



SAPIENZA
Università di Roma

PhD Programme in Life Sciences

XXXIV cycle

**One protein many functions: the non-canonical interactions of
SHMT1.**

**The structural and functional characterization of SHMT1
interactions with RNA and in the de novo thymidylate
synthesis complex.**

Candidate

Sharon Spizzichino

Handwritten signature of Sharon Spizzichino in black ink.

Supervisor

Prof. Francesca Cutruzzolà

Handwritten signature of Prof. Francesca Cutruzzolà in black ink.

Coordinator

Prof. Francesca Cutruzzolà

Handwritten signature of Prof. Francesca Cutruzzolà in black ink.

Sharon Spizzichino

Summary

Several enzymes that were originally characterized to have one defined function in intermediary metabolism, have now been found to participate in variety of other cellular processes [1,2]. In fact, some metabolic enzymes, named moonlighting proteins, have been found to non-canonically interact with other proteins or nucleic acids. Interestingly, not all the newly found RNA binding proteins have been shown to regulate the nucleic acid's function, identifying a novel process known as riboregulation. The catalytic activity of human serine hydroxymethyltransferase (SHMT1) has been found to be controlled by RNA [3]. Moreover, it has been demonstrated that this enzyme takes part in the assembly of the *de novo* thymidylate synthesis complex (dTMP-SC) in the nucleus, together with thymidylate synthase (TYMS) and dihydrofolate reductase (DHFR) [4].

Therefore, the aim of the present thesis is to investigate the details of the non-canonical functions of SHMT1, including the molecular mechanism at the basis of SHMT1 riboregulation. Here I will present the first 3D structure of a metabolic enzyme in complex with RNA, solved using cryo-electron microscopy (cryo-EM), and the identification of the RNA features governing the interaction with SHMT1. In parallel, I will present a deep biochemical and molecular characterization of the dTMP-SC, providing the evidence that the complex can transiently form *in vitro* and in the cytosol as well. An observation that opens a novel scenario on its relevance in different processes.

The present manuscript set the basis to fully understand how metabolic enzymes can be controlled by RNA and strengthen the idea that transient interactions might regulate the organization of biomolecular function in response to the cellular needs.

TABLE OF CONTENT

1. Introduction	1
1.1 The complex world of transient interactions.....	2
1.2 Moonlighting proteins.....	4
1.3 Metabolic enzymes that bind nucleic acids.....	6
1.4 Moonlighting vs riboregulation: Who regulates whom?.....	10
1.5 Beyond the catalytical activity of serine hydroxymethyltransferase (SHMT1).....	13
1.5.1 SHMT1.....	13
1.5.2 SHMT1 is a moonlighting protein but it is also riboregulated.....	18
1.5.3 Identification of the RNA-binding region.....	20
1.5.4 SHMT1 and the dTMP-synthesis complex.....	22
2. Aim of the thesis	26
3. Materials and Methods	28
3.1 The <i>de novo</i> thymidylate synthesis complex.....	28
3.1.1 <i>In situ</i> -Proximity ligation assay (is-PLA).....	28
3.1.2 Immunofluorescence (IF).....	29
3.1.3 Cells fractionation.....	30
3.1.4 Microscopy analysis on fixed samples and quantitative analysis of fluorescent signals.....	31
3.1.5 Chimera rational design.....	32
3.1.6 Proteins expression and purification.....	32
3.1.7 Circular Dichroism (CD) and Thermal Melting (TM) spectroscopic analysis.....	34
3.1.8 Activity assays.....	35
3.1.9 Far-Western blot.....	35

3.1.10 Immunopurification with Protein G-plus agarose beads.....	36
3.1.11 Computational model of dTMP-SC.....	37
3.1.12 Bio-layer interferometry (BLI).....	38
3.1.13 Differential Scanning Fluorimetry (DSF).....	39
3.2 SHMT1-RNA complex.....	40
3.2.1 Cryo-electron microscopy (Cryo-EM).....	40
3.2.1.1 Negative staining.....	41
3.2.1.2 Cryo- EM sample and grid preparation.....	42
3.2.1.3 Data collection.....	44
3.2.1.4 Data processing.....	45
3.2.1.5 Cryo-EM maps resolution estimation and sharpening.....	45
3.2.1.6 Model building.....	46
3.2.2 <i>In vitro</i> experiments.....	47
3.2.2.1 Site-directed mutagenesis.....	47
3.2.2.2 Electrophoretic mobility shift assay (EMSA).....	48
3.2.2.3 Enzymatic assays.....	48
4. Results.....	50
4.1 The <i>de novo</i> thymidylate synthesis complex.....	50
4.1.2 Intracellular localization of the complex.....	50
4.1.3 <i>In vitro</i> assembly.....	58
4.1.3.1 Rational design, expression, and purification of the chimeric bi-functional protein.....	58
4.1.3.2 Far Western blotting and Immuno-Purification..	64
4.1.3.3 Quantification of the dissociation constant of the complex.....	66

4.1.3.4 Effect of substrates and ligands on dTMP-SC formation.....	68
4.2 The SHMT1-RNA complex.....	70
4.2.1 The first structure of the SHMT1-RNA complex by cryo-electron microscopy (cryo-EM).....	71
4.2.1.1 Cryo-EM grids preparation and data collection of the SHMT1-RNA complex.....	71
4.2.1.2 Single Particle Analysis.....	75
4.2.1.2.1 Image processing.....	75
4.2.1.2.2 Particle picking and extraction.....	76
4.2.1.2.3 2D classification.....	76
4.2.1.2.4 3D classification.....	77
4.2.1.2.5 Model building and Refinement.....	80
4.2.2 The Cryo-EM structure of SHMT1 without RNA.....	88
4.2.3 <i>In vitro</i> nucleic acid binding experiments.....	89
4.2.3.1 Validation of the binding region.....	89
4.2.3.2 Effects of the RNA binding on SHMT1 catalytic activity.....	91
4.2.3.3 Identification of the RNA binding determinants.....	95
5. Conclusions.....	97
6. Discussion.....	100
7. References.....	111

1. Introduction

Eighty years ago, George Beadle and Edward Tatum delivered an innovative concept: each gene produces a single enzyme, which affects a single step in a metabolic pathway [5]. This paradigm was then coined by Norman Horowitz as the “one gene-one enzyme hypothesis” [6] and it paved the way to the modern molecular biology field. This once innovative hypothesis has been overthrown by new findings, including the understanding that one gene can specify two or more distinct proteins and that some proteins have more than one function. In fact, in higher organisms the number of genes does not reflect their genetic complexity but is rather related to the genome’s size [7], meaning that there is a versatile use of genes. Moreover, the fact that one protein can perform different tasks is somehow advantageous to the organism, given that, it takes less energy to accomplish two functions with a single protein than synthesizing two proteins that perform one function each.

What it is also interesting to notice is that the cell are now known to be characterized by a crowded environment, in fact, there is a high concentration of macromolecules ($200-400 \text{ g}\cdot\text{L}^{-1}$) [8]. This packed environment modulates macromolecule’s structure, diffusion, association, and function as it promotes intramolecular interactions between neighboring molecules. Many of these interactions are weak and transient, yet sufficiently specific to significantly affect the spatial and temporal distribution of macromolecules and functions in the cell [9].

1.1 The complex world of transient interactions

Proteins perform most of the biological processes that happen within and between cells. Their properties are dynamic and responsive to the cellular needs. These properties are mainly modified and controlled by either interactions with other proteins (protein-protein interactions; PPI) or biomolecules (Chien and Gierasch 2014; Wetie et al. 2013). The whole set of molecular interaction among biological molecules is called the “Interactome” [12] and the nature of these interactions regulates the biological outcome which may lead to the activation or deactivation of a signalling pathway, catalysis, transcription, movements and so on, thus determining the final phenotype. The size of the Interactome in humans has been estimated to be $\approx 650,000$ [13], and the PPI represent a huge part of it [14]. As a consequence of this massive number of proteins interactions, different types of complexes can be formed. For example, the proteasome is a large protein complex responsible for degradation of intracellular proteins that is stable and thus persistent. Instead, other complexes rely on weak transient interactions as in the case of cytochrome c and cytochrome c oxidase in the electron-transfer pathways of respiration [15,16]. Weak interactions, also known as “quinary” interactions [17], lead to the formation of transient complexes, which assemble to perform specific functions, strictly regulated in time and location [11], e.g. the channelling of substrates used to promote the transfer of metabolites required for sequential steps of a metabolic process (for example in the TCA cycle complex) [11,18].

Transient complexes are extremely important for the cell, as a matter of fact, they are involved in a wide range of biochemical processes such as signal transduction, hormone-receptor binding, allostery of enzymes, correction of misfolded proteins and so on [19].

The field of application of the interactome is not limited only to the study of PPI, but it also analyses the interactions involving proteins

that bind nucleic acids, which might affect fundamental processes such as replication, transcription, and repair in case of DNA or translation, splicing and sequencing in case of RNA [20]. Also in this case, the recognition of the nucleic acid by proteins might produce stable assemblies (such as the ribosome) or transient and dynamic complexes (for example the DNA clamps sliding through the DNA) [21].

Weak complexes are now emerging to play a crucial role in the cell, especially because of the tunability of their bound and unbound populations [22]. Due to the nature of these interactions (micromolar to millimolar dissociation constants) and to the limitation in the detection technologies and methodologies, for a long time these interactions were not taken into account [23] and thought to be non-specific interactions. In contrast, they are extremely specific and are believed to be involved also in the organization of biomolecular function in response to the cellular needs [24]. Another interesting theory is that these transient interactions are considered to have a major role in phase transition and in the formation of membrane-less organelles, which are cellular regions that have a different density and composition that may fulfil a specific function within the cell [24,25]. The mechanism through which transient interactions can promote phase separation is yet to be clarified, but all these evidences are suggesting that further studies are needed to shed light into the complex network of interactions taking place in living organisms.

Zuckerandl and Pauling in 1962 had wrote that “Life is a relationship between molecules, not a property of any one molecule. So is, therefore, disease, which endangers life” [26]. In fact, it is getting clearer that inside the cell, biological functions are not performed by individual proteins, but most often by transient complexes they are able to form. For these reasons, many efforts have been made to design inhibitors of selected transient interaction

as it is now clear that transient interactions are a common way for cells to increase the efficiency, tunability and control over crucial metabolic pathways.

For example, nutlin is an inhibitor that targets the transient interaction between murine double minute 2 (MDM2) and p53, in order to stabilise P53. In fact, when complexed with MDM2, p53 loses its ability to bind DNA and it is degraded [19]. Thanks to emerging technologies, the repertoire of transient complexes will surely expand in the future years, opening novel therapeutic routes and contributing to the understanding of how many biological systems work.

In this extremely complex world, it is not surprising that many proteins had been found to non-canonically and transiently interact with many different partners within the cell, and during the last few years these multifunctional proteins, also known as moonlighting proteins, have gained a special attention.

1.2 Moonlighting proteins

In 1999, Costance J. Jeffery has used for the first time the term “moonlighting” when referring to a protein that displays two or more physiological biochemical or biophysical function [27]. By definition, the multiple functions of a moonlighting protein aren't the results of gene fusion, alternative mRNA splicing or post-translational modification. Also, proteins that use multiple substrates or that catalyse multiple steps in the same metabolic pathway are not considered moonlighting proteins [27]. Despite the many restrictions, over 500 proteins have been identified as moonlighting [28].

Examples of such proteins (figure 1) include enzymes that bind also to nucleic acids and regulate translation or transcription, cytosolic enzymes that also function as chaperones, cytokines or receptors on the cell surface and enzymes that interact with other proteins [29].

Many moonlighting proteins can concurrently perform their multiple functions, others instead, are going through changes of function in response to changes in the environment [30]. Moreover, while some proteins can use the active site pocket to process their secondary task, other use a different part of the protein's surface to perform this duty.

In this very complex scenario, the understanding of how the secondary functions evolve remains an open question. Occasionally, proteins had adapted to the secondary function without undergoing substantial structural changes, in other cases modification of unused regions, large conformational changes or transitions between structured and intrinsically unfolded domains have been acquired [29,31].

Given the variety of physical features, activity and cellular localization, the identification of moonlighting proteins wasn't always straightforward [32]. Nevertheless, the number of proteins that were found to moonlight has exponentially grown over the past few years. These new findings helped to better understand many physiological and pathological biochemical pathways. In fact, numerous moonlighting proteins have been found to play a key role in disease. A striking example is phosphoglucose isomerase (PGI), a glycolytic enzyme, that also functions as an extracellular cytokine under the name of autocrine motility factor (AMF) and plays a central role in breast cancer metastasis [33].

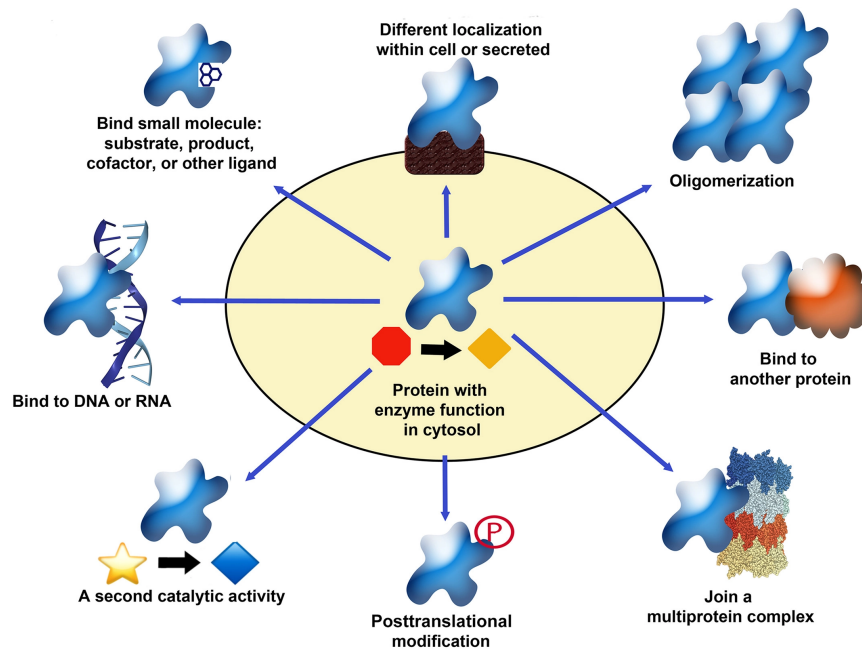


Figure 1. The different functions of a moonlighting protein. An intracellular enzyme can also be a cell surface receptor, a part of a multiprotein complex, a secreted signalling protein, or a transcription factor. The expression of other proteins, concentrations of substrates and other ligands, and post-translational modifications can all affect the role it plays in a given cell type and under different conditions. Picture taken from [34].

As previously mentioned, proteins can moonlight in many different ways, some of which are not mutually exclusive [35]. For example, Enolase, a cytoplasmatic enzyme, has also a role in pathogenic bacteria as a cell surface receptor. Moreover, it has been demonstrated that it also moonlights as an RNA folding chaperone helping the formation of a preMsk1p-tRNA complex [35–37].

1.3 Metabolic enzymes that bind nucleic acids

Another important subset of interactions is represented by protein-nucleic acids interactions. For years, enzymes involved in the intermediary metabolism were thought to have only one defined function, i.e. catalyse a given step of a metabolic pathway. More recently, many of them had been shown to moonlight as nucleic acid binding proteins (figure 2) [38]. The discovery of these new features dates back to the late 1950s, when some glycolytic enzymes were found to be present also in the nucleus [39–41]. After these findings, scientists began to ask what the purpose of those enzymes in the nucleus was. An intriguing possibility was that glycolysis was taking place also in the nucleus, but not enough evidence was produced to demonstrate it [39]. This raised the hypothesis that glycolytic enzymes could moonlight in the nucleus as DNA binding proteins [39]. Strikingly, although lacking the canonical DNA binding domains, several metabolic enzymes were found to regulate gene transcription by either binding to DNA or to transcription factors [1]. By using this non-canonical activity of some metabolic enzymes, cells can adapt and coordinate their metabolic state and gene transcription in response to external cues, a mechanism now known as the metabolism-gene transcription axis [1]. For example, malate dehydrogenase 1 (MDH1), an enzyme involved in the tricarboxylic acid (TCA) cycle, translocates into the nucleus upon glucose deprivation, increasing p53 stability. Indeed, in the nucleus MDH1 binds to p53 directly, facilitating its phosphorylation and acetylation as well as its transcriptional activity [42].

Besides the ability to bind DNA, many metabolic enzymes were also found to interact with RNA. The list of such proteins is constantly expanding, and it has been proposed that metabolic RNA-binding enzymes take part in regulatory network linking a metabolic function to a regulatory mechanism, the “REM (RNA-Enzymes-Metabolites) theory” [38,43]. Again, these moonlighting proteins participate in different metabolic pathways, such as glycolysis, the

tricarboxylic acid (TCA) cycle, lipid metabolism, and nucleotide biosynthesis. For example, thymidylate synthase (TYMS) and dihydrofolate reductase (DHFR), two enzymes involved in the *de novo* thymidylate biosynthesis pathway, can moonlight as RNA-binding proteins (RBPs) and bind to their own mRNA, regulating their own expression [44,45].

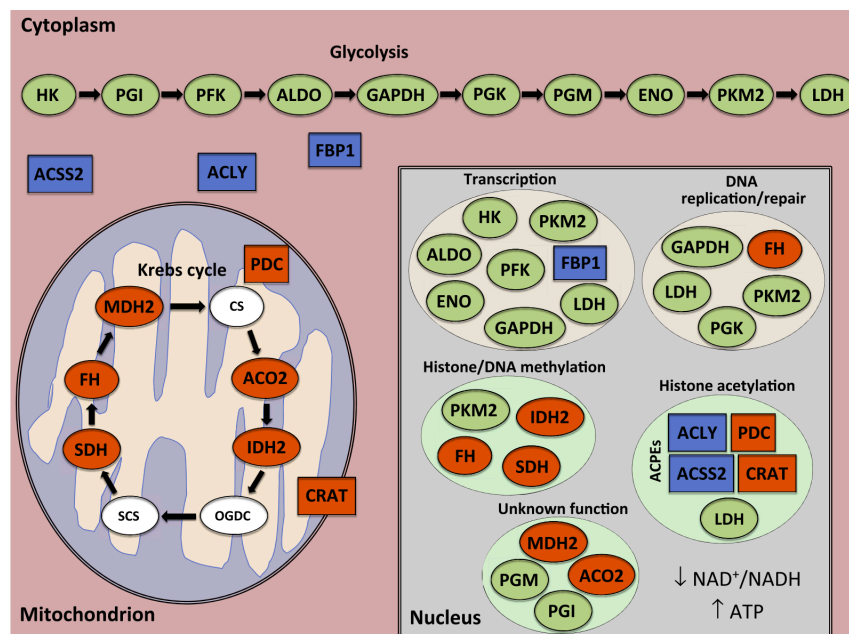


Figure 2. Metabolic Enzymes Moonlighting in the Nucleus. Cytoplasmic [glycolytic (green) and non-glycolytic (blue)] and mitochondrial enzymes found to translocate into the nucleus. While all the essential glycolytic enzymes are found in the nucleus, only some Krebs cycle enzymes have been described in the nucleus (orange) whereas others have not (white filling). In the nucleus, the translocated enzymes have been grouped together according to their proposed nuclear function, whether canonical (i.e., production of a metabolite with secondary effects in nuclear functions) or non-canonical (i.e., a kinase function, participation in a transcription complex). Picture taken from [39].

The development of high throughput studies, such as RNA interactome capture, has allowed to identify hundreds of novel RBPs [2]. Interestingly, most of the thus identified RBPs lack the canonical RNA-binding domains (RBDs) such as the RNA recognition motif (RRM), hnRNP K homology domain (KH) or DEAD box helicase domain [37]. An extensively characterized example of a metabolic enzyme moonlighting as a RBP is glyceraldehyde-3-phosphate dehydrogenase (GAPDH), a glycolytic enzyme that oxidizes glyceraldehyde-3-phosphate to reduce nicotinamide adenine dinucleotide (NAD⁺).

The RNA-binding activity of GAPDH was discovered almost 30 years ago, when it was demonstrated that it preferentially binds to AU-rich RNA elements in the 3' untranslated regions (UTRs) of lymphokine mRNAs [46]. The presence of NAD⁺ interferes with RNA binding but it is required for the glycolytic function, meaning that the Rossmann fold (a domain specialized to bind nucleotide enzymatic cofactors) is involved also in the RNA binding. Physiologically, GAPDH can control cytokine production. In fact, in resting T cells GAPDH is bound to the 3'-UTR of cytokine mRNAs, including interferon γ , inhibiting their translation. When activated, T lymphocytes switch their metabolism from oxidative phosphorylation to aerobic glycolysis and GAPDH disengages from RNA, allowing cytokine production. The switch between GAPDH regulatory and enzymatic activity is due to the competition between the RNA and the NAD⁺ for the binding to the same site, given that in fact, as mentioned before, they both bind to the Rossmann fold. In this way, the metabolic redox state of the cell can direct the activity of GAPDH between its canonical and moonlighting functions [47]. Interestingly, cytokine mRNA sequences can inhibit the enzymatic activity of GAPDH *in vitro*, as RNA and NAD⁺ share the binding site [2]. This modulation of the enzyme activity by RNA has raised an important question: is it only the protein that affects RNA expression or may RNA also affect the function of a metabolic enzymes?

1.4 Moonlighting vs riboregulation: Who regulates whom?

In cases where enzymatic and RNA-binding functions are competitive, like for GAPDH, it is perfectly plausible that the catalytic activity of an enzyme could be regulated by RNA binding. But some of the new proteins identified to bind RNA were not found to have the conventional RNA binding domain nor to have moonlighting properties (figure 3), leading the authors to question which their function as RNA-binding proteins was. Surprisingly, in these cases, RNA can regulate protein function. As shown in figure 4, a moonlighting protein is a protein that binds to RNA and by doing so it affects RNA processing, stability, localization, function, and translation. However, upon RNA binding, the biological function of a protein can be also altered through a phenomenon known as “riboregulation”.

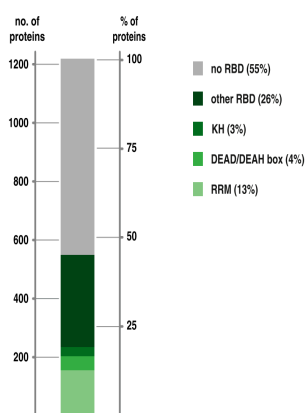


Figure 3. RNA interactome capture discovers many RNA-binding proteins that lack identifiable RNA-binding domains. RNA interactome capture from different human cell lines identified a total of 1218 proteins as RNA binders, most of which do not contain an identifiable RBD (~55%). The remaining proteins harbour domains known to bind RNA, most commonly the RNA recognition motif (RRM, accounting for ~13% of the proteins), DEAD/DEAD box helicase domain (accounting for ~4%) and the K homology domain (KH, accounting for ~3%) [48].

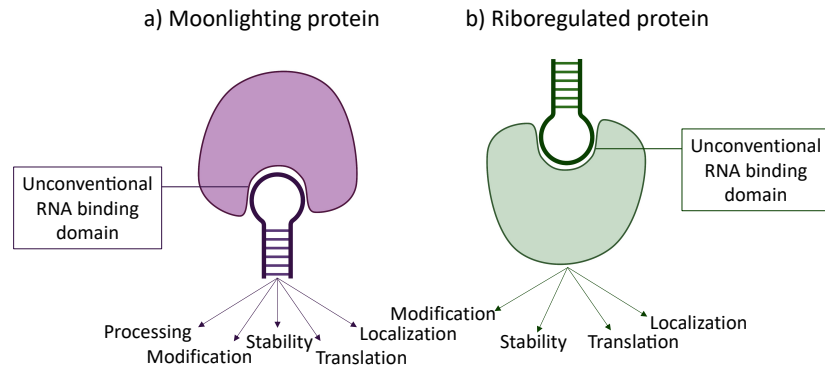


Figure 4. Functional crosstalk between proteins and RNA. a) A Moonlighting protein can interact with an unconventional RNA-binding domains to regulate RNA metabolism and function. B) Mechanism of riboregulation: the RNA can bind to the RBP to affect its fate and function. Image inspired by [37].

Autophagy receptor p62 was identified as an RNA binding protein, and it engages the small non-coding vault RNA1-1 (vtRNA1-1) as a major interacting RNA [49]. Strikingly, in this study the authors had demonstrated that vtRNA1-1 riboregulates p62 by interfering with its multimerization, therefore inhibiting autophagy. The amount of vtRNA1-1 inside a cell varies according to the cell's nutritional status. When is deprived of amino acids, vtRNA1-1 is reduced to enhance autophagy that will refill the pool of amino acids from unnecessary proteins to cover the cellular needs [49]. Enolase 1 (ENO1) was also found to be riboregulated by RNA. ENO1 is a metabolic enzyme that catalyzes the reversible interconversion between 2-phosphoglycerate (2-PG) and phosphoenolpyruvate (PEP). ENO1 had previously been shown to bind RNA in different organisms and that it can interact with a wide range of RNAs in HeLa cells with a preference towards the 5'untranslated region (5'UTR) of mRNAs. RNA interacts with enzyme in a way that is mutually exclusive with substrate binding. Such examples would

represent a novel paradigm for biological regulation, by which enzyme activity and metabolic functions can be regulated by RNAs.

The fact that many of the newly discovered RBPs lack the classical RNA-binding domain has raised the question of how they can interact with RNA. An intriguing hypothesis would be that RNA is the driving force in this interaction. Indeed, RNAs could act as scaffolds to orchestrate protein networks and control their activity, suggesting that nucleic acids might play an important regulatory role in many cellular processes, including metabolism [48]. The available structures of proteins containing multiple RNA binding domain in complex with their ligand unveiled insights into molecular mechanisms and principles of how protein domains cooperate to recognize longer, continuous stretches of RNA to increase both RNA affinity and specificity [50]. However, the number of structures of protein-RNA complexes is still very limited, and to date, no structure is available for the unconventional RBP in complex with RNA. Other techniques, such as RNA interactome capture combined with mass spectrometry, have been developed to have structural clues on how these proteins can interact with RNA. For example, experiments performed in p62 revealed that it preferentially binds with looped regions of the vtRNA central domains without an apparent binding motif [49]. Indeed, alike to proteins, also RNA can fold into elaborate 3D structures that have various functions. Complex RNA secondary structures that promote viral translation can be found in the 5' untranslated region of many different viruses (e.g hepatitis C virus, herpes simplex virus, picornaviruses) suggesting that viral RNA evolved to specifically bind cellular proteins to support viral translation [48]. Also, the discovery of the long non-coding RNAs (lncRNAs) has triggered many efforts in understanding their functions. lncRNAs were found to control the accessibility to genetic information by interacting with chromatin-modifying enzymes and nucleosome-remodeling factors or to organize, scaffold or inhibit protein complexes. Interestingly,

it is thought that lncRNAs are RNAs with protein-binding activity, therefore, they are the driving force of these interactions.

So far, proteins were thought to be regulated by other proteins or by metabolites; all the above-mentioned examples are demonstrating that there is also the possibility in which the known biological function of a protein can be altered through 'riboregulation'. As the list of unconventional RBPs is continuously growing, many more examples of functionally active RNAs will be discovered. Structural studies on these complexes are extremely needed, as they will pave the way for a better understanding of how RNA biology integrates with the protein's functions.

In this thesis, a paradigmatic example of this complexity will be studied, i.e., a metabolic enzyme that is also able to bind nucleic acids and that transiently interacts with several binding partners, yielding diverse macromolecular complexes.

1.5 Beyond the catalytical activity of serine hydroxymethyltransferase (SHMT1)

1.5.1 SHMT1

SHMT1 is a cytosolic metabolic pyridoxal-5'phosphate (PLP)-dependent enzyme that catalyses the reversible conversion of serine and tetrahydrofolate respectively into glycine and 5,10-methyltetrahydrofolate (figure 5). This reaction is the main source of one-carbon (1C) units inside the cell, which are then used to fuel several biosynthetic pathways such as the synthesis of purines and pyrimidines and the methylation of homocysteine in methionine [51]. This enzyme embodies all the characteristic that we have listed so far. In fact, it is a moonlighting enzyme that non-canonically binds RNA and, in this way, regulates the expression of its

mitochondrial counterpart (encoded by a different gene and named SHMT2). Additionally, it is also riboregulated by RNA, in fact, upon RNA binding the catalytical activity of SHMT1 is modulated [3]. In addition, it was suggested that SHMT1 interacts with two other metabolic enzymes (DHFR and TYMS) to form a transient complex in the nucleus. For all these reasons we have decided to focus our attention on this extremely versatile enzyme.

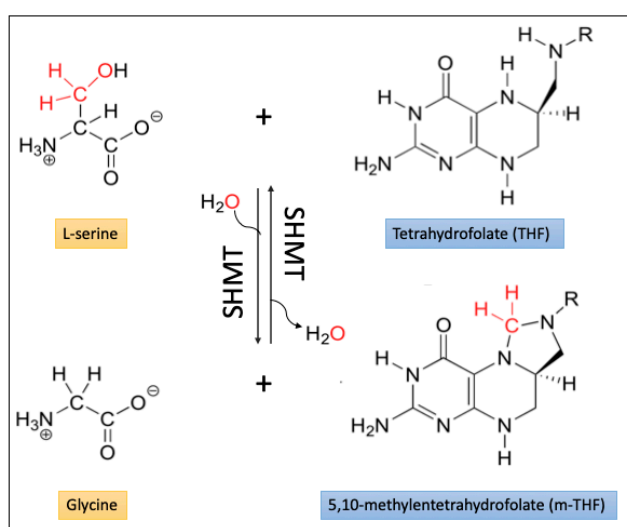


Figure 5. Reaction catalysed by SHMT with L-serine or glycine and mono-glutamylated folates as substrates

As mentioned before, the human genome contains two *shmt* genes, *shmt1* and *shmt2*, encoding three main SHMT isoforms. SHMT1 and SHMT2 α are localized in the cytosol and SHMT2 in the mitochondria.

The isozymes are similar, they both are homotetramers (dimers of obligate dimers) of subunits of ~ 55 kDa, but they are not identical. In fact, although they share 66% of amino acids sequence identity, the quaternary structure of the apo-form (in the absence of PLP) of the enzymes is different. In fact, apo-SHMT1 has a stable

conformation that allows the tetramer assembly, whereas apo-SHMT2 displays a dimeric structure which converts into the holo-tetramer only upon PLP binding. The obligate dimer, also called the “tight dimer”, is the minimal catalytically active unit which form the tetramer (dimers of obligate dimers) in eukaryotes [52].

