The Molecular Basis of Improbable Enzymatic Chemistries

by

Meredith K. Purchal

A dissertation submitted in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
(Chemical Biology)
in the University of Michigan
2023

Doctoral Committee:

Assistant Professor Markos Koutmos, Co-Chair
Assistant Professor Kristin Koutmou, Co-Chair
Professor Neil Marsh
Associate Professor Jayakrishnan Nandakumar
Professor Janet Smith

Meredith K. Purchal

purchalm@umich.edu

ORCID iD: 0000-0001-5919-1057

To Dad, thank you for teaching me to always be exactly who I am.

This is all for you.

Table of Contents

List of Tables

List of Figures

CHAPTER 1

CHAPTER 2

CHAPTER 4

CHAPTER 5

List of Equations

Abstract

Enzymes are Nature's best chemists' and play a vital role in supporting the diverse chemistries fundamental to complex life. Given this central role, protein dysfunction can have serious biological implications. Proteins are defined by the Structure-Function relationship ubiquitous throughout nature, and these relationships can be exploited for detailed enzymatic characterization. Thus, I took an integrative structural and biochemical approach to establish the structural context and molecular basis of challenging chemistries catalyzed by three enzymes involved in mRNA modification and cobalamin-dependent processes, and whose dysfunction result in cancers, developmental- and metabolic disorders.

Pseudouridine (Ψ) is a ubiquitous RNA modification, discovered at hundreds of sites in mRNAs. Pseudouridine synthases (Pus) are responsible for installing Ψ, but exactly how an individual Pus selects a specific target site is unclear. I sought to characterize the basis of Pus-RNA interactions in Pus7 and ultimately determined the contribution of substrate structure and conserved protein elements towards binding and catalysis. Pus7 is one of the predominant mRNA modifying Pus-enzymes, that exhibits distinct diversity in substrate selectivity, as well as increased activity under heat shock. I solved the structure of *Saccharomyces cerevisiae* Pus7 and visualized the architecture of the eukaryotic-specific insertions thought to contribute to expanded substrate scope. Indeed, the largest insertion (Insertion I) contains a nucleic acid binding R3H motif surrounded by positively charged residues. Subsequent analysis demonstrated that Insertion-I

serves to fine-tune Pus7 activity in a substrate-dependent manner both *in vitro* and in cells. Further, this work revealed that Pus7 is extraordinarily promiscuous, modifying every substrate (both natural and non-natural) containing the consensus sequence without regard for structure. My work suggests that Pus7 selectivity is likely governed by additional factors including substrate accessibility and localization, rather than inherent enzyme properties.

B_{12}-dependent enzymes harness the unique organometallic properties of cobalt to catalyze a variety of challenging chemistries integral to single-carbon metabolism in all domains of life. In humans, there are two metabolically essential B_{12}-dependent enzymes: methionine synthase (MS) and methylmalonyl-CoA mutase (MCM). Cobalamin-dependent MS is a multi-modular enzyme that employs remarkable molecular dynamics and domain rearrangements – deemed 'molecular juggling' – to catalyze three difficult methyl-transfer reactions at the site of the cobalt-cofactor. Biochemical challenges have hindered structural and mechanistic characterization of MS catalysis and conformational states. To address this, I describe a *Thermus thermophilus* MS variant that avoids the associated barriers of expression and purification. Using *t*MS as a model, I describe the first full-length structure of *apo*MS – finally visualizing all domains at once and gaining insights into the structural basis of B_{12}-incorporation. Further, we captured MS with the Folate-domain oriented above the B_{12}-domain, and cobalt is within the predicted distance for catalysis, and this likely represents the first catalytic structure captured for any corrinoid protein.

MCM utilizes 5'-deoxyadenosylcobalamin (AdoCbl) to catalyze the interconversion of methylmalonyl-CoA to succinyl-CoA through homolysis of Co-C bond. Here, we determined the structure of *Mycobacterium tuberculous* MCM complexed with the suicide inactivator itaconyl-CoA, a succinyl-CoA analog. Notably, EPR studies confirm that we captured an air-stable biradical comprising a tertiary carbon radical (5'-deoxyadenosyl) coupled to the metal-centered

cob(II)alamin radical *in crystallo*. Thus, in addition to describing the mechanism of I-CoA inhibition, these experiments provide molecular insights into how MCM controls radical trajectories during catalysis.

CHAPTER 1 Introduction

1.1 Structure-Function

Structure begets function, and function defines form. The intimate relationship between structure and function is a central tenant throughout nature and one of the first concepts taught in any biology class. The structure-function relationship is the idea that the arrangement of components in a system are structurally organized to allow the system to perform its function (1). Or, simply, the shape of something allows it to perform its job, which in turn defines its shape.

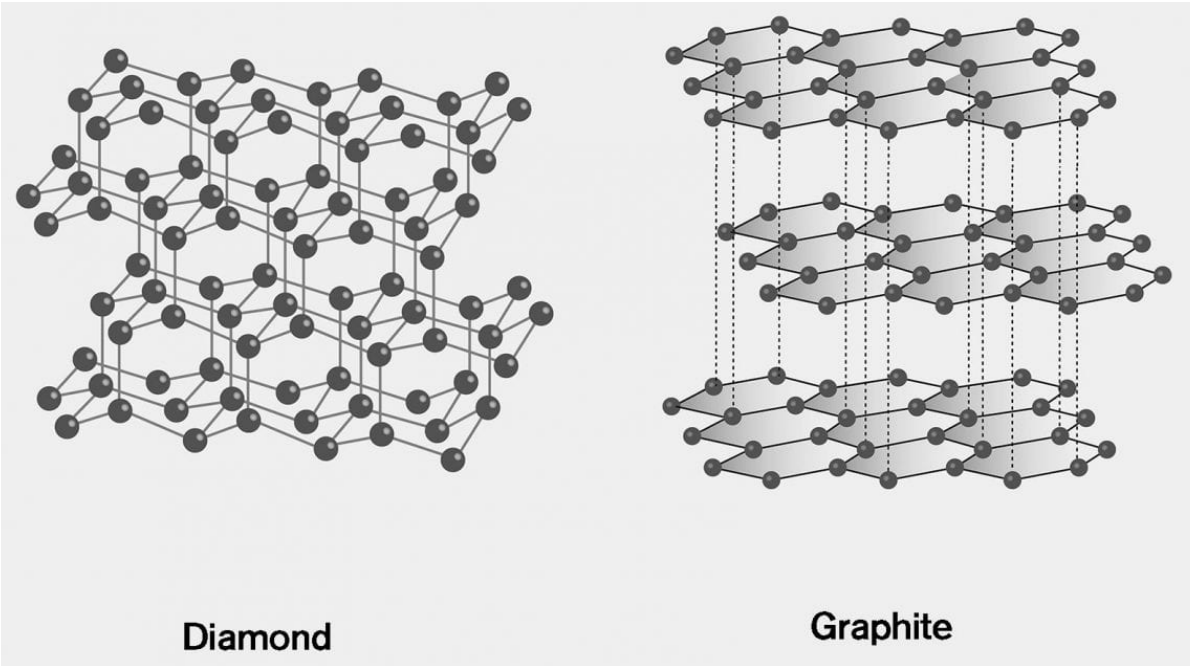


Figure 1.1: Allotropes of carbon in diamond and graphite

(left) Each carbon atom in a diamond is sp^3 hybridized and forms four covalent bonds with neighboring atoms. The tetrahedral arrangement of carbon creates a rigid, three-dimensional lattice structure and accounts for the remarkable

hardness, strength, and durability of diamonds. **(right)** Carbon in graphite is sp^2 hybridized and forms three covalent bonds with adjacent atoms, resulting in parallel layers of hexagons separated by delocalized electrons. The Van der Waals forces separating the hexagonal sheets contribute to the soft and 'slippery' properties of graphite. Taken from Aouf, 2021 (2).

To illustrate this idea, consider the graphite in a pencil and the diamond in a ring. Both graphite and diamond are made of solely of carbon atoms but have radically different physical properties, making both materials uniquely suited for their respective use – just try taking notes with a diamond tipped pen or proposing with a graphite studded ring, in both scenarios you won't find much success. On the molecular level, the distinct physical properties of graphite and diamond arise from different structural geometries adopted by the carbons in each substance, resulting in a unique organization of the atoms (Figure 1.1) (3). In graphite, carbon atoms are arranged as sheets of hexagons, which allows for easy writing with a pencil as the carbon atoms slough off in layers on the page. But in diamonds, carbons are tetrahedrally arranged in a continuous matrix which affords diamonds their prodigious strength and sparkle (refractive properties). In other words, the internal molecular structure of these materials dictates their practical function.

This principle applies to the essential biomolecules that enable complex cellular processes that sustain life. DNA, RNA, and proteins are characterized by their primary, secondary, tertiary, and quaternary structures (Figure 1.2) (4). The primary structure refers to the sequential order of the nucleotides or amino acids in a molecule, while secondary structure describes the local interactions between residues or bases in the chain generating simple motifs and folds. In proteins, secondary structure includes alpha helices and antiparallel beta sheets and for RNAs, this includes (but not limited to) stem loops, multi-stem junctions, double-stranded regions, internal loops, and nucleotide bulges (5). Tertiary structure designates the overall complex folding of an entire biomolecule, which is shaped through a multitude of interactions between secondary structural features (4). Finally, the quaternary structure describes the higher-order complexes formed by

protein and/or nucleic acid complexes – in other words, how completely folded molecules interact with themselves or other biomolecules in the cell (4). To achieve the 'correct' quaternary structure depends on the fidelity and accuracy of each preceding lower-order structural level – without the appropriate primary sequence, higher order structures are not possible, thus precluding adequate function of these biomolecules.

