculated, as previously done for the analysis of CysK. This CI for CysE corresponded to 0.49 Da (indicated as a dotted line in Figure 28).

Significant changes in CysE dynamics (i.e. HDX) following its interaction with CysK were observed across different regions of the protein sequence. In particular, a decrease of HDX upon CS complex formation was induced in the C-terminal tail of CysE (residues 260-275), as well as the N-terminal part of the protein (residues 9-17). On the other side, a significant increase of HDX was noticed for four distinct segments of the protein, including residues 87-105, 159-163, 216 and 228-256.

Regions presenting significant changes in dynamics (i.e. HDX), that will be described, are plotted onto CysE three dimensional structure in Figure 29. Residues showing decreased HDX or increased HDX upon CS complex assembly, compared to the protein alone, are coloured in blue or in red, respectively.

Peptides covering the C-terminal segment of CysE (residues 260-275) showed reduced deuterium uptake upon binding of CysK. The deuterium uptake plotted

as a function of labeling time (i.e., 0.03-720 min) for a representative peptide of this region, is illustrated in Figure 29 (top, left).

The effect of decreased HDX was considered significant for the earliest time point, i.e. 2 seconds, but not for the other exposure times to deuterium. Indeed, this region of the protein is unstructured and flexible and reasonably presents a high exchange rate, reaching the experimental maximum deuterium uptake already at lower sampled timings. These observations are in good overall agreement with previous findings from structural and functional studies [44,49,50], demonstrating that C-terminus of CysE binds CysK active site, stabilizing CS complex formation.

The N-terminal region of the protein (residues 9-17), located at the interface between the two trimers of the hexamer, showed reduced deuterium uptake upon CysE interaction with CysK. The effect started to appear at 1.67 min, extending also to the time point 16.67 min, as it can be observed in the relative deuterium uptake plotted as a function of labeling time (0.03-720 min) for the representative peptide 8-17, illustrated in Figure 29 (bottom, left).

Binding of CysK to CysE C-terminus is likely to be transmitted through an allosteric communication to the other side of the protein, inducing structural stabilization at the trimers interface. These results further confirm the hypothesis that was pointed out from SAXS measurements (see chapter “A model for the solution structure of cysteine synthase complex”), where the S-shape of the assembly suggested a functional connection between the two sides of CysE, interacting with CysK [13].

Furthermore, from the mass spectra, a slight peak broadening for the CS complex with respect to the state of CysE alone was observed (representative peptides are shown in Figure 30).

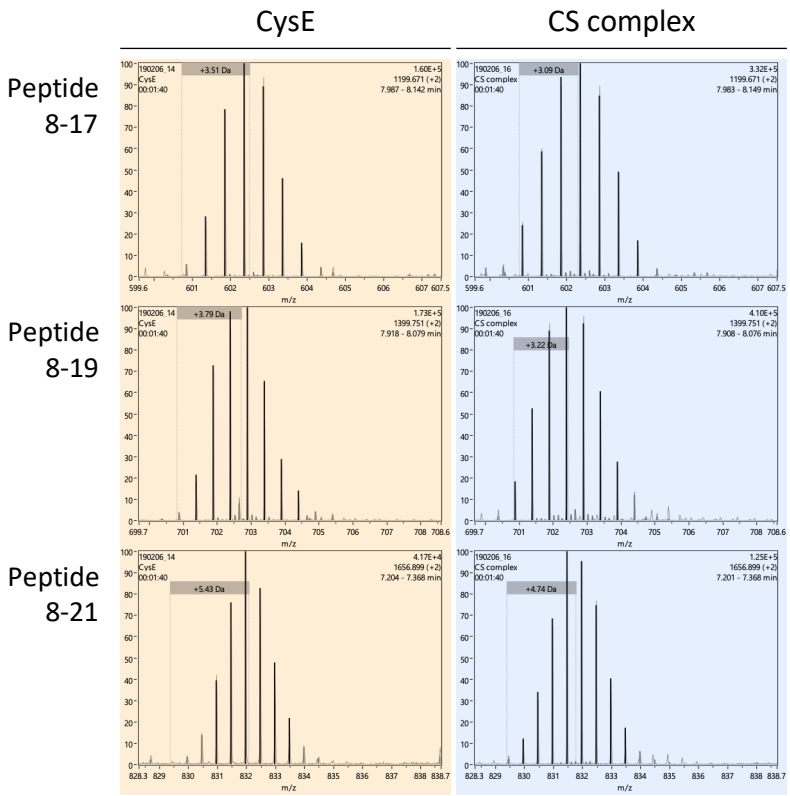


Figure 30. Representative mass spectra for CysE peptides, covering the N-terminal part of the sequence (peptides 8-17, 8-19, 8-21). Spectra at the time point 1.67 min, at which the difference between the deuterium uptake of the protein alone (light orange panel) and assembled with CysK (CS complex, light blue panel) was considered significant, are reported (the charge state illustrated is +2 for all peptides). The isotopic distribution underwent a slight broadening in CS complex spectra, compared to CysE alone.