From a structural point of view, as for many other PLP-dependent enzymes, also the SHMT1 monomer has an overall fold of the α class. The monomer has three domains (figure 6): the N terminus domain, a “large” and a “small” domain. The N terminus domain (residues 1-53) folds into two α helices and one β strand. The “large” domain (residues 53-321) is located at the N terminus and it folds into an $\alpha\beta\alpha$ sandwich containing nine α helices wrapped around a seven-stranded mixed β sheet. This domain binds the PLP cofactor. The “small” C-terminal domain (residues 322-480) folds into three-stranded β -sheets, covered with helices on one side [53,54]. The N terminus is essential for the formation of the tight dimer; indeed, the two N-terminal helices work as clamps that tightly hold together the dimer. Human SHMT1 is an homo-tetramer in solution, and this oligomeric architecture is also observed in the crystals (figure 6) [53].

The active site is located at the interface of the obligate dimer, where PLP is bound to Lys257 forming an internal aldimine and interacting with residues belonging to both subunits of the obligate dimer [52,53].

SHMT has been widely studied given its central role in the one-carbon (1C) metabolism, a KEY metabolic pathway that is reprogrammed in many types of tumors [55].

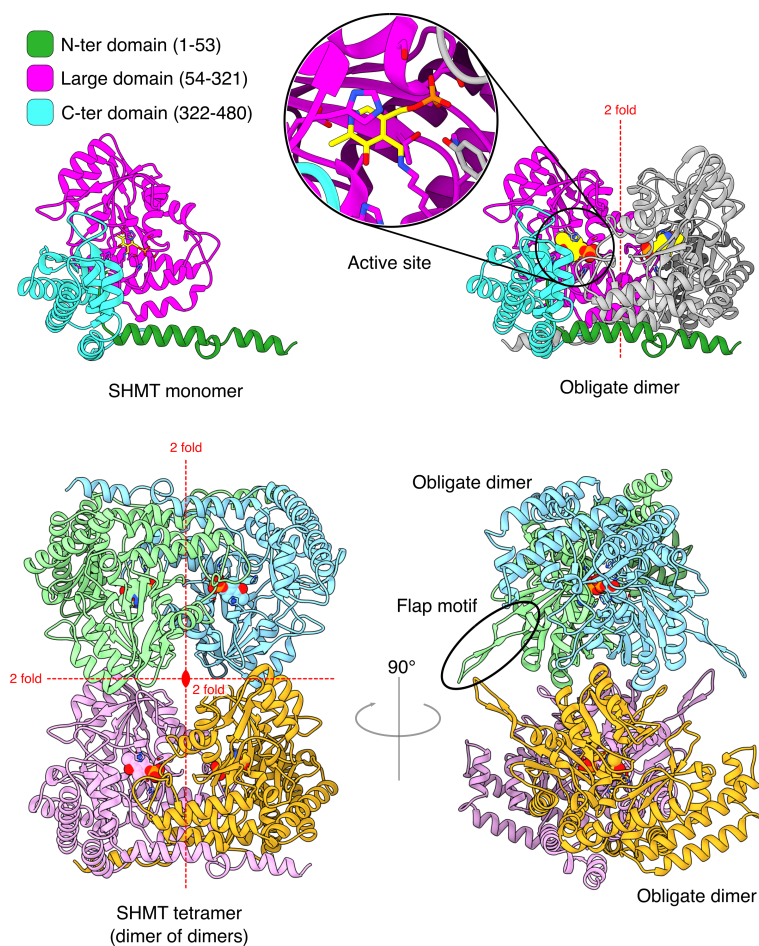


Figure 6. The overall fold of the SHMT tetramer. In the upper left panel, the ribbon diagram of the monomer is shown PDB: 1bj4 [53]. The N terminus is highlighted in green, the large domain in magenta and the C-terminal domain in light blue. In the right upper panel, the obligate dimer is shown, and the active site and the cofactor (PLP) are highlighted. The lower panel shows the tetrameric assembly of SHMT1. Each monomer is represented with a different colour. The obligate dimer is composed of the green and light blue subunits (upper part of the figures), while the other one is formed by the monomers coloured in pink and yellow (lower part of the figures). In the 3D structure the flap motif is highlighted with a black circle.

1C metabolism is compartmentalized within cytosol and mitochondrion with dedicated enzymes (figure 7). Serine metabolism is often dysregulated in tumors [56] and, not surprisingly, cancer cells often overexpress one or more SHMT isoforms. For this reason, both isoforms are considered appealing chemotherapeutic targets [57]. A lot of efforts have been made to develop small molecules able to selectively inhibit the cytosolic or the mitochondrial SHMT isoform; unfortunately, to date, all the identified inhibitors are still far from the clinical application [58].

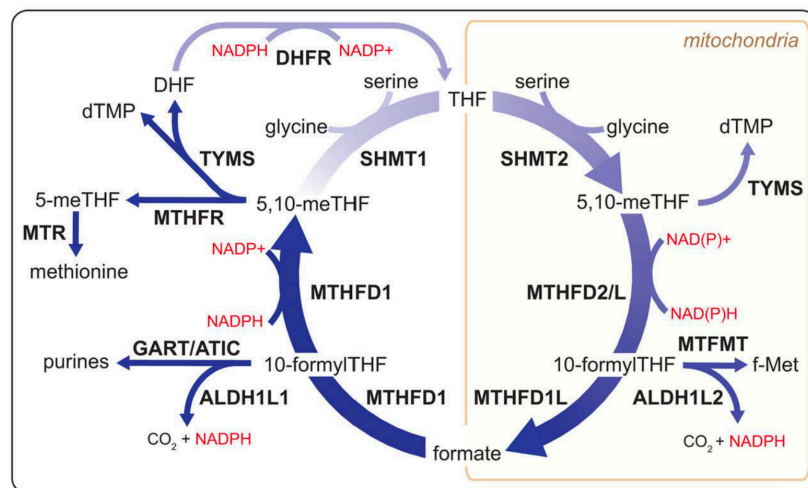


Figure 7. Compartmentalization and products of one-carbon metabolism. Through an interlinked set of mitochondrial and cytosolic reactions, folate metabolism supports 1C anabolic reactions. All abbreviations are standard gene names. Certain descriptions utilize the common protein name for clarity [59].

1.5.2 SHMT1 is a moonlighting protein but it is also riboregulated

In the last two decades, many metabolic enzymes have been identified as non-canonical RNA binding proteins and SHMT1 was recognised to be one of them. Liu et al had performed an *in vitro* analysis showing that the cytosolic isozyme binds to the 5'UTR of its own mRNA, inhibiting its translation [60], proposing a novel mechanism of post-transcriptional autoregulation for SHMT1.

A decade later, a peculiar crosstalk between SHMT1 and SHMT2 expression was observed in normal and tumour cells by our group. As a matter of fact, experiments performed on H1299 human non-small cell lung adenocarcinoma cell line showed that SHMT1 knockdown triggered apoptosis and induced a compensatory increase in the SHMT2 protein expression. On the other hand, the knockdown of SHMT2 did not have any effects on the expression levels of the cytosolic counterpart in the same system [57]. Moreover, SHMT1 overexpression decreased total SHMT2 levels [61]. We then hypothesised that the crosstalk mechanism between the cytosolic and mitochondrial isoform could be related to SHMT1 moonlighting ability to bind RNA. Therefore, it was investigated whether SHMT1 could regulate the expression of SHMT2 by binding to its transcripts via 5'UTR recognition. Indeed, SHMT1 was found to bind specifically and with high affinity to the 5'UTR of SHMT2 (5' UTR2), affecting the expression and the translation of the corresponding transcript [3]. These data suggested that SHMT1 moonlights as an RNA binding protein, regulating SHMT2 expression levels. Surprisingly, SHMT1 is also riboregulated by RNA as penta-glutamylated folate competes with UTR2 for binding to SHMT1, an effect that is enhanced by increasing the glycine concentration. Keeping in mind that SHMT1 catalyzes the reversible conversion of serine and tetrahydrofolate (THF) into glycine and 5,10-methylenetetrahydrofolate (CH₂-THF), our studies have shown that, upon RNA binding, serine cleavage is inhibited by the

presence of UTR2 RNA ($IC_{50} \sim 33$ nM), whereas the opposite reaction is not affected [3]. Previous evidences had shown that SHMT1 is able to switch its activity by producing serine from glycine and vice versa, depending on the metabolic needs of the cell [62]. Now, these new evidence suggests that RNA can selectively riboregulate SHMT1 activity, taking part to the control of serine consumption by the cytosolic SHMT.

A putative mechanistic interpretation of how RNA can inhibit only one of the two semi-reaction is illustrated in figure 8. The SHMT-serine and SHMT-glycine complexes have a different conformational status. In fact, L-serine binding induces an open-to-close conformational change that rearranges a protein region close to the active site, while the binding of glycine doesn't induce such change [63]. Hence it is plausible that RNA has a different affinity towards those dissimilar conformations or that RNA binding could differentially affect the binding of folates. In this scenario, the binding of RNA to SHMT1-serine complex could hamper the binding of folate substrate more than the binding of the RNA to the SHMT-glycine complex. Coherently, a similar behavior was reported for small molecules inhibitors of SHMT1, which showed to have a stronger inhibitory effect on the serine cleavage reaction with respect to the serine synthesis one [58]. Structural studies are therefore urgently needed to unveil how this riboregulation mechanism could work.

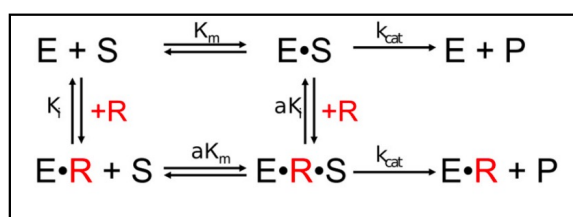


Figure 8. Putative mechanism of inhibition of SHMT1 catalysis by RNA

Equilibria describing hyperbolic competitive inhibition between SHMT1 and RNA. In this particular case, a saturating amino acid substrate concentration (either L-serine or glycine) is present and therefore E corresponds to the enzyme–amino acid complex. E•S is the ternary complex with the folate substrate (either THF or CH₂THF). If RNA (R) is present it binds to both E and E•S forming E•R and E•R•S complexes, respectively [3].

1.5.3 Identification of the RNA-binding region

In order to identify the residues involved in the binding of RNA, site directed mutagenesis experiments had been performed on SHMT1 in a region close to the folate binding site, which represent a positively charged cleft of the enzyme. In fact, it is known that the formation of a protein-nucleic acid complex is often mediated by positively charged side chains (such arginine and/or lysine) and by aromatic residues [64]. Therefore, lysine 157 and 158 were mutated to serine in a double mutant (K157S-K158S). These residues flank the folate binding site and notably, one of the two lysins makes a salt bridge with the glutamic end of the folate (figure 9A) [3]. Also, tyrosine 82, that forms an aromatic stacking interaction with the aminobenzoic moiety of the folate, was mutated in alanine in the single mutant (Y82A) (figure 9B). This residue is located on the other subunit of the obligate dimer with respect to the lysine residues.

As shown in figure 9, the double mutant K157S-K158S strongly affects the RNA binding, whereas the mutation of tyrosine 82 has only a slight effect (2-fold). These data confirmed the hypothesis of a partial overlapping between the folate and the RNA binding sites, highlighting an important role of the positively charged region surrounding the folate binding cleft in the RNA binding event [3].

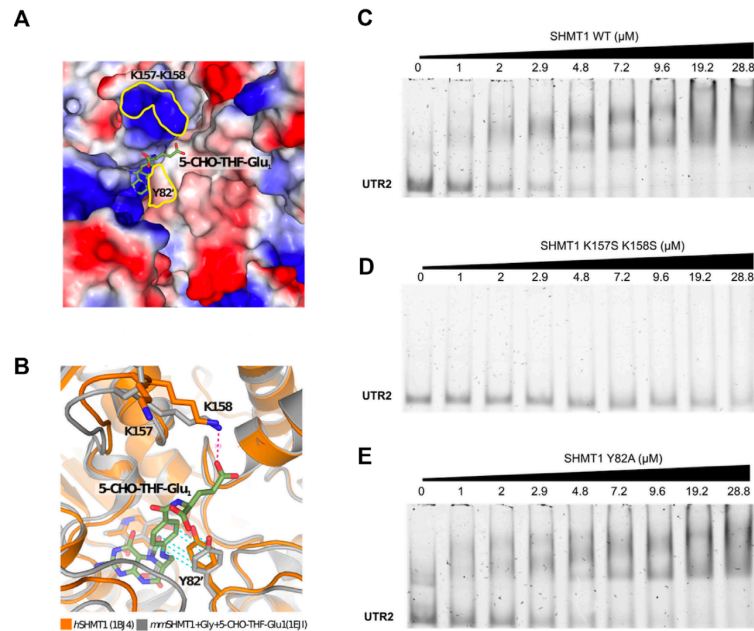


Figure 9. SHMT1 amino acid residues involved in folate and RNA binding. (A) Surface representation of the folate-binding site of human SHMT1 (hSHMT, PDB id: 1BJ4) coloured by electrostatic potential (in red are the acidic and in blue the basic residues). The substrate (5-CHO-THF-Glu1) from mouse SHMT1 (mmSHMT1, PDB id: 1EJ1) is shown as green sticks and superposed to the human enzyme. The location of K157, K158 and Y82A is highlighted by the yellow contours. (B) Superposition of the human and mouse crystal structures (in orange and grey, respectively) showing the interaction of K158 and Y82A with the folate substrate. (C-E) EMSA assays carried out by incubating UTR2 with the indicated concentrations of SHMT1 wild-type (WT) (C), K157S-K158S (D) and Y82A (E) mutants [3].

1.5.4 SHMT1 and the dTMP-synthesis complex

As previously mentioned, the reaction catalysed by SHMT provides the largest part of the one-carbon units available to the cell, which can then be used in different metabolic pathways. Among them, the *de novo* nucleotide synthesis pathways, and in particular

deoxythymidine monophosphate (dTMP) synthesis is active in many tissues and is critical for the proliferation of different tumours. Maintenance of the physiological dNTP levels is critical for genome stability and alteration in their levels have complex consequences. dTMP is synthesized starting from deoxyuridine monophosphate (dUMP) through the thymidylate cycle, which involves three enzymes: thymidylate synthase (TYMS), dihydrofolate reductase (DHFR) and serine hydroxymethyltransferase (SHMT). In this cycle, SHMT catalyses the formation of 5,10-methylenetetrahydrofolate and glycine from tetrahydrofolate (THF) and serine; TYMS is responsible for the reductive methylation of dUMP to dTMP by oxidizing 5,10-methylenetetrahydrofolate to dihydrofolate (DHF). Lastly, DHFR restores the cycle catalysing the NADPH-dependent reduction of DHF to THF [65].

The dTMP synthesis seems compartmentalized to the mitochondria [4] and the nucleus. In fact, recent studies had highlighted that TYMS, DHFR and SHMT are post-translationally modified by small ubiquitin-like-modifier (SUMO) and translocate into the nucleus during the S and G2M phases. Moreover, it was demonstrated that the *de novo* thymidylate synthesis enzymes are present in the nucleus as a metabolic complex associated with the nuclear lamina, and that SHMT1 works as scaffold proteins during the assembly of this nuclear supercomplex [66]. Complex formation in the nucleus is assumed to be responsible for dTMP synthesis and to prevent genome uracil misincorporation [57,66,67] (figure 10).

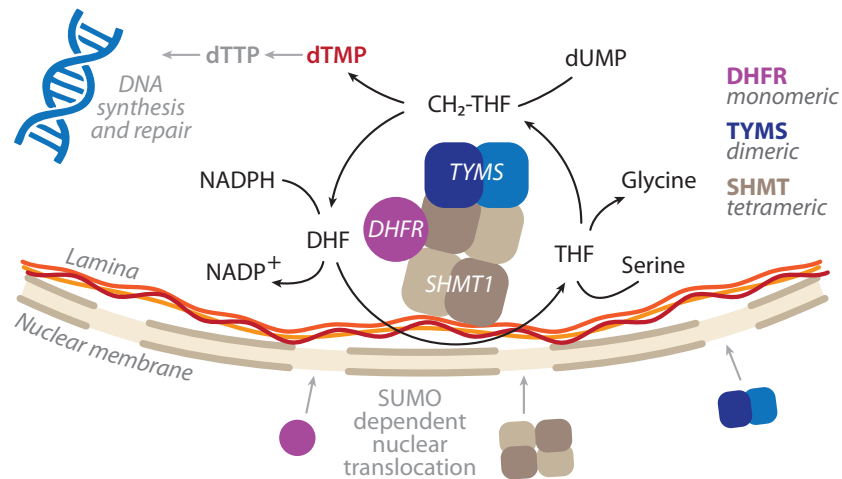


Figure 10. Scheme of the nuclear thymidylate synthesis complex (dTMP-SC) catalytic cycle. SHMT, DHFR and TYMS are SUMOylated and translocate to the nucleus during G1/S-phase, where they are proposed to assemble to form the dTMP synthesis complex (dTMP-SC), anchored to the nuclear lamina. The oligomeric state of the three enzymes is also reported [68].

The first evidence that one carbon metabolism was taking place also in the nucleus was reported already in 1980 in a study that had spotted the activity of DNA polymerase, Thymidine kinase 1 (TK1), DHFR, TYMS, nucleoside 5'-phosphate kinase, and ribonucleotide reductase (RNR) in the S-phase nuclei of the Chinese hamster embryo fibroblast cell line CHEF/18 [69].

This multienzyme complex was named the “replitase”, and it was thought to be composed of both the enzymes involved in the dTMP synthesis and of the ones involved in DNA replication (replisome) in order to deliver directly at the DNA replication fork the precursors needed for the DNA synthesis. The investigation into the role of the nuclear folate-dependent dTMP synthesis had lost momentum after the 1980s, but the interest in such pathway was renewed when SHMT1 was found to be SUMOylated and to translocate into the nucleus [70,71]. Tandem affinity purification of SHMT1-binding

partners from S-phase and UV-treated HeLa cells indicated that SHMT1 physically interacted with the enzymes of the *de novo* dTMP synthesis complex, as well as with numerous enzymes required for DNA replication and repair, many of which were already identified to be part of the replitase complex [70].

Recently, it was also highlighted SHMT1 nuclear function is crucial for cell survival in lung cancer cell lines (A549; H1299) [57], and that *de novo* thymidylate synthesis requires SHMT1 to be active, regardless of its oligomeric state [61]. Furthermore, we have recently found out that SHMT1 is also a DNA binding protein and when transfecting SHMT1 mutants that impaired the catalytic activity from the nucleic acid-binding ability (K157S/K158S) in lung cancer cells, cell viability is reduced to 50%. Remarkably, this effect could only be rescued by supplementation of exogenous thymidylate, indicating that SHMT1 is essentially working both as a scaffolding protein for the assembly of the tertiary complex with TYMS and DHFR, and as an enzyme providing *in situ* the substrates needed for the *de novo* thymidylate synthesis [61].

The *de novo* dTMP synthesis was described to be a crucial metabolic pathway in many tumours [72], in fact, both TYMS and DHFR are already well-known target of chemotherapeutic drugs (antifolates and antimetabolites). But, given their role in DNA replication, the side effects of these treatments are severe and in many cases cancer cells can reprogramme their metabolism by exploiting the moonlighting activity of many of its enzymes. Indeed, both TYMS and DHFR in the absence of substrates bind to their own mRNA repressing proteins' expression [38]. To overcome the salvage pathways activated by cancer cells when treated with antifolates or antimetabolites, conformational drugs could be used. These compounds would alter the conformational equilibrium or disrupt protein-nucleic acids and protein-protein interactions (PPIs), rather than targeting the active site of the enzyme.

The idea would be to use these conformational drugs to prevent the assembly of the dTMP synthesis complex (dTMP-SC), or to inhibit SHMT binding to nucleic acids. Of course, in order to rationally design such compounds, structural studies on this complex are mandatory.

2. Aim of the thesis

It has become evident that cells and their components are dynamic machines that integrate and adapt a wide array of biochemical and mechanical information continuously. Many of the paradigms that were for a long time considered to be true are now being questioned again. For example, the “one gene-one enzyme hypothesis” has been overthrown by new findings, including the understanding that one gene can specify two or more distinct proteins and that some proteins have more than one function. Also, the idea that there is a net boundary between the metabolic network and the RNA world within the cell is now becoming blurrier. In fact, several enzymes that were originally characterized to have one defined function in intermediary metabolism are now shown to participate in several other cellular processes. Among them, they are being found to interact with RNA and either control the nucleic acid’s function (moonlighting protein) or being themselves regulated by the RNAs themselves (ribo-regulation). The new challenges in this field would be to try to re-define these blurred boundaries and to better understand these novel interactions and to which extent the metabolic network is overlapped with other cellular networks via RNA molecules.

In this context, we have found out that SHMT1, a pivotal metabolic enzyme, is also a multitasking protein which interacts with nucleic acids and as well with other metabolic enzymes.

The present work aims to investigate the details of the non-canonical functions of SHMT1, including the molecular mechanism at the basis of SHMT1 riboregulation. Here I will present the first 3D structure of a metabolic enzyme in complex with RNA, solved using Cryo-EM, and the identification of the RNA features governing the interaction with SHMT1.

In parallel, I will present a deep biochemical and molecular characterization of the SHMT1-TYMS-DHFR complex, which is

involved in the *de novo* synthesis of thymidylate, providing a reliable estimate of the affinity of these binding partners, together with the evidence that this complex can transiently form also in the cytosol, an observation which opens a novel scenario on its relevance in different processes.

Overall, the present work fits well in the emerging need of understanding how intermediary metabolism tightly interconnects with the RNA world through its main players: metabolic enzymes, nucleic acids, and metabolites.

3. Materials and Methods

3.1 The *de novo* thymidylate synthesis complex

3.1.1 *In situ*-Proximity ligation assay (is-PLA)

is-PLA was performed in both synchronised and asynchronous A549 and HeLa cell lines, by using the Duolink PLA kit, according to manufacturer's instruction. A549 lung cancer cell lines were purchased from ATCC (Manassas, VA, USA). The cells were grown in RPMI-1640 medium (Corning), supplemented with 100 IU/ml penicillin/streptomycin (P 4458, Sigma

Aldrich) and 10% foetal bovine serum (FBS, Corning). HeLa cells purchased from ATCC (CCL-2), cultured in DMEM and supplemented with 2% penicillin/streptomycin, 2% L-glutamine, 2,5% HEPES and 10% foetal bovine serum. All experiments were run in triplicate and in separate biological sets. Cells were synchronised in S-phase by a single thymidine block. Cells were treated as follows: a) 2mM thymidine (T1895, Sigma-Aldrich) for 24 hours at 37°C, to block DNA synthesis; b) release in a thymidine-free medium; c) after 4 hours cells were fixed in 3.7% paraformaldehyde/30 mM sucrose for 10 min and processed either for the *is*-PLA experiment by using the Duolink PLA kit (DUO92007, SigmaAldrich) or for the immunofluorescence (IF).

The primary antibodies pair to detect the interaction among the three proteins were: i) mouse anti-DHFR (1:50)/ rabbit anti-SHMT1 (1:50) ii) rabbit anti-TYMS (1:20)/ mouse anti-DHFR (1:50) iii) rabbit anti-SHMT1 (1:50)/ mouse anti-TYMS (1:50).

Cells and antibodies were incubated o.n. in a dark humidity chamber at 4°C. Subsequently, the incubation with the PLA probes was performed in a pre-heated humidity chamber for 1 hour at 37°C followed by the ligation addition. If the target proteins (SHMT1-DHFR-TYMS) are interacting among each other or closer than 40 nm, this step will produce a circular DNA. The amplification time

was 100 min and was performed at 37°C in a dark humidity chamber. Negative controls: PLA experiment was performed without one of the primary antibodies and by silencing the expression of SHMT1 by RNA interference. To do so, cells were seeded at the density of 10⁵ cells/well on six-well culture plate in RPMI. 24 h after seeding, cells were transfected with 15 nmol/l siRNA with Qiagen AllStars RNAi. Controls scrambled sequences, or siRNA sequences against shmt1 (Qiagen, Hilden, Germany) using specific siRNA sequences as described in [57] using the JetPRIME reagent (PolyPlus), were used according to the manufacturer's instructions. After 48 h of transfection, the cells were detached and used for PLA experiments.

3.1.2 Immunofluorescence (IF)

IF staining was performed to check whether the cells were synchronised in S-phase, to assess the differential cellular localization of the single proteins (SHMT1, TYMS and DHFR) according to the cellular phase, and to co-localize the proteins. Asynchronous and S-phase synchronised A549 (lung cancer cell lines) were grown on coverslips and fixed in 3.7% paraformaldehyde/30 mM sucrose for 10 min. For the co-localization samples, asynchronous A549 were fixed by using ice cold methanol and incubating the coverslips at -20°C for 6 minutes. Afterwards, cells were permeabilized in 0.1% Triton X-100 and blocked in 3% bovine serum albumin in PBS/0.05% Tween-20 for 1h at room temperature. The incubation with primary antibodies was performed at 4°C o.n. in a dark humidity chamber. Subsequently, the fluorescently labelled secondary antibodies were added in a PBS containing 0.05% Tween 20 and 3% BSA solution and incubated at RT for 30 min. Cells were counterstained with 4,6-diamidino-2-

phenylindole (DAPI, 0.1 µg/mL) and mounted by using the mounting media.

Primary antibodies used: i) rabbit anti-SHMT1 (1:50), ii) mouse anti-DHFR (1:50), iii) rabbit anti-TYMS (1:20), iv) anti-PCNA Alexa Fluor® 488-conjugated antibody (1:1000), v) anti-tubulin-FITC (1:300), vi) anti-chicken tubulin (1:50), vii) mouse anti-TYMS (1:50).

Secondary antibodies used: i) anti-mouse FITC (1:100), ii) anti-rabbit FITC (1:50), iii) anti-mouse Rhodamine (1:50) and iv) anti-rabbit CY3 (1:500) v) anti-chicken alexa fluor 647(1:100).

3.1.3 Cells fractionation

Cell fractionation was performed to assess the cellular localization of the single proteins (SHMT1, TYMS and DHFR) according to the cellular phase.

Nuclear and cytosolic fractionation of A549 cell lines (both asynchronous and S-phase enriched) was performed by treating with trypsin and washing the cells twice in ice-cold PBS. Cells were lysed in 200 µl of cytoplasmic lysis buffer (Tris Ph 7,8 10mM, MgCl₂ 1,5 mM, KCl 10mM, 1mM PMSF) and kept on ice for 10 min. Nuclei were then pelleted at 1000xg for 7 minutes at 4°C, and the supernatant containing the cytoplasmic fraction was removed. Nuclei were resuspended in 60 µl of SDS 10% incubated for 30 min on ice, boiled for 10 min and the DNA fraction was sedimented by centrifugation at 1000xg for 5 minutes at 4°C. Protein concentrations were determined by the BCA assay. Cytosolic and nuclear fractions (20 µg) were separated by SDS-PAGE and transferred into a nitrocellulose membrane. Western blot analysis was performed as usual, and the primary antibodies diluted 1:1000. WB was performed by using mouse anti-SHMT1 (1:1000, Cell signalling), mouse anti-DHFR (1:1000), rabbit anti-TYMS (1:1000). Anti-tubulin (cytosolic marker, 1:1000) and anti-Histone H3 (nuclear marker, 1:5000) were used as controls.

3.1.4 Microscopy analysis on fixed samples and quantitative analysis of fluorescent signals

Images were acquired using a Nikon Eclipse 90i microscope (Nikon Instruments S.p.A., Campi Bisenzio Firenze, IT, USA) equipped with 40× (N.A. 0.75) and 100× (oil immersion, N.A. 1.3) objectives and a Qicam Fast 1394 CCD camera (QImaging) or with an inverted microscope (Eclipse Ti, Nikon) using a 60× (oil immersion, N.A. 1.4) objective and the Clara camera (ANDOR technology). Images were acquired using NIS-Elements AR 3.2 (Nikon) or Nis-Elements H.C. 5.11 using the JOBS module for automated acquisitions; elaboration and processing was performed using NIS-Elements HC 5.02 (Nikon) and Adobe Photoshop CS 8.0. For the quantification of the of fluorescence intensity signals of IF experiments, images of interphase cells of asynchronous and S-enriched populations were acquired using 60x or 40x objectives, along the z axis every 0.4 μm for 5 μm . Signals were measured using Nis Elements H.C. 5.02 (nd2 file format) as follows: sum intensity values of the nuclear fluorescence respect to sum intensity values of the cytoplasmic fluorescence. To avoid differences due to different cell sizes, for each cell the same square mask was used to calculate the nuclear and the cytoplasmic fluorescence. Images were corrected for external background. For is-PLA spots of interaction counts images were acquired with 60x objective, along the z axis every 0.4 μm for 8 μm . The “general analysis” module of Nis Elements H.C. 5.11 was then used for automatic spot count; is-PLA signals were identified based on fixed parameters in all images and those inside the nuclei (defined by DAPI signal) were counted.

Images for quantification fluorescence signals were Maximum Intensity Projections from z-stacks (0.4 μm for a range of 5-8 μm). Statistical analyses were performed with GraphPad Prism 8.

3.1.5 Chimera rational design

The Chimera rational design was performed in collaboration with Professor Alessandro Paiardini. Structural analysis necessary to design and model the two constructs of Chimera was largely conducted within PyMOL [The PyMOL Molecular Graphic System, Version 1.8, Schrödinger, LLC] taking advantage of the PyMod2.0 plugin [73] that allows to perform sequence and structural alignments as well as homology modelling by integrating BLAST, Clustal Omega, MUSCLE, CEalign and MODELLER [74] into PyMOL. The structure of DHFR-TYMS from *B. bovis* (PDB ID 3I3R) was used as a reference to orient the structures of human DHFR and TYMS (PDB ID 1DRF; 1HZW [75,76] the linker regions were modelled using the *de novo* loop modelling function in MODELLER and the N-terminal portion of hTYMS was modelled based on the closest homologous structure (TYMS from *M. musculus* - PDB ID: 4EB4 [77]). The final Chimera models were further refined using the sculpting function in PyMOL and MOE [Molecular Operating Environment (MOE), 2019.01; Chemical Computing Group ULC, 1010 Sherbooke St. West, Suite #910, Montreal, QC, Canada, H3A 2R7, 2020] using the conjugate gradients method for energy minimization. Surface electrostatic potentials were calculated using APBS (Adaptive Poisson-Boltzmann Solver) [77], also available as PyMOL plugin. Sequence conservation analysis was performed using CAMPO [78].