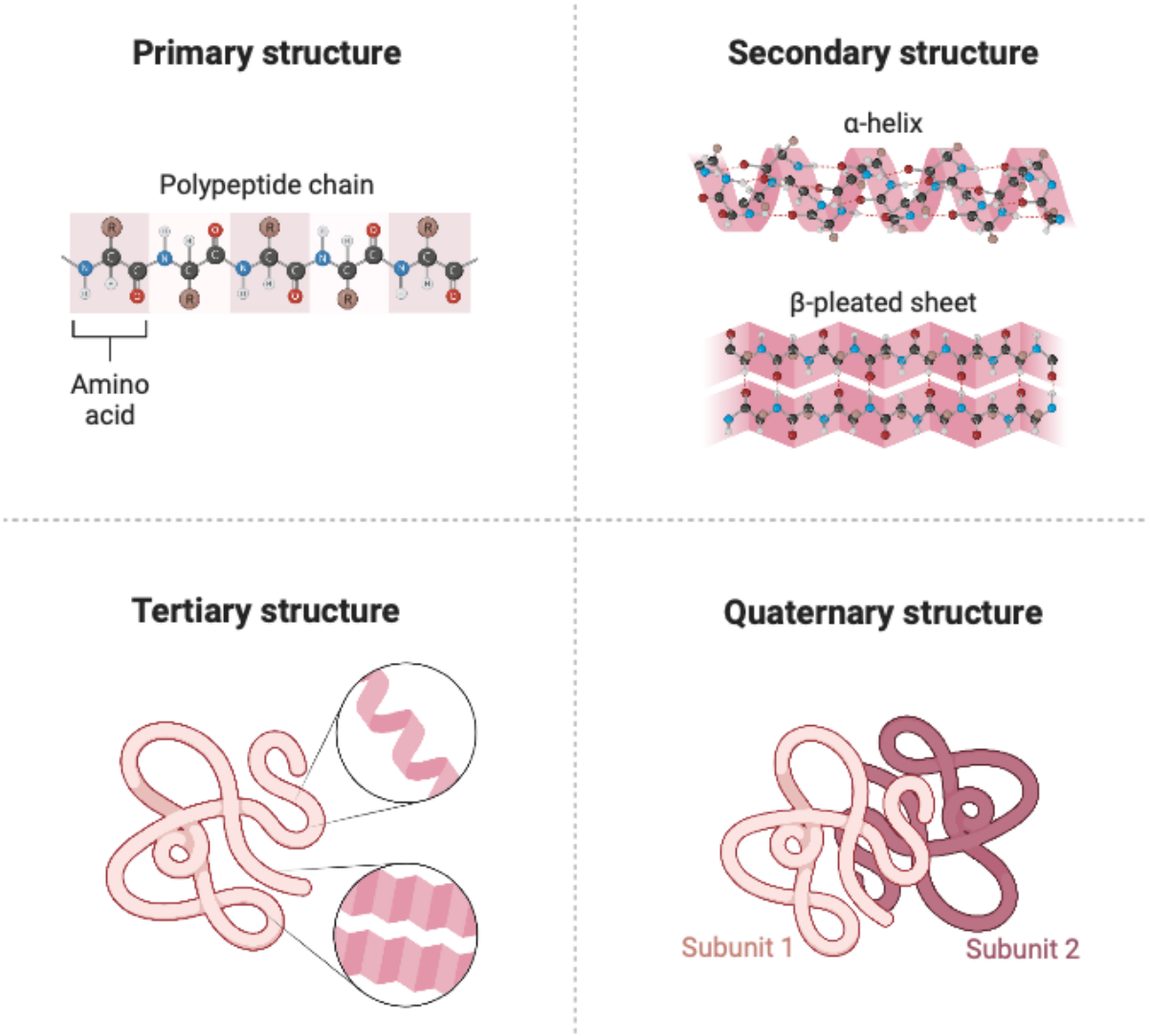


Figure 1.2: Levels of organization in proteins and other biomolecules

Multiple levels of structural organization describe DNA, RNA, and proteins. Though this figure is specific to proteins, DNA and RNA also form local secondary structure features, in addition to higher-order tertiary and quaternary structures.

The inter-dependence of sequence, structure, and function is the fundamental basis of many areas of scientific research (6–9). Everything is defined by its structure which gives us insight into how it might function, and scientists leverage this relationship in a variety of ways. Drug development, for example, is largely based in the structure-function relationship (10). Lead compounds are selected for how they interact with the target protein and then systematically altered at the atomic level to garner the desired effect, be it inhibition of the target or stimulation. It is a powerful approach that has led to a multitude of therapeutics and a deeper understanding of the innerworkings of our biology (11,12).

Detailed structural information is a prerequisite to be able to target a protein or other biomolecule strategically and successfully for therapeutic purposes (13,14). Proteins and nucleic acids are at the center of all biological processes and the ability to visualize the fine details of their structures can be extremely informative. Thus, a variety of techniques have been developed to structurally characterize essential biomolecules (15,16). The most powerful of these techniques – including X-ray Crystallography (17), Cryo Electron Microscopy (CryoEM) (18), and NMR spectroscopy (19) – can produce a 3D molecular model, displaying not only the overall shape of the molecule, but the precise orientation of primary and secondary structural features in space – allowing us to 'see' at atomic resolution. These techniques can be used to address significant questions like why a pathogenic genetic mutation causes protein dysfunction, or how a known inhibitor exerts its function. However, due to intrinsic biological or technical barriers some biomolecules are unable to be visualized using X-ray crystallography or CryoEM (19–21). RNA, for example, is notoriously challenging to crystallize, though it has been done (22,23). RNA is quite dynamic and highly charged, both of which pose a challenge to forming an ordered crystalline structure. CryoEM has size limitations and is insufficient for smaller proteins (10-40 kDa) or

characterizing small-molecule interactions (21,24). To overcome these limitations, the application other 'low-resolution' biochemical techniques and enzymatic assays can be used to create a biochemical-model and gain insight into structural basis of observed behavior. Higher-order protein structures can be studied using hydroxyl radical protein foot-printing (HRPF) or hydrogen-deuterium exchange mass spectrometry (HDX-MS) to identify solvent accessible residues and map protein interactions (25–27). Notably, hydroxyl-radical foot printing can also be used to identify DNA conformations and structural features, and has been applied to study DNA-protein interactions (28,29). Techniques including selective 2'hydroxyl acylation with primer extension (SHAPE) and dimethyl sulfate (DMS) mapping can be applied to probe elements of RNA secondary structure and the reactivity of individual bases (30,31). And together, the combination of high-resolution structural studies with systematic biochemical and enzymatic characterization constitutes a powerful tool to interrogate the molecules responsible for the fundamental processes that sustain life.

In this work, I have explored the structure-function relationship in three different enzymes that perform challenging chemistries. In proteins, the precise shape and orientation of the domains, as well as the structure of the active site, dictate the chemistries that it can perform. Different conformations are adopted at different times, which enable the enzyme to enact its ultimate function. I studied the RNA modifying enzyme pseudouridine synthase 7 (Pus7) to understand the basis of selectivity in RNA substrates. There are many pseudouridine synthases, but Pus7 is remarkably promiscuous, and it is unclear how Pus7 modifies so many different types of RNAs. I also worked with two cobalamin-dependent enzymes, Methionine synthase (MS) and Methyl-malonyl CoA Mutase (MCM). Methionine synthase is a large, multi-modular enzyme that performs three different methyl-transfers and adopts a variety of different conformations to carry

out these reactions. Studying these orientations has been a significant challenge due to the inherent conformational flexibility of the protein. I worked to characterize the catalytic conformations of MS to gain further insight into how the protein is able to catalyze unfavorable reactions with ease and efficiency. Finally, MCM catalyzes the reversible isomerization of M-CoA to Succinyl-CoA but is irreversibly inhibited by I-CoA. I worked to characterize the mechanism of inhibition of Itaconyl-CoA in MCM and gained significant insights into how MCM is able to perform its standard reaction.

In total, the research presented here is characterized as an exploration of a variety of different facets of the structure-function relationship. From substrate selection, to catalysis, to inhibition – the structure of these enzymes plays a huge part in determining what each one is able to accomplish. Studying these relationships is an important way to characterize these proteins and is foundational to understanding their ultimate purpose in the cell and exploiting and manipulating these enzymes for future development (therapeutics, biochemical tools, etc.).

1.2 Pseudouridine and Pseudouridine Synthases

1.2.1 The landscape of mRNA modifications

Chemical modification of the three major biomolecules in the cell (DNA, RNA, Proteins) constitutes an added layer of complexity and customization, allowing cells to adapt to rapidly changing environments by modulating the structure, function, stability, expression, and localization of these molecules (32–34). In RNA, over 170 unique post-transcriptional modifications have been identified at specific sites across the transcriptome distributed throughout all domains of life (Figure 1.3) (35). Modifications serve to dramatically expand the chemical landscape of RNA, imparting functional, topological, and chemical diversity otherwise inaccessible to the four canonical RNA bases. Post-transcriptional RNA modifications range from small structural changes (like the isomerization of uridine to pseudouridine) and seemingly-minute chemical additions (the methylation of adenosine to make m^6A), to the addition of large chemical groups sequentially installed over multiple enzymatic steps (Figure 1.3) (35). Though the precise functions of most of these modifications are largely unknown, their broad distribution throughout all kingdoms of life suggests modifications represent an evolutionarily conserved molecular toolbox that may serve to modulate the flow of genetic information or allow for adaptation to environmental challenges (34,36–39).

1.2.1.1 Modifications in mRNA

Except for the 5' m^7G cap in eukaryotic mRNAs, post-transcriptional RNA modifications were largely thought to be limited to high-abundance non-coding RNAs (tRNA, rRNA, snRNA, etc.), and less abundant mRNAs, which make up ~1-2% of the total RNA in the cell, were thought to be unmodified. However, recent advancements in detection methods and sequencing technologies

have fostered the discovery of a dynamic landscape of post-transcriptional modifications in protein-coding messenger RNAs (mRNA) (38,40–43). The discovery of mRNA modifications including methyl-6-adenosine (m^6A) (44), methyl-5-cytosine (m^5C) (45), ribose-methylation (2'-O-Me) (46), and pseudouridine (Ψ) (40–43), amongst others, in mRNAs invigorated the field of RNA modification research. mRNA serves as the blueprint for protein synthesis and modifications could serve as sophisticated, yet uncharacterized mechanism of regulation of gene expression (38,47–53).

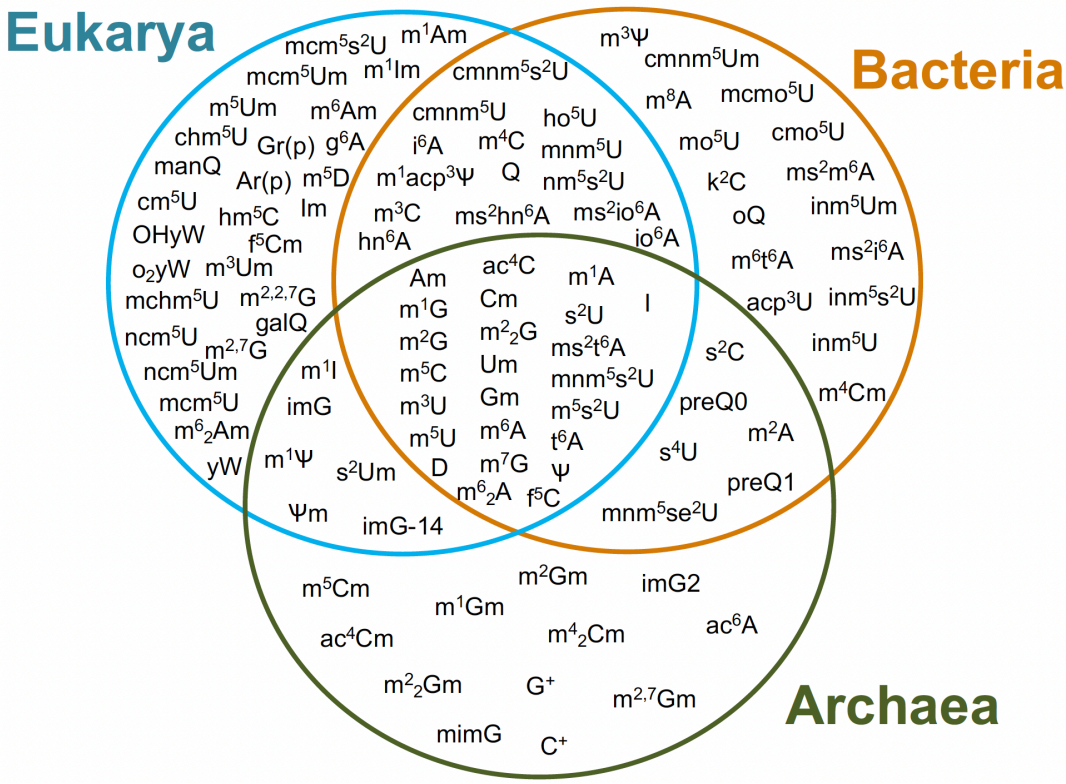


Figure 1.3: Diversity of post-transcriptional RNA modifications

The set of known RNA modifications classified by their reference organism.