This effect, involving peptides that cover the N-terminal region of the protein, could indicate that the subunits of CysE trimers are not equivalent, but only one or two out of three are engaged in interaction with CysK active site, as already suggested from the SAXS envelope of CS complex.

Finally, an increase of HDX upon CysE binding to CysK, i.e. a structural destabilization, was noticed in different segments of the protein sequence, involving residues 87-105, 159-163, 216, and becoming more extended towards C-terminus, where it concerned the region 228-256. This effect generally started to appear at the longer time points studied, as it can be observed from the deuterium uptake plots reported in Figure 29 for representative peptides, covering these regions (panels displayed on the right).

Together these data suggest that CysK binding to the C-terminus of the protein develops in a communication pathway across CysE sequence, concluding with an allosteric effect at the N-terminus, where the interface between the two trimers of the hexamer is stabilized, as a consequence of CS complex formation.

No peak broadening was noticed for these regions of the protein showing structural destabilization.

Moreover, all the peptides showing significant differences in HDX (i.e. dynamics) in CysE following association with CysK, underwent EX2 kinetics, as indicated by the presence of a single isotopic distribution in the mass spectra, shifting at higher m/z, following deuterium uptake.

CONCLUSIONS

Cysteine is intimately related to bacterial fitness, virulence and pathogenicity, therefore enzymes involved in its biosynthetic pathway have been targeted for the development of new antibiotics or antibiotic enhancers [6]. Among these, CysE and CysK, the two enzymes that conclude the reductive sulfate assimilation pathway (RSAP), leading to the formation of cysteine, associate within the so-called Cysteine Synthase (CS) complex, that was the topic of this PhD thesis. Many aspects of CS assembly, including its regulation, the exact biological role played by CysK/CysE interaction in cysteine biosynthesis, the three dimensional organization of the complex and the intriguing possibility to exploit it as a pharmaceutical target itself are still questionable topics.

In this work, several experimental approaches were applied to the CS system, in order to gain insights into structural organization and dynamic changes upon binding of the two constituent enzymes.

The "protein painting" technique was exploited to reveal protein-protein interaction hotspots of CS complex. The assay was validated on a complex with a known three-dimensional structure [79], formed by CysK with a bacterial toxin (CdiA-CT), that shares the same interaction mechanism and structural features also adopted by CysE to assemble with CysK. Afterwards, the experimental workflow was applied to CS complex, whose crystal structure is still unavailable. As a result, a common residue in both interprotein complexes was found to be part of the interaction interface, supporting the similarity of the mechanism by which the two partner proteins bind to CysK. Moreover, the results from protein painting experiments also supported previous findings indicating that the interaction with CysE stabilizes a "closed" conformation of the enzyme. Additionally, other two sites were spotted at CysK intermonomer interface becoming buried only upon complex formation. Hence, the existence of an allosteric communication between the two monomers of the enzyme was suggested to take place.

Considering the quaternary geometry of CS complex, two different scenarios have been proposed, according to the determined stoichiometry of CysE binding to CysK [49]. According to these alternative hypotheses, either both the subunits of each CysK dimer, or only one per dimer, interact with one or two subunits, respectively, of CysE. SAXS measurements helped to create a model of CS assembly. The envelope obtained from the ab initio modelling was further refined, using as constraints the experimental contact points retrieved from the protein painting assay. The S-shape of the complex suggested that only one subunit of CysK dimer is bound at each side of CysE dimer of trimers. The occupation of the active site of CysK is hence allosterically communicated to the second monomer, as supported by protein painting findings, and leads to the almost full inhibition of CysK activity, as already suggested by functional data [44,45]. Furthermore, binding of one CysK dimer to a trimer of CysE seems to dictate the orientation of the interaction of the second CysK on the other side of CysE, through an allosteric cross-talk between the two dimer of trimers.

HDX-MS data provided a global view on how CS complex formation modulates the dynamics of both CysK and CysE in solution. Several segments of CysK presented significant changes in HDX upon binding to CysE. In particular, residues located next to CysK active site were confirmed to interact with CysE C-terminal tail. Moreover, overlapping peptides belonging to a subdomain of CysK N-terminus, that undergoes a conformational rearrangement towards a closed conformation upon binding with CysE, showed stabilization following CS assembly formation. Decreased HDX in the CS complex compared to the component proteins alone, was observed also for a conserved loop of CysK, that has been found to contain residues that are crucial for complex formation.