3.1.6 Proteins expression and purification

Long and Short Chimera genes were synthesized and optimized for expression in *Escherichia coli* by GeneArt (Life Technologies), cloned into a pET28a expression vector (Invitrogen), using NdeI and XhoI restriction sites. Small scale expression tests were made on two different *Escherichia coli* expression strains varying different conditions. The best results, in terms of solubility and yield of

protein, were obtained by basal expression of the protein in BL21(DE3) cells, grown at 25° for 24 h, without IPTG, and cell lysis by sonication. Cells were then harvested by centrifugation at 4°C, resuspended in the lysis buffer (0,1M K-phosphate pH 7,5; 1mM PMSF; 4 mg/mL Dnase; 5mM MgCl₂; 5% Glycerol) and lysed by ultrasonic disruption on ice. The supernatant, supplemented with 10 mM imidazole, was loaded on a Ni²⁺-His-Trap column (GE Healthcare, Chicago, IL, USA) equilibrated in 0,1M K-phosphate pH 7,5, 10 mM imidazole. The fractions containing the protein, eluting at 150-200mM imidazole, were concentrated by ultrafiltration (Amicon, Millipore) and then injected onto a Superdex 200 10/300 FPLC column (GE Healthcare, Chicago, IL, USA) equilibrated with buffer K-phosphate 0,1M pH 7,5 and eluted with a flow rate of 0,9 mL/min. The column was calibrated using the following protein standards; aprotinin, carbonic anhydrase, ovalbumin, conalbumin, aldolase, ferritin (Sigma): R² was 0.994. The purified protein eluted with an apparent Mw corresponding to a dimeric state. The purity of the protein was assessed by SDS PAGE. The molar extinction coefficient at 280 nm was determined experimentally by BCA assay (Sigma) according to the manufacturer's specifications and was 1,51 (1 mg/ml) or 91355 M⁻¹cm⁻¹. The purified proteins were aliquoted, frozen in liquid nitrogen and stored at -20° C until use.

Wild-type SHMT1 gene was cloned into a pET22b(+) vector (Novagen) and expressed as N-terminal histidine-tagged fused protein in Escherichia coli (BL21-DE3). Bacterial cultures were grown at 37 °C in Luria-Bertani (LB) liquid medium supplemented with 30 µg/ml kanamycin. When OD600 reached about 0.8, the temperature was lowered to 22 °C and the protein expression was induced by adding 0.1 mM IPTG (Isopropil-β-D-1-tiogalattopiranoside). Cells were harvested 20 hours later by centrifugation and stored at -20 °C. Bacterial pellets were then resuspended in lysis buffer (20mM Hepes pH 7.2; 150 mM NaCl;

5% glycerol) and lysed by ultrasonic disruption on ice. After centrifugation, the supernatant was purified by IMAC on a Ni²⁺-His-Trap column (GE Healthcare, Chicago, IL, USA) in buffer A (20mM Hepes pH 7.2; 150 mM NaCl; 5% glycerol) mixed with different concentrations of buffer B (20mM Hepes pH 7,2 ; 150 mM NaCl ; 5% glycerol, 500mM imidazole). SHMT1 eluted with 300 mM imidazole. Fractions containing pure protein were pooled, and imidazole was removed with desalting columns PD10 (GE Healthcare), the histidine tag was then removed by proteolytic digestion with 1U/mg of thrombin at 16 °C for o.n.. The digestion mix was then loaded again on the IMAC on a Ni²⁺-His-Trap column (GE Healthcare, Chicago, IL, USA) and thrombin was eluted in the flow thought, whereas SHMT1, which now binds with low affinity, was eluted with 100 mM imidazole. The protein was then concentrated and injected into a Superdex 200 column (16/600; GE Healthcare) and eluted with buffer A. The protein elutes as a tetramer. Protein concentration was determined measuring the absorbance at 280 nm and applying the Beer-Lambert law (SHMT1 ϵ_{280} 47565 M⁻¹cm⁻¹). Samples were frozen in liquid nitrogen and stored at -20 °C.

3.1.7 Circular Dichroism (CD) and Thermal Melting (TM) spectroscopic analysis

For both Short and Long Chimera, the CD spectra were collected at a 10 μ M protein concentration in a 0.1 cm quartz cuvette (Hellma) using a JASCO J-710 spectropolarimeter equipped with a Peltier temperature control unit. The thermal stability of the two proteins was measured by increasing the temperature from 30°C to 90 °C with a 1 °C/min rise in T, and monitoring the dichroic signal at 220 nm every 0.5 °C.

3.1.8 Activity assays

Activity assays were carried out at 20°C in 20 mM K-phosphate pH 7.2, containing 75 mM 2-mercaptoethanol, using a Hewlett-Packard 8453 diode-array spectrophotometer (Agilent Technologies, Santa Clara, CA) and a 1 cm pathlength cuvette. For TYMS activity assay, substrate concentration was 0.2 mM CH₂-THF, 0.1 mM dUMP and varying concentration of Chimera (0.3, 0.6 and 1.1 μM). Saturation curves, from which TYMS kinetic parameters were calculated, were obtained using 1 μM Chimera keeping one substrate at fixed concentration (0.1 mM dUMP or 0.05 mM CH₂-THF) and varying CH₂-THF from 0 to 0.2 mM or dUMP from 0 to 0.1 mM. In order to calculate kinetic parameters, the saturation curve shown in figure 18 Cleft panel was fitted to an equation describing the substrate inhibition [58], whereas the saturation curves in the right panel was fitted to the Michaelis-Menten equation. Concerning TYMS-DHFR activity, substrates concentration was 0.1 mM CH₂-THF, 0.1 mM dUMP, 0.1 mM NADPH, whereas Chimera was 0.5 μM. In the activity assay of dTMP-SC substrates concentration was 10 mM L-serine and 16 μM THF, 0.1 mM dUMP and 0.1 mM NADPH. Chimera and SHMT1 concentration was 0.5 μM.

3.1.9 Far-Western blot

The prey (Chimera (4 μg), SHMT1 (4 μg)) and the control protein - a truncated construct of RmcA from *P. aeruginosa* (4 μg) [79], SHMT1 or SHMT1-dimeric mutant (H135N-R137A) (0.5 μg), Chimera (0.5 μg) were resolved by SDS-PAGE electrophoresis, and electro-blotted onto PVDF membranes (Biorad: 100 V, 2h, 4°C). The transferred proteins were then renatured by progressively reducing the guanidine HCl concentration, starting from the following stock solution: AC buffer (10% glycerol; 5M NaCl; 1M Tris pH 7.5; 0.5 M EDTA; 10% Tween-20; 8M Guanidine-HCl; 2% Milk; 1mM DTT) as described by Wu et al. [80]. Proteins were re-

naturated by incubating the membranes for 30 min at R.T. in AC buffer containing decreasing concentrations of guanidine-HCl (6M and 3M), then for 30 min at 4°C in 0.1M guanidine-HCl and finally o.n. at 4°C in AC buffer containing no guanidine-HCl. The following day the membranes were blocked with 5% milk in PBS-Tween for 1h at RT and incubated with purified bait proteins (SHMT1 or SHMT1-dimeric mutant and CHIMERA 50 ng/ml) at 4°C overnight. Membranes were then washed and incubated with primary antibody against the protein used for the incubation step. Subsequently, membranes were incubated with the appropriate secondary antibody and detected with Luminata™ Crescendo Western HRP Substrate (Millipore) according to manufacturer's instructions.

Antibodies used: mouse anti-DHFR (1:1000) , mouse anti-TYMS (1:1000) and rabbit anti-SHMT1 (1:1000; Cell Signalling) - o.n. at 4°C.

3.1.10 Immunopurification with Protein G-plus agarose beads

SHMT1 and Chimera Long were mixed in a 1:2 ratio in 50 mM NaCl, 20 mM Hepes pH 7.5. Each sample was 1 ml in volume and contained SHMT1 (40 µg); Chimera (92 µg) or the mixed proteins. A precleaning step of the samples was performed by adding 50 µl of Protein G-Plus agarose beads suspension (Santa Cruz Biotechnology, INC) to 1 ml of each sample (Chimera, SHMT1 and Mix) and incubating for 2 h at 8°C on a rocking platform. The beads were then centrifuged at 12.000 xg for 20 seconds and the supernatant was transferred to fresh tubes. Afterwards, 1 µg of anti-SHMT1 (Santa Cruz Biotechnology, INC) or anti-TYMS (Santa Cruz Biotechnology, INC) was added to each sample, incubating for 1 h at 8°C rocking. 50 µl of beads were then added to the mixture and the final solution was left rocking at 8°C o.n.. Samples were centrifuged at 12.000g for 20 seconds and the supernatant was discarded. Two washes were performed by resuspending the beads

in 500 µl of 50 mM NaCl, 20 mM Hepes pH 7.5 and rocking for 30 min at 8°C. The supernatant was then discarded. 50 µl of SDS-page loading buffer 2x were added to the beads-protein mixture. Samples were then heated at 95°C for 1h, resolved by SDS-PAGE electrophoresis and subsequently electro-blotted onto a nitrocellulose membrane following a standard western blot protocol. For the detection of the prey proteins mouse anti-DHFR (1:1000), mouse anti-TYMS (1:1000) and rabbit anti-SHMT1 (1:1000; Cell Signalling) antibodies were used.

3.1.11 Computational model of the dTMP-SC

Since the results obtained with the BLI suggested that that SHMT1 may bind two Chimera dimers. We have therefore performed, in collaboration with Professor Paiardini, a molecular docking starting from the available crystal structures of the three proteins. The computational model was obtained as follows.

Protein Structures preparation and docking: human TS (PDB: 1JU6 [81]), DHFR (PDB: 2DHF [82]) and SHMT1 (PDB:1BJ4 [53]) were downloaded from the PDB. TS and DHFR were initially superposed to Chimera. MOE [83] and PyMod3 [84] were used to add missing hydrogens and partial charges. The structures were energy minimized using default values in MOE. To gain insights into their interaction mode, the DHFR-TS complex was docked to SHMT1 using the HADDOCK server (<https://wenmr.science.uu.nl/haddock2.4/>) [85], a widely used in protein-protein docking server. The server allows backbone conformational modification during complex formation. Default parameters were kept for docking. Ten different starting orientations of the complex were submitted, and the best one was assessed by HADDOCK score.

3.1.12 Bio-layer interferometry (BLI)

BLI was performed using the BLItz platform (FortéBio). Protein-A probes (ProA; 18-5010, FortéBio) was used as biosensors, Chimera was immobilised using mouse anti-DHFR (SigmaAldrich) and SHMT1 was used as the analyte at different concentrations.

Several experimental conditions were tried to minimize the side processes interfering with the binding kinetics; initially experiments were carried out by changing the biosensor at each measurement. This approach was discarded due to the low reproducibility caused by both the avidity effect and the variability of the amount of immobilized ligand on the sensor in each step. Thus we followed the “kinetic titration series” protocol described by Frenzel D et al. [86], which uses the same biosensor for all the titration steps. The experiment was performed as follows: Protein-A biosensor was hydrated in Kinetic Buffer (18-1105; FortéBio) for 10 minutes. Afterwards, a mixture of anti-DHFR ($5 \mu\text{g mL}^{-1}$) and Chimera ($4,5 \mu\text{M}$) was loaded for 120 s at 2200 r.p.m shaking speed. The unbound antibody-protein complex was washed away for 240 s with Kinetic Buffer. The anti-DHFR-Chimera loaded sensor was then incubated with SHMT1 at increasing concentration (1, 3, 7.5, 15, 30 μM); 120 s of association and 120s of dissociation in Kinetic Buffer at 2200 r.p.m were collected. This setup led to a biphasic behaviour on both the association and dissociation kinetics; since the vendor software includes automatic global fitting only for monophasic events, data were manually fitted with a two-exponential equation with Igor Pro (Wavemetrics). For the dissociation time course, a very fast phase (including the first 10 s of kinetics) was occasionally observed at high SHMT1 concentrations. This is likely due to a small percentage of loosely immobilised ligand that may dissociate from the biosensor after binding the analyte; if present, this part was not included in the fit. The observed k_{off} in the dissociation kinetics were used as constrain for k_{on} determination. The values reported in table 2 were obtained by linear fit of the k_{obs} , obtained by two series

of independent experiments, each series including at least 4 different SHMT1 concentrations. A representative series is shown in figure 9B. Reference binding experiments were carried out titrating SHMT1 with a sensor loaded with the anti-DHFR in the absence of Chimera. Operating temperature was maintained at 25 °C.

3.1.13 Differential Scanning Fluorimetry (DSF)

DSF assays were performed using a Real Time PCR Instrument (CFX Connect Real Time PCR system, Bio-Rad). In a typical experiment, 0.5 µM SHMT1 and 0.5 µM Chimera (Long or Short) in 20 mM Na HEPES pH 7.2, containing 50 mM NaCl, were incubated o.n. at 4°C (in a total volume of 25 µl) in a 96-well PCR plate. When substrates were added, this solution also contained 0.1 mM CHO-THF, 1 mM dUMP, 0.1 mM NADPH. After the incubation, Sypro Orange (4x, Thermo Scientific) was added and fluorescence was measured from 25°C to 95°C in 0.4°C/30 sec steps (excitation 450-490 nm; detection 560-580 nm). All samples were run in triplicate. Denaturation profiles were analysed using the PRISM Graph Pad software, after removal of points representing quenching of the fluorescent signal due to post-peak aggregation of protein-dye complexes. All curves were normalised and fitted to the following sigmoidal equation 1 to obtain the melting temperatures (T_m).

Equation 1

$$Fluorescence = F_1 + \frac{(F_2 - F_1)}{1 + e^{\frac{T_m - X}{s}}}$$

where X is the temperature (°C), F_1 is the fluorescence at low temperature and F_2 the maximal fluorescence at the top of the

truncated dataset, while s describes the steepness of the curves. Alternatively, T_m values were obtained by plotting the first derivative of the fluorescence emission as a function of temperature ($-dF/dT$) by using the Bio Rad CFX manager software.

3.2 SHMT1-RNA complex

3.2.1 Cryo-electron microscopy (Cryo-EM)

Cryo-EM is the new emerging method for structural analysis of macromolecular assemblies. Joachim Frank, Jacques Dubochet, and Richard Henderson received the Nobel prize in Chemistry 2017 for their contribution to the current “resolution revolution” in microscopy. Cryo-EM has been developed starting from electron crystallography of 2D crystals, no more than few unit cells thick. The use of cryogenic temperatures reduces both specimen damage and motion, therefore, a method for plunge-freezing protein samples on the grid while preserving their native states was implemented by Dubochet and colleagues [87]. Joachim Frank studies paved the way of single particle analysis [88], in fact, when a large assemble of molecules is imaged by EM, all possible orientations may be found, allowing a 3D reconstruction from the orientations of individual particles. However, the real breakthrough was due to the development of direct detectors able to detect single electrons, which led to the current “resolution revolution” in cryo-EM. The advantage of cryo-EM over X-ray crystallography and NMR is that huge protein complexes can be studied and small impurities and disordered regions of the protein sample, which usually hamper the crystallization process, are not an issue. Indeed, this new technique can provide structural information of different conformations of proteins in solution [89]. The only limitation for cryo-EM is the size of the target protein, as structures below 100 kDa are still very difficult to solve.

In our case, we decided to proceed with cryo-EM over X-ray crystallography, as the RNA oligonucleotide is extremely flexible and therefore the SHMT1-RNA complex would have been difficult to crystalize.

3.2.1.1 Negative staining

Before starting a cryo-EM experiment, the most conventional and fast procedure to assess sample quality is by visualising the sample using negative staining EM (NS-EM) at room temperature. We have performed this experiment in collaboration with Dr. Matteo Ardini at University of L'Aquila.

For the NS-EM experiment we preprepared the samples by mixing SHMT1 and UTR2 1-50 in a ratio of 1:1 with a final concentration of 25 nM and a volume of 40 μ l. We used a Carbon film 400 mesh Copper CF400-CU grid and prepare it as follows; 5 μ l of the mixture were deposited on the grid and the solution was allowed to equilibrate for 1 minute. The excess of solution grid was then removed by using Whatman filter paper and 5 μ l of the stain (uranyl acetate 2%) were then applied and allowed to equilibrate for 20 seconds. Afterwards, the grid was washed with deionized water and dried with filter paper. The grid was screened using a TEM Philips CM100 (0.34 nm resolution) equipped with a Tungsten (W) filament and a high-contrast objective lens. Micrographs have been acquired using a Kodak MegaPlus camera, Model 1.6i. In NS-EM the stain does not penetrate the biological material, but remains in the solvent surrounding the proteins, that becomes dark due to the high electron density of the stain, in this way, the contrast (signal-to-noise ration) of the macromolecules is increased and this allows to see the proteins of the sample as clear particles.

Although uranyl ions contained in the stain may precipitate when in contact with phosphate groups of the backbone of nucleic acids, we were able to see a good particles distribution, which encouraged us to proceed with the cryo-EM technique.

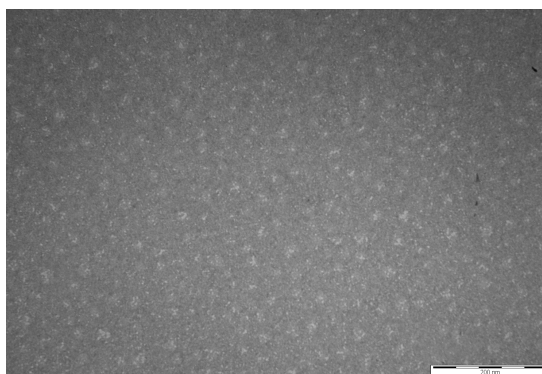


Figure 11. Negative stain EM of the SHMT1-RNA complex. The image was acquired by using a Tem Philips CM 100 (0.34 nm resolution) equipped with a tungsten (W) filament and a high-contrast objective lens. Micrographs have been acquired using a Kodak Megaplug camera, model 1.61. Scalebar 200nm.

3.2.1.2 Cryo- EM sample and grid preparation

In order to guarantee the highest degree of homogeneity of the protein particles, purified SHMT1 was further injected into a Superdex 200 10/300 equilibrated with 20mM HEPES pH 7.2 and 150 mM NaCl prior to the experiment. The RNA UTR2 1-50 was ordered from Eurofin genomics at a HPLC grade of purity, it was dissolved in DEPC water to prevent RNase contamination, aliquoted and stored at -80°C.

For the sample preparation we tried different ratios of SHMT1:RNA. First, we tried a 1:1 ratio, with a final concentration of SHMT1 of 5.7 μM , that resulted in a good particle distribution. We performed a data collection of the best grid. We then increased RNA concentration to a ratio of 1:5 for SHMT1:RNA. To do so, we have diluted SHMT1 to at a final concentration of 5,7 μM mg/ml in 20mM HEPES pH 7.2 and 150 mM NaCl was mixed with RNA in denoised water to a final concentration of 28.5 μM . We did not dilute

RNA in DEPC water since we noticed that DEPC favored particle aggregation and resulted in high background of vitrified grids.

Before adding the sample, the grid must be carefully prepared. We used copper grids covered by a layer of holey carbon of the Quantifoil type (Quantifoil™ R1.2/1.3, 300-mesh). To facilitate a homogeneous spreading of aqueous solutions and retaining particles on the grid surface, the grids were treated with low-energy plasma (glow discharge; GloQube Plus). Low discharging employs the reactivity of ionized water (present in the normal atmosphere, but also specific mixtures that can be used) over the surface of the grid to be cleaned. Hydroxy and Hydride ions react with the exposed grid surface to form aldehydes, alcohols, carboxylic acids and other types of hydrophilic moieties, also depending on the plasma composition. In this way, the surface of the carbon layer on the grid becomes intensively hydrophilic, to allow the distribution of the sample in aqueous solution equally over the grid squares.

Ideally, proteins in cryo-EM samples need to be trapped in a very thin layer of vitreous ice, this minimizes multiple scattering events and maximize sample contrast in the microscope [90]. Unfortunately, optimal ice thickness is difficult to achieve, due to different factors (i.e the uneven surface of the blotting paper). To minimize ice thickness, we placed 3-3.5 µl of sample mixture on the discharged grid and tried different waiting times (time of the incubation on the grid) and blotting times. We used Vitrobot Mk IV (Thermo Fischer Scientific) operating at 4 °C and 100% humidity, which after blotting, plunged the grid in liquid ethane cooled at the liquid nitrogen temperature, around -180°C. for vitrification.

Plunge-frozen grids were screened on a Falcon 3EC direct electron detector (Thermo Fischer Scientific) mounted on a Thermo Fischer Talos Arctica transmission electron microscope operating at 200 kV at the cryo-EM facility of University of Milan. Once the grids were loaded onto the stable microscope, we selected the holes onto which promising images could be collected. Best grid squares and holes were selected among those without contaminants, ice crystals, and

scratches, kinks or cuts on the grid support. The grid presenting the best ice quality was obtained with the following parameters: 3 μl of sample mix volume, 20 sec of waiting time and 4 sec of blotting time.

3.2.1.3 Data collection

Data collection was performed at the cryo-EM facility of University of Milan on a Falcon 3EC direct electron detector (Thermo Fischer Scientific) mounted on a Thermo Fischer Talos Arctica transmission electron microscope operating at 200 kV. The first two data collections, performed on the same grid (SHMT1-RNA ratio 1:1) with the following parameters. In the first one a total number of 1881 micrographs were acquired. Each image was acquired with an exposure time of 1.5 sec and a total accumulated dose of 40 electrons per \AA^2 equally distributed over 39 frames per movie. Images were acquired at nominal magnification of x 73,000, corresponding to a pixel size of 1.43 $\text{\AA}/\text{pixel}$ at the specimen level, with applied defocus between -0.5 to -1.75 μm . In the second one a total number of 1849 images were collected using an exposure time of 60 seconds and a total accumulated dose of 40 electrons per \AA^2 distributed over 39 movie frames. This time the nominal magnification was of 120,000 corresponding to a pixel size 0.889 $\text{\AA}/\text{pixel}$.

The last data collection was performed on the grid having a SHMT1-RNA ratio of 1:5. In this case, the collection lasted for seven days and a total number of 5450 movie frames were collected with an exposure time of 60 seconds and a total accumulated dose of 40 electrons per \AA^2 equally distributed over 40 frames per movie. Images were acquired directly in the electron counting mode with a nominal magnification of 120,000, corresponding to a pixel size of 0.889 $\text{\AA}/\text{pixel}$ at the specimen level, with applied defocus between -0.5 to -2.5 μm .

3.2.1.4 Data processing

All the collected data were imported in RELION 3.1 [91], which was used for all the following steps. Movies were corrected for anisotropic magnification distortion and motion using MOTIONCOR2 [92]. Then contrast transfer function (CTF) estimation was performed on aligned, unweighted sum images using CTFIND4.1 [93]. From here, only micrographs presenting a resolution estimated around 5 Å or better were selected for further analysis and the ones showing ice contamination were removed as well. After various attempts of using a RELION'S Gaussian blob as reference [91], we decided to proceed with a manual particle picking in small subset of micrographs. Proper particle picking and alignment are essential to obtain high-resolution data, as these processes are usually performed in Fourier space, with low-pass filtering. The particles manually picked were then extracted and subjected to a 2D classification. The 2D classified particles were then used to perform an automatic particle picking, resulting in almost 5 million particles. Multiple rounds of 2D classifications were performed, in these steps particles are rotated, translated in all possible orientations, and matched to each other, then averaged together to form a certain number of classes containing different views of the specimen. Junk particles were then removed, resulting in 900,000 clean SHMT1 particles. The thus obtained 24 2D classes were used to generate a first 3D model. Multiple rounds of 3D classification yielded 8 different classes, among which two main classes were identified: one of the SHMT1-RNA complex and one belonging to SHMT1 alone.

3.2.1.5 Cryo-EM maps resolution estimation and sharpening

To estimate the resolution in cryo-EM, the Fourier shell correlation (FSC) has become a standard. The FSC measures the correlation between Fourier coefficients within every resolution shell of two

independently determined half-maps. A threshold value must be selected to report a resolution value of structures based on FSC curves [94]. Rosenthal and Henderson 2003, proposed a fixed threshold of 0.143 which is widely used for resolutions better than 10 Å [95]. In our case, as the resolution was lower than 10 Å, we fixed the threshold at 0.5, which resulted with an overall resolution of 3.63 Å. The cryo-EM map was then sharpened using Autosharpen tool in Phenix [96], yielding a final map with a resolution of 3.52 Å. UCSF Chimera [97] and Coot [98] were used for the graphical visualization of the maps.

3.2.1.6 Model building

The initial model of the SHMT1-RNA complex, but also of SHMT1, was generated performing a rigid body fit of SHMT1 crystal structure (PDB: 1bj4 [53]) by using UCSF Chimera [97]. Each subunit was fitted independently. The initial model was then manually modified and rebuild in Coot [98] to fit the sharpened map or, in the more dynamic regions, the unsharpened one. The final models were then subjected 3 cycles of real-space refinement using PHENIX [99].

The 6 nucleotide hairpin RNA (CUACAA, belonging to the sequence of UTR2 1-50 AUA AAG AAA AAA GCG GUG AGU GGG CGA ACU ACA AUU CC CAA AAG GCC ACA, was fitted into the extra electron density in Coot as well [98]. The initial geometry of the RNA hairpin was modelled based on a bacterial RNA structure (pdb: 2jr4 [100]) and then refined in the electron density in Coot. The full atomic models of SHMT1-RNA complex, and SHMT1 unbound were subjected to multiple rounds of real-space refinement in PHENIX, including global minimization and refinement of atomic displacement parameters, and applying secondary structures, Ramachandran, and nucleic acids restraints [101].

3.2.2 In vitro experiments

3.2.2.1 Site-directed mutagenesis

SHMT mutants were produced by site-directed mutagenesis using the QuikChange Lightning Site-Directed Mutagenesis Kit (Agilent Technologies). The pET28b vector containing the wild-type SHMT1 gene was used as template DNA. Two complementary oligonucleotides, synthesized by Eurofin genomics (Eurofins Scientific Group, Luxemburg) were used as primers (table x) for the mutagenesis reactions. The DNA was then transformed in *E. coli* Neb5 α cells for amplification, plasmids were purified with QIAprep Spin Miniprep kit (Qiagen). Mutagenesis was then verified by sequencing. The mutants were then growth and purified as described in section 3.1.6.

SHMT1 K282S	Fw GGA TCC CAA GAC TGG CTC AGA GAT TCT GTA CAA CC Rv GGT TGT ACA GAA TCT CTG AGC CAG TCT TGG GAT CC
SHMT1 R393S- R397S	Fw GAC AGC AGC GCT CTG AGC CCC AGT GGA CTG Rv CAG TCC ACT GGG GCT CAG AGC GCT GCT GTC

Table 1. Primers used in the mutagenesis reactions (written 5' to 3').

3.2.2.2 Electrophoretic mobility shift assay (EMSA)

EMSA assays were used to assess the binding affinity of different RNAs towards SHMT1, and the ability of the mutants to bind RNA. These experiments were conducted by incubating a fixed amount of

RNA with either fixed or increasing concentrations of purified SHMT1 (or mutants).

All the components were incubated at 16°C for 30 minutes in 12 µl of binding buffer (20 mM HEPES pH 7.4, 100 mM NaCl) containing 20 µg/ml bovine serum albumin (BSA) and 8% (v/v) glycerol. The reaction mixtures are then subjected to electrophoresis under native conditions using a non-denaturing 4% polyacrylamide gel in 0.5× TBE buffer (45 mM Tris-Borate, 1 mM ethylenediaminetetraacetic acid pH 8.6). For the visualization, gels were stained with SYBR Safe (Invitrogen) in 30 ml of 0.5× TBE and images were acquired using Chemidoc MP Imaging System (Bio-Rad). Densitometric measurements of the free RNA bands were transformed into percentages, the amount of bound RNA (% bound) was then calculated by subtracting the percentage of free RNA from the total, the result was plotted as a function of protein concentration. The apparent dissociation constant (K_D) was estimated fitting the data with Equation 2, in which B_{max} corresponds to the maximum binding (100%) and $[P]$ corresponds to the concentration of SHMT1 in the reaction mixture.

Equation 2

$$\% Bound = \frac{B_{max} \cdot [P]}{K_d + [P]}$$

3.2.2.3 Enzymatic assays

All assays were carried out at 30 °C in 20 mM K-phosphate buffer at pH 7.2, treated with diethyl pyrocarbonate (DEPC). The serine cleavage reaction was measured using 0.2 µM of purified SHMT1 with fixed 10 mM L- serine and 50 µM THF as substrates while varying the RNA concentration. A spectrophotometric coupled assay, in which the 5,10-CH₂-THF produced by the reaction was oxidized by the NAD-dependent *E. coli* 5,10-CH₂-tetrahydrofolate dehydrogenase was used [102].

All obtained inhibition curves were fitted to Equation 3, in which [I] corresponds to the RNA concentration.

Equation 3

$$\% \textit{Activity} = 100 \cdot \left(1 - \frac{[I]}{[I] + IC50} \right)$$

All spectrophotometric measurements were performed using a Hewlett- Packard 8453 diode-array spectrophotometer. Fitting of data to equations was carried out with the PRISM software (GraphPad).

4. Results

4.1 The *de novo* thymidylate synthesis complex

The *de novo* dTMP synthesis was described to be a crucial metabolic pathway in normal and in cancer cells [72], which involves the combined action of SHMT1, DHFR and TYMS.

Both TYMS and DHFR are already well-known targets of chemotherapeutic drugs (antifolates and antimetabolites), but, given their role in DNA replication, the side effects of these treatments are often severe. To overcome the salvage pathways activated by cancer cells when treated with antifolates or antimetabolites, conformational drugs could be used. These compounds would alter the conformational equilibrium or disrupt protein-protein interactions (PPIs), rather than targeting the active site of the enzyme. The idea would be to use these conformational drugs to prevent the assembly of the dTMP synthesis complex (dTMP-SC). Therefore, structural studies on these complexes are needed and here we report the intracellular dynamics of the complex in cancer cells and the first evidence of complex formation *in vitro*.

4.1.2 Intracellular localization of the complex

To monitor the formation of the dTMP-SC and to obtain novel information on its dynamic, we performed an in-situ proximity ligation assay (is-PLA).

Is-PLA uses ordinary primary antibodies to detect the proteins of interest, which are then revealed by using secondary antibodies conjugated to DNA oligonucleotides which produce a fluorescent signal in the form of a spot, given by the ligation and amplification of the oligonucleotides that occurs only when the two proteins are interacting or closer than 40 nm (figure 12 A) (Di Francesco *et al.*, 2021).

We performed the experiment on both non-synchronised and S-phase synchronised lung adenocarcinoma cells (A549) to determine the interactions occurring between all the possible couples in the ternary complex SHMT1-TYMS-DHFR. As shown in figure 12 B we could detect the positive PLA signals for all the combinations of the three proteins in both synchronous and asynchronous A549 cells. Interestingly, the complex was more abundant in the cytoplasm than in the nucleus, and nuclear localization appears to be enriched in synchronised S-phase cells only to some extent (figure 12 C).