To date, >15-20 different modifications have been identified in eukaryotic mRNAs in yeast, mouse, human, and plants (35,54–56), though that number will likely increase in the coming years with further advancements in detection methodologies and the explosive growth of the field of epitranscriptomics (57). Seminal investigations suggest that mRNA modification is a dynamic

process, providing a means to fine-tune cellular response and enabling organisms to react to a rapidly changing environment (39,53,58,59). Moreover, modification of mRNA has the potential to affect every stage of the mRNA lifecycle, including cellular localization, maturation, structure, half-life, translational fidelity, and mRNA stability (60–62). It follows then that aberrant mRNA modification may be particularly destructive and, indeed, changes in modification patterns have been implicated in a variety of diseases (37,63–66), including obesity (67,68), cervical cancer (69), ovarian cancer(70), glioblastoma (71,72), type 2 diabetes (67), intellectual disability and developmental disorders (73–76), and major depressive disorder (77).

1.2.2 Pseudouridine

Pseudouridine (Ψ) is an isomer of uridine and commonly referred to as the fifth nucleoside (Figure 1.4). Ψ is one of the most ubiquitous RNA modifications – found throughout all domains of life (78,79). The moniker stems from seminal investigations in 1951 (80) that identified an unknown 'fifth nucleoside', distinct from the four canonical bases, comprised ~4% of total nucleosides in yeast tRNAs (81). Subsequent chemical characterization unambiguously identified the unknown base as pseudouridine (82–85). Accordingly, Ψ is now recognized as one of the most abundant and phylogenetically conserved modifications found across all domains of life and present in many, if not all, species of noncoding RNAs [e.g., tRNA (86), rRNA (87), snRNA (88), miRNA (89), lncRNA (90), etc.], and has recently been identified throughout eukaryotic protein-coding messenger RNAs (mRNA) (35,55,56).

1.2.2.1 Chemical properties of pseudouridine

Pseudouridine is the C-5 glycosidic isomer of uridine, formed through an internal transglycosylation wherein the N1-C1' bond between the ribose sugar and the uracil base is

broken. After subsequent rotation by 180° on the N3-C6 axis, the base is reattached through the C5-C1' glycosidic bond, resulting in an extra hydrogen bond donor at N1, positioned on the non-Watson-Crick edge (Figure 1.4).

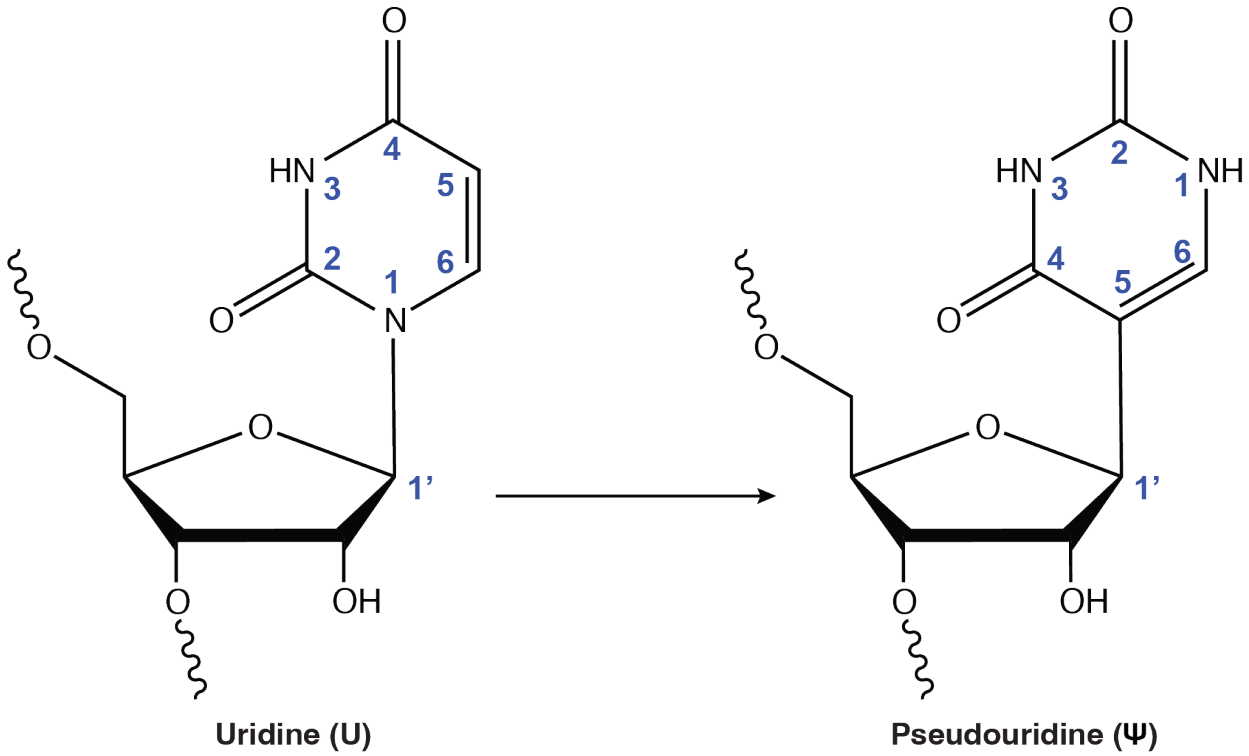


Figure 1.4: Pseudouridine is the C-5 glycosidic isomer of uridine

In solution, Ψ has a slight preference for the *syn* glycosyl conformation, though Ψ has only been found in the *anti*-conformation within the context of a nucleotide chain (91–93). The *anti*-conformation of Ψ in RNA places the novel N1-H donor in the appropriate geometry and distance to coordinate a water molecule between the 5' phosphates of Ψ and the preceding residue (94,95). The additional hydrogen bonding potential and water coordination through the extra imido group serve the basis for the structural stability Ψ imparts on RNAs (91,93–99). Ψ adds rigidity to the phosphate backbone of RNAs as water-coordination limits the conformational flexibility of this modification in both single and double stranded regions (91,93,94,96). Further, Ψ improves base

stacking interactions which propagates to neighboring nucleosides to cooperatively increase stacking around the site of modification in single stranded and duplexed regions (93,98); this is thought to be the most significant contribution of Ψ to the stabilization of RNA structures (91,93,96–98).

1.2.2.2 Pseudouridine in noncoding RNA

Ψ has traditionally been studied in the context of non-coding RNAs, where it plays an integral role in the structure and biological function of said RNAs (78,100). Initial evidence for Ψ playing a functional role in RNA stems from the fact that Ψ modifications are often clustered in evolutionally conserved and functionally important regions of RNAs (35). Subsequent investigations have shown Ψ plays a variety of important biological roles, including maintaining the proper folding and function of ribosomal RNA, translational fidelity, RNA-RNA interactions, and RNA stability and decay (48–53,101,102).

tRNA - Ψ is present in almost all tRNAs (35), widely distributed across the molecule where it stabilizes the specific structural motif in which it occurs (103,104). In tRNAs across all domains of life, Ψ is found in the anti-codon stem loop (ASL), the D-loop, and the TΨC loop, named for the invariant Ψ55 universally conserved across tRNAs (105–107). Ψ is found less frequently at many other sites in tRNAs, often distributed in a domain-specific manner (94). In addition to stabilizing the L-shaped tRNA fold (103,104), modifications in tRNA may also play an important role in the speed and accuracy of translation (108–110), decoding (111), and modulating interactions with the ribosome (112,113).

rRNA - Ψ are present throughout rRNA (35,114); there are ~95 predicted sites of Ψ in human ribosomal subunits, making up 1.4% of all bases (115–118). There are comparatively less

Ψs in bacterial and yeast rRNAs, 36 in *E. coli* and 46 in *S. cerevisiae* (35,119), suggesting the level of rRNA modification scales with the complexity of the organism (120,121). The Ψ modifications in the ribosome are found in (or very near to) functionally important sites (122,123), including the peptidyl transferase center (PTC) and the decoding center (DC) (120,122–124). Broadly, Ψ is important for ribosome biogenesis and overall ribosome function. Ψ enhances conformational stability and assembly (125) of the ribosome, enabling proper folding and maintaining interactions with ribosomal proteins (87,122,126,127). Further, the cryo-EM structure of a pseudouridine-free ribosome demonstrated the critical role Ψ plays in maintaining proper ribosomal function (Figure 1.5) (128,129). The pseudouridine-free ribosome exhibited conformations reflective of abnormal inter-subunit movements and displayed erroneous swiveling and ratcheting motions (129). In total, these findings provide the structural basis for the observed deficiencies in translation upon loss of Ψ and underscore the importance of Ψ to maintaining ribosomal function (128–130).

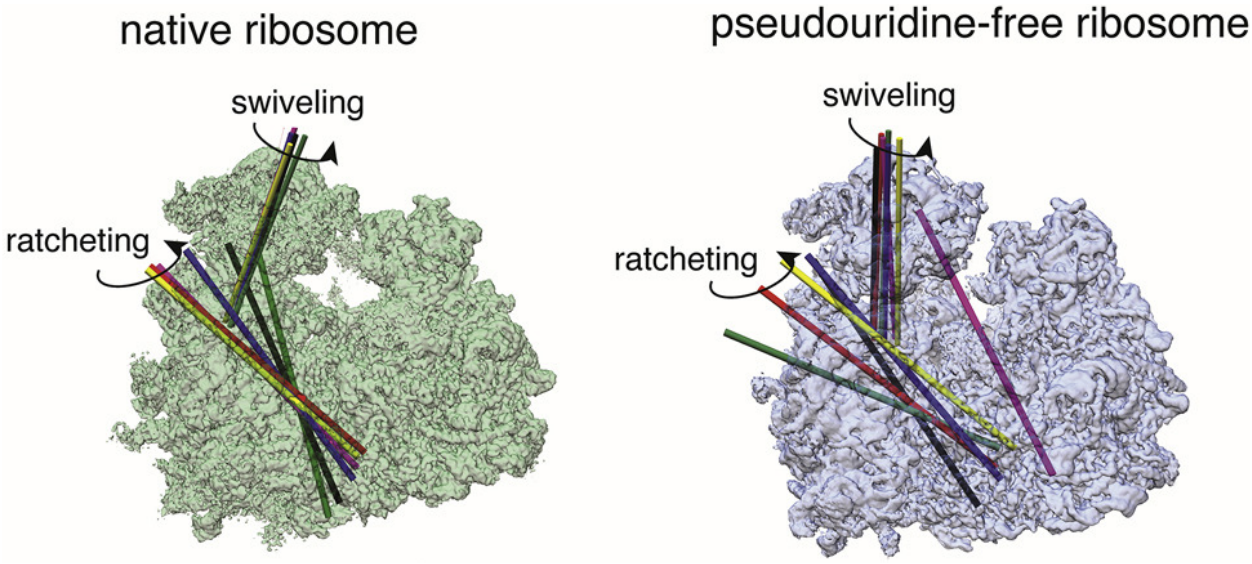


Figure 1.5: Pseudouridine in the ribosome

(left) Fully modified native ribosome with rotation axes shown for swiveling and ratcheting motions undergone during translation. **(right)** Pseudouridine-free ribosome. Rotation axes display disordered ribosome motions. Figure taken from Zhao, 2022 (129).

snRNA - Ψ is found in the major spliceosomal snRNAs (U1, U2, U4, U5, and U6) (131) and in the variants responsible for AU/AC intron splicing (U12, U4atac, and U6atac) (132). The presence and location of Ψ in snRNAs are often phylogenetically conserved in functionally important regions of snRNAs where Ψ affects intermolecular RNA-RNA or RNA-protein interactions critical to the assembly and overall function of the spliceosome (132). For example, in U2snRNA a Ψ-A base pair adjacent to the intronic branch site in pre-mRNA triggers a structural change that results in the 'bulging out' of the adenosine nucleophile that serves as the first step of the splicing reaction (133–136).