On the other side, CysE partner of CS assembly interacts with CysK inserting its C-terminal region into the ligand binding site, as already hypothesized, and further confirmed by observation of an HDX decrease in that region upon

complex formation. This binding led also to an increased dynamics of many segments of the protein sequence, spanning all the way from the region interacting with CysK to the N-terminal part of CysE monomers, ending with an allosterically induced stabilization at the interface between the two trimers.

Taken together these results led to the elucidation of the quaternary structure of CS complex and changes in structural dynamics affecting the constituent enzymes upon their association. These findings could help to better understand the regulatory effect and the biological role of CysE assembly with CysK, therefore driving the selection and development of molecules able to affect CysK interaction with CysE, with a potential as innovative antimicrobial agents.

BIBLIOGRAPHY

1. Beinert, H. A tribute to sulfur. *Eur. J. Biochem* **267**, 5657-5664 (2000).
2. Kessler, D. Enzymatic activation of sulfur for incorporation into biomolecules in prokaryotes. *FEMS Microbiol. Rev.* **30**, 825–840 (2006).
3. Mihara, H. & Esaki, N. Bacterial cysteine desulfurases: Their function and mechanisms. *Applied Microbiology and Biotechnology* **60**, 12–23 (2002).
4. Sekowska, A., Kung, H.-F. & Danchin, A. Sulfur Metabolism in *Escherichia coli* and Related Bacteria: Facts and Fiction. JMMB Review. Sulfur Metabolism in Enterobacteria *J. Mol. Microbiol. Biotechnol* **2**, 145-177 (2000).
5. Turnbull, A. L. & Surette, M. G. Cysteine biosynthesis, oxidative stress and antibiotic resistance in *Salmonella typhimurium*. *Res. Microbiol.* **161**, 643–650 (2010).
6. Campanini, B. *et al.* Inhibitors of the Sulfur Assimilation Pathway in Bacterial Pathogens as Enhancers of Antibiotic Therapy. *Curr. Med. Chem.* **22**, 187–213 (2014).
7. Spyrakis, F. *et al.* Isozyme-Specific Ligands for *O*-acetylserine sulfhydrylase, a Novel Antibiotic Target. *PLoS One* **8**, (2013).
8. Spyrakis, F. *et al.* Fine tuning of the active site modulates specificity in the interaction of *O*-acetylserine sulfhydrylase isozymes with serine acetyltransferase. *Biochim. Biophys. Acta - Proteins Proteomics* **1834**, 169–181 (2013).
9. Amori, L. *et al.* Design and synthesis of trans-2-substituted-cyclopropane-1-carboxylic acids as the first non-natural small molecule inhibitors of *O*-acetylserine sulfhydrylase. *Med. Chem. Commun.* **3**, 1111–1116 (2012).
10. Jean kumar, V. U. *et al.* Discovery of novel inhibitors targeting the

Mycobacterium tuberculosis O-acetylserine sulfhydrylase (CysK1) using virtual high-throughput screening. *Bioorg. Med. Chem. Lett.* **23**, 1182–1186 (2013).

11. Schnell, R. & Schneider, G. Structural enzymology of sulphur metabolism in *Mycobacterium tuberculosis*. *Biochem. Biophys. Res. Commun.* **396**, 33–38 (2010).
12. Nagpal, I., Raj, I., Subbarao, N. & Gourinath, S. Virtual Screening, Identification and In Vitro Testing of Novel Inhibitors of *O*-Acetyl-L-serine Sulfhydrylase of *Entamoeba histolytica*. *PLoS One* **7**, 1–7 (2012).
13. Rosa, B. *et al.* Combination of SAXS and Protein Painting Discloses the Three-Dimensional Organization of the Bacterial Cysteine Synthase Complex, a Potential Target for Enhancers of Antibiotic Action. *Int. J. Mol. Sci.* **20**, (2019).
14. Kredich, N. M. The enzymic synthesis of L-cysteine in *Escherichia coli* and *Salmonella typhimurium*. *J Biol Chem* **241**, 4955–4966 (1966).
15. Cook, P. F. & Wedding, R. T. A reaction mechanism from steady state kinetic studies for *O*-acetylserine sulfhydrylase from *Salmonella typhimurium* LT 2. *J. Biol. Chem.* **251**, 2023–2029 (1976).
16. NAKAMURA, T. & SATO, R. Synthesis from sulphate and accumulation of S-sulphocysteine by a mutant strain of *Aspergillus nidulans*. *Biochem. J.* **86**, 328–335 (1963).
17. Nakatani, T. *et al.* Enhancement of thioredoxin/glutaredoxin-mediated L-cysteine synthesis from S-sulfocysteine increases L-cysteine production in *Escherichia coli*. *Microb. Cell Fact.* **11**, (2012).
18. Chattopadhyay, A. *et al.* Structure, mechanism, and conformational dynamics of *O*-acetylserine sulfhydrylase from *Salmonella typhimurium*: Comparison of A and B isozymes. *Biochemistry* **46**, 8315–8330 (2007).
19. Kredich, N. M., Becker, M. A. & Tomkins, G. M. Purification and