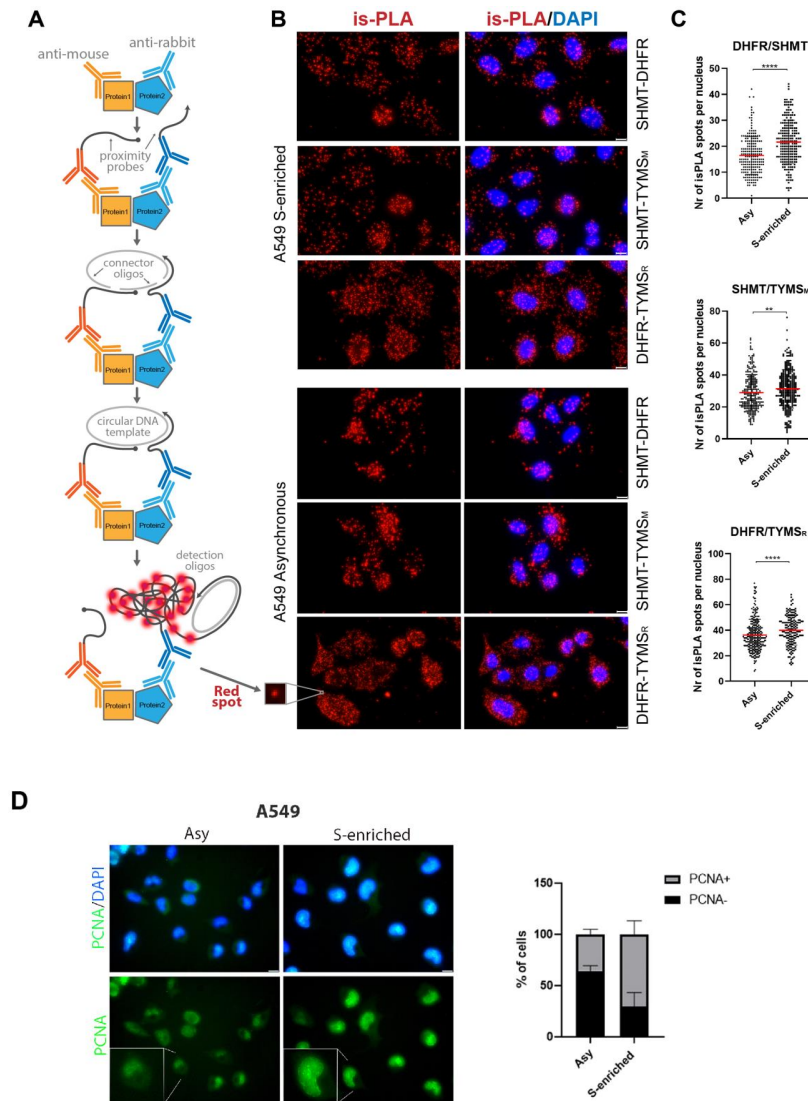


Figure 12. Identification of proteins proximity in cancer cell lines. A) Schematic representation of the in situ Proximity Ligation Assay (is-PLA). B) is-PLA signals (red dots), corresponding to DHFR/SHMT, SHMT/TYMS and DHFR/TYMS interactions. TYMSM and TYMSR both indicate antibodies against TYMS but

*anti-mouse and anti-rabbit, respectively. The is-PLA spots of interaction of proteins are shown in A549 cells synchronised in the S-phase after a 24 h single thymidine block (upper panel) or in A549 asynchronous cells (bottom panel). Scale bars: 10 μm . The merge with DAPI signal is shown on the right. C) Nuclear localization of the is-PLA signal (number of spots within the normalized nucleus area) In all the experiments the number of PLA spots increase in the S-enriched cells (T-test: $**P < 0.005$; $****P < 0.0001$). D) On the left: representation of PCNA positive cells, corresponding to S- phase of the cell cycle. Scale bar: 10 μm . On the right: immunofluorescence signal in an asynchronous population or after a single thymidine block of A549. The histogram on the right shows the percentage of PCNA positive or negative cells after and before synchronisation. At least 500 cells were counted per conditions, from three independent experiments, s.d. are shown. After synchronisation, the cells in S-phase are about the 80% of cellular population [68].*

To confirm these results, we performed a co-localization analysis (figure 13) and independently detected the presence of the single proteins (SHMT1, TYMS and DHFR) in the cytoplasm and in the nucleus of both synchronous and asynchronous A549 cells by immunofluorescence (figure 14 A to D) and by western blot after cell fractionation (figure 14 E). The three proteins were always present in the cytoplasm, whereas nuclear localization was more abundant in S-enriched cells only for TYMS and SHMT1 (figure 14 D). This was particularly evident for the latter, suggesting that nuclear translocation of SHMT1 may drive the increase of dTMP-SC formation in the nucleus during the S-phase.

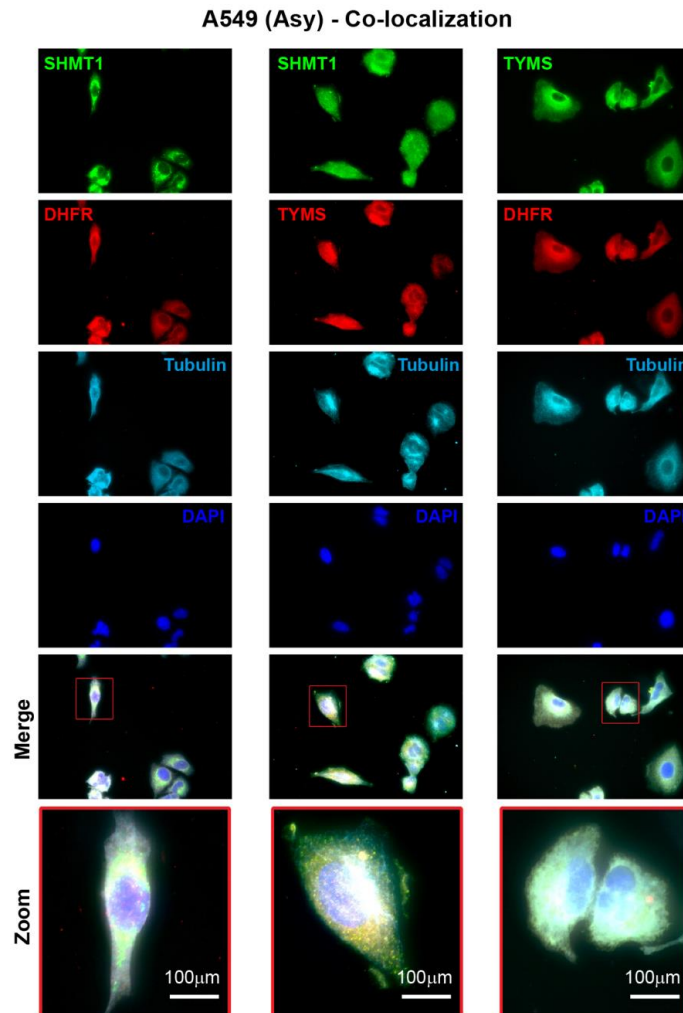


Figure 13. Immunofluorescence Co-localization. Co-localization of DHFR/SHMT, SHMT/TYMS and TYMS/DHFR in asynchronous A549 cells. Co-localization is evident both in the cytosol (tubulin alpha – cyan) and in the nucleus (DAPI - blue) for all the three couples. The original red fluorescence signal of tubulin was changed to cyan for a better visualization of the merged images. The merge of green, red and cyan pixels yields white pixels [68].

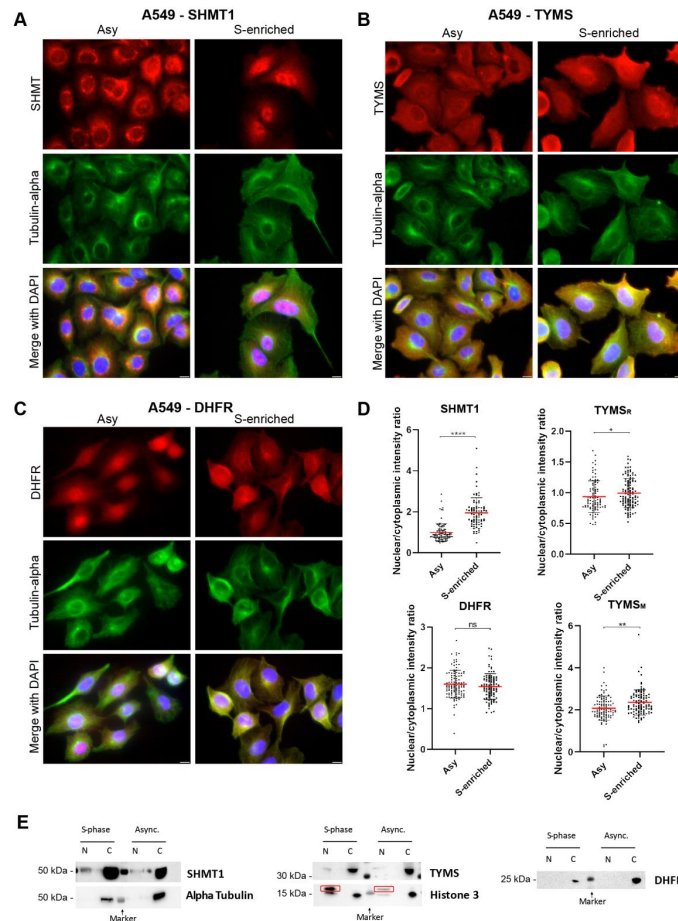


Figure 14. Immunofluorescence analysis. A-B-C) Compartmentalization of the three proteins with respect to the cellular phase by Immunofluorescence analysis. SHMT1 (A), DHFR (B) and TYMS (C) show a cytoplasmatic localization both in the asynchronous and S-phase synchronised cells. Scale bar: 10 μ m. D) Nuclear localization (as deduced by the normalized intensity of the fluorescence signal) is clearly more abundant in the S-enriched cells for SHMT1, a slight increase is observed for TYMS (with both the antibodies used in the PLA experiments; rabbit - TYMS_R and mouse - TYMS_M). No change was

observed for DHFR. (T-test: * $P < 0.05$; ** $P < 0.005$; **** $P < 0.0001$). E) Western blot of subcellular fractionation of both asynchronous and S-phase enriched A549 cell lines. WB analysis was performed using a 1:1000 dilution of the primary antibodies. For Histone H3 the correct band is boxed in red. The other bands detected in the cytosol, have a molecular weight lower than 15kDa and are also present in the nuclear fractions. Giving that the Mw of Histone H3 is 17kDa, and the bands recur in all the four lanes, it is plausible that they are detected because of non-specific interactions of the primary or secondary antibody. In this experiment the quantitative analysis was not performed, but it is still possible to detect the bands of SHMT1 and TYMS only in the nucleus of the S-phase enriched cells. As shown in the IF experiments, it was not possible to detect an increase of the presence of DHFR in the nucleus of the synchronised cells [68].

Overall, the PLA results suggest that, under these conditions, the dTMP-SC is located in the cytoplasm as well as in the nucleus. This distribution was further confirmed in asynchronous HeLa cells (figure 15 A).

As negative controls we either used only one out of the two antibodies previously used in combination and we have used RNA interference (RNAi) to lower SHMT1 levels. In both cases, no PLA signal was observed, indicating that the one we observed is specific (figure 15 B to E).

These results indicate that the three enzymes are able to assemble the dTMP-SC complex even in the absence of DNA or lamina proteins as binding partners [66], suggesting that it is possible to assemble the dTMP-SC *in vitro* starting from the purified proteins. Our strategy to help achieving this goal was to obtain a binary interaction, by assembling SHMT1 with a TYMS-DHFR fusion protein.

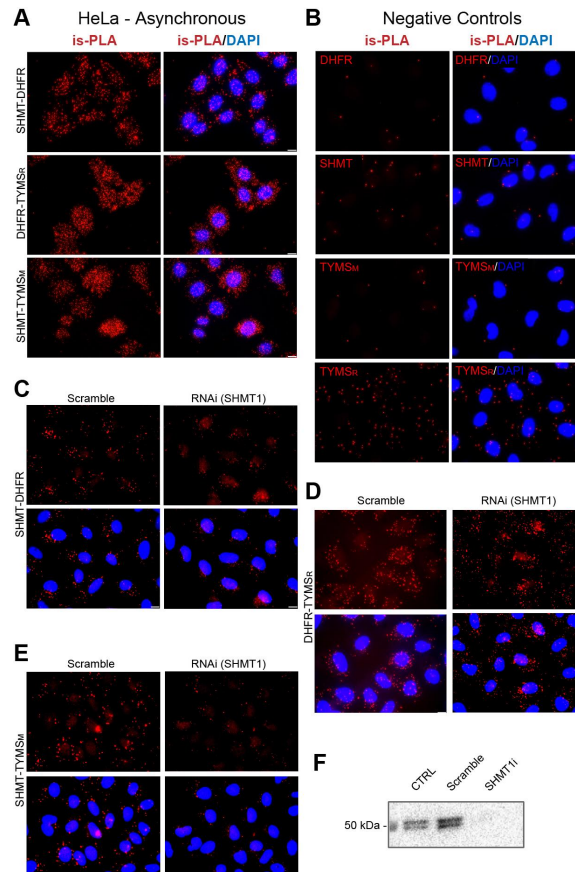


Figure 15. is-PLA analysis on HeLa cells and is-PLA negative controls on A549 cells. A) PLA signal of protein interaction is shown in HeLa non-synchronized cells. B) Negative controls of is-PLA signals. PLA experiment was performed without one of the primary antibodies (A549 cells). In these conditions there is no or little PLA signal. TYMSM and TYMSR both indicate antibodies against TYMS but anti-mouse and anti-rabbit, respectively. C-D-E) is-PLA for the three protein-protein interaction performed on A549 cells 48h after transfection with scrambled or siRNA for *shmt1*. For the DHFR/SHMT1 and TYMS/SHMT1 interactions the PLA signal is significantly lower in the RNAi cells, whereas for the DHFR/TYMS

interaction the signal of the scramble and RNAi samples is similar (D). For all panels the scale bar in the bottom right corner is 10 μm . F) SHMT1 expression control of RNAi [68].

4.1.3 In vitro assembly

4.1.3.1 Rational design, expression, and purification of the chimeric bi-functional protein

In order to facilitate the complex formation in vitro, we downgraded the assembly from a ternary to a binary complex by linking together DHFR and TYMS in a single chimeric construct (hereinafter referred to as Chimera). To do so, in collaboration with Professor Paiardini, we have used as a starting point the observation that these bifunctional enzymes are endogenously expressed in several Protozoa.

Two different classes of DHFR-TYMS bifunctional enzymes can be distinguished, which differ mainly in the length of the linker between the two domains and consequently in the orientation with which the DHFR domain contacts the TYMS domain [103]. Due to the higher identity with the human DHFR sequence, the crystallographic structure of DHFR-TYMS from *Trypanosoma cruzi* (PDB ID: 2H2Q [104]) was chosen as a representative for the short linker class, while the structure from *Babesia bovis* (PDB ID: 3I3R [105]) was selected among several DHFR-TYMS structures with long linkers from different organisms. To identify the possible interaction interfaces, the crystallographic structures of human DHFR and TYMS were aligned to the two representatives. An analysis of the evolutionarily conserved residues at the interfaces showed that in both cases none of the residues involved in the stabilisation of the DHFR-TYMS complex was conserved. Following this evidence, the surface potential at the interface was evaluated and a perfect complementarity of the surface charge of human proteins was revealed at the level of the interaction interfaces originating after a structural alignment with *B. bovis* DHFR-TYMS

(figure 16 A-B). As previously suggested, this indicates that the contacts between DHFR and TYMS are nonspecific and that the interaction interface on TYMS provides only an attractive force for DHFR and not an orienting one [103]. Starting from this conclusion, two different linkers were designed, a short 10 aa linker and a long 20 aa one. The linkers were designed to be flexible but with a higher content of Ser residues compared to the canonical (GGGS)_n flexible module [104], because bioinformatic analysis showed that the linker would be completely exposed to the solvent. Moreover, a sequence sensitive to digestion by PreScission protease was introduced to allow the *in vitro* separation of the two enzymes when needed. The two final linkers were GSSGGGSLFQGPSGGSSGG for Chimera-Long and GSLFQGPSGG for Chimera-Short (figure 16 C).

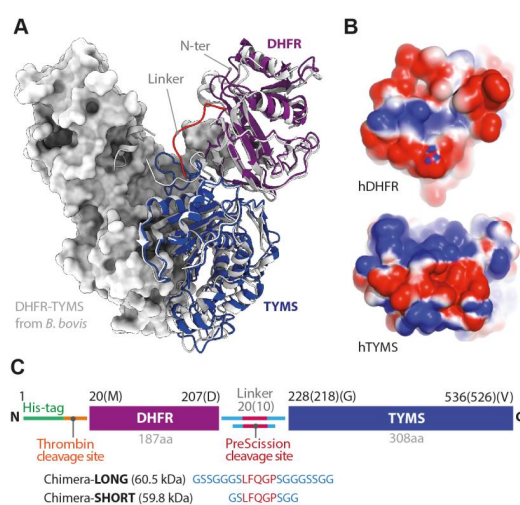


Figure 16. Designing human DHFR-TYMS Chimera. A) Cartoon representation of the chimeric model of human DHFR-TYMS (in purple and blue respectively; the linker is shown in red; the C-terminal position is also indicated)

superposed with the structure of the bifunctional enzyme from B. bovis (PDB ID: 3I3R [104]) in light grey. The enzyme is dimeric, the partner subunit is shown as surface representation. B) Electrostatic surface potentials - calculated using APBS (Adaptive Poisson-Boltzmann Solver) [77] the modelled interface between hDHFR and hTYMS, showing a perfect complementarity. Partially positive or negative regions are indicated in blue and red, respectively. C) Scheme of the final constructs differing only for the linker length and consequently named Chimera-Long and Chimera-Short [68].

Both Chimera constructs, with the shorter and the longer linker, were expressed with an N-terminal histidine-tag and were purified by immobilized metal affinity chromatography (IMAC), eluting between 150 mM and 200 mM imidazole.

They showed to proteolyse in the linker region, as highlighted by the low molecular weight species detected by SDS PAGE and western blot (WB) of the purified fractions (figure 17 A). This phenomenon was more evident at pH higher than 7.5 and was more pronounced for Chimera-Short. Size Exclusion Chromatography (SEC) was therefore used to remove the proteolyzed protein. Both Chimera-Long and Short eluted as dimers (figure 17 B), suggesting that at least TYMS was correctly folded in the Chimera, as human TYMS is dimeric. After purification, no further proteolysis occurred, and the proteins were stable for at least one week at 4 °C (figure 17 C). Also, the UV spectra of Chimera showed a shoulder around 325 nm, likely due to NADPH bound to DHFR (figure 17 D). This result suggested that also DHFR was correctly folded in the Chimera. Analysis by CD spectroscopy confirmed that both Chimera-Short and Long are folded and stable up to 45° C, with an apparent melting Temperature (T_m) of 53.5° C (figure 17 E-F).

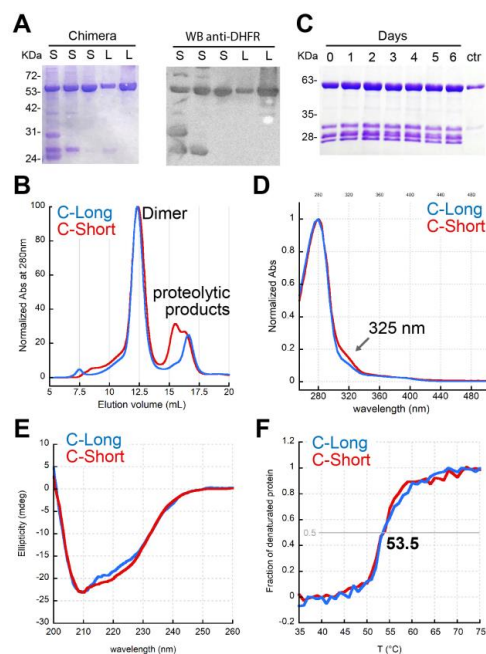


Figure 17. Purification and spectroscopic characterization of Chimera. A) SDS-PAGE and WB of IMAC elution peaks (150-200mM imidazole) of different preparations of Chimera Short (S) and Long (L). B) Typical SEC chromatogram of IMAC fractions of Chimera Constructs. The proteolyzed proteins are well separated from the main peak, containing the full-length protein eluting as a dimer. The signal is normalized for a better comparison (column: Superdex 200 10/300). C) Auto proteolysis assay: after IMAC, a small amount of Chimera Short was kept at 4°C for 6d, to test whether the purified protein undergoes proteolysis in the purification buffer. Ctr = protein after SEC. No further proteolysis was observed. Typical final yields were between 6 and 4 mg/L of culture for Chimera-Long and Short, respectively. D) Normalised UV spectra of Chimera-constructs in 100 mM potassium phosphate pH 7.4. The shoulder at 325 nm is likely due to bound NADPH. E) Dichroic spectra of 10 μ M Chimera constructs at 20 °C in a 1 mm quartz cuvette. F) Normalized thermal denaturation profiles of 10 μ M Chimera constructs [68].

The last step was to observe if the two fused enzymes were active. To do so, we analysed the complete thymidylate cycle as described in figure 6. The reaction catalysed by TYMS was assessed following the increasing absorbance at 340 nm due to the formation of DHF from CH₂-THF (reaction 1 – figure 18 A). By increasing the enzyme concentration at constant substrates concentrations, a linear increase of TYMS activity is observed for both Chimera-Short and Long (figure 18 B). Since it has been demonstrated that mutation on the N-terminal domain of TYMS affects catalysis [106,107], and in our construct DHFR is fused to that domain, we have therefore performed a complete characterization of TYMS activity, that yielded kinetic parameters ($K_{cat} = 0.5 \text{ s}^{-1}$; K_m for CH₂-THF = $2.9 \pm 0.5 \text{ }\mu\text{M}$; K_m for dUMP = $7.1 \pm 1.0 \text{ }\mu\text{M}$) very similar to those reported for wild type TYMS [108](figure 18 C). These results indicate that TYMS is fully functional in the designed constructs. Following, DHFR activity of Chimera was assessed by reducing the DHF produced by TYMS activity to THF by DHFR, with concomitant NADPH oxidation (reactions 1 and 2 – figure 18 A). The decrease of absorbance at 340 nm due to NADPH oxidation was followed to observe this reaction. The time course of the reaction is shown in figure 18 D. With this experimental setup both constructs were able to catalyse the coupled reactions at similar rates, indicating that also DHFR is functional.

Finally, the complete thymidylate cycle was tested *in vitro*. The initial assay mix contained dUMP, serine and NADPH; then THF, Chimera and finally SHMT1 were added (figure 18 E). Chimera activity can only start when SHMT1 produces the substrate CH₂-THF, the reaction proceeds until dUMP and NADPH are consumed, while the folate species cycles, thus reconstituting the functionality of the thymidylate cycle *in vitro*.

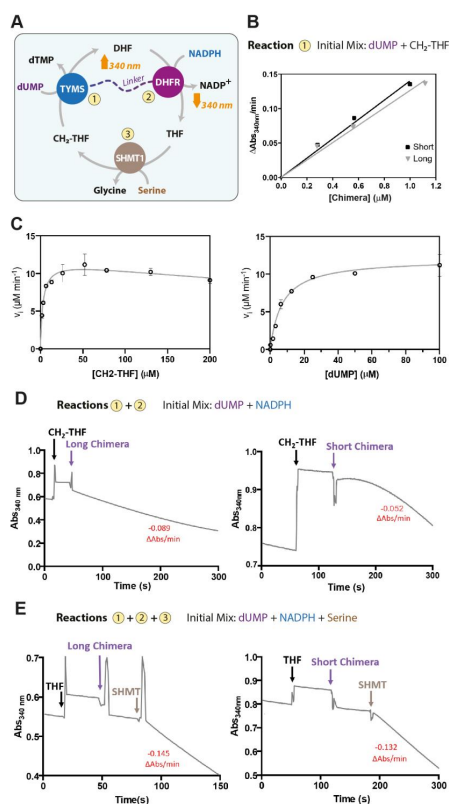


Figure 18. Activity assays. A) Scheme of the reactions assayed to test the catalytic activity of Chimera. All the reactions were performed at 20°C in 20 mM K-Phosphate pH 7.2, 75 mM β-Mercaptoethanol. B) Reaction 1; plot of the initial rate of dTMP formation as a function of Chimera concentration at constant substrate (0.1 mM dUMP, 0.2 mM CH₂-THF); Chimera-Short (black squares), Chimera-Long (gray triangles). C) Plot of initial rates of reaction 1 as a function of CH₂-THF concentration (left panel; 0.1 mM dUMP fixed concentration) and of dUMP concentration (right panel; 0.05 mM CH₂-THF fixed concentration). D) Reactions 1 + 2; time course of the coupled reactions of TMYS and DHFR. The observed rates ($\Delta\text{Abs}@340\text{ nm}/\text{min}$) are also reported in red. To 0.1 mM dUMP and 0.1 mM NADPH, was sequentially added 0.1 mM CH₂-THF and 0.5 μM Chimera-Long (left) or Chimera-Short (right). E) Reactions 1 + 2 + 3; to the reaction mixture containing 0.1 mM dUMP, 10 mM serine and 0.1 mM NADPH, were added 16 μM THF, 0.5 μM Chimera-Long (left) or Chimera-Short (right)

and finally 0.5 μ M SHMT1. The thymidylate cycle can only start after the addition of SHMT1 that is needed to convert THF to CH₂-THF [68].

4.1.3.2 Far Western blotting and Immuno-Purification

The Far Western Blot (FWB) analysis was used to evaluate complex formation between SHMT1 and the Chimera *in vitro*. In this assay, proteins can form a complex in native conditions on the membrane that can be detected by specific antibodies (figure 19 A). Both purified Chimera constructs and SHMT1 were resolved by SDS-PAGE and electro-blotted on a PVDF membrane. The proteins were then renatured directly on the membrane and incubated with the bait protein (either SHMT1 or Chimera).

As shown in figure 19 A, the antibodies directed against DHFR and SHMT1 detected the presence of Chimera-Long in the lane of SHMT1 and *vice versa*, suggesting that Chimera-Long and SHMT1 interact with each other, but not with the control protein. The same results were not observed when using the Chimera-Short construct (figure 19 B), suggesting that the shorter linker might prevent the optimal orientation of TYMS and DHFR with respect to SHMT1. Due to these results, we chose to proceed using only this construct for further analysis of complex formation. Therefore, here and after, if not otherwise specified, I will refer to the Long construct as Chimera.

Then we have investigated the stoichiometry of the dTMP-SC complex, and to do so, we performed the FWB using Chimera-Long and a dimeric variant of SHMT1 (H135N-R137A) [61]. The inability of dimeric SHMT1 to form the complex under the same experimental conditions (figure 19 C), indicates that the tetrameric state of SHMT1 is crucial for complex formation.

The formation of dTMP-SC was also investigated by using Immuno-purification (IP) assay, which confirmed that SHMT1 and Chimera-Long are interacting *in vitro* (figure 19 D). In this case, the proteins

were bound to the protein G-agarose beads by using either anti-SHMT1 or anti-TYMS antibodies. Chimera and SHMT1 were detected only in the SHMT1 plus Chimera mixture (MIX sample), using respectively anti-DHFR or anti-TYMS and with anti-SHMT1 antibodies, but not in the individual SHMT1 or Chimera samples (figure 19 D).

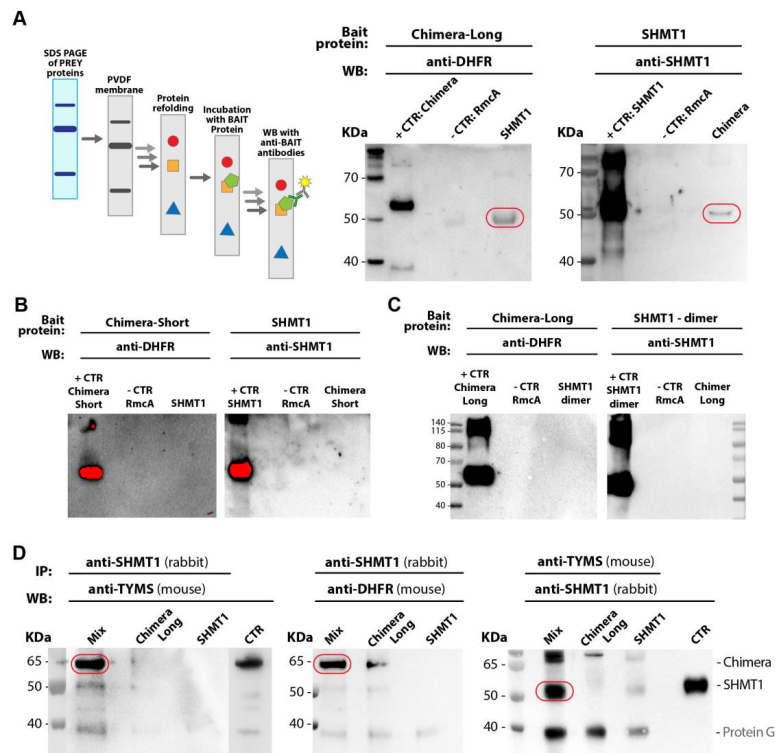


Figure 19. **Complex formation in vitro** A) FWB: The prey proteins (SHMT1 or Chimera-Long) were resolved alone by SDS-PAGE, electro-blotted on a PVDF membrane, refolded and then incubated with 50 ng/ml of the bait protein at 4°C o.n. In the left image the band detected at SHMT1 lane and height represents Chimera, and the band detected at Chimera lane and height, on the right image

represents SHMT1. The bait proteins were used in place of the prey proteins as positive controls, whereas RmcA (a bi-domain bacterial protein construct from *P. aeruginosa* whose two domains have a molecular weight comparable to DHFR and TYMS) was used as a negative control [79]. The formation of the complex between SHMT1 and Chimera-Short (panel C) or SHMT1 dimeric mutant and Chimera-Long (panel D) was tested by following the experimental set-up previously described. Nevertheless, and despite over exposition of the membranes, in both cases formation of the complex was not observed. D) IP experiment: SHMT1 and Chimera-Long were mixed in a 1:2 ratio at a final concentration of 18 μM for SHMT1 and 36 μM for Chimera-Long. The proteins were attached to the Protein G-agarose beads by using specific antibodies (anti-TYMS or anti-SHMT1). The samples were incubated at 4°C o.n. In the first two images, on the left, the detected band represents Chimera-Long whereas in the right image the detected band refers to SHMT1 [68].

4.1.3.3 Quantification of the dissociation constant of the complex

Given all the above mentioned evidence of complex formation *in vitro* we have used Bio-layer interferometry (BLI) to evaluate the binding affinity. In this experiment Chimera was immobilised using anti-DHFR antibodies and a Protein-A conjugated biosensor and SHMT1 was the analyte. To minimize background effects due to the multiple binding partners (Protein-A, anti-DHFR, Chimera), and keep the amount of immobilized ligand as constant as possible, we have decided to use the kinetic titration series experimental setup as described by Frenzel and co-workers [86]. Both the association and the dissociation kinetics were biphasic, with the association rates dependent on SHMT1 concentration. To fit the data, we used a two-exponential equation for each kinetics, assuming a general heterogeneous binding model in which two different binding events may occur (figure 20 A). This setup gave a good fit and yielded the parameters reported in table 2. The two binding events show comparable affinity ($K_D \approx 14 \mu\text{M}$) suggesting that a symmetric bidentate model is likely the best interpretation for these data (figure 20 A). These results suggest that SHMT1 may bind two Chimera dimers. We have therefore performed, in collaboration with

Professor Paiardini, a molecular docking starting from the available crystal structures of the three proteins that confirmed that this stoichiometry is actually feasible (figure 20 C), and that SHMT1 can indeed act as a bidentate analyte.