1.2.2.3 Pseudouridine in mRNA

Recent studies mapping Ψ across the transcriptome of yeast, mice, humans identified hundreds of unique positions of Ψ in protein-coding messenger RNAs (40–43). The occurrence, location, and distribution of Ψ in mRNA transcripts appears to be dynamically regulated – cellular-stress conditions, like heat shock, result in a differential pattern of Ψ compared to normal growth conditions (40,41). Notably, Ψ in mRNAs are not randomly distributed. Rather, Ψs are primarily found within the coding region and the 3' UTR, and comparatively underrepresented in the 5'UTR (41). Given this finding, it is apparent that the ribosome will encounter this modification in mRNAs, and as Ψ can alter the structures and base pairing abilities of RNAs (98,137,138), this suggests Ψ may play a much broader role in influencing protein production and gene expression than previously thought – though the mechanism of how mRNA pseudouridylation could affect these processes is still being untangled. Indeed, Ψ in stop codons has been found to suppress translation termination of nonsense codons (139,140), but the overall effect of pseudouridylated stop codons remains unresolved (141). Further, reporter-based studies in human and bacterial

lysates investigating the role of Ψ in mRNA have conflicting findings: some studies suggest Ψ in mRNAs increases protein production (142), while others report decreased protein output (143,144). Finally, studies examining how Ψ in mRNA impacts the kinetics of translation found that Ψ alters both the speed of translation and mRNA decoding by the ribosome (87,111,145). In total, these findings clearly indicate that Ψ in mRNAs affect translation and interactions with the ribosome, however the exact contributions and consequences of Ψ, especially in the context of gene expression, are still being deconvoluted.

1.2.3 Pseudouridine synthases

1.2.3.1 RNA-guide dependent and guide-independent pseudouridine synthases

Post-transcriptional pseudouridine formation is catalyzed by a family of enzymes called pseudouridine synthases, which fall into two categories based on how they recognize their RNA substrates (117,146–148). The guide-RNA dependent pseudouridine synthases, called H/ACA sRNPs (small ribonucleoproteins), rely on a guide RNA that is complementary to the site of modification for target recognition in conjunction with a protein (Cbf5 in yeast, dyskerin in humans) to catalyze pseudouridine formation (149–151). Absent in prokaryotes but conserved in eukaryotes and archaea, H/ACA sRNPs are primarily responsible for the modification of uridines in rRNAs and snRNAs (35,151). For example, all Ψs in the yeast ribosome are installed by H/ACA sRNPs located in the nucleolus, collectively referred to as small nucleolar RNPs (snoRNPs). In addition to installing Ψ, H/ACA sRNPs play an important role in processing pre-rRNA and stabilization of the mammalian telomerase RNA component (TERC) (152,153). Defects in dyskerin (154,155), the protein component in human H/ACA sRNPs, arising from mutations of the *DKC1* gene and overexpression, result in severe pathologies including the premature aging

syndrome, X-linked dyskeratosis congenita (X-DC) (156,157), Chronic Hypersensitivity Pneumonitis (158), and cancers (159–161). The H/ACA sRNPs have been extensively reviewed (149–151) and this work centers on the other category of pseudouridine synthases.

In the second category, and the focus of this work, are the guide-independent Pseudouridine Synthases (Ψ-synthase) that are found throughout all domains of life (Figure 1.6) (146,148). As the name would suggest, these are 'stand-alone' enzymes that do not require a guide-RNA to recognize their targets[1]. Instead, these enzymes install Ψ through a guide-independent mechanism where the recognition elements required for modification, which include sequence and/or secondary structural elements, are entirely contained within the RNA itself. Pseudouridine synthases are recognized for their role installing Ψ in tRNA, snRNA, and rRNA – influencing tRNA biogenesis (162), splicing (136), translation fidelity (145,163), mutagenesis (164), and other fundamental RNA processes in eukaryotes, prokaryotes and archaea (39,78,146,148). Pseudouridine synthases are also responsible for Ψ55 (165–167), universally conserved in tRNAs, suggesting a fundamental evolutionarily role of pseudouridine synthases in translation (35). The recent efforts to characterize the occurrence of Ψ in protein-coding messenger RNA identified eukaryotic pseudouridine synthases as the enzymes largely responsible for the dynamic Ψ landscape in eukaryotic mRNAs (40–42). Thus, pseudouridine synthases have come under renewed scrutiny because this could signify a more prominent role of these enzymes in regulating essential cellular processes (39).

1.2.3.2 Significance of stand-alone pseudouridine synthases in the cell

[1] Nor do they require ATP, cofactors, or accessory proteins to catalyze Ψ formation – truly 'stand-alone'.

Pus enzymes are broadly important, with identified roles in rRNA processing (168), mi-RNA processing (169,170), translational control in stem cells (171), translational fidelity (163), apoptosis (172,173), RNA folding (174), and RNA and quality control (175). Pseudouridine synthases play a role in plant development (176), nutrient sensing (177), and thermal adaptation. Knockout studies demonstrate that pseudouridine synthases are not essential under normal growth conditions (178–180), but are associated with slow growth in mouse embryonic fibroblast cells (181). Under conditions of cellular stress, pseudouridine synthases help the cell recover (182) and often display increased activity towards RNAs (40,41).

Loss or dysfunction of Pus enzymes impacts neuronal development (183), and morphological plasticity in *drosophila* (168). In humans, dysfunction of Pus enzymes leads to intellectual disability and neurodevelopmental defects (73–76), mitochondrial and metabolic disorders (179,184–186), and overexpression of Pus enzymes has been linked to a variety of cancers (187–189). Both Ψ levels and the enzymes themselves can be used as a biomarkers to track disease progression in cancers (70,189–192) and targeting pseudouridine synthases may have therapeutic potential for cancer and disease treatment (188,189,193–195).

1.2.3.3 Organization of the pseudouridine synthase family

All pseudouridine synthases can be classified into six distinct families based on sequence, structural, and functional considerations (Figure 1.6) (146). The TruA, TruB, TruD, RluA, and RsuA families are so named after their well-studied *E. coli* representatives but include pseudouridine synthases from archaea and eukaryotes as well as the prokaryotic members (146). The PUS10 family was more recently defined and is unique to some archaea and eukaryotes but absent in prokaryotes (196–198). In eukaryotes there are ten pseudouridine synthases, referred to as Pus enzymes (Pus1 - Pus10, in yeast), though some organisms (humans) have several paralogs

of a specific Pus enzyme (Figure 1.6) (39,148). There are no eukaryotic representatives of the RsuA family, whose bacterial members are largely responsible for modifying the small subunit of the ribosome (199–201), and pseudouridylation of the ribosome in eukaryotes is primarily carried out by the RNA-dependent H/ACA pseudouridine synthases (151).

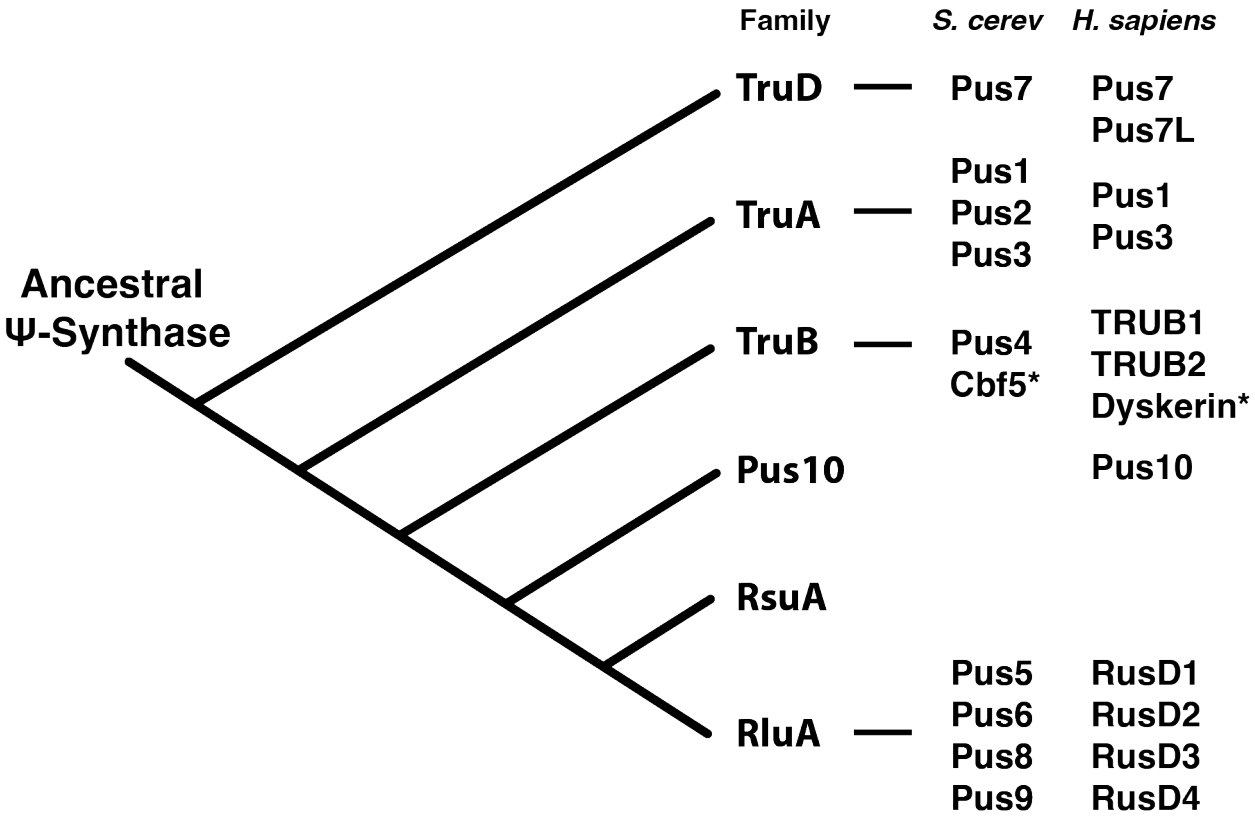


Figure 1.6: The family of pseudouridine synthases

Hypothetical evolutionary relationship between pseudouridine synthase families based on sequence and structure. Each family (except Pus10) is named after the *E. coli* representative of said family.

1.2.3.4 Conserved catalytic domain

Despite minimal sequence similarity (9 – 23%) (146) between the families, Ψ-synthases are united by a structurally conserved catalytic domain (PUS domain) that appears to be exclusively utilized by this superfamily (Figure 1.7). The PUS domain is a unique eight-stranded β-sheet fold that houses the universally conserved catalytic aspartate residue, the only residue conserved throughout

the entire pseudouridine synthase superfamily (146,202). Together, the structural conservation of the catalytic domain and the universally conserved catalytic residue suggest a common evolutionary origin of these enzymes, as well as a shared mechanism to catalyze Ψ formation.