characterization of cysteine synthetase, a bifunctional protein complex, from *Salmonella typhimurium*. *J. Biol. Chem.* **244**, 2428–2439 (1969).

20. Campanini, B. *et al.* Moonlighting *O*-acetylserine sulfhydrylase: New functions for an old protein. *Biochim. Biophys. Acta - Proteins Proteomics* **1854**, 1184–1193 (2015).
21. Kredich, N. M. The molecular basis for positive regulation of cys promoters in *Salmonella typhimurium* and *Escherichia coli*. *Mol. Microbiol.* **6**, 2747–2753 (1992).
22. Kredich, N. M. Regulation of L-cysteine Biosynthesis in *Salmonella typhimurium*. *Biol. Chem.* **246**, 3474–3484 (1971).
23. Tyrrell, R. *et al.* The structure of the cofactor-binding fragment of the LysR family member, CysB: a familiar fold with a surprising subunit arrangement. *Structure* **5**, 1017–1032 (1997).
24. Hindson, V. J. & Shaw, W. V. Random-order ternary complex reaction mechanism of serine acetyltransferase from *Escherichia coli*. *Biochemistry* **42**, 3113–3119 (2003).
25. Johnson, C. M., Huang, B., Roderick, S. L. & Cook, P. F. Kinetic mechanism of the serine acetyltransferase from *Haemophilus influenzae*. *Arch. Biochem. Biophys.* **429**, 115–122 (2004).
26. Olsen, L. R., Huang, B., Vetting, M. W. & Roderick, S. L. Structure of serine acetyltransferase in complexes with CoA and its cysteine feedback inhibitor. *Biochemistry* **43**, 6013–6019 (2004).
27. Burkhard, P., Tai, C.-H., Jansonius, J. N. & Cook, P. F. Identification of an allosteric anion-binding site on *O*-acetylserine sulfhydrylase: structure of the enzyme with chloride bound. *J. Mol. Biol.* **303**, 279–286 (2000).
28. Droux, M., Ruffet, M.-L., Douce, R. & Job, D. Interactions between serine acetyltransferase and *O*-acetylserine (thiol) lyase in higher plants. *Eur. J. Biochem.* **255**, 235–245 (1998).

29. Pye, V. E., Tingey, A. P., Robson, R. L. & Moody, P. C. E. The structure and mechanism of serine acetyltransferase from *Escherichia coli*. *J. Biol. Chem.* **279**, 40729–40736 (2004).
30. Hindson, V. J., Moody, P. C. E., Rowe, A. J. & Shaw, W. V. Serine acetyltransferase from *Escherichia coli* is a dimer of trimers. *J. Biol. Chem.* **275**, 461–466 (2000).
31. Tai, C. H., Nalabolu, S. R., Jacobson, T. M., Minter, D. E. & Cook, P. F. Kinetic mechanisms of the A and B isozymes of *O*-acetylserine sulfhydrylase from *Salmonella typhimurium* LT-2 using the natural and alternate reactants. *Biochemistry* **32**, 6433–6442 (1993).
32. Burkhard, P. *et al.* Three-dimensional structure of *O*-acetylserine sulfhydrylase from *Salmonella typhimurium*. *J. Mol. Biol.* **283**, 121–133 (1998).
33. Burkhard, P., Tai, C.-H., Ristroph, C. M., Cook, P. F. & Jansonius, J. N. Ligand binding induces a large conformational change in *O*-acetylserine sulfhydrylase from *Salmonella typhimurium*. *J. Mol. Biol.* **291**, 941–953 (1999).
34. Claus, M. T., Zocher, G. E., Maier, T. H. P. & Schulz, G. E. Structure of the *O*-Acetylserine Sulfhydrylase Isoenzyme CysM from *Escherichia coli*. *Biochemistry* **44**, 8620–8626 (2005).
35. Hulanicka, M. D., Hallquist, S. G., Kredich, N. M. & Mojica, T. Regulation of *O*-acetylserine sulfhydrylase B by L-cysteine in *Salmonella typhimurium*. *J. Bacteriol.* **140**, 141–146 (1979).
36. Filutowicz, M., Wiater, A. & Hulanicka, D. Delayed Inducibility of Sulphite Reductase in *cysM* Mutants of *Salmonella typhimurium* Under Anaerobic Conditions. *J. Gen. Microbiol.* **128**, 1791–1794 (1982).
37. Kröger, C. *et al.* An infection-relevant transcriptomic compendium for Salmonella enterica Serovar Typhimurium. *Cell Host Microbe* **14**, 683–