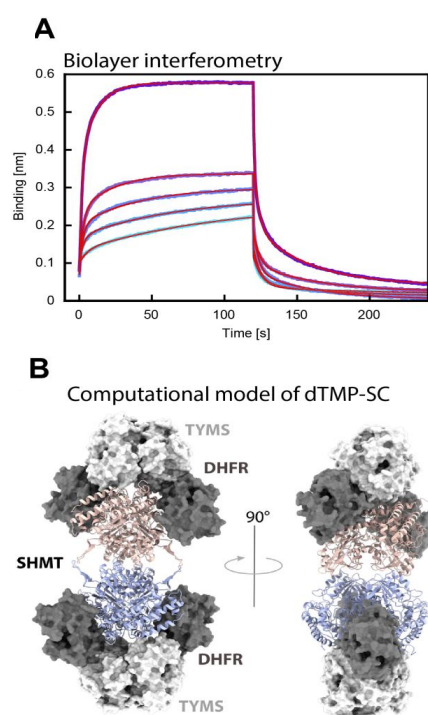


Figure 20. Quantification of the dissociation constant of the complex and proposed model. A) BLI: aligned showing association and dissociation steps of SHMT1 (1, 3, 7.5, 15 and 30 μM) to Chimera, immobilized using mouse anti-DHFR and a Protein-A coated biosensor (in blue); time courses of each step were fitted with a two-exponential equation (in red). The kinetic binding parameters - affinity (K_D) and rate constants (k_{on} , k_{off}) - for the SHMT1-Chimera interaction calculated from the BLI experiments are reported in Table 2. B) Computational model of the dTMP-SC: predicted interaction mode between human TS-DHFR (light and dark grey surface) and SHMT1 tetramer (each dimer is depicted as pink and slate cartoons) [68].

	BLI - binding 1	BLI - binding 2
K_{on} ($\text{M}^{-1}\text{s}^{-1}$)	$(1.4 \pm 0.1) \cdot 10^4$	$(1.6 \pm 0.3) \cdot 10^3$
K_{off} (s^{-1})	$(2 \pm 0.1) \cdot 10^{-1}$	$(2 \pm 0.2) \cdot 10^{-2}$
K_d (μM)	16.4 ± 1.7	13.5 ± 3.6

Table 2. Kinetic binding parameters calculated by BLI [68].

4.1.3.4 Effect of substrates and ligands on dTMP-SC formation

Differential Scanning Fluorimetry (DSF) assay was used to assess the effect of substrate/ligands (such as 5-formyl-tetrahydrofolate (CHO-THF), dUMP, and NADPH) on *in vitro* complex formation. Purified SHMT1 and Chimera were mixed and incubated o.n.. The melting temperature of SHMT1 alone is 57.1 ± 0.1 °C, instead Chimera's one is 48.5 ± 0.5 °C (table 3). When the two proteins were mixed, the observed T_m was 50.2 ± 0.1 °C (figure 21 A). The change of the denaturation profile is likely due to the interaction taking place between SHMT1 and Chimera. Among the different substrate/ligands tested, only dUMP stabilized Chimera and the Chimera-SHMT1 complex, leaving the T_m of SHMT1 unaffected (figure 21 B and table 3). Interestingly, the slope of the transition increases for the complex, indicating that the denaturation process is more cooperative in the presence of dUMP. To assess whether the change in the denaturation profile observed in the presence of dUMP is due to complex formation or not, we performed the same experiment using the Chimera-Short construct that fails to form the complex. Strikingly, in this case the presence of dUMP stabilises Chimera-Short (which as expected displayed an overall lower stability) but has no significant effect on the stability of the complex (figure 21 C).

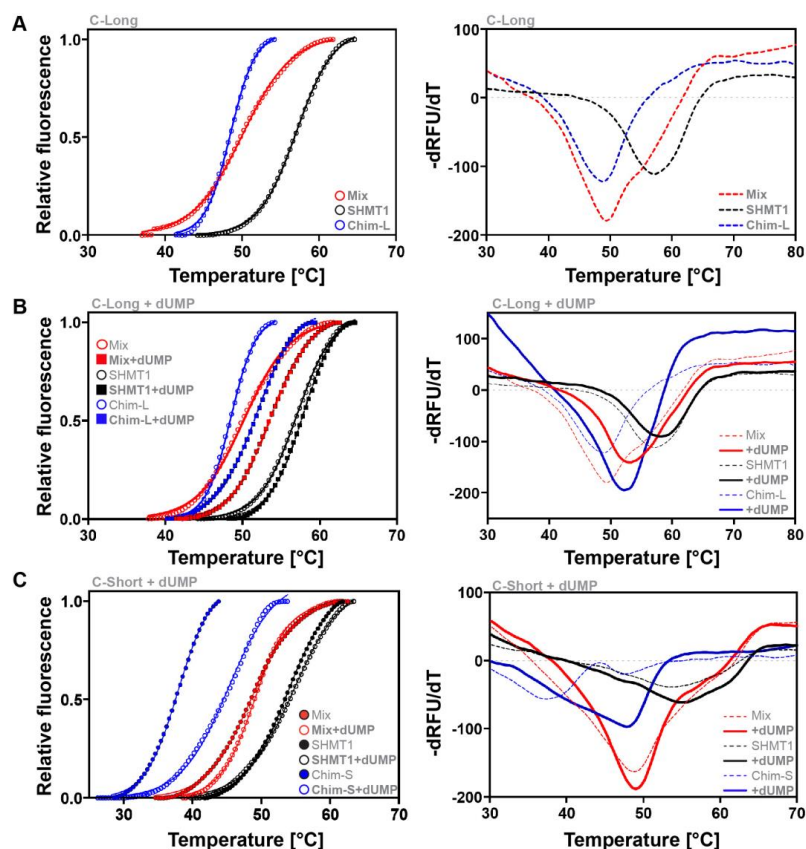


Figure 21. Differential scanning fluorimetry. SHMT1 and Chimera constructs were mixed at a final concentration of $0.5 \mu\text{M}$. When present, the concentration of CHO-THF and NADPH were 0.1 mM , dUMP was 1 mM . All samples were incubated *o.n.* at $4 \text{ }^\circ\text{C}$ in $20 \text{ mM Na HEPES - pH 7.5}$ and 50 mM NaCl , before starting the analysis. A) denaturation profiles of SHMT1, Chimera-Long and of the mixed proteins in the absence of ligands or substrates. B) Denaturation profiles of SHMT1 and Chimera-Long in the presence of dUMP (square markers and continuous lines) compared to the signals obtained with no ligands/substrates (circle markers and dashed lines). C) Same profiles as in panel B with the Short instead of Long construct. For panels all panels, plots in the left column show the change in fluorescence as a function of temperature, while in the right column the denaturation profiles are plotted as the first derivative of the fluorescence

emission as a function of temperature. The calculated T_m for all the substrates assayed are reported in table 3 [68].

	SHMT1	Chimera-Long	Mix
Buffer	57.1 ± 0.1	48.5 ± 0.5	50.2 ± 0.1
CHO-THF	57.7 ± 0.2	48.2 ± 0.5	50.5 ± 0.2
NADPH	57.3 ± 0.5	48.3 ± 0.1	50.7 ± 0.1
dUMP	57.8 ± 0.7	51.8 ± 0.3	53.6 ± 0.1
	SHMT1	Chimera Short	Mix
Buffer	53.8 ± 0.7	37.9 ± 0.2	48.5 ± 0.2
CHO-THF	54.0 ± 1.0	40.6 ± 0.3	48.0 ± 0.2
NADPH	53.6 ± 0.4	39.5 ± 0.2	47.0 ± 0.5
dUMP	54.7 ± 0.2	45.2 ± 0.2	48.9 ± 0.3

Table 3. Substrate effect on complex thermal stability T_m (°C) [68].

4.2 SHMT1-RNA complex

SHMT1, a pivotal enzyme in one carbon metabolism, has been identified as a non-canonical RNA binding protein (Liu *et al.*, 2000). Moreover, a peculiar crosstalk between SHMT1 and SHMT2 expression was observed in normal and tumour cells by our group [57,61]. This mechanism was then related to SHMT1 moonlighting ability to bind RNA. Indeed, SHMT1 was found to bind specifically and with high affinity to the 5'UTR of SHMT2 (5' UTR2), affecting the expression and the translation of the corresponding transcript [3]. Surprisingly, SHMT1 is also riboregulated by RNA as the serine cleavage is inhibited by the presence of UTR2 ($IC_{50} \sim 33$ nM), whereas the opposite reaction is not affected [3].

To unveil how RNA can riboregulate SHMT1, we have characterized the structure of the SHMT1-RNA complex by Cryo-electron microscopy technique (Cryo-EM) and here I present you the first structure of a metabolic enzyme in complex with RNA.

4.2.1 The first structure of the SHMT1-RNA complex by cryo-electron microscopy (cryo-EM)

4.2.1.1 Cryo-EM grids preparation and data collection of the SHMT1-RNA complex

To unveil the structure of SHMT1-RNA complex we employed the single particle cryo-electron microscopy (cryo-EM) technique in collaboration with Professor Bolognesi's group at the University of Milan (Italy). This method uses 2D images of single particles to reconstruct an average high-resolution density map of the analysed protein from which the 3D structure can be modelled. In brief, protein samples are loaded onto a copper grid that support a carbon foil containing holes. The excess of the sample is then removed (blotting), and the grid is flash frozen in liquid helium in order to achieve vitrification of the protein solution (figure 22).

Before collecting the data, many preliminary optimization steps were carried out to adjust sample buffer and concentration, as well as grid preparation and freezing conditions parameters. In our case, we first assessed the purity of the sample by using negative staining EM (NS-EM) at room temperature in collaboration with Dr. Matteo Ardini at University of L'Aquila, as described in section 3.2.2.1. Once we understood that the sample was pure enough to proceed with the cryo-EM, we had to screen several samples concentration as RNA is prone to aggregate on the carbon foil of the grid. We incubated purified SHMT1 with an RNA oligo corresponding to the first 50 nucleotides of the SHMT2 5'UTR transcript (UTR2 1-50), which showed to have a good binding affinity in solution towards the protein. Moreover, we observed that the DEPC water in which RNA was dissolved was causing particle aggregation and a high background of vitrified grids. We therefore decided to first dilute RNA in distilled water and then mix it with SHMT1. We tested that, in these conditions, RNA was stable for at least 30 min. A final concentration of 5.7 μM of SHMT1 and RNA was selected to have

an adequate number of particles in the grid holes; the optimal buffer concentration was 10mM Hepes pH 7.2, 75mM NaCl in order to minimize the background signal. By using a Vitrobot system six different vitrification conditions were then tested. The best ice quality was obtained with the following parameters: 3.5 μ l of sample mix volume, 30 sec of waiting time and 5 sec of blotting time. Grid screening was performed using Falcon 3EC direct electron detector (Thermo Fisher Scientific) mounted on a Thermo Fisher Talos Arctica transmission electron microscope operating at 200 kV. In the screening step, in order to evaluate the best grid, the main parameters that were taken into account were ice thickness and particle distribution. Once the best grids were identified (figure 22) we proceeded with the data collection.

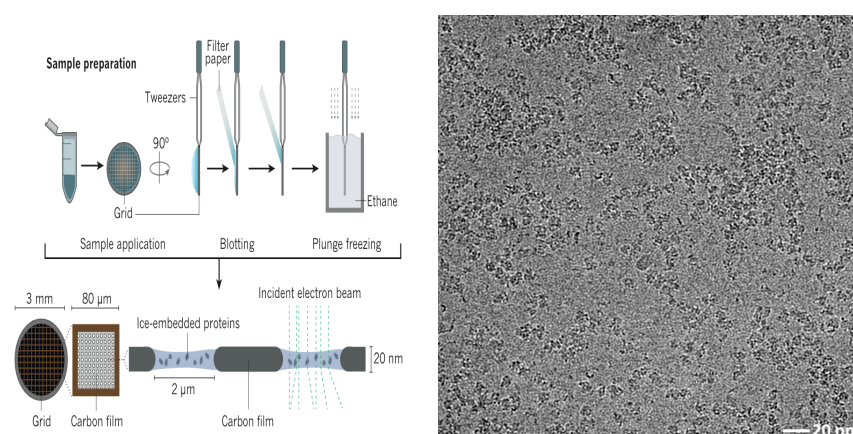


Figure 22. Cryo-EM grid preparation and micrograph. On the left, the scheme of cryo-EM grid preparation. On the right, inspection of cryo-EM micrograph of SHMT1-RNA complex. A representative micrograph is shown, from the screening obtained by mixing 5.7 μ M of SHMT1 and 5.7 μ M RNA, showing the well-distributed particles of SHMT1-RNA complex. Scalebar 20 nm.

We have performed two sets of data collection on the same grid. In the first one a total number of 1881 micrographs were acquired. Each image was acquired with an exposure time of 1.5 seconds and

a total accumulated dose of 40 electrons per \AA^2 equally distributed over 39 frames per movie. Images were acquired at nominal magnification of $\times 73,000$, corresponding to a pixel size of $1.43 \text{ \AA}/\text{pixel}$ at the specimen level, with applied defocus between -0.5 to -1.75 \mu m . For a typical cryo-EM experiment, images are collected in a range of defocus values, in order not to lose too much high-resolution information, while at the same time retaining the capability of visually detecting the individual particles.

From this first data set we were able to account for the presence of an extra-density, besides that of the SHMT1 protein, likely belonging to the RNA. But we did not have enough particles of the SHMT1-RNA complex to move forward. Therefore, to have the possibility to reconstruct the complex at high resolution, we decided to perform a second data collection on the same grid. In this case, a total number of 1849 images were collected using an exposure time of 60 seconds and a total accumulated dose of 40 electrons per \AA^2 distributed over 39 movie frames. This time the nominal magnification was of $120,000$ corresponding to a pixel size $0.889 \text{ \AA}/\text{pixel}$. With this data collection we were able to reach a 3.8 \AA overall resolution for the complex (figure 23), but the density of the RNA was not well defined, and it was not possible to perform the model building on that area.

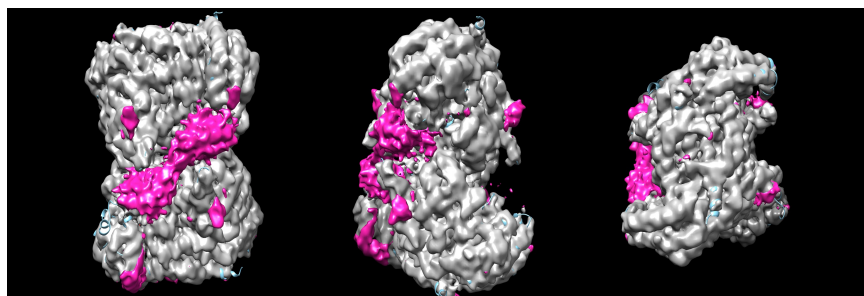


Figure 23. 3D refinement on a low-pass (60 \AA) filtered pdb map as a reference (PDB ID: 6FL5).

A final reconstruction with 18k particles using C2 symmetry, and 360 pix box was obtained at 3.8Å resolution.

Even if the secondary structures are well resolved, and the side chains are clearly visible after sharpening, the extra density is not good enough to model the RNA.

Therefore, the concentration of the RNA on the grid was increased, in order to increase the number of SHMT1-RNA particles and enhance the resolution for the RNA region.

We tested different sample concentrations, but the higher RNA concentration induced protein aggregation. However, we obtained a very good grids by mixing 5.7 μ M SHMT1 and 28.5 μ M RNA and by using the following blotting parameters; 3 μ l of sample mix volume, 20 sec of waiting time, 4 sec of blotting time (figure 24).

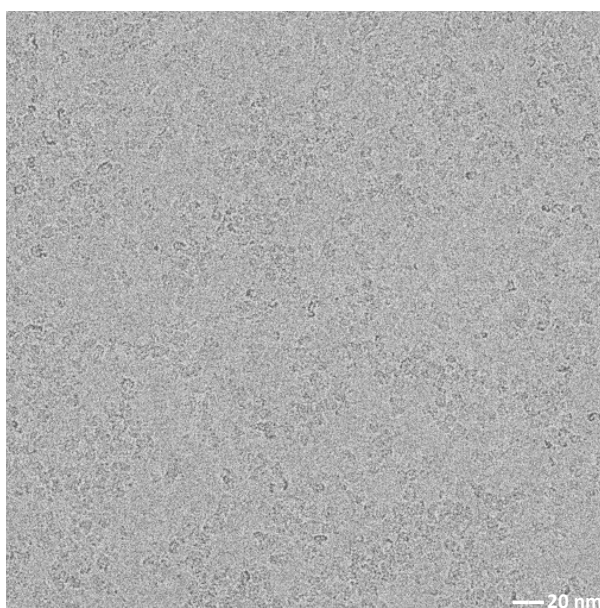


Figure 24. Inspection of cryo-EM micrograph of SHMT1-RNA complex obtained by increasing RNA concentration. A representative micrograph from the screening obtained by mixing 0.3 mg/ml of SHMT1 and 1.5 mg/ml RNA, showing the well-distributed particles of SHMT1-RNA complex. Scalebar 20nm.

We eventually performed an extensive data collection on this grid, which resulted in a huge number of total micrographs, but again, this was indeed a strategic choice: collecting as many particles as possible to increase the number of RNA-bound particles that we expected to be about 20 % of the total.

A full dataset of 5450 movie frames were collected with an exposure time of 60 seconds and a total accumulated dose of 40 electrons per Å^2 equally distributed over 40 frames per movie. Images were acquired directly in the electron counting mode with a nominal magnification of 120,000, corresponding to a pixel size of 0.889 $\text{Å}/\text{pixel}$ at the specimen level, with applied defocus between -0.5 to -2.5 μm .

4.2.1.2 Single particle analysis

4.2.1.2.1 Image processing

Prior to image processing, the 5450 movies collected were corrected for anisotropic magnification distortion and motion using MOTIONCOR2 [92]. Then contrast transfer function (CTF) estimation was performed on aligned, unweighted sum images using CTFFIND4.1 [93]. The CTF is the Fourier transform of the Point Spread Function of the experimental image and describes how the information present on the image is affected by interference with the unscattered wave, defocusing and spherical aberrations. CTF correction of the images is necessary to recover those negative contributions of the wave carrying the image of the object potential, to the detected signal, and correct the image accordingly. At this point only the micrographs that were reporting a resolution limit better than 5 Å were selected for further analysis, yielding a total number of 4977 images that were imported in RELION 2.1 [91,109]. The overall quality of the micrographs was very good.

4.2.1.2.2 Particle picking and extraction

The identification and selection of protein particles in cryo-electron micrographs is an important step in single particle analysis. This step is based on a machine learning approach, in which human eyes play a pivotal role in selecting the first subset of particles that are needed to train the software. Particle picking might be extremely challenging as the micrographs usually have very low signal-to-noise ratio (SNR) and there might be other interferences, such as ice contamination, background noise, amorphous carbon and particle overlap. Usually, in order to have a high-resolution reconstruction of the protein of interest an extensive particle quantification (> 100,000) is required [110]. In our case, particle picking was a bit tricky, as Relion's automated picking, using a Gaussian blob as a reference [91] did not give good results. We therefore launched autopicking first on a very small subset of 10 micrographs, then 2D classified the results and selected the only classes not significantly affected by background noise, trying to avoid any bias. At this point, with a better guess of the particle's shape, we launched an overall 2D-reference based autopicking process, which resulted in almost 5 million particles.

4.2.1.2.3 2D classification

An initial reference-free 2D classification was performed on the obtained particles. This procedure allows classification of the particles based on their shape and spatial orientation, which generates several 2D class averages, containing views of the particle in different orientations.

Multiple rounds of 2D classification followed by visual inspection were then performed, which resulted in a set of 24 2D classes, containing almost 900,000 clean SHMT1 particle (figure 25).

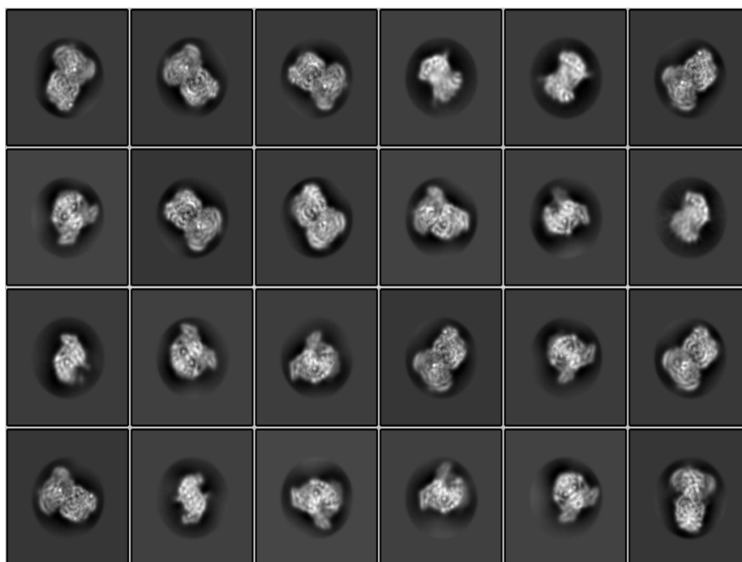


Figure 25. 2D classes of SHMT1 particles. Examples of best 2D classes resulted by different round of 2D classification.

The selection was carried out following these parameters: only particles with the best signal-to-noise ratio were selected, that correspond to those classes showing strong, white density for the macromolecular complex (with visible secondary structure elements) surrounded by a black and featureless background [109]. The distribution (how many particles were averaged in each class) and particle shape were taken into account as well.

At this stage we were not able to distinguish by eye whether the RNA was present or not and further analysis was carried out by the 3D classification.

4.2.1.2.4 3D classification

A total number of 900,000 SHMT1 clean particles were selected for subsequent 3D classification and refinement steps. The very first round of 3D classification removed 200K junk particles, reducing to

700K the final particle number. The second round immediately revealed the presence of some extra-density located near one of the four nucleotide-binding sites of SHMT1, identified in the previous data collection. Therefore, despite the symmetry of the SHMT1 tetramer is 222 we decided to perform our processing without applying any kind of symmetry axis.

Starting from the second round of 3D classification we ended up with 8 different classes, from which we realized that we could isolate, by performing multiple 3D classification runs, the two most significant classes: one containing the bound RNA density, and one containing only the SHMT1 protein (figure 26).

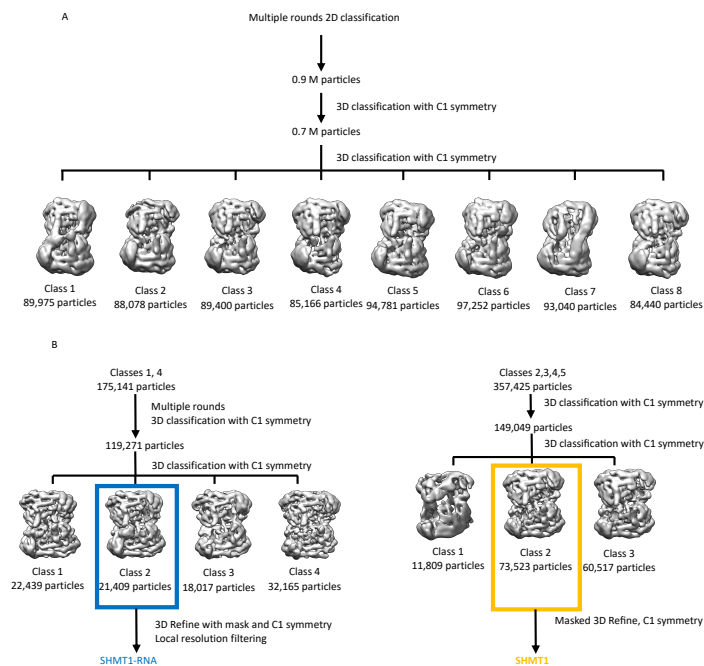


Figure 26. Schematic representation of the 3D classification. A) The first round of the 3D Classification was performed to remove junk particles. From the second round it was possible to distinguish 8 different classes, some showing the RNA

extradensity and others not. B) The best classes were selected and, by performing multiple rounds of 3D classification, two most significant classes were identified: the one containing a few bases of RNA, and one containing SHMT1 protein alone. Class 2 of the SHMT1-RNA complex was the one selected for further analysis, as it showed to be the most stable.

It is interesting to note that there were some other, lower-resolution, classes showing extra-densities at the expected oligonucleotide binding site. This indicates a relatively high degree of flexibility of the RNA oligo, that becomes more rigid only in the main docking site observed in the selected class 2. Therefore, the observed extra density accounts only for the RNA region that is in close contact with SHMT1. The overall resolution estimate of the final map of the SHMT1-RNA complex was 3.63 Å, calculated from Fourier shell correlations at 0.5 (figure 27), the estimation of the local resolution was performed using RELION 2.1.

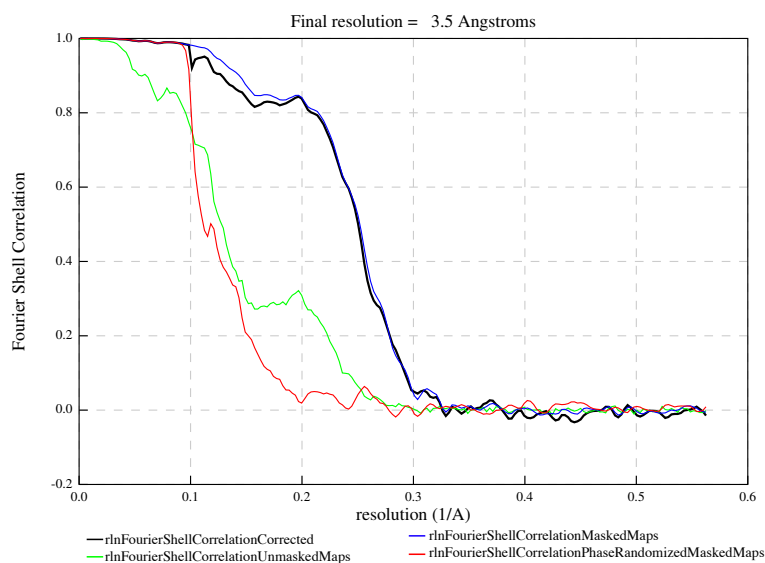


Figure 27. FSC curves calculated for independently refined half-maps. FSC curves of the SHMT1-RNA complex. The final resolution is at 3.5 Å.

4.2.1.2.5 Model building and Refinement

Prior to model fit, the cryo-EM map was sharpened using Autosharpen tool in Phenix [96], yielding a final map with a resolution of 3.52 Å (figure 28). This procedure adjusts the resolution dependence of the map to maximize its clarity.

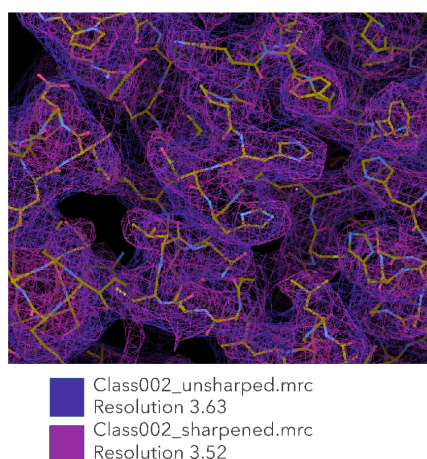


Figure 28. Comparison between unsharpened and sharpened maps. The sharpening procedure was performed by using the Autosharpen tool in Phenix [96]. The unsharpened map is coloured in purple and has an overall resolution of 3.63 Å, whereas the sharpened one is in magenta and the resolution is 3.52 Å.

An initial model for the density map corresponding to the SHMT1-RNA complex was generated using a rigid body fit of SHMT1 crystal structure (PDB: 1bj4 [53]) by using UCSF Chimera [97]. Each of the four subunits was fitted independently, to account for eventual differences with respect to the crystallographic structure due to RNA binding. The fitted model was then visualized and further improved by manual rebuilding in Coot [98]. The reconstruction showed a clear density attributable to the RNA only on one out of the four possible binding sites located on SHMT1,

meaning that we have a 1:4 stoichiometry, or at least that this stoichiometry was the most representative of the complex on the grid. The RNA binding site is in front of the active site, and it is in contact with the C-terminal domain of the chain to which is attached to, and with a long loop (flap) belonging to a monomer of the other obligate dimer of SHMT1 (figure 29).

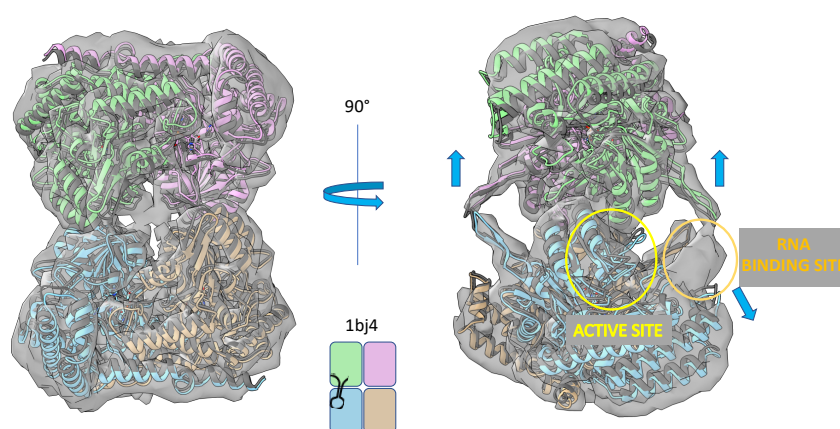


Figure 29. Preliminary Rigid body FIT of SHMT1 Single chains in cryo-EM map. The Cryo-EM map is shown here superimposed with manually fitted monomers (represented in green, pink, brown and light blue) of SHMT1 deposited to PDB 1bj4 [53]. The active site and the RNA binding site are highlighted. From a preliminary comparison it seems like a conformational change is happening.

Also, from a comparison with the SHMT1 crystallographic map (PDB: 1bj4 [53]), the cryo-EM map unveils a conformational change upon binding. In fact, it appears that two chains are moved slightly apart to accommodate the RNA molecule.

Strikingly, even though the RNA binding site is not in contact with the active site, we have observed a loss of structural details in the active site of the monomer to which the RNA is predominantly bound to and to some extent also of the active site of the upper monomer. The degree of disorder of each chain may be deduced

from the inspection of the real space correlation coefficient (CC) reported in figure 29. The CC accounts for the agreement between the model and the electron density. If a region is particularly dynamic, the corresponding electron density will be blurred or absent and the CC of the model will be low.