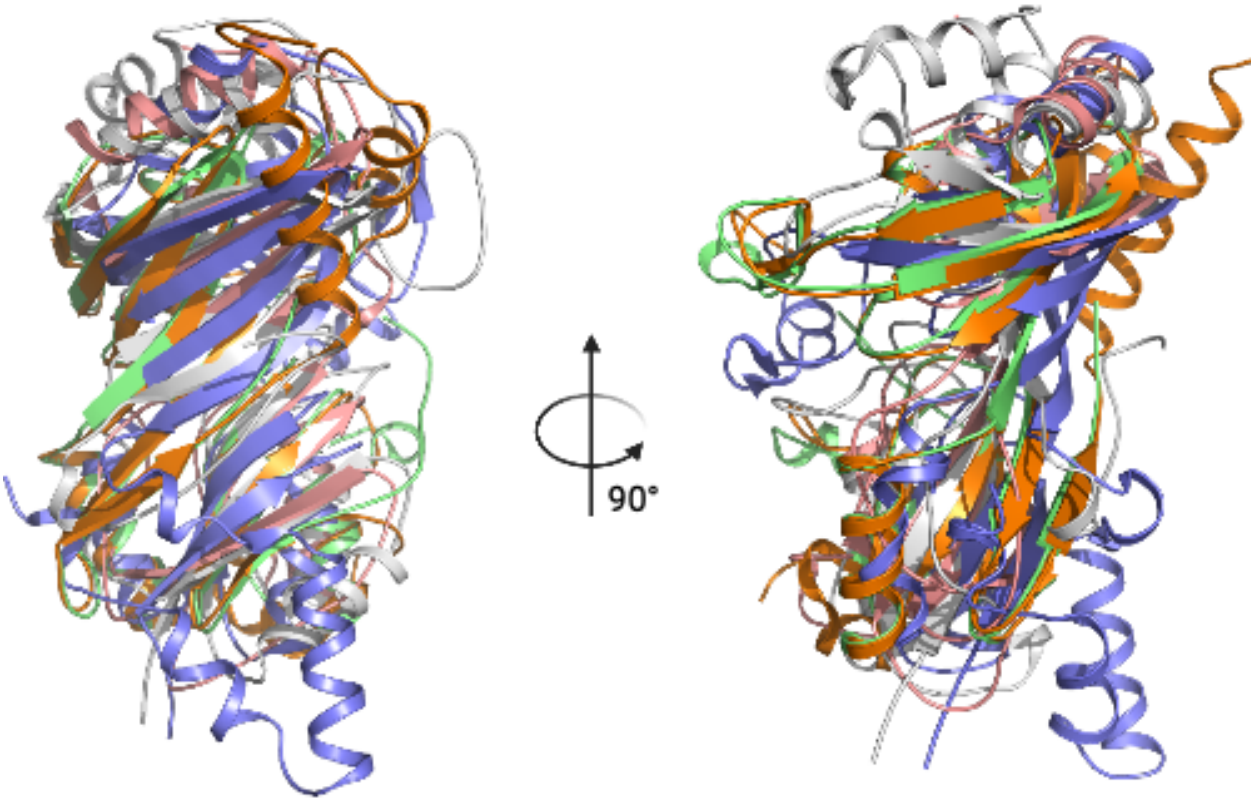


Figure 1.7: Structural conservation of PUS catalytic domain

(A) Superposition of ribbon representations of the structures of TruD (203) (salmon, PDB: 1SB7), TruA (204) (gray, PDB: 1DJ)0), TruB (205) (purple, PDB: 1K8W), RluD (206) (orange, PDB: 1PRZ), and RluA (207) (green, PDB: 2I82) all from *E. coli.* The core domain of all enzymes is very similar. In this front view, the active site cleft bisects the core domain horizontally. **(B)** View rotated 90° along the vertical axis.

1.2.3.5 The TruA, TruB, and TruD family in Eukaryotes

TruA – Pus1, Pus2, and Pus3 are part of the TruA family (Figure 1.6). All three enzymes share significant sequence homology, though Pus1 and Pus2 are paralogs that emerged from a gene duplication event and have since evolved different functions and different cellular localizations (Pus1 resides in the nucleus, Pus2 in the mitochondria). In yeast, Pus2 modifies two adjacent sites

(U26 and U27) in mitochondrial tRNAs and Pus3 modifies U38 and U39 in both cytoplasmic and mitochondrial tRNAs. In comparison, Pus1 has a much larger substrate scope - modifying eight distinct sites in cytoplasmic tRNAs, including U1/26/27/28/34/36/65/67, one site in pre-tRNA (U35), and U44 and U28 in U2 snRNA and U6 snRNA, respectively.

Pus1 is the most prominent member of the TruA family with distinct roles in nuclear export of specific tRNAs (162) and tRNA biogenesis (162) in yeast (162,208). Pus1 in *Toxoplasma gondii* is required for cellular differentiation (209), with many Pus1 mRNA targets enriched in developmentally regulated transcripts (210). A variety of genetic mutations of Pus1 are associated with mitochondrial myopathy and sideroblastic anemia (MLASA) (211), a rare autosomal recessive disorder affecting skeletal muscle and bone marrow (186,212–216). Further, Pus3 mutations are associated with intellectual disability and neurodevelopmental disorders (75,76).

TruB – The TruB family is represented by Pus4 in yeast (217), found in both the nucleus and the cytoplasm (166), and by nuclear TRUB1 and mitochondrial TRUB2 in humans/mammals (Figure 1.6). These enzymes are responsible for the universally conserved Ψ55 in the TΨC loop in tRNAs (166) – notably, mammalian Pus10 also installs Ψ55 in some cytoplasmic tRNAs and this functional redundancy (197) underscores the importance of Ψ55 in tRNA (167).

In addition to its role in tRNA modification, TRUB1 has been shown to play an important role in micro-RNA (miRNA) maturation (169). TRUB1 binds to the stem loop structure of pri-let-7 miRNA to selectively enhance let-7 interactions with the microprocessor DGCR8 (169). Furthermore, yeast Pus4 has been shown to bind Brome Mosiac Virus (BMV), an RNA virus that mainly infects plants but can replicate in yeast (218). The Pus4-BMV interaction inhibits encapsulation of the BMV RNA and prevents viral spread. Interestingly, Pus4's effect on BMV is

independent of its catalytic ability (218). The prokaryotic *E. coli* homolog of Pus4, TruB, has been suggested to act as a chaperone, serving to help RNAs fold correctly – also independent of its catalytic activity (174). Thus, eukaryotic members of the TruB family, including Pus4, TRUB1 and TRUB2, may serve additional roles in the cell that are independent of pseudouridylation activity, which may include aiding tRNA folding or chaperone-like activity, as has been observed in other RNA modifying enzymes (219).

TruD – Pus7 is the sole member of the divergent (220,221) TruD family in eukaryotes and modifies a variety of different RNA substrates (222). Known for its multi-site substrate specificity, Pus7 in yeast modifies Ψ13 in cytoplasmic tRNAs, Ψ35 in pre-tRNA^{Tyr} (222), Ψ50 in 5S rRNA, and Ψ35 in U2snRNA (223). Remarkably, Pus7 activity is inducible and under heat-shock Pus7 modifies an additional site (Ψ56) in U2snRNA (224).

Though Pus7 has long been recognized for its diverse substrate selectivity, recent findings indicate that dysfunction of Pus7 activity or expression can have significant consequences to cellular function. A genetic frameshift mutation resulting in Ser282Cys substitution in Pus7 is associated with intellectual developmental disorders with abnormal behavior, microcephaly and short stature (IDDABS) (73,74). Further, dysregulation or overexpression of Pus7 is associated with metastasis of colorectal cancer (188), ovarian cancer (70), and glioblastoma tumorigenesis (71).

1.2.3.6 mRNA modification and regulation

A host of transcriptome-wide sequencing techniques used to map the position of Ψ in mRNAs, including Pseudo-seq (Psi-seq) and CeU-seq, have identified thousands of Ψs throughout mRNAs

in yeast, mice, and humans (40–43). Subsequent deletion studies demonstrate that in addition to modifying non-coding RNAs, stand-alone pseudouridine synthases from the TruA, TruB, and TruD families are responsible for the majority of these mRNA pseudouridylation events (Figure 1.8) (40,41). Since the original set of papers mapping Ψ across mRNAs using Pseudo-seq (225) or CeU-seq (43), a variety of novel approaches have been developed for detection and mapping pseudouridylation (226–229) and, though we know of many mRNA targets, it is likely that we will discover many additional sites of mRNA pseudouridylation in the future.

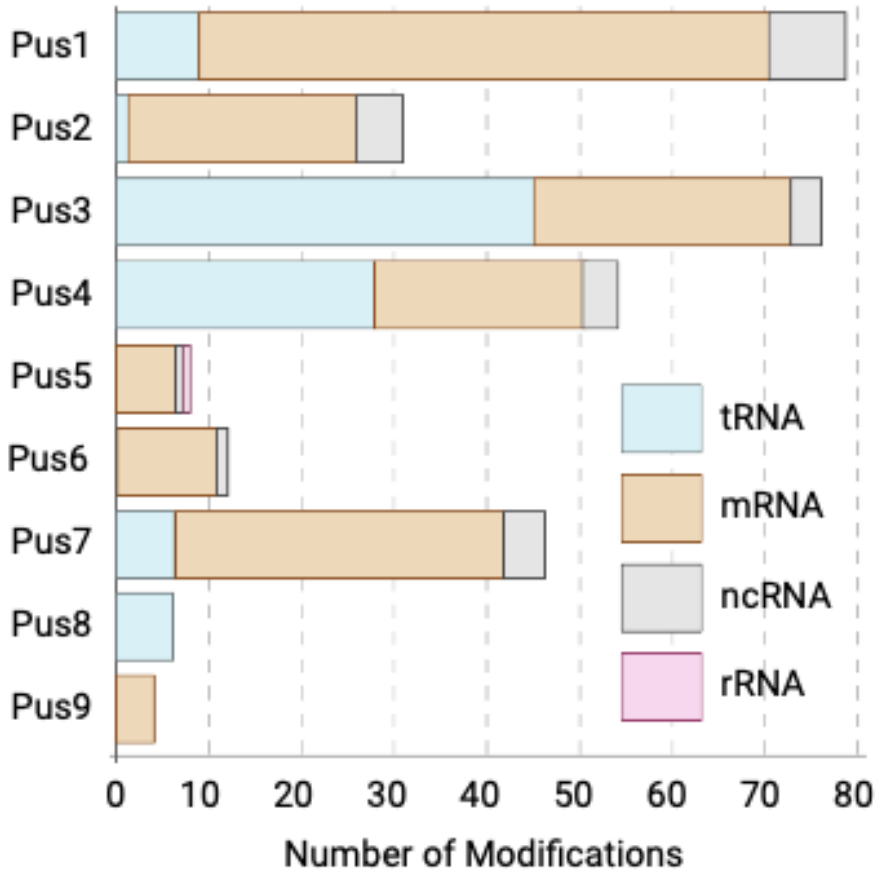


Figure 1.8: Types of substrates modified by Pus enzymes in yeast

General types of RNAs that are modified by each Pus enzyme under standard growth conditions. These data are based on the transcriptome-wide mapping studies by Carlile et al. (41) and Schwartz et al. (40) using the Pseudoseq (or c-seq) technique for the identification of pseudouridines present in *S. cerevisiae*. The combined number of each type of RNA (tRNA, mRNA, ncRNA and rRNA) modified by each of the Pus enzymes is shown as a colored bar. Note, that alteration of growth conditions such as heat stress and nutrient deprivation leads to additional pseudouridylation events by some Pus enzymes, for example by Pus7. These additional events are not included in this figure. Adapted from Rintala-Dempsey, 2017 (148).