95 (2013).

38. Sweetlove, L. J. & Fernie, A. R. The role of dynamic enzyme assemblies and substrate channelling in metabolic regulation. *Nature Communications* **9**, (2018).
39. Schmitt, D. L. & An, S. Spatial Organization of Metabolic Enzyme Complexes in Cells. *Biochemistry* **56**, 3184–3196 (2017).
40. Scott, D. E., Bayly, A. R., Abell, C. & Skidmore, J. Small molecules, big targets: Drug discovery faces the protein-protein interaction challenge. *Nature Reviews Drug Discovery* **15**, 533–550 (2016).
41. Shepard, W. *et al.* Insights into the Rrf2 repressor family – the structure of CymR, the global cysteine regulator of *Bacillus subtilis*. *FEBS J.* **278**, 2689–2701 (2011).
42. Diner, E. J., Beck, C. M., Webb, J. S., Low, D. A. & Hayes, C. S. Identification of a target cell permissive factor required for contact-dependent growth inhibition (CDI). *Genes Dev.* **26**, 515–525 (2012).
43. Ma, D. K., Vozdek, R., Bhatla, N. & Horvitz, H. R. CYSL-1 interacts with the O2-sensing hydroxylase EGL-9 to promote H2S-modulated hypoxia-induced behavioral plasticity in *C. elegans*. *Neuron* **73**, 925–940 (2012).
44. MINO, K. *et al.* Effects of Bienzyme Complex Formation of Cysteine Synthetase from *Escherichia coli* on Some Properties and Kinetics. *Biosci. Biotechnol. Biochem.* **64**, 1628–1640 (2000).
45. Benoni, R. *et al.* Modulation of *Escherichia coli* serine acetyltransferase catalytic activity in the cysteine synthase complex. *FEBS Lett.* **591**, 1212–1224 (2017).
46. Kumada, Y. *et al.* Protein–protein interaction analysis using an affinity peptide tag and hydrophilic polystyrene plate. *J. Biotechnol.* **128**, 354–361 (2007).

47. Zhao, C. *et al.* On the interaction site of serine acetyltransferase in the cysteine synthase complex from *Escherichia coli*. *Biochem. Biophys. Res. Commun.* **341**, 911–916 (2006).
48. Salsi, E. *et al.* A two-step process controls the formation of the bienzyme cysteine synthase complex. *J. Biol. Chem.* **285**, 12813–12822 (2010).
49. Campanini, B. *et al.* Interaction of serine acetyltransferase with *O*-acetylserine sulfhydrylase active site: Evidence from fluorescence spectroscopy. *Protein Sci.* **14**, 2115–2124 (2005).
50. Huang, B., Vetting, M. W. & Roderick, S. L. The active site of *O*-acetylserine sulfhydrylase is the anchor point for bienzyme complex formation with serine acetyltransferase. *J. Bacteriol.* **187**, 3201–3205 (2005).
51. Feldman-Salit, A., Wirtz, M., Hell, R. & Wade, R. C. A Mechanistic Model of the Cysteine Synthase Complex. *J. Mol. Biol.* **386**, 37–59 (2009).
52. Feldman-Salit, A. *et al.* Allosterically Gated Enzyme Dynamics in the Cysteine Synthase Complex Regulate Cysteine Biosynthesis in *Arabidopsis thaliana*. *Structure* **20**, 292–302 (2012).
53. Francois, J. A., Kumaran, S. & Jez, J. M. Structural Basis for Interaction of *O*-Acetylserine Sulfhydrylase and Serine Acetyltransferase in the Arabidopsis Cysteine Synthase Complex. *Plant Cell* **18**, 3647–3655 (2006).
54. Jez, J. M. & Dey, S. The cysteine regulatory complex from plants and microbes: what was old is new again. *Curr. Opin. Struct. Biol.* **23**, 302–310 (2013).
55. Wirtz, M., Berkowitz, O., Droux, M. & Hell, R. The cysteine synthase complex from plants. Mitochondrial serine acetyltransferase from *Arabidopsis thaliana* carries a bifunctional domain for catalysis and

protein-protein interaction. *Eur. J. Biochem.* **268**, 686–693 (2001).