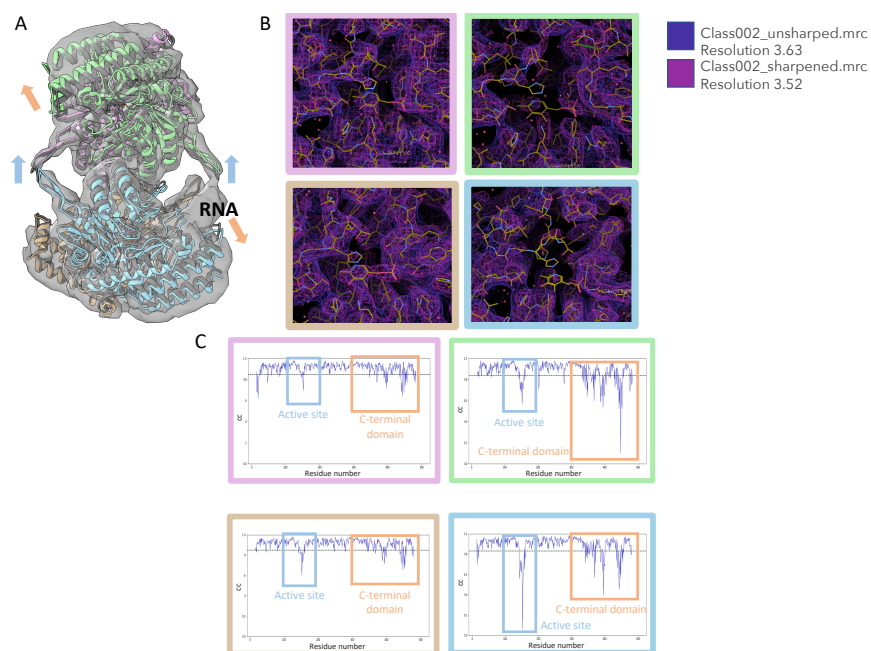


Figure 29. RNA binding affects the active site structure. A) Ribbon representation of the structure of SHMT1 (PDB: 1bj4 [53]) superimposed with the Cryo-EM map. B) Unsharpened (purple) and sharpened (magenta) maps of the active site of each monomer (purple, green, brown, and light blue) C) Real space correlation coefficient plot as a function of residue number. Plot of the active site is highlighted in light blue, whereas the C-terminal domain is highlighted in light orange.

Interestingly, we have also noticed a loss of structural details on the C-terminal domains of the above-mentioned monomers (figure 29 C). It is known that, in fold type I PLP dependent enzymes, the

transition of the active site from an ordered to a disordered conformation correlates to a higher dynamic of the C-terminal domains [111–113], therefore our current hypothesis is that the RNA might work as a conformational switch.

Overall, we had a very well-defined density throughout the model except for the mobile regions, including the RNA and the protein residues interacting with RNA. In these regions, not only the density is blurred, but the sharpening process fails to enhance the fine details and produces a lack of density. Therefore, the RNA modelling was performed at a lower resolution that does not allow for identification of the position and the nature of the single nucleobases. To overcome this problem, we are trying to increase the resolution in these regions by performing a sharpening of the cryo-EM density maps that takes into account the local temperature factors of the model, as a measure of the local dynamics [114]. This procedure (still ongoing) optimizes the contrast of Cryo-EM densities by adjusting the amplitude of the signal against the radially averaged local falloff estimated from the B-factors of a reference model. However, we tried to fit both a ssRNA or dsRNA inside the RNA density pocket and they appeared to be either too small or too big for the observed density (figure 30).

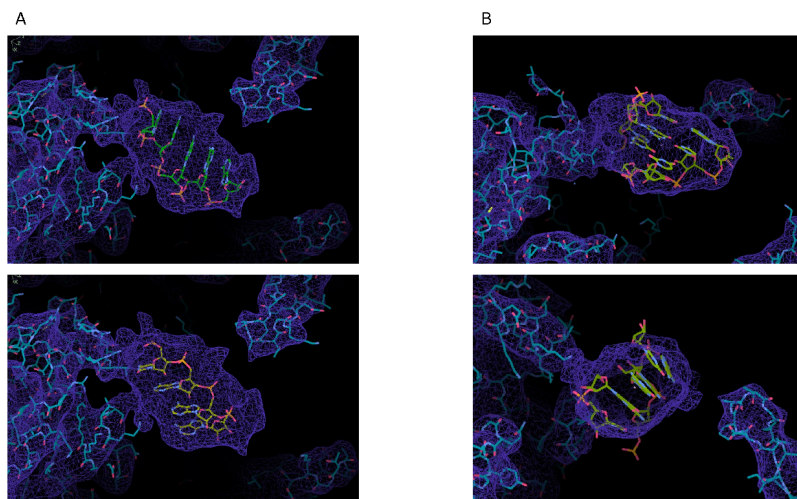


Figure 30. Density map improvement and RNA model fit. A) Model of SHMT1 and 6 nucleotides long ssRNA modelled into the density map, which appears to be too small. B) Model of SHMT1 and 7 nucleotides long dsRNA modelled into the density map, which appears to be too big for the given density. All the maps are visualized in Coot.

To help model building we then predicted the UTR2 1-50 possible structure by using the ViennaRNA package [115], which showed the presence of a harpin (figure 31). Therefore, and considering also previous results obtained by the Hentze group [49,116], in which they showed that both p62 and enolase 1 (two riboregulated proteins) bind to looped RNA regions, we envisaged the possibility that the observed density could belong to a hairpin.

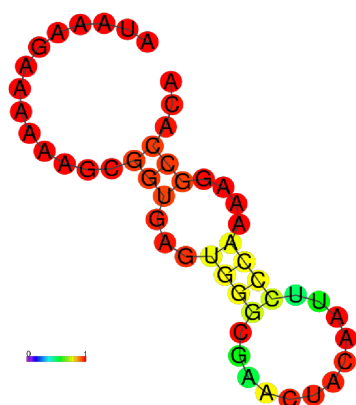


Figure 31. RNA secondary structure prediction. UTR2 1-50 structural prediction was performed by using the ViennaRNA package [115], shows that oligonucleotide has looped regions. UTR2 1-50 sequence: AUA AAG AAA AAA GCG GUG AGU GGG CGA ACU ACA AUU CC CAA AAG GCC ACA

Therefore, we tried to fit this structure in our density, and it appeared to have just the right size. To do that, we first fitted the NMR structure of a bacterial hairpin RNA (pdb: 2jr4 [100]), and then we mutated the bases followed by a real space refinement. The refinement process did not significantly alter the RNA geometry, as observed when we previously tried to refine ss or dsRNA, suggesting that the hairpin was indeed compatible with the observed density. A detail of the final fit is shown in figure 31. In our case this structure is stabilized by the OH group of the ribose which performs 2 H-bonds (figure 31).

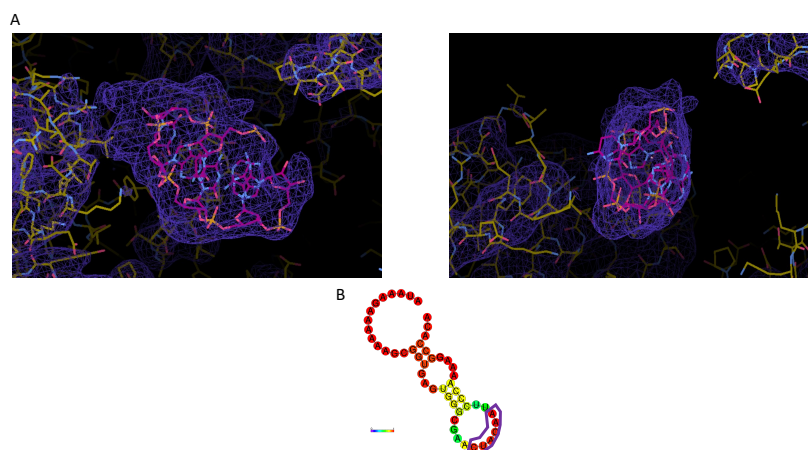


Figure 31. Density map improvement and hairpin RNA model fit. A) Model of SHMT1 and a 6 nucleotides long fragment of the hairpin region of UTR2 1-50 (CUACAA) modelled into the density map. B) UTR2 1-50 structural prediction was performed by using the ViennaRNA package [115]. Highlighted in purple the nucleotides that were fitted into the RNA density pocket.

Overall, the proposed model of the SHMT1-RNA complex was reliably fitted in the respective maps after manual building and refinement, even where the amino acid side chains were not positioned unambiguously in the density, the correlation coefficient was still around 80%.

4.2.2 The Cryo-EM structure of SHMT1 without RNA

As mentioned before, during the 3D classification we were also able to isolate the classes corresponding to the unbound enzyme. After the sharpening step, performed as described before, we obtain a final map with a resolution of 3.3 Å (figure 32).

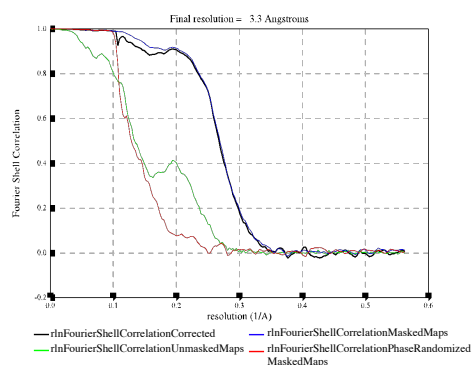


Figure 32. *FSC curves calculated for independently refined half-maps. FSC curves of the unbound SHMT1 enzyme. The final resolution is at 3.3 Å.*

Also, in this case an initial model was generated by rigid body fit of the SHMT1 monomer crystal structure (PDB: 1bj4 [53]) in the cryo-EM maps using UCSF Chimera, and the fitted model was further improved by manual rebuilding in Coot. Surprisingly also the unbound SHMT1 protein is much more mobile than in the crystal structure (figure 33). The active sites two out of the four chains appear to be disordered, similarly to the conformation of the subunits contacting the RNA in the bound protein. However, this time the chains that appear to be disordered are in a diagonal with respect to the tetramer centre. Also in this case, it is probably the enhanced C-terminal domain dynamics that leads to a deconstruction of the active site.

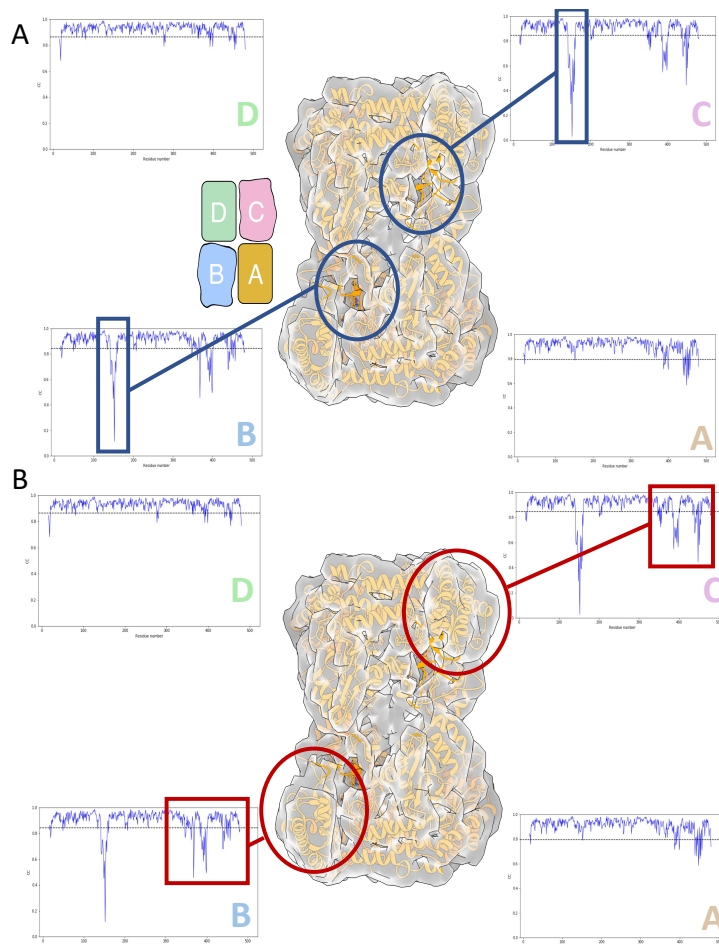


Figure 33. Rigid body fit of SHMT1 single chains in cryo-EM map. The Cryo-EM map is shown here superimposed with manually fitted monomers of the SHMT1 crystallographic structure (PDB 1bj4 [53]). The active site and the RNA binding site are highlighted in blue (A) and red (B) respectively. From a preliminary comparison it seems like a conformational change is taking place. A) In the real space correlation coefficient plot as a function of residue number the loss of structural details of the active site of chains B and C is highlighted. B) The real space correlation coefficient plot as a function of residue number shows the disordered C-terminal regions of chains B and C.

4.2.3 In vitro nucleic acid binding experiments

4.2.3.1 Validation of the binding region

To validate the RNA binding region, we have performed site directed mutagenesis in two different regions of the protein, including the amino acids in the C-terminal domain and in the loop that is in contact with RNA. The formation of a protein-nucleic acid complex is often mediated by positively charged side chains (i.e., arginine and/or lysine side chains) and by aromatic residues that make stacking interactions with the nucleobases [64]. We have therefore decided to mutate two arginines (R393 and R397), located in the C-terminal domain, into serine (double mutant R393S-R397S), and to mutate a lysine residue (K282) into a serine (single mutant K282S). This latter residue is situated in the loop that connects two monomers (figure 34).

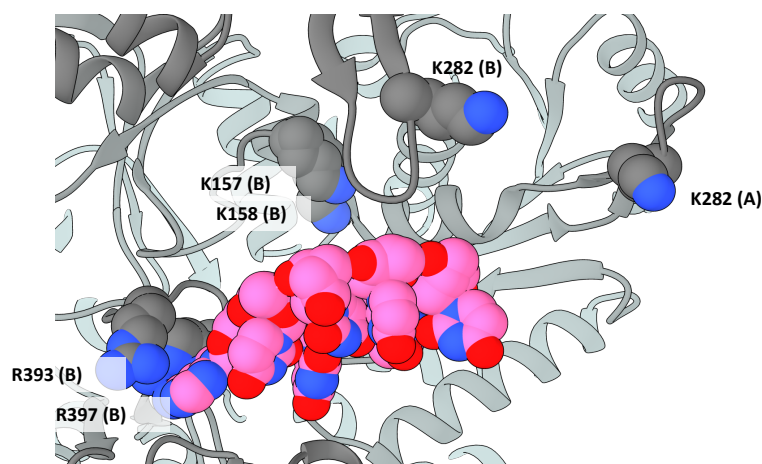


Figure 34. RNA binding residues. Ribbon representation of the folate-binding site of human SHMT1 (hSHMT, PDB id: 1BJ4 [53]). The substrate (5-CHO-THF-Glu1) is shown as sticks. The location of K157-K158 (which was previously reported to be involved in RNA binding [3]), K282 and R393-R397 is highlighted.

After successfully purifying the mutants, we analysed the RNA binding affinity of the R393S-R397S double and the K282S single SHMT1 mutants by performing electrophoretic mobility shift assays (EMSA). In this assay, a solution of different concentrations of SHMT1 (or the mutants) and a set concentration of RNA (UTR2 1-50, 0.4 μM) is prepared and incubated at 16°C for 30 minutes. Afterwards, the mixture is subjected to electrophoresis under native conditions using a 4% polyacrylamide gel. After electrophoresis, the distribution of species containing nucleic acid is determined, and the remaining unbound RNA is taken into account for the quantification of the binding affinity. As shown in figure 35, both the mutations almost completely abolished the binding of SHMT1 to UTR2 1-50. The apparent dissociation constant (K_D) of the protein-UTR2 1-50 complex was measured for the WT ($K_D 0.178 \pm 0.047 \mu\text{M}$), whereas the one of the mutants could not be calculated given the low affinity for UTR2 1-50 in the concentration range explored. These results confirmed the RNA binding region previously identified with the Cryo-EM experiments.

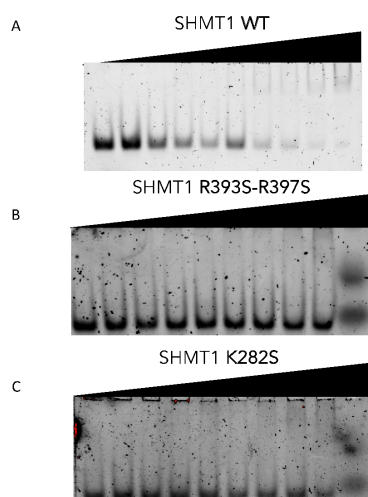


Figure 35. SHMT1 amino acid residues involved in RNA binding. A) EMSA assays carried out by incubating 0.4 μ M of UTR2 1-50 with increasing concentrations of SHMT1 (from 0 μ M to 4 μ M) B) EMSA assays with SHMT1 R393S-R397S mutant, carried out as described in A. C) EMSA assays with SHMT1 K282S mutant, carried out as described in A.

4.2.3.2 Effects of the RNA binding on SHMT1 catalytic activity

The effect of the RNA binding on SHMT1 enzymatic activity was evaluated employing biochemical assays in the presence of the selected RNA sequences. To do so, we have designed different RNAs and predicted their structure by using the RNA Vienna package [115], (summarized in table 4). The serine cleavage activity of the enzyme, whereby L-serine and THF are converted into glycine and Me-THF, was measured by incubating SHMT1 (0.2 μ M) with a saturating concentration of serine (10 mM) and 80 μ M of mono-glutamylated THF while varying the concentration of the RNAs. The ability of SHMT1 to catalyse this reaction is inhibited by all the RNAs, although to a different extent, as summarized in table 4.

By performing these experiments, we observed that the inhibition efficacy of the RNA is not sequence dependent, but rather structure dependent. In fact, when comparing UTR2 1-50 with an oligonucleotide having the same structure but a different sequence (UTR2 1-50 Rv) we noticed that the inhibition efficacy was similar (figure 36).

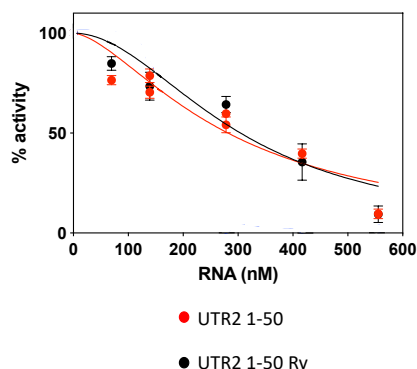


Figure 36. RNA-mediated inactivation of the serine cleavage reaction. The inactivation of SHMT1 by the RNAs was followed as described in section 3.2.2.3. The figure shows the inhibitory efficacy of UTR2 1-50 S (red) and UTR2 1-50 Rv are compared, showing a similar inhibition pattern towards SHMT1.

Furthermore, we observed that the length of the oligonucleotides appears to influence its inhibitory efficacy. As reported in table 4, the longer oligonucleotide (UTR2 loop length 28 nt) showed the lower IC₅₀ value which increased when the oligonucleotide length was shortened.

Interestingly, we noticed that the presence of a loop structure is essential. In fact, when comparing oligonucleotides having the same length, we observed that the ssRNA (Poly U) had decreased the inhibitory efficacy compared to UTR2 1-50.

Interestingly, we noticed that the presence of a secondary looped structure is essential. In fact, when comparing oligonucleotides having the same length, we observed that the ssRNA (Poly U) had decreased the inhibitory efficacy compared to UTR2 1-50.

From these preliminary results, we could understand that the presence of the looped region is essential for the inhibition of the serine cleavage activity of SHMT1, but also the length of the RNA also plays a crucial role in this context.

Moreover, when performing these experiments, we noticed a heterogeneity in the inhibitory effect of the oligonucleotides having

the same sequence but synthesized from different companies. This is probably due to a different grade of purity or to the presence of different excipients that might interfere with the properties of the oligonucleotide.

The comparison between the inhibitory efficiency between UTR2 1-50 S, purchased from Sigma Aldrich and UTR2 1-50 E, purchased from Eurofins Genomics, is reported in figure 37, in which is shown that UTR2 1-50 E has a slightly higher inhibitory effect.

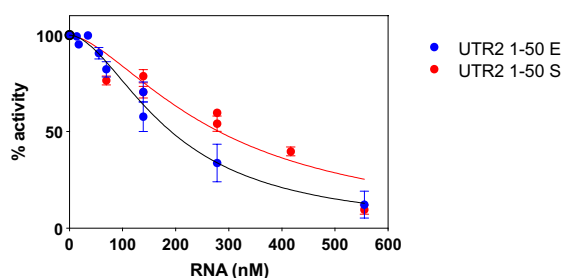


Figure 37. RNA comparison. The figure shows the comparison between the same oligonucleotide ordered from different companies UTR2 1-50 E (blue), purchased from Eurofins genomics, and UTR2 1-50 S (red) purchased from Sigma Aldrich, with UTR2 1-50 E showing a higher inhibitory efficacy.

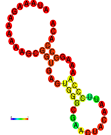
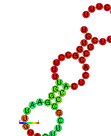
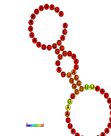
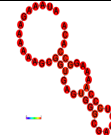


RNA name	Sequence	Length	Structure prediction	IC ₅₀ (nM)	Binding affinity
UTR2 1-50 (loop length 14 nt)	AUA AAG AAA AAA GCG GUG AGU GGG CGA ACU ACA AUU CCC AAA AGG CCA CA	50		281.8 ± 42.41	100%
UTR2 1-50 Rv (loop length 14 nt)	UGU GGC CUU UUG GGA AUU GUA GUU CGC CCA CUC ACC GCU UUU UUC UUU AU	50		307.7 ± 62.71	78,5%
UTR2 loop length 28 nt	AUA AAG AAA AAA GCG GUG AGU GGG CGA ACU AAA AAA AAA AAA ACA AUU CCC AAA AGG CCA CA	62		114.3 ± 21.64	47,5 %
UTR2 loop length 8nt	AUA AAG AAA AAA GCG GUG AGU GGG CGA AUU CCC AAA AGG CCA CA	44		221.6 ± 34.84	Not performed
UTR2 loop length 12 nt	CCC CCC CCC CGA ACU ACA AAA CGG GCC CCC CCC CC	35		348.6 ± 39.50	69,1%
dsRNA	Fw- AUA AAG AAA AAA GCG GUG AGU GGG CGA ACU ACA AUU CCC AAA AGG CCA CA Rv- UGU GGC CUU UUG GGA AUU GUA GUU CGC CCA CUC ACC GCU UUU UUC UUU AU	50		Not performed	14,09%
PolyU	UUU UUU UUU UUU UUU UUU UUU UUU UUU UUU UUU UUU UUU UUU UUU UUU UU	50		401.7 ± 69.46	Not performed

Table 4. Sequence, length, structure prediction, IC₅₀ and binding affinity of the tested RNAs.

4.2.3.3 Identification of the RNA binding determinants

To check whether our decision to fit 6 nucleotides of the hairpin region of UTR2 1-50 inside the RNA density pocket in the Cryo-EM structure was right and to see if there is a correlation between the inhibitory efficacy and the binding affinity of the RNAs, we have assessed by EMSA the binding affinity *in vitro* of SHMT1 with RNAs having different structures. In this case, we also wanted to check whether the dsRNA was able to bind *in vitro* SHMT1, therefore we decided to anneal UTR2 1-50 and UTR2 1-50 Rv.

In this experimental setup the different oligonucleotides (2 μM) were incubated with an excess of SHMT1 (4-fold). As shown in figure 38., the enzyme can retard the migration of all the oligonucleotides except for the dsRNA. The quantification of the binding affinity was performed by analysing the remaining unbound RNA. Even if SHMT1 can interact with all the hairpin-containing oligonucleotides, the best affinity is obtained with UTR2 1-50. The reverse oligonucleotide (UTR2 1-50 Rv) still has a good binding affinity, but it binds 20% less than the UTR2 1-50. When scaling down the looped region (from 14 oligonucleotides to 12) we observed a loss of almost 30% of the binding affinity, instead when enlarging the loop, the capability of binding SHMT1 was halved (table 4).

All these results strongly suggest that the looped region of UTR2 1-50 is the one responsible for the binding to SHMT1, supporting the choice of this RNA structural motif to fit in the RNA density pocket. Moreover, the fact that the UTR2 bigger loop shows a reduced affinity towards SHMT1 but has the best inhibitory efficacy might suggest that the two events, binding and riboregulation, are not fully superimposable, and that other factors might influence the inhibitory efficacy of the RNA.

Furthermore, all these results strongly suggest that the looped region of UTR2 1-50 is the one responsible for the binding to SHMT1,

supporting the choice of this RNA structural motif to fit in the RNA density pocket of the cryo-EM structure.

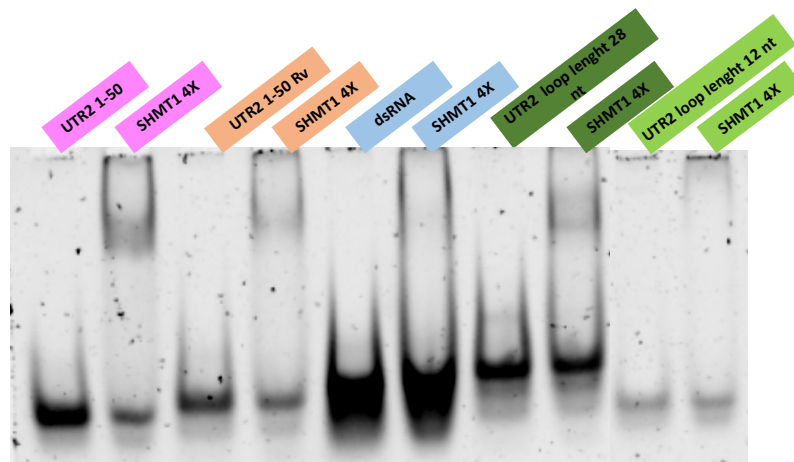


Figure 38 Comparison between the different RNA oligonucleotides. Electrophoretic mobility shift assays. Migration of 2 μ M RNA (UTR2 1-50 pink, UTR2 1-50 Rv orange, dsRNA light blue, UTR2 loop length 28nt dark green, UTR2 loop length 12nt light green) in the presence of the indicated concentrations of SHMT1.

5. Conclusions

Conclusions in brief:

- ◆ In this work, we report the intracellular dynamics of the thymidylate synthesis complex (dTMP-SC) in cancer cells by *in situ* proximity ligation assay, showing that it is also detected in the cytoplasm [68] and not only in the nucleus as previously reported in literature [4,66]. This results also indicated that the assembly of the dTMPC-SC might go beyond the mere dTMP synthesis.
- ◆ The fact that DNA was not strictly required for the assembly of the dTMPC-SC allowed us to successfully assemble the complex *in vitro*, employing tetrameric SHMT1 and a bifunctional chimeric enzyme comprising human TYMS and DHFR.
- ◆ We have estimated that K_D of the complex is $\approx 14 \mu\text{M}$ by employing the BLI technique and we unveiled an avidity effect which depends on the ability of SHMT1 to act as bidentate binder and therefore it binds two Chimera dimers. We confirmed this hypothesis modelling the interaction.
- ◆ We showed that the SHMT1 tetrameric state is required for efficient complex assembly, indicating that this aggregation state is evolutionary selected in eukaryotes to optimize protein-protein interactions.
- ◆ At this stage we cannot exclude the idea that other proteins or binding partners may interact with the dTMP-SC and modulate its assembly. Finding the missing component(s) will allow to further characterize the complex and proceed with the structural analysis.
- ◆ In the present work we also unveiled for the first time the structure of SHMT1 bound to RNA, by cryo-EM. We determined that the RNA is bound only at one out of the possible binding sites of SHMT1. We observed that the tetrameric state of SHMT1 is necessary for binding RNA, as the two monomers involved in the binding belong to different obligate dimers.

- ◆ The RNA binding led to a conformational change in the protein, as the two chains that are in contact with RNA are moved apart to accommodate the RNA molecule. Also, we observed a loss of structural details in both the active sites and the C-terminal domains of the chains that are in contact with RNA.
- ◆ The transition of the active site from an ordered to a disordered conformation might be related to a higher dynamic of the C-terminal domains.
- ◆ The overall resolution of the cryo-EM map was of 3.52 Å, but for the regions of the protein that were in contact with RNA and for the RNA itself, the resolution was blurred. This was probably due to the high degree of dynamics of these regions.
- ◆ As the resolution of the RNA density pocket was blurred, we decided to fit 6 nucleotides belonging to the hairpin region of the UTR2 1-50. We also performed binding assays in solution, that confirmed that SHMT1 preferentially binds to hairpin RNAs.
- ◆ In the cryo-EM experiment, we were also able to identify the particles belonging to the unbound SHMT1, and to 3D reconstruct them, obtaining a final map with a resolution of 3.3 Å.
- ◆ Surprisingly also the unbound SHMT1 protein is much more mobile than in the crystal structure. The active sites of two out of the four chains appears to be disordered, similarly to the conformation of the subunits contacting the RNA in the bound protein. However, we observed that this time the chains that appear to be disordered are in a different with respect to the RNA bound SHMT1.
- ◆ Thanks to the cryo-EM technique we were able to unveil the dynamism of SHMT1. The dynamics of the c-terminal domains have never been observed before because of the constraint induced by crystal packing. This result shows how cryo-EM and x-ray crystallography are complementary techniques, that can be used to reveal different aspects of the analysed structure.
- ◆ The fact that SHMT1 is extremely dynamic was already speculated in the literature [63], and this result is in agreement

with solution data suggesting that SHMT1 is a conformationally dynamic and sophisticated enzyme.

- ◆ Our hypothesis is that RNA can riboregulate SHMT1 serine cleavage activity by working as an allosteric regulator, locking the protein in a conformation that does not allow the formation of glycine but still allows the reverse reaction. This might be related to the fact that each reaction occurs with different conformational steps and the RNA works as a conformational switch.

6. Discussion

Transient interactions, which involve protein-protein or protein-biomolecules interactions that are formed and broken easily, are now emerging to be essential for many biological processes [8]. These interactions can be further divided into weak and strong. Weak transient complexes are characterized by a dissociation constant (K_D) in the micromolar range and last only few seconds, whereas strong transient interactions have a lower K_D , typically in the nanomolar range [117].

For many years transient interactions were not taken into account, due to the limitation in the detection technologies and methodologies [23]. They were therefore thought to be non-specific interactions, but as a matter of fact, transient interactions are extremely specific and they are believed to regulate the organization of biomolecular function in response to the cellular needs [24].

In the past few years, many metabolic proteins have been found to non-canonically and transiently interact with other partners within the cell, often playing other roles besides their primary function and were therefore included in the group of proteins with “moonlighting functions” [27]. More recently, moonlighting proteins are emerging as a crucial linker between different pathways, for example, offering unforeseen possibilities to connect intermediary metabolism to the RNA regulation.