With the datasets currently available, it is evident that Pus enzymes tend to target more modification sites in mRNAs compared to their traditional ncRNA targets (Figure 1.8). Pus1 is responsible for the most sites of mRNA pseudouridylation in growing yeast cells, modifying over 60 sites in mRNAs. Pus2 and Pus3 are each responsible for 25-30 unique mRNA targets. Given that Pus4 modifies over 20 mRNA targets and Pus7 installs Ψ at ~35 sites in mRNAs, Pus enzymes belonging to the TruA family seem to be responsible for a disproportionate amount of mRNA pseudouridylation under normal growth conditions (40,41).

However, under heat-shock conditions in yeast, Pus7 catalyzes Ψ formation at over 100 additional sites in mRNAs, though overall both transcript and protein levels of Pus7 decrease (40). Enhanced activity under these conditions is correlated with relocalization of Pus7 from the nucleus to the cytosol where additional Ψ modifications could contribute to mRNA stability under temperature stress (Figure 1.9) (40). Pus1 modifies more mRNA sites under post-diauxic growth in *S. cerevisiae* (41) and multiple Pus enzymes show changes in mRNA targets under nutrient deprivation (41,230). Together, these findings suggest the involvement of some sort of regulatory control over mRNA pseudouridylation events (39).

1.2.4 Substrate selectivity and mRNA modification by tRNA modifying Pus

For decades, eukaryotic pseudouridine synthases were overlooked as non-essential enzymes that installed a seemingly simple modification in a handful of noncoding RNAs, mainly tRNAs and snRNAs. Thus, studies of these enzymes were limited to the identification and basic characterization of pseudouridine synthases in yeast. However, following the discovery of a dynamic network of Ψ in mRNAs, has fostered renewed interest in these enzymes. Genetic and bioinformatic analysis, combined with biochemical deletion studies indicate that the tRNA-

modifying pseudouridine synthases (mainly Pus1, Pus4, and Pus7) are responsible for the majority of mRNA pseudouridylation events in yeast and human cells (40–42). Because the distribution and occurrence of Ψ in mRNAs changes in response to cellular conditions (40,41) – suggesting a potential regulatory mechanism for mRNA pseudouridylation – the basis of substrate selectivity in Pus1, Pus4, and Pus7 has come under renewed scrutiny.

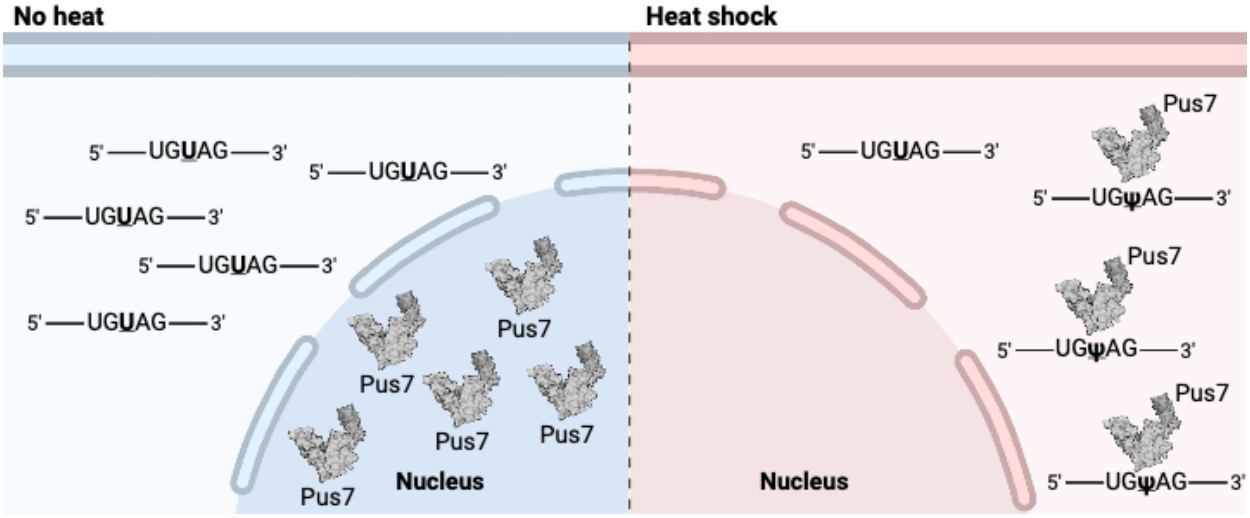


Figure 1.9: Pus7 relocalization and activity under heat-shock

Pus7 may orchestrate the yeast mRNA pseudouridylation program in heat shock. At 30°C, Pus7 is primarily localized in the nucleus, and its localization into the cytoplasm upon heat shock may induce mRNAs pseudouridylation. Adapted from Schwartz, 2014 (40).

1.2.4.1 Eukaryotic Insertions: potential structural basis of substrate selectivity in Pus

Pus enzymes are known for their remarkable substrate specificity, recognizing the target uridine within the context of the RNA substrate. Thus, to achieve site-specific pseudouridine incorporation, Pus enzymes recognize sequence and/or structural motifs within RNAs for modification without the aid of cofactors or accessory proteins (146). Some pseudouridine synthases have strict substrate selectivity and modify only one position or adjacent positions on one type of RNA. Pus4, for example, installs the universally conserved pseudouridine at U55 in the tRNA TΨC loop of tRNA (166). Alternatively, other Pus enzymes are more versatile and install

pseudouridine at structurally distinct positions on different types of RNAs, e.g., Pus1 and Pus7 are both known for modifying substrates displaying diversity in the secondary structure and sequence around the site of modification (208,222). Thus, eukaryotic pseudouridine synthases display a remarkable range of substrate selectivity, each primed for a specific role (148).

Given the structural conservation of the catalytic domain and catalytic residue shared across all pseudouridine synthases, deconvoluting *how* a particular pseudouridine synthase recognizes and selects diverse RNA substrates remains a persistent question. Pus enzymes must select their substrates from the myriad of RNAs in the cell, but exactly how this happens for many individual Pus enzymes is still unclear, especially in the context of mRNA. With the discovery of a dynamically regulated landscape of pseudouridine modifications in mRNA, understanding the molecular mechanisms underlying substrate recognition and selection are especially important to define if/how these modifications contribute to gene expression and cellular adaptation.

The diversity in the functionality of Ψ-synthases arises from structural elaborations of the core catalytic domain, which can be unique to each enzyme (39,146,148). These structural features range from additional α-helices and loops that decorate the exterior of the catalytic domain, to insertions of entire auxiliary domains and may contribute to substrate specificity (221). Indeed, structures of TruB from *E. coli* and *Thermotoga maritima* complexed with RNA stem-loop substrates identified a disordered thumb-loop flanking the catalytic cleft that becomes reordered upon substrate binding, inserting into the groves of the RNA stem, essentially pinching it into place (205,231,232). Furthermore, *E. coli* TruD has a unique TRUD domain (221) in addition to the PUS domain (Figure 1.10). Crystal structures of *E. coli* TruD demonstrate that the catalytic domain and TRUD domain fold together in an overall V-shape and active site cleft, where residues from both domains are required for catalysis (205,220,233,234). At the bottom of the active site cleft, the

TRUD domain forms extensive interactions with the PUS domain forming a hinge-like region. Indeed, comparison of various TruD structures show that the two domains can flex ~18° towards /away from each other (205). Similar 'hinge' regions have been recorded (235,236), suggesting clamping down on an RNA substrate may be a common mechanism supporting substrate recognition.

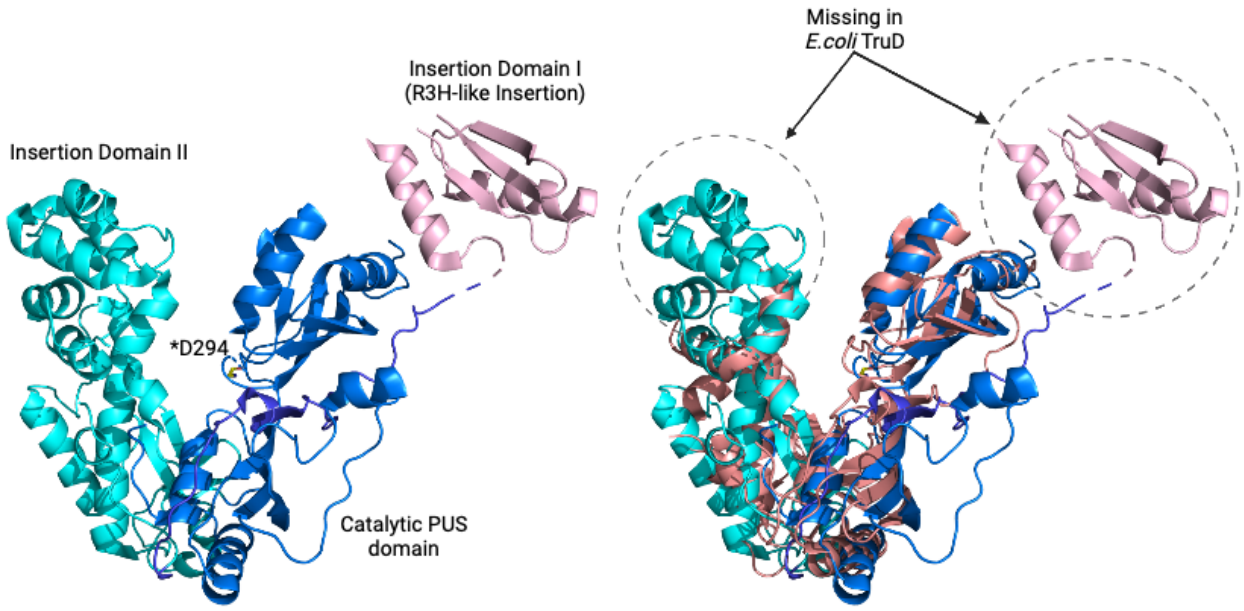


Figure 1.10: Eukaryotic insertion domains in human Pus7

(left) Crystal structure of human Pus7 is shown in cartoon representation and color-coded according to its domain architecture as labeled (PDB: 5KKP) (237). The catalytic D294 residue is shown in yellow sticks and marked with an asterisk. PUS domain shown in blue with TRUD and Insertion II shown in cyan. Insertion I, the largest, is shown in light pink. **(right)** Overlay of human Pus7, color-coded same as (left), and *Escherichia coli* TruD (PDB: 1SB7), shown in salmon, lacking the extra insertion domains.

Further, eukaryotic Ψ-synthases carry additional auxiliary domains affixed to the catalytic PUS domain that are not represented in their prokaryotic counterparts, often referred to as 'insertion' domains (Figure 1.10) (148). The deposited structure of human Pus7 (PDB: 5KKP), a eukaryotic representative of the TRUD family, exhibits three additional structural domains (insertions) that are not present in *E. coli* TruD (Figure 1.10) (203,237,238). Pus7 from *S. cerevisiae* is also predicted to have three insertions at the same location as the additional domains

found in human Pus7 (222). This trend carries across the pseudouridine synthase super family, with the size of these insertions scaling with the complexity of the organism (39). This makes sense as eukaryotic Pus enzymes typically, though not always, have larger substrate scopes and different substrate selectivity in comparison to their prokaryotic counterparts (35,146,148,236).