56. Wirtz, M. & Hell, R. Functional analysis of the cysteine synthase protein complex from plants: Structural, biochemical and regulatory properties. *J. Plant Physiol.* **163**, 273–286 (2006).
57. Wirtz, M. *et al.* Structure and function of the hetero-oligomeric cysteine synthase complex in plants. *J. Biol. Chem.* **285**, 32810–32817 (2010).
58. MINO, K. *et al.* Increase in the Stability of Serine Acetyltransferase from *Escherichia coli* against Cold Inactivation and Proteolysis by Forming a Bienzyme Complex. *Biosci. Biotechnol. Biochem.* **65**, 865–874 (2001).
59. Kumaran, S., Yi, H., Krishnan, H. B. & Jez, J. M. Assembly of the cysteine synthase complex and the regulatory role of protein-protein interactions. *J. Biol. Chem.* **284**, 10268–10275 (2009).
60. Hell, R. & Hillebrand, H. Plant concepts for mineral acquisition and allocation. *Curr. Opin. Biotechnol.* **12**, 161–168 (2001).
61. Mino, K. *et al.* Characteristics of serine acetyltransferase from *Escherichia coli* deleting different lengths of amino acid residues from the C-terminus. *Biosci. Biotechnol. Biochem.* **64**, 1874–1880 (2000).
62. Salsi, E. *et al.* Design of *O*-acetylserine sulfhydrylase inhibitors by mimicking nature. *J. Med. Chem.* **53**, 345–356 (2010).
63. Bonner, E. R., Cahoon, R. E., Knapke, S. M. & Jez, J. M. Molecular basis of cysteine biosynthesis in plants: Structural and functional analysis of *O*-acetylserine sulfhydrylase from *Arabidopsis thaliana*. *J. Biol. Chem.* **280**, 38803–38813 (2005).
64. Wang, T. & Leyh, T. S. Three-stage assembly of the cysteine synthase complex from *Escherichia coli*. *J. Biol. Chem.* **287**, 4360–4367 (2012).
65. Ronda, L., Bruno, S. & Bettati, S. Tertiary and quaternary effects in the allosteric regulation of animal hemoglobins. *Biochim. Biophys. Acta - Proteins Proteomics* **1834**, 1860–1872 (2013).

66. Changeux, J.-P. & Edelstein, S. J. Allosteric Mechanisms of Signal Transduction. *Science* **308**, 1424–1428 (2005).
67. Guo, J. & Zhou, H.-X. Protein Allostery and Conformational Dynamics. *Chem. Rev.* **116**, 6503–6515 (2016).
68. Chandramohan, A. *et al.* Predicting Allosteric Effects from Orthosteric Binding in Hsp90-Ligand Interactions: Implications for Fragment-Based Drug Design. *PLOS Comput. Biol.* **12**, 1–17 (2016).
69. Liu, J. & Nussinov, R. Allostery: An Overview of Its History, Concepts, Methods, and Applications. *PLOS Comput. Biol.* **12**, 1–5 (2016).
70. Poyraz, Ö. *et al.* Structure-Guided Design of Novel Thiazolidine Inhibitors of *O*-Acetyl Serine Sulfhydrylase from *Mycobacterium tuberculosis*. *J. Med. Chem.* **56**, 6457–6466 (2013).
71. Brunner, K. *et al.* Inhibitors of the Cysteine Synthase CysM with Antibacterial Potency against Dormant *Mycobacterium tuberculosis*. *J. Med. Chem.* **59**, 6848–6859 (2016).
72. Mori, M. *et al.* Discovery of Antiamebic Compounds That Inhibit Cysteine Synthase From the Enteric Parasitic Protist *Entamoeba histolytica* by Screening of Microbial Secondary Metabolites. *Front. Cell. Infect. Microbiol.* **8**, 409 (2018).
73. Brunner, K. *et al.* Profiling of in vitro activities of urea-based inhibitors against cysteine synthases from *Mycobacterium tuberculosis*. *Bioorg. Med. Chem. Lett.* **27**, 4582–4587 (2017).
74. Palde, P. B. *et al.* First-in-Class Inhibitors of Sulfur Metabolism with Bactericidal Activity against Non-Replicating *M. tuberculosis*. *ACS Chem. Biol.* **11**, 172–184 (2016).
75. Berkowitz, O., Wirtz, M., Wolf, A., Kuhlmann, J. & Hell, R. Use of Biomolecular Interaction Analysis to Elucidate the Regulatory Mechanism of the Cysteine Synthase Complex from *Arabidopsis*

thaliana . *J. Biol. Chem.* **277**, 30629–30634 (2002).