In this context, serine hydroxymethyltransferase (SHMT1), a pivotal metabolic enzyme involved in one-carbon metabolism, has been found to be a multitasking protein able to transiently interact with nucleic acids as well as with other metabolic enzymes [4,60,66]. In fact, SHMT1 takes part in the *de novo* synthesis of thymidine together with two other enzymes, i.e., thymidylate synthase (TYMS) and dihydrofolate reductase (DHFR). Deoxythymidine monophosphate (dTMP) is essential for genomic stability and to avoid dUMP misincorporation in the DNA [57].

It was demonstrated that the three enzymes involved in the dTMP synthesis are forming the dTMP synthesis complex (dTMP-SC), which was detected in the nucleus attached to the nuclear lamina [4,66].

Our results on the assembly of the dTMP-SC *in vitro*, allowed to obtain relevant information on the spatial-temporal localization of the complexes. We observed that three proteins can assemble in the dTMP-SC in cytoplasm as well, and not only after nuclear translocation. Interestingly, we also observed only a slight increase in the nuclear complex localization for the S-phase synchronized A549.

These results open a new prospective for the role of this transient complex, with many implications that will require further investigations. For example, the formation of the dTMP-SC may affect not only the dTMP pool, but the whole carbon metabolism as well, as in the cytoplasm both SHMT1 and DHFR are known to participate in the folate cycle [70]. Moreover, when assembled in the complex, the surface accessibility of the single enzymes is very different with respect of the unbound states, and this might affect or control post translational modifications such as SUMOylation or affect the ability of these enzymes to bind RNA [3,60,118–120].

The evidence of the cytosolic localization also suggested that the complex can assemble in the absence of DNA, allowing us to start an *in vitro* characterization, which is an essential prerequisite for future structural characterization and rational drug design.

Complex formation *in vitro* was achieved by using a chimeric construct linking TYMS and DHFR into a single enzyme named Chimera. We designed two different linker, a shorter one (Chimera Short) and a longer one (Chimera Long). We observed complex formation by using Far western blot analysis (FWB) and immunoprecipitation (IP) analysis (figure 19). As Chimera short failed to form the complex under these conditions, we hypothesized that the longer linker is likely necessary to orient TYMS and DHFR in the right position with respect to SHMT1. The K_D of the Chimera

Long and SHMT1 complex was estimated to be $\approx 14 \mu\text{M}$ by biolayer interferometry (BLI), indicating that the dTMP-SC is a weak transient complex. When performing the BLI experiment, we also observed an avidity effect, which might depend on the ability of SHMT1 to act as bidentate binder and therefore binding two Chimera dimers, a hypothesis that we confirmed also by modelling the interaction (figure 20 B).

Remarkably, the tetrameric assembly of SHMT1 seems to be crucial in the binding process, as shown by the lack of complex formation when a dimeric mutant of SHMT1 was used (figure 19 C).

It is interesting to recall that the minimal catalytic unit of SHMT1 is the dimer [61] and only in higher organisms SHMT1 is found to be tetrameric [121]. Therefore, these results strongly suggest that the higher oligomeric state of SHMT1 may have evolved to favour novel non-canonical interactions, above and beyond the catalytic role of the protein, thus increasing the complexity of the cellular regulation network.

Strikingly, as discussed below, we obtained a similar result also from the structural analysis of the SHMT1-RNA complex.

We started the structural characterization of the latter complex because we wanted to further investigate how RNA could inhibit SHMT1 serine cleavage activity [3], a mechanism known as riboregulation. We unveiled for the first time by cryo-EM the structure of a metabolic enzyme that non-canonically binds RNA, in complex with an RNA molecule. From the comparison between the crystallographic map of SHMT1 (PDB: 1bj4 [53]) and the cryo-EM map of the SHMT1-RNA complex, we could detect the presence of an extra-density attributable to the RNA. The RNA oligonucleotide was bound only to one of the four possible binding sites located on SHMT1, with a 1:4 stoichiometry. Although there is only one RNA molecule bound to SHMT1, the tetrameric state is required to form the RNA binding site. In fact, the RNA binding site is placed in front of the active site of one monomer, contacting the C-terminal domain of this monomer and a long loop (flap) [122] belonging to a

monomer of the other obligate dimer of SHMT1 (figure 29). Therefore, the binding of RNA would not be possible without the tetrameric assembly of SHMT1. Intriguingly, recent studies on the bacterial SHMT1 have shown that it does not have a RNA binding ability [120]. Bacterial SHMT1 functions as dimer [123] further supporting the idea that the quaternary assembly in eukaryotes evolved also to allow non-canonical interactions that could enlarge the repertoire of different biochemical roles of this protein.

The comparison of the cryo-EM structure of the SHMT1-RNA complex and the crystallographic structure of the unbound enzyme, highlights that a conformational change has occurred, involving the two chains that are in contact with RNA that are slightly moved apart to accommodate the RNA molecule. Interestingly, many proteins undergo conformational changes upon transient interactions, and the most extreme cases involve disorder-to-order transitions [124]. Although the RNA binding site is not in contact with the active site, we could detect a loss of structural details in the active site near the bound RNA molecule, but also in the active site of the monomer that provides the flap loop.

With respect to the RNA binding site, the active site of this monomer is located on the opposite side of SHMT1. This suggests that the active site conformation may be influenced by even slight changes in peripheral regions of each monomer. This transition of the active site from an ordered to a disordered conformation was observed in other PLP containing enzymes, and has been related to the dynamics of the C-terminal domain [112,125,126]. Therefore, the conformational change induced by RNA binding at the level of the C-terminal domain may be responsible for the loss of structural details of the active site in closer contact to the RNA. Yet, the conformational transition of the other subunit participating to RNA binding is more difficult to explain.

During the cryo-EM analysis, we could also isolate the particles of the unbound SHMT1, leading to a final map at 3.3 Å resolution. Surprisingly, from the cryo-EM particle analysis, SHMT1 appears to be a very dynamic enzyme, and this could explain why during the 3D classification of the protein, we detected many different classes. In the final map, we observed that two out of four active sites were disordered and that the corresponding C-terminal domains had a more dynamic conformation, confirming that the structural fluctuation of this region influences the active site conformation. However, this time the disordered active sites belong to different subunits with respect to the RNA bound SHMT1 (figure 33). Such a dynamics of the C-terminal domain has never been observed before, as the only available SHMT1 structure to date was solved by x-ray crystallography, and the constraints induced by crystallographic contacts and the high symmetry of the space group (P6₂22; 1 monomer/asymmetric unit - pdb: 1bj4 [53]) does not allow a comparison of the conformation of each subunit of the tetramer [53]. Therefore, our new findings, shedding light on the actual protein conformation in solution, can explain, in agreement with the literature data, why the enzyme can show allosteric behaviour and that substrate binding implied a transition between an open and a closed (active) conformation [63]. In addition, this could also explain why, the substrate binding stoichiometry in the crystallographic structure of rabbit SHMT1 in complex with THF is 1:2 (1 tetramer and 2 THF molecules; figure 39) [127] and not 1:1 as could be expected for a non-allosteric enzyme.

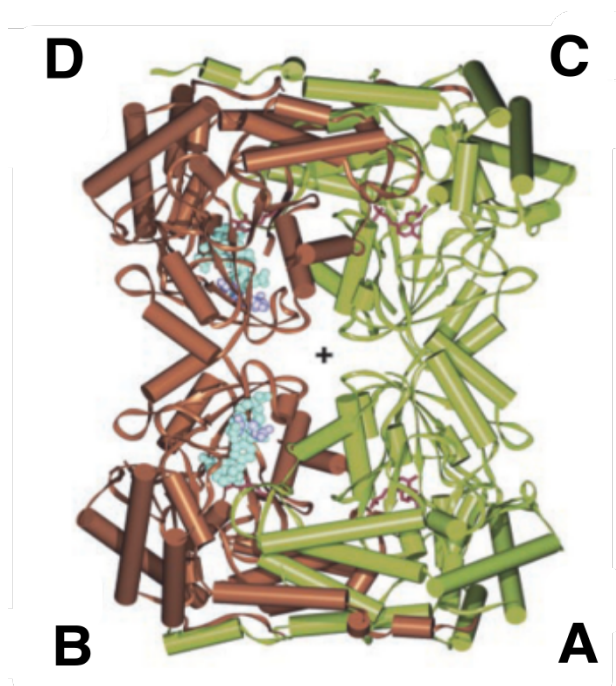


Figure 39. Orthogonal views of the rcSHMT tetramer complexed with 5-CHO-H4PteGlu 3. Subunits A + B and C + D (green and gold, respectively, in each pair) make up each tight dimer. The 5-CHO-H4PteGlu3 is rendered in subunits B and D as space-filling models. The two different conformations of the polyglutamate chain are indicated by magenta and cyan space-filling models [127].

The dynamic behaviour of this enzyme, whose structural details were unveiled in this study, might explain why, despite the great efforts made so far, there are still no effective inhibitors targeting SHMT1 [58,113,128,129]. All the tested inhibitors were designed based on the ordered (closed) structure of the active site, whereas it is likely that the folate binds to a flexible region and then induces the conformational transition to the closed conformation. If this is the case, then a molecule designed to bind to the ordered active site will bind with a much lower affinity to a flexible one. Interestingly,

if we think that the initial binding must occur in a flexible region, then polyglutamylation of THF may be a strategy to increase the affinity of SHMT1 for the substrate, while still allowing the active site to be more dynamic [121,130]. Therefore, the present result suggests that while an increased conformational mobility might have evolved to promote novel non-canonical interactions, at the same time evolution also found a way to maintain the enzymatic function efficient.

Overall, we may speculate that the weak interaction between SHMT1 and RNA, is enough to affect the allosteric equilibrium of the enzyme and result in inhibition of the enzymatic activity. Nevertheless, we are observing the inhibition of the serine cleavage activity of the enzyme, but not of the reverse reaction. The resolution for the chains that are in contact with RNA and for the RNA itself, is blurred, as there is an intrinsic mobility of these regions. Our hypothesis is that RNA might work as a conformational switch, allosterically regulating the enzyme and locking it in a conformation that is not favourable for the binding of serine, that has a ten-fold higher K_M than glycine [131].

Moreover, since each reaction could involve different conformational steps (figure 40), RNA binding may indeed inhibit only one direction of the reaction catalysed by SHMT1. It must be also recalled that SHMT1 undergoes substrate inhibition in both catalytic directions [131]. This substrate inhibition is not observed in dimeric SHMTs and is correlated to the presence of the flap motif and the tetrameric assembly [132]. Since RNA interacts with the flap motif that was shown to affect substrate binding, product release and substrate inhibition equilibrium [122] may also be correlated to the ability of RNA to affect only the serine conversion.

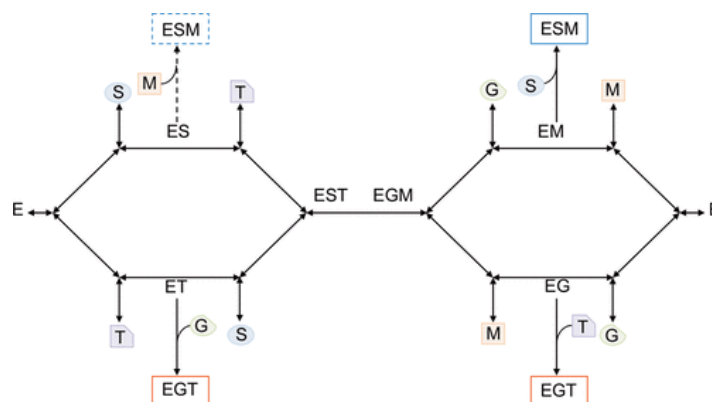


Figure 40. Scheme of the sequential random mechanism of addition of substrates and release of products followed by SHMT1. All steps in the mechanism are considered to be reversible. Abbreviations: E, enzyme; G, glycine; S, L-serine; T, tetrahydrofolate; M, 5,10- methylene-THF; EST and EGM, productive complexes; EGT and ESM, unproductive complexes [131].

As the resolution of the RNA density pocket was blurred, we decided to fit a region of 6 nucleotides belonging to the hairpin structure formed by the RNA oligonucleotide (UTR2 1-50). This decision was based also on previous literature data, which reported that other riboregulated enzymes were interacting with the looped regions of the RNAs [49,116]. We have therefore performed some *in vitro* binding assays to validate this hypothesis. We observed that the looped region of the RNA is indeed necessary to bind SHMT1 and that its optimal length is around 14 nucleotides. When using oligonucleotides having the same structure and the same length of the looped region, but with a different nucleotide sequence, we observed that a good binding affinity towards SHMT1 is still maintained, and the inhibitory efficacy is unchanged. But when using a dsRNA the binding affinity is drastically reduced and when using a ssRNA (poly U) the IC_{50} is increased with respect to the one of UTR2 1-50. These results suggest that the 3D structure, more than

its sequence is essential for binding and inhibiting SHMT1, forming the driving force of the interaction. Alike proteins, also RNA can fold into elaborate 3D structures that have various functions and can favour the interaction with proteins also RNA can fold into elaborate 3D structures that have various functions and can favour the interaction with proteins [37,48]. What emerges from both our data and those by Hentze groups, is that the hairpin motif is involved in weak interactions and riboregulation [49,116]. We have seen that also the length of the RNA plays a crucial role in the inhibition efficacy of SHMT1. In fact, the longer oligonucleotide (62 nt) has showed the higher inhibition efficacy although it isn't as effective in binding SHMT1. The fact that a longer RNA has a higher inhibitory effect was already reported by our group, indeed, a UTR2 transcript that is 200 nt long has showed a IC_{50} of 33 ± 5 nM [3].

Interestingly, with this set of preliminary experiments we might also have unveiled a different effect of the RNA binding affinity and inhibitory effect. In fact, these two events might not be strictly correlated, and the structure of the SHMT1-RNA complex might provide other possible factors, such as the conformational changes induced by RNA binding, that could participate in the inhibitory efficacy of the RNA. Therefore, an RNA oligonucleotide that has a good binding affinity towards SHMT1 might not have an optimal inhibitory efficacy because it might not work as a conformational switcher. This aspect will require a more detailed investigation in the future.

In the present work we have gained insights in the mechanism through which SHMT1 is riboregulated by RNA. We are convinced that is just the tip of the iceberg, and that more examples of these phenomena will arise in the future, also considering that many metabolic enzymes are RNA-binding proteins. We are now entering an era in which an increasing number of entangled regulatory mechanism are emerging as central for the cellular networking. In this context, weak interactions are playing a crucial role in shaping

a regulatory network that is far more sophisticated than predicted. It is becoming clearer that there isn't a net boundary between the metabolic networks and the RNA world within the cell. The new challenges in this field would be to try to re-define these blurred boundaries and to better understand to which extent the metabolic network is regulated and controlled via RNA molecules.

Therefore, in the future we will focus our attention to validate the hypothesis that the RNA has an allosteric control over SHMT1, and we will further investigate the complex physiological role of SHMT1 riboregulation.

From the structural viewpoint, we are now focusing our attention in increasing the resolution in the blurred regions by trying to perform a local sharpening procedure [114] of the cryo-EM density maps, an approach which might unveil other precious details of SHMT1-RNA interaction. Given the unsuccess of small molecules in inhibiting SHMT1, these investigations will further help in understanding how to effectively design a class of innovative SHMT1 inhibitors based on nucleic acids, targeting the accessory transient interactions performed by the enzyme and acting as allosteric drugs. In this context, a deep biochemical characterisation on the optimal length of the RNA oligonucleotide is needed to design effective inhibitors. Interestingly, all the above considerations might be applied also on SHMT2, the mitochondrial isoform of the enzyme, which was recently found to bind and be inhibited by RNA (unpublished results). Therefore, selected oligonucleotides might be strategically used to inhibit only one of the two isoforms.

In conclusion, the list of metabolic enzymes that non-canonically bind RNA is continuously growing. As already mentioned, both TYMS and DHFR have been found to be RNA-binding proteins [44,118] but is yet to be understood if they are only moonlighting proteins or if they are riboregulated as well. Also, given the fact that the major drawback of chemotherapy based on targeting TYMS and DHFR is that cancer cells can rewire their metabolism in response

to the lack of THF and dTMP by increasing the expression of both the enzymes, the idea of targeting them by using inhibitors based on nucleic acids might be extremely useful.

7. References

- [1] P. Huangyang, M.C. Simon, Hidden features: Exploring the non-canonical functions of metabolic enzymes, *DMM Dis. Model. Mech.* 11 (2018). <https://doi.org/10.1242/dmm.033365>.
- [2] A. Castello, M.W. Hentze, T. Preiss, Metabolic Enzymes Enjoying New Partnerships as RNA-Binding Proteins, *Trends Endocrinol. Metab.* 26 (2015) 746–757. <https://doi.org/10.1016/j.tem.2015.09.012>.
- [3] G. Guiducci, A. Paone, A. Tramonti, G. Giardina, S. Rinaldo, A. Bouzidi, M.C. Magnifico, M. Marani, J.A. Menendez, A. Fatica, A. MacOne, A. Armaos, G.G. Tartaglia, R. Contestabile, A. Paiardini, F. Cutruzzol, The moonlighting RNA-binding activity of cytosolic serine hydroxymethyltransferase contributes to control compartmentalization of serine metabolism, *Nucleic Acids Res.* 47 (2019) 4240–4254. <https://doi.org/10.1093/nar/gkz129>.
- [4] D. Anderson, C. Quintero, P. Stover, Identification of a de novo thymidylate biosynthesis pathway in mammalian mitochondria, *Proc. Natl. Acad. Sci. U. S. A.* 108 (2011) 15163–15168. <https://doi.org/10.1073/PNAS.1103623108>.
- [5] G.W. Beadle, E.L. Tatum, Genetic Control of Biochemical Reactions in *Neurospora*, *Proc. Natl. Acad. Sci.* 27 (1941) 499–506. <https://doi.org/10.1073/pnas.27.11.499>.
- [6] N.H. Horowitz, One-gene-one-enzyme: remembering biochemical genetics., *Protein Sci.* 4 (1995) 1017–1019. <https://doi.org/10.1002/pro.5560040524>.
- [7] G.M. Cooper, *The Complexity of Eukaryotic Genomes*, Cell A Mol. Approach. 2nd Ed. Underl. Sinauer Assoc. (2000). <https://www.ncbi.nlm.nih.gov/books/NBK9846/> (accessed October 4, 2021).
- [8] S. Ribeiro, S. Ebbinghaus, J.C. Marcos, Protein folding and quinary interactions: creating cellular organisation through functional disorder, *FEBS Lett.* 592 (2018) 3040–3053. <https://doi.org/10.1002/1873-3468.13211>.
- [9] M.M. Rickard, Y. Zhang, M. Gruebele, T. V Pogorelov, In-Cell Protein–Protein Contacts: Transient Interactions in the Crowd, *J. Phys. Chem. Lett.* 10 (2019) 2021. <https://doi.org/10.1021/acs.jpcclett.9b01556>.
- [10] A.G. Ngounou Wetie, I. Sokolowska, A.G. Woods, U. Roy, J.A. Loo, C.C. Darie, Investigation of stable and transient protein-protein interactions: Past, present, and future, *Proteomics.* 13 (2013) 538–557. <https://doi.org/10.1002/pmic.201200328>.
- [11] P. Chien, L.M. Gierasch, Challenges and dreams: Physics of weak interactions essential to life, *Mol. Biol. Cell.* 25 (2014) 3474–3477. <https://doi.org/10.1091/mbc.E14-06-1035>.
- [12] C. Sanchez, C. Lachaize, F. Janody, B. Bellon, L. Röder, J. Euzenat, F.

Rechenmann, B. Jacq, Grasping at molecular interactions and genetic networks in *Drosophila melanogaster* using FlyNets, an Internet database, *Nucleic Acids Res.* 27 (1999) 89–94. <https://doi.org/10.1093/nar/27.1.89>.

[13] T. Hazbun, S. Fields, E. de Silva, R. Stewart, H.J. An, M. Lappe, C. Wiuf, Networking proteins in yeast, *PNAS.* 98 (2008) 4277–4278. <https://doi.org/10.1073/pnas.091096398>.

[14] S. Ohlson, Designing transient binding drugs: A new concept for drug discovery, *Drug Discov. Today.* 13 (2008) 433–439. <https://doi.org/10.1016/j.drudis.2008.02.001>.

[15] C. Gómez-Moreno, M. Martínez-Júlvez, M. Medina, J.K. Hurley, G. Tollin, Protein-protein interaction in electron transfer reactions: The ferredoxin/ flavodoxin/ferredoxin:NADP⁺ reductase system from *Anabaena*, *Biochimie.* 80 (1998) 837–846. [https://doi.org/https://doi.org/10.1016/S0300-9084\(00\)88878-1](https://doi.org/https://doi.org/10.1016/S0300-9084(00)88878-1).

[16] D.G.G. McMillan, S.J. Marritt, M.A. Firer-Sherwood, L. Shi, D.J. Richardson, S.D. Evans, S.J. Elliott, J.N. Butt, L.J.C. Jeuken, Protein–Protein Interaction Regulates the Direction of Catalysis and Electron Transfer in a Redox Enzyme Complex, *J. Am. Chem. Soc.* 135 (2013) 10550–10556. <https://doi.org/10.1021/ja405072z>.

[17] E.H. McConkey, Molecular evolution, intracellular organization, and the quinary structure of proteins., *Proc. Natl. Acad. Sci. U. S. A.* 79 (1982) 3236–3240. <https://doi.org/10.1073/pnas.79.10.3236>.

[18] S.A.B. D K Srivastava, Metabolite Transfer via Enzyme-Enzyme Complexes, *Science* (80-.). 243 (1986) 1081–1086.

[19] S.E. Acuner Ozbabacan, H.B. Engin, A. Gursoy, O. Keskin, Transient proteinprotein interactions, *Protein Eng. Des. Sel.* 24 (2011) 635–648. <https://doi.org/10.1093/protein/gzr025>.

[20] L. Di Francesco, A. Verrico, I. Asteriti, P. Rovella, P. Cirigliano, G. Guarguaglini, M. Schininà, P. Lavia, Protein–DNA/RNA Interactions: An Overview of Investigation Methods in the-Omics Era, (2021). <https://doi.org/10.1021/acs.jproteome.1c00074>.

[21] F.J. Blanco, G. Montoya, Transient DNA / RNA-protein interactions, *FEBS J.* 278 (2011) 1643–1650. <https://doi.org/10.1111/j.1742-4658.2011.08095.x>.

[22] S. Sukenik, P. Ren, M. Gruebele, Weak protein-protein interactions in live cells are quantified by cell-volume modulation, *Proc. Natl. Acad. Sci. U. S. A.* 114 (2017) 6776–6781. <https://doi.org/10.1073/pnas.1700818114>.

[23] C.H. Yong, L. Wong, From the static interactome to dynamic protein complexes: Three challenges, *J. Bioinform. Comput. Biol.* 13 (2015). <https://doi.org/10.1142/S0219720015710018>.

[24] S. Ribeiro, S. Ebbinghaus, J. Ao, C. Marcos, J.C. Marcos, Protein folding

and quinary interactions: creating cellular organisation through functional disorder, (2018). <https://doi.org/10.1002/1873-3468.13211>.

[25] A.J. Wirth, M. Gruebele, Quinary protein structure and the consequences of crowding in living cells: Leaving the test-tube behind, *BioEssays*. 35 (2013) 984–993. <https://doi.org/10.1002/bies.201300080>.

[26] E. Zuckerkandl, L.B. Pauling, Molecular disease, evolution, and genetic heterogeneity., *Horizons Biochem.* Kasha, M., Pullman, B., Eds. Academic P (1962) 189–225.

[27] C.J. Jeffery, Moonlighting proteins, 0004 (1999) 8–11.

[28] C. Chen, H. Liu, S. Zabad, N. Rivera, E. Rowin, M. Hassan, S.M. Gomez De Jesus, P.S. Llinás Santos, K. Kravchenko, M. Mikhova, S. Ketterer, A. Shen, S. Shen, E. Navas, B. Horan, J. Raudsepp, C. Jeffery, MoonProt 3.0: An update of the moonlighting proteins database, *Nucleic Acids Res.* 49 (2021) D368–D372. <https://doi.org/10.1093/nar/gkaa1101>.

[29] C.J. Jeffery, Why study moonlighting proteins?, *Front. Genet.* 6 (2015). <https://doi.org/10.3389/fgene.2015.00211>.

[30] C.J. Jeffery, An introduction to protein moonlighting, *Biochem. Soc. Trans.* 42 (2014) 1679–1683. <https://doi.org/10.1042/BST20140226>.

[31] Shelley D. Copley, An Evolutionary Perspective on Protein Moonlighting, *Biochem. Soc. Trans.* 42 (2014) 1648–1891. <https://doi.org/10.1042/BST20140245>.

[32] G. Sriram, J.A. Martinez, E.R.B. McCabe, J.C. Liao, K.M. Dipple, REVIEW ARTICLE Single-Gene Disorders: What Role Could Moonlighting Enzymes Play?, *Am. J. Hum. Genet.* 76 (2005) 911–924.

[33] Y. Li, Y. Jia, Q. Che, Q. Zhou, K. Wang, X.P. Wan, AMF/PGI-mediated tumorigenesis through MAPK-ERK signaling in endometrial carcinoma, *Oncotarget*. 6 (2015) 26373–26387. <https://doi.org/10.18632/oncotarget.4708>.

[34] C.J. Jeffery, An enzyme in the test tube, and a transcription factor in the cell: Moonlighting proteins and cellular factors that affect their behavior, *Protein Sci.* 28 (2019) 1233–1238. <https://doi.org/10.1002/PRO.3645>.

[35] C.J. Jeffery, Moonlighting proteins - An update, *Mol. Biosyst.* 5 (2009) 345–350. <https://doi.org/10.1039/b900658n>.

[36] I. Brandina, J. Graham, C. Lemaitre-Guillier, N. Entelis, I. Krashennikov, L. Sweetlove, I. Tarassov, R.P. Martin, Enolase takes part in a macromolecular complex associated to mitochondria in yeast, *Biochim. Biophys. Acta - Bioenerg.* 1757 (2006) 1217–1228. <https://doi.org/https://doi.org/10.1016/j.bbabi.2006.07.001>.

[37] M.W. Hentze, A. Castello, T. Schwarzl, T. Preiss, A brave new world of RNA-binding proteins, (2018). <https://doi.org/10.1038/nrm.2017.130>.

[38] J. Cieśla, Metabolic enzymes that bind RNA: Yet another level of cellular regulatory network?, *Acta Biochim. Pol.* 53 (2006) 11–32.

https://doi.org/10.18388/abp.2006_3360.

[39] A.E. Boukouris, S.D. Zervopoulos, E.D. Michelakis, Metabolic Enzymes Moonlighting in the Nucleus: Metabolic Regulation of Gene Transcription, *Trends Biochem. Sci.* 41 (2016) 712–730. <https://doi.org/10.1016/j.tibs.2016.05.013>.

[40] J. Kim, C. V. Dang, Multifaceted roles of glycolytic enzymes, *Trends Biochem. Sci.* 30 (2005) 142–150. <https://doi.org/10.1016/J.TIBS.2005.01.005>.

[41] G. Siebert, [Determination of glycolytic metabolites in isolated cell nuclei], *Experientia*. 14 (1958) 449–449. <https://doi.org/10.1007/BF02327367>.

[42] S.M. Lee, J.H. Kim, E.J. Cho, H.D. Youn, A nucleocytoplasmic malate dehydrogenase regulates p53 transcriptional activity in response to metabolic stress, *Cell Death Differ.* 2009 165. 16 (2009) 738–748. <https://doi.org/10.1038/cdd.2009.5>.

[43] M.W. Hentze, T. Preiss, The REM phase of gene regulation, *Trends Biochem. Sci.* 35 (2010) 423–426. <https://doi.org/10.1016/j.tibs.2010.05.009>.

[44] E. Chu, C. Takimoto, D. Voeller, J. Grem, C. Allegra, Specific binding of human dihydrofolate reductase protein to dihydrofolate reductase messenger RNA in vitro, *Biochemistry.* 32 (1993) 4756–4760. <https://doi.org/10.1021/BI00069A009>.

[45] C. E. A. CJ, The role of thymidylate synthase as an RNA binding protein, *Bioessays*. 18 (1996) 191–198. <https://doi.org/10.1002/BIES.950180306>.

[46] E. Nagy, W.F.C. Rigby, Glyceraldehyde-3-phosphate Dehydrogenase Selectively Binds AU-rich RNA in the NAD⁺-binding Region (Rossmann Fold) (*), *J. Biol. Chem.* 270 (1995) 2755–2763. <https://doi.org/10.1074/JBC.270.6.2755>.

[47] C.H. Chang, J.D. Curtis, L.B. Maggi, B. Faubert, A. V. Villarino, D. O’Sullivan, S.C.C. Huang, G.J.W. Van Der Windt, J. Blagih, J. Qiu, J.D. Weber, E.J. Pearce, R.G. Jones, E.L. Pearce, Posttranscriptional control of T cell effector function by aerobic glycolysis, *Cell.* 153 (2013) 1239. <https://doi.org/10.1016/j.cell.2013.05.016>.

[48] B.M. Beckmann, A. Castello, J. Medenbach, A. Castello AlfredoCastellopalomares, The expanding universe of ribonucleoproteins: of novel RNA-binding proteins and unconventional interactions, *Pflügers Arch. - Eur. J. Physiol.* 468 (2016) 1029–1040. <https://doi.org/10.1007/s00424-016-1819-4>.

[49] R. Horos, M. Büscher, R. Kleinendorst, A.M. Alleaume, A.K. Tarafder, T. Schwarzl, D. Dziuba, C. Tischer, E.M. Zielonka, A. Adak, A. Castello, W. Huber, C. Sachse, M.W. Hentze, The Small Non-coding Vault RNA1-1 Acts as a Riboregulator of Autophagy, *Cell.* 176 (2019) 1054–1067.e12. <https://doi.org/10.1016/j.cell.2019.01.030>.

[50] J. Hennig, M. Sattler, Deciphering the protein-RNA recognition code:

Combining large-scale quantitative methods with structural biology, *BioEssays*. 37 (2015) 899–908. <https://doi.org/10.1002/BIES.201500033>.

[51] V. Schirch, D.M.E. Szebenyi, Serine hydroxymethyltransferase revisited, *Curr. Opin. Chem. Biol.* 9 (2005) 482–487. <https://doi.org/10.1016/j.cbpa.2005.08.017>.

[52] G. Giardina, P. Brunotti, A. Fiascarelli, A. Cicalini, M.G.S. Costa, A.M. Buckle, M.L. di Salvo, A. Giorgi, M. Marani, A. Paone, S. Rinaldo, A. Paiardini, R. Contestabile, F. Cutruzzolà, How pyridoxal 5'-phosphate differentially regulates human cytosolic and mitochondrial serine hydroxymethyltransferase oligomeric state, *FEBS J.* 282 (2015) 1225–1241. <https://doi.org/10.1111/febs.13211>.

[53] S.B. Renwick, K. Snell, U. Baumann, The crystal structure of human cytosolic serine hydroxymethyltransferase: A target for cancer chemotherapy, *Structure*. 6 (1998) 1105–1116. [https://doi.org/10.1016/S0969-2126\(98\)00112-9](https://doi.org/10.1016/S0969-2126(98)00112-9).