While the precise function of each insertion in specific Pus enzymes is still being explored, the structural elaborations of the catalytic domain may allow for specific Pus enzymes to recognize and modify their unique target substrates (239). For example, the N-terminal THUMP domain in Pus10 is thought to be involved in facilitating binding of tRNA substrates (240,241). Human Pus1, a member of the TruA family, has a C-terminal extension that folds into a small helical bundle (236,242). A similar extension is also found in yeast Pus1 where it serves to constrain RNA targets to gain access to the target uridine (230). This C-terminal helical bundle is absent in TruA (204,243) and co-crystal structures of TruA bound to an RNA substrate demonstrate that the position and presence of the Pus1 helical bundle would impede a similar mode of substrate binding (244). Further, the structure of human Pus1 demonstrated that Pus1 cannot access the same binding mode as TruA (236,242), consistent with different target selectivity in each enzyme.

1.2.4.2 Substrate selectivity in Pus1, Pus4, and Pus7

Given the involvement of select Pus enzymes in mRNA pseudouridylation, researchers have worked to determine what dictates substrate selectivity in the context of mRNAs. Pus1 and Pus7 are, perhaps, the most intriguing Pus enzymes in this regard as both are known for their multi-site substrate specificity and highly active towards mRNA targets. Pus4 is also intriguing as it displays strict substrate specificity in noncoding RNAs, but now also modifies a variety of new sites in mRNAs (40–42). Therefore, the question of *how* Pus1, Pus4, and Pus7 achieve site-specific

modification of both ncRNA and mRNAs – the molecular and structural basis of these interactions – is of significant interest (39,148).

TruB – One of the first attempts to characterize Pus substrate selectivity in *E. coli* TruB, determined that all of the recognition elements for pseudouridine formation at U55 are contained within the T-arm stem loop (217). *In vivo*, mutations disrupting tRNA tertiary D and T loop interactions did not affect Ψ formation at U55 (245). However, mutations that disrupt the structure of the TΨC loop are not modified by TruB (217,246). Additionally, a loose consensus sequence was noted around the site of modification - suggesting TruB recognizes both the conformation of the T-arm structure and specific bases in the loop for modification (245). Subsequent co crystal structure of TruB in complex with T-loop substrate (205) confirmed that TruB recognized the performed three-dimensional structure of the T loop through shape complementarity - this shape complementarity serves to disrupt the tertiary structure of the stem loop and allow the target uridine residue to be flipped into the active site (205).

In the context of mRNA, Pus4 selectivity is relatively well understood. Sites of Ψ in mRNA installed by Pus4 are broadly united by a loose recognition motif surrounding the site of modification (Figure 1.11) (41). The GUUCNANYCY motif is shared between tRNA and mRNA substrates, suggesting Pus4 may employ a similar mechanism of recognition in both tRNA and mRNA (41,247). Further, a stem-loop structure resembling the TΨC loop of tRNAs may be required for recognition and modification of mRNA targets by human TRUB1 (41,165,247). Thus, Pus4 and TRUB1 likely utilize both primary sequence and secondary structural elements for selective substrate recognition in mRNAs.

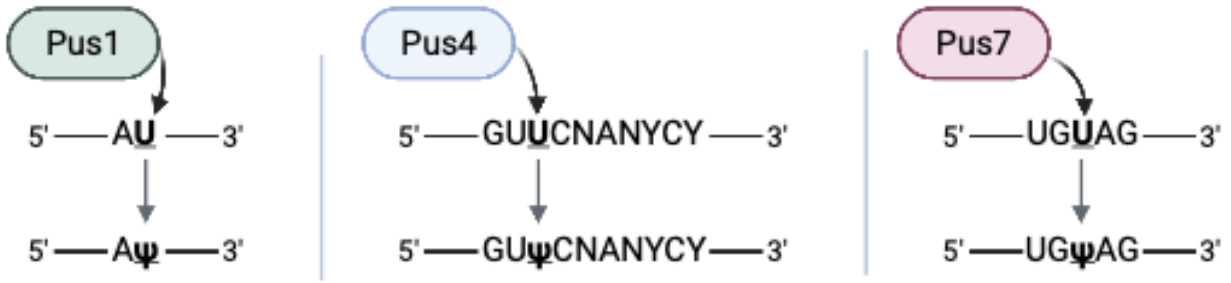


Figure 1.11: Recognition (consensus) sequences for Pus1, Pus4, and Pus7 in mRNAs

Analysis of modified mRNA targets identified Pus-specific recognition sequences that cluster around sites of modification. mRNA pseudouridylation in eukaryotes is mediated by at least four conserved pseudouridine synthases, some site specific (Pus1p, Pus4p, and Pus7p). Adapted from Schwartz, 2014 (40).

TruD – Early interrogations into the basis of substrate specificity in Pus7 suggested a potential requirement of a stem-loop structure proximal to the target uridine (222,224), though given the diversity in secondary structure of Pus7 substrates, RNA structure is unlikely to be the sole determinant of Pus7 specificity. Instead, analysis of the sequence surrounding the site of modification in Pus7-depent mRNA targets suggests Pus7 may utilize a recognition sequence (consensus sequence) for target recognition (40,41,247). In some cases, Pus7 recognizes the 5-mer consensus sequence UG**U**AR (R = G > A) to distinguish target substrates from non-target substrates (Figure 1.11) (41). This makes sense as a strategy for substrate recognition as direct interactions with primary sequence is a proposed method of substrate recognition other pseudouridine synthases (248). However, Ψ-mapping studies demonstrate that the UG**U**AR sequence, which occurs many times throughout the entire transcriptome, is only found to be modified ~2% of the time (41). The 5-mer consensus sequence is insufficient to entirely explain Pus7 specificity. When taken with the dramatic increase in Pus7 pseudouridylation activity under heat-shock (40), it is apparent that additional recognition elements contribute to Pus7 substrate selectivity and a complete understanding of the determinates of Pus7 substrate selection remains to be defined.

TruA – Pus1 has long been recognized for its multi-site substrate specificity – modifying a variety of structurally diverse positions in noncoding RNAs (132,208,249,250). Pus1 requires a coordinated zinc ion for activity and substrate binding (251), and initial characterization of tRNA-substrate selection (determining elements required for high affinity recognition) demonstrated that Pus1 generally has an affinity for RNA and tightly binds (K_D=15-150nM) to both substrates and non-substrates alike (252,253). Interestingly, Pus1 binding is impacted by a G-A base pair near the site of modification (252). The rate of association for Pus1-tRNA is decreased 100-fold by loss of a G26-A44 base pair in *S. cerevisiae* tRNA, though the overall K_D remained unchanged (252). The structure of human Pus1 suggest a different substrate binding orientation compared to TruA, potentially to support the differences in target selectivity (35).

A weak HRU recognition sequence was initially proposed to contribute to Pus1 selectivity in mRNAs (Figure 1.11) (41). Subsequent high-throughput investigation of Pus1-dependent mRNA pseudouridylation using computational and mutational analysis suggested a structure-dependent mode of substrate recognition for Pus1. Many sites of Pus1 modification in mRNA place the target uridine in a similar structural context located at the base of a stem-loop structure (39). This structural motif was demonstrated to be necessary and sufficient for modification by Pus1, but further validation is necessary to delineate the precise determinants of Pus1 substrate selectivity.

1.2.5 Conclusions

Pseudouridine and pseudouridine synthases are undoubtedly important to maintain adequate cellular functions, however the underlying mechanisms that connect Pus enzymes to essential biochemical processes remain poorly defined. Prokaryotic pseudouridine synthases are relatively

well characterized in comparison to their eukaryotic counterparts, but the discovery of Ψ in mRNA revitalized the field of pseudouridine synthase research and placed renewed focus on the enzymes responsible. Therefore, only a select subset of eukaryotic pseudouridine synthases have really been explored and significant questions regarding substrate selectivity, the role of eukaryotic insertions, mechanisms of regulation, and modification activity remain. Studies that aim to characterize the molecular-level details of Pus activity and selectivity are required to begin to untangle the role of pseudouridine synthases and the biological consequences of Ψ in the cell.

1.3 B_{12}-dependent Enzymes

Vitamin B_{12} is a complex organometallic cofactor used by three classes of enzymes to catalyze improbable chemistries, including isomerization, single-carbon transfers, and dehalogenases. In humans, there are only two B12-dependent enzymes, Methionine Synthase (MS) and Methylmalonyl-CoA Mutase (MCM). Both enzymes are essential, playing a central role in vital metabolic processes, including amino acid biosynthesis and single-carbon metabolism. To achieve this, MS and MCM leverage different aspects of the powerful reactivity of B_{12}-cofactors, controlling the surrounding chemical environment to enable and direct the chemistries happening at the cobalt center (254,255).

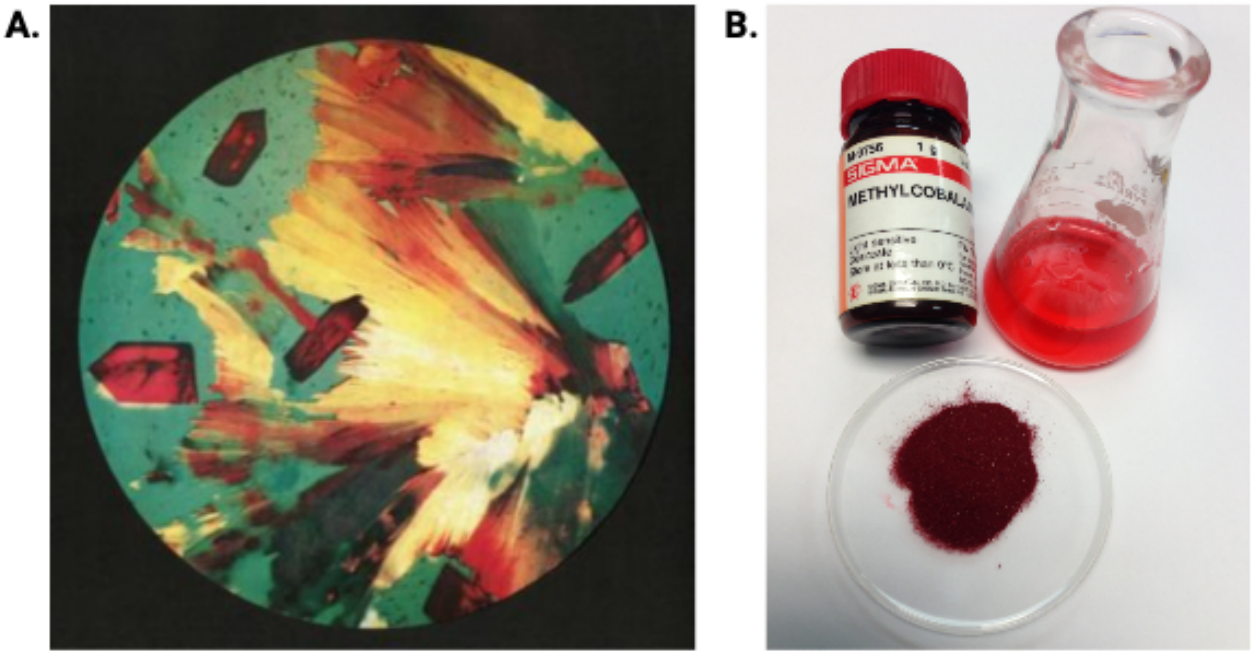


Figure 1.12: Nature's most beautiful cofactor

(A) A photomicrograph showing pure red crystals of vitamin B_{12} (cobalamin). The isolation of vitamin B_{12} by researchers at Merck & Co., Inc., capped a two-decade long initiative to isolate, determine structures, and synthesize vitamins. Adapted from (256). **(B)** Methylcobalamin (shown) is a form of vitamin B_{12}. Physically it resembles the other forms of vitamin B_{12}, occurring as dark red crystals that freely form cherry-colored transparent solutions in water. Adapted from (257).