76. Aoki, S. K. *et al.* Contact-Dependent Inhibition of Growth in *Escherichia coli*. *Science (80-.).* **309**, 1245–1248 (2005).

77. Aoki, S. K. *et al.* A widespread family of polymorphic contact-dependent toxin delivery systems in bacteria. *Nature* **468**, 439–442 (2010).

78. Hayes, C. S., Aoki, S. K. & Low, D. A. Bacterial Contact-Dependent Delivery Systems. *Annu. Rev. Genet.* **44**, 71–90 (2010).

79. Johnson, P. M. *et al.* Unraveling the essential role of CysK in CDI toxin activation. *Proc. Natl. Acad. Sci.* **113**, 9792–9797 (2016).

80. Benoni, R. *et al.* Activation of an anti-bacterial toxin by the biosynthetic enzyme CysK: mechanism of binding, interaction specificity and competition with cysteine synthase. *Sci. Rep.* **7**, 8817 (2017).

81. Burkhard, P. *et al.* Three-dimensional structure of *O*-acetylserine sulfhydrylase from *Salmonella typhimurium*. *J. Mol. Biol.* **283**, 121–133 (1998).

82. Peterson, E. A. & Sober, H. A. Preparation of Crystalline Phosphorylated Derivatives of Vitamin B6. *J. Am. Chem. Soc.* **76**, 169–175 (1954).

83. Gaitonde, M. K. A spectrophotometric method for the direct determination of cysteine in the presence of other naturally occurring amino acids. *Biochem. J.* **104**, 627–633 (1967).

84. Luchini, A., Espina, V. & Liotta, L. A. Protein painting reveals solvent-excluded drug targets hidden within native protein–protein interfaces. *Nat. Commun.* **5**, 4413 (2014).

85. Kobe, B. *et al.* Crystallography and protein–protein interactions: biological interfaces and crystal contacts. *Biochem. Soc. Trans.* **36**, 1438–1441 (2008).

86. Haymond, A. *et al.* Protein painting, an optimized MS-based technique, reveals functionally relevant interfaces of the PD-1/PD-L1 complex and

the YAP2/ZO-1 complex. *J. Biol. Chem.* **294**, 11180–11198 (2019).

87. Laskowski, R. A. *et al.* PDBsum: a web-based database of summaries and analyses of all PDB structures. *Trends Biochem. Sci.* **22**, 488–490 (1997).

88. Lipfert, J. & Doniach, S. Small-Angle X-Ray Scattering from RNA, Proteins, and Protein Complexes. *Annu. Rev. Biophys. Biomol. Struct.* **36**, 307–327 (2007).

89. Svergun, D. I. Determination of the regularization parameter in indirect-transform methods using perceptual criteria. *J. Appl. Crystallogr.* **25**, 495–503 (1992).

90. Chacón, P., Morán, F., Díaz, J. F., Pantos, E. & Andreu, J. M. Low-Resolution Structures of Proteins in Solution Retrieved from X-Ray Scattering with a Genetic Algorithm. *Biophys. J.* **74**, 2760–2775 (1998).

91. Volkov, V. V & Svergun, D. I. Uniqueness of ab initio shape determination in small-angle scattering. *J. Appl. Crystallogr.* **36**, 860–864 (2003).

92. Grishaev, A., Wu, J., Trewhella, J. & Bax, A. Refinement of Multidomain Protein Structures by Combination of Solution Small-Angle X-ray Scattering and NMR Data. *J. Am. Chem. Soc.* **127**, 16621–16628 (2005).

93. Takamoto, K. & Chance, M. R. Radiolytic protein footprinting with mass spectrometry to probe the structure of macromolecular complexes. *Annu. Rev. Biophys. Biomol. Struct.* **35**, 251–276 (2006).

94. Petoukhov, M. V & Svergun, D. I. Global rigid body modeling of macromolecular complexes against small-angle scattering data. *Biophys. J.* **89**, 1237–1250 (2005).

95. Bergmann, A., Fritz, G. & Glatter, O. Solving the generalized indirect Fourier transformation (GIFT) by Boltzmann simplex simulated

annealing (BSSA). *J. Appl. Crystallogr.* **33**, 1212–1216 (2000).