[54] R. Florio, M.L. Di Salvo, M. Vivoli, R. Contestabile, Serine hydroxymethyltransferase: A model enzyme for mechanistic, structural, and evolutionary studies, *Biochim. Biophys. Acta - Proteins Proteomics*. 1814 (2011) 1489–1496. <https://doi.org/10.1016/j.bbapap.2010.10.010>.

[55] D. Hanahan, R.A. Weinberg, Hallmarks of Cancer: The Next Generation, *Cell*. 144 (2011) 646–674. <https://doi.org/10.1016/j.cell.2011.02.013>.

[56] K.R. Mattaini, M.R. Sullivan, M.G. Vander Heiden, The importance of serine metabolism in cancer, *J. Cell Biol.* 214 (2016) 249–257. <https://doi.org/10.1083/jcb.201604085>.

[57] A. Paone, M. Marani, A. Fiascarelli, S. Rinaldo, G. Giardina, R. Contestabile, A. Paiardini, F. Cutruzzolà, SHMT1 knockdown induces apoptosis in lung cancer cells by causing uracil misincorporation, *Cell Death Dis.* 5 (2014) 1–11. <https://doi.org/10.1038/cddis.2014.482>.

[58] A. Tramonti, A. Paiardini, A. Paone, A. Bouzidi, G. Giardina, G. Guiducci, M.C. Magnifico, S. Rinaldo, L. McDermott, J.A. Menendez, R. Contestabile, F. Cutruzzolà, Differential inhibitory effect of a pyrazolopyran compound on human serine hydroxymethyltransferase-amino acid complexes, *Arch. Biochem. Biophys.* 653 (2018) 71–79. <https://doi.org/10.1016/J.ABB.2018.07.001>.

[59] G.S. Ducker, J.D. Rabinowitz, One-Carbon Metabolism in Health and Disease, *Cell Metab.* 25 (2017) 27–42. <https://doi.org/10.1016/J.CMET.2016.08.009>.

[60] X. Liu, B. Reig, I. Nasrallah, P. Stover, Human cytoplasmic serine hydroxymethyltransferase is an mRNA binding protein, *Biochemistry*. 39 (2000) 11523–11531. <https://doi.org/10.1021/BI000665D>.

[61] G. Giardina, A. Paone, A. Tramonti, R. Lucchi, M. Marani, M.C. Magnifico, A. Bouzidi, V. Pontecorvi, G. Guiducci, C. Zamparelli, S. Rinaldo, A.

- Paiardini, R. Contestabile, F. Cutruzzolà, The catalytic activity of serine hydroxymethyltransferase is essential for de novo nuclear dTMP synthesis in lung cancer cells, *FEBS J.* 285 (2018) 3238–3253. <https://doi.org/10.1111/febs.14610>.
- [62] G. Ducker, L. Chen, R. Morscher, J. Ghergurovich, M. Esposito, X. Teng, Y. Kang, J. Rabinowitz, Reversal of Cytosolic One-Carbon Flux Compensates for Loss of the Mitochondrial Folate Pathway, *Cell Metab.* 23 (2016) 1140–1153. <https://doi.org/10.1016/J.CMET.2016.04.016>.
- [63] V. Schirch, K. Shostak, M. Zamora, M. Guatam-Basak, The origin of reaction specificity in serine hydroxymethyltransferase., *J. Biol. Chem.* 266 (1991) 759–764. [https://doi.org/https://doi.org/10.1016/S0021-9258\(17\)35237-7](https://doi.org/https://doi.org/10.1016/S0021-9258(17)35237-7).
- [64] A. Cléry, M. Blatter, F.H.T. Allain, RNA recognition motifs: boring? Not quite, *Curr. Opin. Struct. Biol.* 18 (2008) 290–298. <https://doi.org/10.1016/J.SBI.2008.04.002>.
- [65] J. Chon, P. Stover, M. Field, Targeting nuclear thymidylate biosynthesis, *Mol. Aspects Med.* 53 (2017) 48–56. <https://doi.org/10.1016/J.MAM.2016.11.005>.
- [66] D.D. Anderson, C.F. Woeller, E.P. Chiang, B. Shane, P.J. Stover, Serine hydroxymethyltransferase anchors de Novo thymidylate synthesis pathway to nuclear lamina for DNA synthesis, *J. Biol. Chem.* 287 (2012) 7051–7062. <https://doi.org/10.1074/jbc.M111.333120>.
- [67] M. AJ, L. X, P. CA, F. P, A. RH, S. SP, S. PJ, Cytoplasmic serine hydroxymethyltransferase regulates the metabolic partitioning of methylenetetrahydrofolate but is not essential in mice, *J. Biol. Chem.* 283 (2008) 25846–25853. <https://doi.org/10.1074/JBC.M802671200>.
- [68] S. Spizzichino, D. Boi, G. Boumis, R. Lucchi, F.R. Liberati, D. Capelli, R. Montanari, G. Pochetti, R. Piacentini, G. Parisi, A. Paone, S. Rinaldo, R. Contestabile, A. Tramonti, A. Paiardini, G. Giardina, F. Cutruzzolà, Cytosolic localization and in vitro assembly of human de novo thymidylate synthesis complex, *FEBS J.* (2021). <https://doi.org/10.1111/FEBS.16248>.
- [69] G. Prem veer Reddy, A. Pardee, Multienzyme complex for metabolic channeling in mammalian DNA replication, *Proc. Natl. Acad. Sci. U. S. A.* 77 (1980) 3312–3316. <https://doi.org/10.1073/PNAS.77.6.3312>.
- [70] M. Field, E. Kamynina, J. Chon, P. Stover, Nuclear Folate Metabolism, *Annu. Rev. Nutr.* 38 (2018) 219–243. <https://doi.org/10.1146/ANNUREV-NUTR-071714-034441>.
- [71] D.D. Anderson, P.J. Stover, SHMT1 and SHMT2 Are Functionally Redundant in Nuclear De novo Thymidylate Biosynthesis, *PLoS One.* 4 (2009) e5839. <https://doi.org/10.1371/journal.pone.0005839>.
- [72] J. Locasale, Serine, glycine and one-carbon units: cancer metabolism in full circle, *Nat. Rev. Cancer.* 13 (2013) 572–583. <https://doi.org/10.1038/NRC3557>.

- [73] E. Bramucci, A. Paiardini, F. Bossa, S. Pascarella, PyMod: sequence similarity searches, multiple sequence-structure alignments, and homology modeling within PyMOL, *BMC Bioinformatics*. 13 (2012) S2. <https://doi.org/10.1186/1471-2105-13-S4-S2>.
- [74] A. Šali, T.L. Blundell, Comparative Protein Modelling by Satisfaction of Spatial Restraints, *J. Mol. Biol.* 234 (1993) 779–815. <https://doi.org/10.1006/JMBI.1993.1626>.
- [75] C. OEFNER, A. D'ARCY, F.K. WINKLER, Crystal structure of human dihydrofolate reductase complexed with folate, *Eur. J. Biochem.* 174 (1988) 377–385. <https://doi.org/10.1111/J.1432-1033.1988.TB14108.X>.
- [76] R. Almog, C.A. Waddling, F. Maley, G.F. Maley, P. Van Roey, Crystal structure of a deletion mutant of human thymidylate synthase Δ (7–29) and its ternary complex with Tomudex and dUMP, *Protein Sci.* 10 (2001) 988. <https://doi.org/10.1110/PS.47601>.
- [77] N. Baker, D. Sept, S. Joseph, M. Holst, J. McCammon, Electrostatics of Nanosystems: Application to Microtubules and the Ribosome, *Proc. Natl. Acad. Sci.* 98 (2001) 10037–10041. <https://doi.org/10.1073/pnas.181342398>.
- [78] A. Paiardini, F. Bossa, S. Pascarella, CAMPO, SCR_FIND and CHC_FIND: a suite of web tools for computational structural biology, *Nucleic Acids Res.* 33 (2005) W50. <https://doi.org/10.1093/NAR/GKI416>.
- [79] F. Mantoni, A. Paiardini, P. Brunotti, C. D'Angelo, L. Cervoni, A. Paone, L. Cappellacci, R. Petrelli, M. Ricciutelli, L. Leoni, G. Rampioni, A. Arcovito, S. Rinaldo, F. Cutruzzolà, G. Giardina, Insights into the GTP-dependent allosteric control of c-di.GMP hydrolysis from the crystal structure of PA0575 protein from *Pseudomonas aeruginosa*, *FEBS J.* 285 (2018) 3815–3834. <https://doi.org/10.1111/febs.14634>.
- [80] Y. Wu, Q. Li, X.-Z. Chen, Detecting protein–protein interactions by far western blotting, *Nat. Protoc.* 2007 212. 2 (2007) 3278–3284. <https://doi.org/10.1038/nprot.2007.459>.
- [81] P.H. Sayre, J.S. Finer-Moore, T.A. Fritz, D. Biermann, S.B. Gates, W.C. MacKellar, V.F. Patel, R.M. Stroud, Multi-targeted antifolates aimed at avoiding drug resistance form covalent closed inhibitory complexes with human and *Escherichia coli* thymidylate synthases, *J. Mol. Biol.* 313 (2001) 813–829. <https://doi.org/10.1006/JMBI.2001.5074>.
- [82] I. Jay F. Davies, T.J. Delcamp, N.J. Prendergast, V.A. Ashford, J.H. Freisheim, J. Kraut, Crystal structures of recombinant human dihydrofolate reductase complexed with folate and 5-deazafolate, *Biochemistry.* 29 (2002) 9467–9479. <https://doi.org/10.1021/BI00492A021>.
- [83] S. Vilar, G. Cozza, S. Moro, Medicinal Chemistry and the Molecular Operating Environment (MOE): Application of QSAR and Molecular Docking to Drug Discovery, *Curr. Top. Med. Chem.* 8 (2008) 1555–1572.

<https://doi.org/10.2174/156802608786786624>.

[84] G. Janson, A. Paiardini, PyMod 3: a complete suite for structural bioinformatics in PyMOL, *Bioinformatics*. 37 (2021) 1471–1472. <https://doi.org/10.1093/BIOINFORMATICS/BTAA849>.

[85] G.C.P. Van Zundert, J.P.G.L.M. Rodrigues, M. Trellet, C. Schmitz, P.L. Kastritis, E. Karaca, A.S.J. Melquiond, M. Van Dijk, S.J. De Vries, A.M.J.J. Bonvin, The HADDOCK2.2 Web Server: User-Friendly Integrative Modeling of Biomolecular Complexes, *J. Mol. Biol.* 428 (2016) 720–725. <https://doi.org/10.1016/J.JMB.2015.09.014>.

[86] D. Frenzel, D. Willbold, Kinetic Titration Series with Biolayer Interferometry, *PLoS One*. 9 (2014) e106882. <https://doi.org/10.1371/JOURNAL.PONE.0106882>.

[87] M. Adrian, J. Dubochet, J. Lepault, A.W. McDowell, Cryo-electron microscopy of viruses, *Nature*. 308 (1984) 32–36. <https://doi.org/10.1038/308032A0>.

[88] J. Frank, A. Verschoor, M. Boublik, Computer Averaging of Electron Micrographs of 40S Ribosomal Subunits, *Science* (80-.). 214 (1981) 1353–1355. <https://doi.org/10.1126/SCIENCE.7313694>.

[89] E.H. Egelman, The Current Revolution in Cryo-EM, (2016). <https://doi.org/10.1016/j.bpj.2016.02.001>.

[90] G.G. Sgro, T.R.D. Costa, Cryo-EM Grid Preparation of Membrane Protein Samples for Single Particle Analysis, *Front. Mol. Biosci.* 0 (2018) 74. <https://doi.org/10.3389/FMOLB.2018.00074>.

[91] S.H.W. Scheres, RELION: Implementation of a Bayesian approach to cryo-EM structure determination, *J. Struct. Biol.* 180 (2012) 519–530. <https://doi.org/10.1016/J.JSB.2012.09.006>.

[92] S.Q. Zheng, E. Palovcak, J.-P. Armache, K.A. Verba, Y. Cheng, D.A. Agard, MotionCor2: anisotropic correction of beam-induced motion for improved cryo-electron microscopy, *Nat. Methods* 2017 144. 14 (2017) 331–332. <https://doi.org/10.1038/nmeth.4193>.

[93] A. Rohou, N. Grigorieff, CTFFIND4: Fast and accurate defocus estimation from electron micrographs, *J. Struct. Biol.* 192 (2015) 216–221. <https://doi.org/10.1016/J.JSB.2015.08.008>.

[94] M. Beckers, C. Sachse, Permutation testing of Fourier shell correlation for resolution estimation of cryo-EM maps, *J. Struct. Biol.* 212 (2020) 107579. <https://doi.org/10.1016/J.JSB.2020.107579>.

[95] P.B. Rosenthal, R. Henderson, Optimal Determination of Particle Orientation, Absolute Hand, and Contrast Loss in Single-particle Electron Cryomicroscopy, *J. Mol. Biol.* 333 (2003) 721–745. <https://doi.org/10.1016/J.JMB.2003.07.013>.

[96] D. Lieschner, P.V. Afonine, M.L. Baker, G. Bunkóczi, V.B. Chen, T.I.

- Croll, B. Hintze, L.-W. Hung, S. Jain, A.J. McCoy, N.W. Moriarty, R.D. Oeffner, B.K. Poon, M.G. Prisant, R.J. Read, J.S. Richardson, D.C. Richardson, M.D. Sammito, O.V. Sobolev, D.H. Stockwell, T.C. Terwilliger, A.G. Urzhumtsev, L.L. Videau, C.J. Williams, P.D. Adams, IUCr, *Macromolecular structure determination using X-rays, neutrons and electrons: recent developments in Phenix*, *Urn:Issn:2059-7983*. 75 (2019) 861–877. <https://doi.org/10.1107/S2059798319011471>.
- [97] E.F. Pettersen, T.D. Goddard, C.C. Huang, G.S. Couch, D.M. Greenblatt, E.C. Meng, T.E. Ferrin, UCSF Chimera—A visualization system for exploratory research and analysis, *J. Comput. Chem.* 25 (2004) 1605–1612. <https://doi.org/10.1002/JCC.20084>.
- [98] P. Emsley, B. Lohkamp, W. Scott, K. Cowtan, Features and development of Coot, *Acta Crystallogr. D. Biol. Crystallogr.* 66 (2010) 486–501. <https://doi.org/10.1107/S0907444910007493>.
- [99] P.D. Adams, P. V Afonine, G. Bunkó, V.B. Chen, I.W. Davis, N. Echols, J.J. Headd, L.-W. Hung, G.J. Kapral, R.W. Grosse-Kunstleve, A.J. McCoy, N.W. Moriarty, R. Oeffner, R.J. Read, D.C. Richardson, J.S. Richardson, T.C. Terwilliger, P.H. Zwart, Biological Crystallography PHENIX: a comprehensive Python-based system for macromolecular structure solution, *Res. Pap. Acta Cryst.* 66 (2010) 213–221. <https://doi.org/10.1107/S0907444909052925>.
- [100] F.A.P. Vendeix, A. Dziergowska, E.M. Gustilo, W.D. Graham, B. Sproat, A. Malkiewicz, P.F. Agris, Anticodon Domain Modifications Contribute Order to tRNA for Ribosome-Mediated Codon Binding†‡, *Biochemistry*. 47 (2008) 6117–6129. <https://doi.org/10.1021/BI702356J>.
- [101] P.V. Afonine, B.K. Poon, R.J. Read, O.V. Sobolev, T.C. Terwilliger, A. Urzhumtsev, P.D. Adams, IUCr, Real-space refinement in PHENIX for cryo-EM and crystallography, *Urn:Issn:2059-7983*. 74 (2018) 531–544. <https://doi.org/10.1107/S2059798318006551>.
- [102] T. Fu, J. Rife, V. Schirch, The role of serine hydroxymethyltransferase isozymes in one-carbon metabolism in MCF-7 cells as determined by ¹³C NMR, *Arch. Biochem. Biophys.* 393 (2001) 42–50. <https://doi.org/10.1006/ABBI.2001.2471>.
- [103] R.H. O’Neil, R.H. Lilien, B.R. Donald, R.M. Stroud, A.C. Anderson, Phylogenetic Classification of Protozoa Based on the Structure of the Linker Domain in the Bifunctional Enzyme, Dihydrofolate Reductase-Thymidylate Synthase *, *J. Biol. Chem.* 278 (2003) 52980–52987. <https://doi.org/10.1074/JBC.M310328200>.
- [104] D.W. Begley, T.E. Edwards, A.C. Raymond, E.R. Smith, R.C. Hartley, J. Abendroth, B. Sankaran, D.D. Lorimer, P.J. Myler, B.L. Staker, L.J. Stewart, Structural Biology and Crystallization Communications Inhibitor-bound complexes of dihydrofolate reductase-thymidylate synthase from *Babesia bovis*,

- Struct. Commun. Acta Cryst. (2011) 67. <https://doi.org/10.1107/S1744309111029009>.
- [105] N. Schormann, O. Senkovich, K. Walker, D.L. Wright, A.C. Anderson, A. Rosowsky, S. Ananthan, B. Shinkre, S. Velu, D. Chattopadhyay, Structure-based approach to pharmacophore identification, in silico screening, and three-dimensional quantitative structure–activity relationship studies for inhibitors of *Trypanosoma cruzi* dihydrofolate reductase function, *Proteins Struct. Funct. Bioinforma.* 73 (2008) 889–901. <https://doi.org/10.1002/PROT.22115>.
- [106] R.T. Reilly, A.M. Forsthoefel, F.G. Berger, Functional Effects of Amino Acid Substitutions at Residue 33 of Human Thymidylate Synthase, *Arch. Biochem. Biophys.* 342 (1997) 338–343. <https://doi.org/10.1006/ABBI.1997.0116>.
- [107] A. Calascibetta, F. Contino, S. Feo, G. Gulotta, M. Cajozzo, A. Antona, G. Sanguedolce, R. Sanguedolce, Analysis of the Thymidylate Synthase Gene Structure in Colorectal Cancer Patients and Its Possible Relation with the 5-Fluorouracil Drug Response, *J. Nucleic Acids.* 2010 (2010). <https://doi.org/10.4061/2010/306754>.
- [108] Z. Islam, I. Gurevic, T.S. Strutzenberg, A.K. Ghosh, T. Iqbal, A. Kohen, Bacterial versus human thymidylate synthase: Kinetics and functionality, *PLoS One.* 13 (2018) e0196506. <https://doi.org/10.1371/JOURNAL.PONE.0196506>.
- [109] S. He, S.H.W. Scheres, Helical reconstruction in RELION, *J. Struct. Biol.* 198 (2017) 163–176. <https://doi.org/10.1016/J.JSB.2017.02.003>.
- [110] M. Serna, Hands on Methods for High Resolution Cryo-Electron Microscopy Structures of Heterogeneous Macromolecular Complexes, *Front. Mol. Biosci.* 0 (2019) 33. <https://doi.org/10.3389/FMOLB.2019.00033>.
- [111] S. Spizzichino, G. Pampalone, M. Dindo, A. Bruno, L. Romani, F. Cutruzzolà, T. Zelante, M. Pieroni, B. Cellini, G. Giardina, Crystal structure of *Aspergillus fumigatus* AroH, an aromatic amino acid aminotransferase, *Proteins Struct. Funct. Bioinforma.* (2021). <https://doi.org/10.1002/PROT.26234>.
- [112] G. Giardina, R. Montioli, S. Gianni, B. Cellini, A. Paiardini, C.B. Voltattorni, F. Cutruzzolà, Open conformation of human DOPA decarboxylase reveals the mechanism of PLP addition to Group II decarboxylases, *Proc. Natl. Acad. Sci. U. S. A.* 108 (2011) 20514. <https://doi.org/10.1073/PNAS.1111456108>.
- [113] A. Paiardini, A. Fiascarelli, S. Rinaldo, F. Daidone, G. Giardina, D.R. Koes, A. Parroni, G. Montini, M. Marani, A. Paone, L.A. McDermott, R. Contestabile, F. Cutruzzolà, Screening and In Vitro Testing of Antifolate Inhibitors of Human Cytosolic Serine Hydroxymethyltransferase, *ChemMedChem.* 10 (2015) 490–497. <https://doi.org/10.1002/CMDC.201500028>.
- [114] A.J. Jakobi, M. Wilmanns, C. Sachse, Model-based local density sharpening of cryo-EM maps, *Elife.* 6 (2017).

<https://doi.org/10.7554/ELIFE.27131>.

[115] I.L. Hofacker, Vienna RNA secondary structure server, *Nucleic Acids Res.* 31 (2003) 3429–3431. <https://doi.org/10.1093/nar/gkg599>.

[116] I. Huppertz, J.I. Perez-Perri, P. Mantas, T. Sekaran, T. Schwarzl, L. Dimitrova-Paternoga, J. Hennig, P.A. Neveu, M.W. Hentze, RNA regulates Glycolysis and Embryonic Stem Cell Differentiation via Enolase 1, *BioRxiv.* (2020) 2020.10.14.337444. <https://doi.org/10.1101/2020.10.14.337444>.

[117] I.M.A. Nooren, J.M. Thornton, Structural characterisation and functional significance of transient protein-protein interactions, *J. Mol. Biol.* 325 (2003) 991–1018. [https://doi.org/10.1016/S0022-2836\(02\)01281-0](https://doi.org/10.1016/S0022-2836(02)01281-0).

[118] E. Chu, C.J. Allegra, The role of thymidylate synthase as an RNA binding protein, *BioEssays.* 18 (1996) 191–198. <https://doi.org/10.1002/BIES.950180306>.

[119] N.D. Brunn, S.M. Dibrov, M.B. Kao, M. Ghassemian, T. Hermann, Analysis of mRNA recognition by human thymidylate synthase, *Biosci. Rep.* 34 (2014) 168. <https://doi.org/10.1042/BSR20140137>.

[120] M. Monti, G. Guiducci, A. Paone, S. Rinaldo, G. Giardina, F.R. Liberati, F. Cutruzzolà, G.G. Tartaglia, Modelling of SHMT1 riboregulation predicts dynamic changes of serine and glycine levels across cellular compartments, *Comput. Struct. Biotechnol. J.* 19 (2021) 3034–3041. <https://doi.org/10.1016/J.CSBJ.2021.05.019>.

[121] ‡ Doletha M. E. Szebenyi, § Xiaowen Liu, † Irina A. Kriksunov, *, § and Patrick J. Stover, D.J. Thiel‡, Structure of a Murine Cytoplasmic Serine Hydroxymethyltransferase Quinonoid Ternary Complex: Evidence for Asymmetric Obligate Dimers†, *Biochemistry.* 39 (2000) 13313–13323. <https://doi.org/10.1021/BI000635A>.

[122] S. Ubonprasert, J. Jaroensuk, W. Pornthanakasem, N. Kamonsutthipajit, P. Wongpituk, P. Mee-udorn, T. Rungrotmongkol, O. Ketchart, P. Chitnumsub, U. Leartsakulpanich, P. Chaiyen, S. Maenpuen, A flap motif in human serine hydroxymethyltransferase is important for structural stabilization, ligand binding, and control of product release, *J. Biol. Chem.* 294 (2019) 10490–10502. <https://doi.org/10.1074/JBC.RA119.007454>.

[123] G. Giardina, P. Brunotti, A. Fiascarelli, A. Cicalini, M.G.S. Costa, A.M. Buckle, M.L. Di Salvo, A. Giorgi, M. Marani, A. Paone, S. Rinaldo, A. Paiardini, R. Contestabile, F. Cutruzzolà, How pyridoxal 5'-phosphate differentially regulates human cytosolic and mitochondrial serine hydroxymethyltransferase oligomeric state, *FEBS J.* 282 (2015) 1225–1241. <https://doi.org/10.1111/febs.13211>.

[124] J.R. Perkins, I. Diboun, B.H. Dessailly, J.G. Lees, C. Orengo, Transient Protein-Protein Interactions: Structural, Functional, and Network Properties, *Structure.* 18 (2010) 1233–1243. <https://doi.org/10.1016/j.str.2010.08.007>.

- [125] S. Spizzichino, G. Pampalone, M. Dindo, A. Bruno, L. Romani, F. Cutruzzolà, T. Zelante, M. Pieroni, B. Cellini, G. Giardina, Crystal structure of *Aspergillus fumigatus* AroH, an aromatic amino acid aminotransferase, *Proteins*. (2021). <https://doi.org/10.1002/prot.26234>.
- [126] A. Paiardini, G. Giardina, G. Rossignoli, C. Voltattorni, M. Bertoldi, New Insights Emerging from Recent Investigations on Human Group II Pyridoxal 5'-Phosphate Decarboxylases, *Curr. Med. Chem.* 24 (2017) 226–244. <https://doi.org/10.2174/0929867324666161123093339>.
- [127] T.-F. Fu, J.N. Scarsdale, G. Kazanina, V. Schirch, H.T. Wright, Location of the Pteroylpolyglutamate-binding Site on Rabbit Cytosolic Serine Hydroxymethyltransferase *, *J. Biol. Chem.* 278 (2003) 2645–2653. <https://doi.org/10.1074/JBC.M210649200>.
- [128] G. Ducker, J. Ghergurovich, N. Mainolfi, V. Suri, S. Jeong, S. Hsin-Jung Li, A. Friedman, M. Manfredi, G. Z. K. H, J. Rabinowitz, Human SHMT inhibitors reveal defective glycine import as a targetable metabolic vulnerability of diffuse large B-cell lymphoma., *Proc. Natl. Acad. Sci. U. S. A.* 114 (2017) 11404–11409. <https://doi.org/10.1073/PNAS.1706617114>.
- [129] M. Marani, A. Paone, A. Fiascarelli, A. Macone, M. Gargano, S. Rinaldo, G. Giardina, V. Pontecorvi, D. Koes, L. McDermott, T. Yang, A. Paiardini, R. Contestabile, F. Cutruzzolà, A pyrazolopyran derivative preferentially inhibits the activity of human cytosolic serine hydroxymethyltransferase and induces cell death in lung cancer cells, *Oncotarget*. 7 (2016) 4570. <https://doi.org/10.18632/ONCOTARGET.6726>.
- [130] V. Besson, F. Rebeille, M. Neuburger, R. Douce, E.A. Cossins, Effects of tetrahydrofolate polyglutamates on the kinetic parameters of serine hydroxymethyltransferase and glycine decarboxylase from pea leaf mitochondria., *Biochem. J.* 292 (1993) 425. <https://doi.org/10.1042/BJ2920425>.
- [131] A. Tramonti, C. Nardella, M.L. di Salvo, A. Barile, F. Cutruzzolà, R. Contestabile, Human Cytosolic and Mitochondrial Serine Hydroxymethyltransferase Isoforms in Comparison: Full Kinetic Characterization and Substrate Inhibition Properties, *Biochemistry*. 57 (2018) 6984–6996. <https://doi.org/10.1021/ACS.BIOCHEM.8B01074>.
- [132] W. Amornwatcharapong, S. Maenpuen, P. Chitnumsub, U. Leartsakulpanich, P. Chailyen, Human and Plasmodium serine hydroxymethyltransferases differ in rate-limiting steps and pH-dependent substrate inhibition behavior, *Arch. Biochem. Biophys.* 630 (2017) 91–100. <https://doi.org/10.1016/J.ABB.2017.07.017>.

Publications

- 1. S. Spizzichino**, D. Boi, G. Boumis, R. Lucchi, F.R. Liberati, D. Capelli, R. Montanari, G. Pochetti, R. Piacentini, G. Parisi, A. Paone, S. Rinaldo, R. Contestabile, A. Tramonti, A. Paiardini, G. Giardina, F. Cutruzzolà, Cytosolic localization and in vitro assembly of human de novo thymidylate synthesis complex, *FEBS J.* (2021). <https://doi.org/10.1111/FEBS.16248>.
- 2. S. Spizzichino**, G. Pampalone, M. Dindo, A. Bruno, L. Romani, F. Cutruzzolà, T. Zelante, M. Pieroni, B. Cellini, G. Giardina, Crystal structure of *Aspergillus fumigatus* AroH, an aromatic amino acid aminotransferase, *Proteins Struct. Funct. Bioinforma.* (2021). <https://doi.org/10.1002/PROT.26234>.
- F. Cutruzzolà, A. Paiardini, C. Scribani Rossi, **S. Spizzichino**, A. Paone, G. Giardina, S. Rinaldo, A conserved scaffold with heterogeneous metal ion binding site: the multifaceted example of HD-GYP proteins, *Coord. Chem. Rev.* 450 (2021). <https://doi.org/10.1016/j.ccr.2021.214228>.
- F. Cutruzzolà, A. Bouzidi, F.R. Liberati, **S. Spizzichino**, G. Boumis, A. Macone, S. Rinaldo, G. Giardina, A. Paone, The emerging role of amino acids of the brain microenvironment in the process of metastasis formation, *Cancers (Basel)*. 13 (2021). <https://doi.org/10.3390/CANCERS13122891>.
- 5. S. Spizzichino**, G. Mattedi, K. Lauder, C. Valle, W. Aouadi, B. Canard, E. Decroly, S.J.F. Kaptein, J. Neyts, C. Graham, Z. Sule, D.J. Barlow, R. Silvestri, D. Castagnolo, Design, Synthesis and Discovery of N,N'-Carbazoyl-aryl-urea Inhibitors of Zika NS5 Methyltransferase and Virus Replication, *ChemMedChem*. 15 (2020) 385–390. <https://doi.org/10.1002/CMDC.201900533>.

Acknowledgments

I would like to express my deepest and most sincere gratitude to my tutor, Professor Francesca Cutruzzolà, for giving me the opportunity to perform my research in her laboratory and for her continuous support throughout this period. She is a role model for all of us and a source of inspiration. I promise that from now on I will try to say “solo cose belle”.

I would like to pay a special regard to my supervisor Dr. Giorgio Giardina, for having taught me everything I know so far, for always being there with a piece of advice but also for making me laugh really hard during these years and for always believing in me. My sincere thanks go also to Professor Serena Rinaldo, for being an extremely good mentor but also a friend when needed and to Dr. Alessio Paone and Dr. Giovanna Boumis, it was a great privilege working with you and I am deeply thankful for all the patience you had in explaining me things.

I am extremely grateful to my fellow lab mates: Chiara, Francesca, Federica, Amani and Elisabetta. We have created a bond that goes beyond being simple colleagues, we are friends, we support each other, laugh like crazy, but most importantly we are always ready to eat something ☺. It was a pleasure working with you, and the future holds just the best for us. Please, don't ever change, be always yourself even if it means being weird. I love you girls and the best is yet to come!

Last but not the least, I would like to thank my family: I am extremely grateful to my Mirky for having always supported me, for making me laugh even when I was at my deepest, for sharing the same passions with me and for believing in me. I would like to thank my parents for their love and for educating and preparing me for my

future. I wish to thank also my sister and my grandmother for supporting me throughout this period.