96. Franke, D. & Svergun, D. I. DAMMIF, a program for rapid ab-initio shape determination in small-angle scattering. *J. Appl. Crystallogr.* **42**, 342–346 (2009).
97. Kozin, M. B., Svergun, D. I. & Embl, B. Automated matching of high- and low-resolution structural models. *J. Appl. Cryst* **34**, 33–41 (2001).
98. Panjkovich, A. & Svergun, D. I. SASpy: a PyMOL plugin for manipulation and refinement of hybrid models against small angle X-ray scattering data. *Bioinformatics* **32**, 2062–2064 (2016).
99. Svergun, D., Barberato, C. & Koch, M. H. J. CRYSOL– a Program to Evaluate X-ray Solution Scattering of Biological Macromolecules from Atomic Coordinates. *J. Appl. Crystallogr.* **28**, 768–773 (1995).
100. Burkhard, P. *et al.* Three-dimensional structure of *O*-acetylserine sulfhydrylase from *Salmonella typhimurium*. *J. Mol. Biol.* **283**, 121–133 (1998).
101. Mylonas, E. & Svergun, D. I. Accuracy of molecular mass determination of proteins in solution by small-angle X-ray scattering. *J. Appl. Crystallogr.* **40**, s245-s249 (2007).
102. Wales, T. E. & Engen, J. R. Hydrogen exchange mass spectrometry for the analysis of protein dynamics. *Mass Spectrom. Rev.* **25**, 158–170 (2006).
103. Pan, H., Raza, A. S. & Smith, D. L. Equilibrium and Kinetic Folding of Rabbit Muscle Triosephosphate Isomerase by Hydrogen Exchange Mass Spectrometry. *J. Mol. Biol.* **336**, 1251–1263 (2004).
104. Lanman, J. *et al.* Identification of Novel Interactions in HIV-1 Capsid Protein Assembly by High-resolution Mass Spectrometry. *J. Mol. Biol.* **325**, 759–772 (2003).
105. Engen, J. R. Analysis of protein complexes with hydrogen exchange and

mass spectrometry. *Analyst* **128**, 623–628 (2003).

106. Lee, T. *et al.* Docking Motif Interactions in MAP Kinases Revealed by Hydrogen Exchange Mass Spectrometry. *Mol. Cell* **14**, 43–55 (2004).
107. Skinner, J. J., Lim, W. K., Bédard, S., Black, B. E. & Englander, S. W. Protein dynamics viewed by hydrogen exchange. *Protein Sci.* **21**, 996–1005 (2012).
108. Skinner, J. J., Lim, W. K., Bédard, S., Black, B. E. & Englander, S. W. Protein hydrogen exchange: Testing current models. *Protein Sci.* **21**, 987–995 (2012).
109. Zhang, Z. & Smith, D. L. Determination of amide hydrogen exchange by mass spectrometry: a new tool for protein structure elucidation. *Protein Sci.* **2**, 522–531 (1993).
110. Dyson, H. J. & Wright, P. E. Unfolded Proteins and Protein Folding Studied by NMR. *Chem. Rev.* **104**, 3607–3622 (2004).
111. Deng, Y., Zhang, Z. & Smith, D. L. Comparison of continuous and pulsed labeling amide hydrogen exchange/mass spectrometry for studies of protein dynamics. *J. Am. Soc. Mass Spectrom.* **10**, 675–684 (1999).
112. Englander, S. W. & Mayne, L. Protein Folding Studied Using Hydrogen-Exchange Labeling and Two-Dimensional NMR. *Annu. Rev. Biophys. Biomol. Struct.* **21**, 243–265 (1992).
113. Konermann, L., Pan, J. & Liu, Y.-H. Hydrogen exchange mass spectrometry for studying protein structure and dynamics. *Chem. Soc. Rev.* **40**, 1224–1234 (2011).
114. Engen, J. R. Analysis of Protein Conformation and Dynamics by Hydrogen/Deuterium Exchange MS. *Anal. Chem.* **81**, 7870–7875 (2009).
115. Svejdal, R. R., Dickinson, E. R., Sticker, D., Kutter, J. P. & Rand, K. D. Thiol-ene Microfluidic Chip for Performing Hydrogen/Deuterium Exchange of Proteins at Subsecond Time Scales. *Anal. Chem.* **91**, 1309–

1317 (2019).

116. Miranker, A., Robinson, C. V, Radford, S. E., Aplin, R. T. & Dobson, C. M. Detection of transient protein folding populations by mass spectrometry. *Science (80-.).* **262**, 896–900 (1993).