1.3.1 Nature's most beautiful cofactor(s)

Vitamin B_{12} first earned the moniker of 'nature's most beautiful cofactor' due to its vibrant coloration (Figure 1.12), though flavins are undeniably a close second. In the years following its discovery (258,259), vitamin B_{12} and related corrinoid derivatives have captured the interest of scientists from a multitude of disciplines due to its exceptional organometallic properties (260–263) and ability to catalyze a variety thermodynamically challenging reactions as enzymatic cofactors (255,264).

Innovative x-ray crystallographic studies performed by Hodgkin et al. in 1955 (265,266) described the structure of Vitamin B_{12} and identified the unique corrin core that characterizes this class of organometallic cofactors (267). B_{12}-derivatives are composed of a cobalt ion equatorially suspended in a corrin-macrocycle with a dimethylbenzimidazole (DBI/DMB) nucleotide tail that coordinates the cobalt in the lower (α) axial position. Several different chemical groups are found in the upper (β) axial position, giving rise to a range of B_{12}-derivatives with varied structures and catalytic capabilities (268,269). Vitamin B_{12} is characterized by a cyanide ligand (cyanocobalamin, CNCbl) in the upper axial position. Though CNCbl was first isolated as extrinsic anti-pernicious anemia factor (258,259), no direct biological role has ever been identified for CNCbl (270). Instead, Vitamin B_{12} is processed in the cell, and the resultant B12-derivatives participate in the requisite biological processes. The physiologically relevant B_{12}-deriviatves (cobalamins) include 5'deoxyadenosyl-cobalamin (AdoCbl), methylcobalamin (MeCbl), and aquocobalamin (H_2OCbl^+) (Figure 1.13) (271,272).

1.3.1.1 Cobalamin structure

A combination of techniques, including small-molecule crystallography and solution Nuclear Magnetic Resonance (NMR), have been used to characterize the structure and dynamics of many cobalamin analogs. Broadly, these studies identified the precise chemical structure (bond

saturation), as well as the non-planar nature of the corrin ring (268,269). The precise angle and orientation of the pucker/folding across the macrocycle is variable between cobalamins (268), and dependent on coordination and chemical environments – potentially serving to fine-tune the reactivity of the Co center (273).

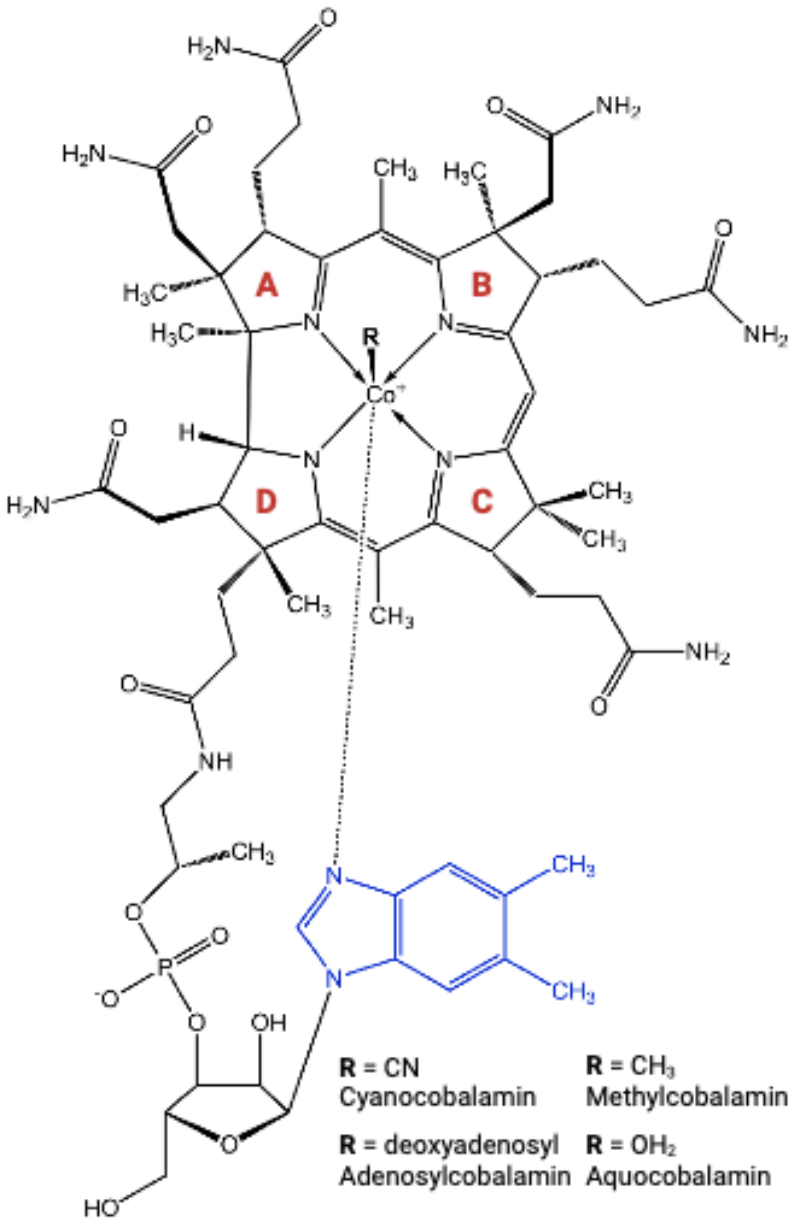


Figure 1.13: Chemical structures of vitamin B_{12} and its derivatives

General chemical structure of B_{12} in the base-on conformation with cobalt formally in the Co(I) state, the lower axial dimethylbenzimidazole ligand below, and the upper axial ligand denoted by "R". Selected upper axial ligands and the corresponding B_{12} forms are shown.

The first structure of B_{12} captured in an enzymatic context, solved by Drennan et al. in 1994 (274), depicted methylcobalamin bound in a novel conformation by the B_{12}-binding domain of *E. coli* MS. Notably, the DBI nucleotide tail was no longer coordinated to the lower-axial position of Co, and instead, inserted deep into a cleft in the center of the protein. In the lower Co coordination site, the DBI tail was replaced with a protein-derived histidine in the α-axial position. Following this discovery in MS, and later identified in MCM (275), B_{12}-dependent enzymes were classified as binding the cofactor in either the 'base-on' or 'base-off/his-on' state, indicating coordination of the lower axial position by the DBI tail, or a protein-derived histidine, respectively (Figure 1.14).

The novel 'base-off/his-on' conformation was intriguing. Attempts to establish the structural rational for the displacement of the DMB-base by histidine, in absence of the enzyme, a host of unnatural biosynthetic imidazolyl-cobamides (ImCba) were crystallized to evaluate the effect of the imidazole ligand on B_{12}-derivitaves (276–278). The DBI tail was thought to have a strong 'ligand-folding' influence on the corrin ring and contribute to the variability in degree and distribution of ring non-planarity observed in different states (268). Indeed, the degree of ring-folding in cyano-imidazolyl-cobalamin (CN-ImCba) was half that of CNCbl with α-DBI coordination, but this is simply explained by smaller steric bulk and greater nucleophilicity of imidazole compared to the nucleotide tail (279). In total, these experiments demonstrate that the structural-rational for DBI-displacement could not be described based solely on the ligand-dependent electronic/chemical changes observed in the isolated cofactor. It is likely that direct interactions between the cofactor and protein binding partners affect DBI-displacement, and binding involves destabilizing the 'base-on' state in favor of histidine coordination. Thus, following this discovery in MS (and MCM soon after), B_{12}-dependent enzymes began to be

classified by the mode of cofactor binding employed - either binding the cofactor in either the 'base-on' or 'base-off/his-on' state, indicating coordination of the lower axial position by the DBI tail, or a protein-derived histidine, respectively (Figure 1.14). Furthermore, though this example only pertained to the lower axial ligands, these experiments suggest that the chemical and structural properties of the cofactor are highly sensitive to the surrounding chemical and electronic environment, and hints towards the unique power of enzymes to manipulate the chemical reactivity and electronics throughout the cofactor.

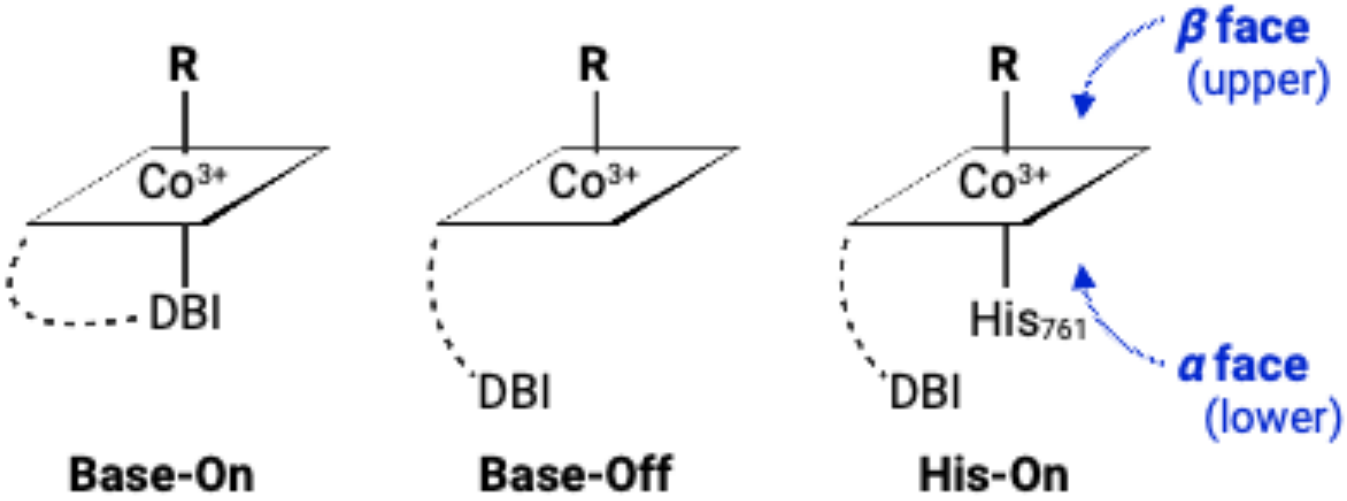


Figure 1.14: Coordination of lower axial ligand

The free cofactor can exist in the base-on (or Dmb-on) or base-off (or Dmb-off) conformations, with the former predominating at physiological pH. The His-on conformation in which the endogenous ligand, dimethylbenzimidazole, is replaced by an active site histidine is seen is some B_{12}-dependent enzymes.

1.3.1.2 Redox chemistry and cobalt coordination

The impressive catalytic versatility/potential of B_{12}-cofactors is enabled by the unique organometallic properties and redox-capabilities of the Cobalt (Co) center. Under physiological conditions, cobalamins access the Co(III), Co(II), and Co(I) oxidation states (Figure 1.15), and as each oxidation state correlates with distinct reactivities and coordination properties (262,280), it is unsurprising that redox-based processes form the basis of many cobalamin chemistries. Indeed,

redox-transitions in B_{12}-derivatives were analyzed using electrochemical methods in 1983 (281) and further investigation showed that the oxidation state of Co influences, if not dictates, the axial coordination environment (268,280–282). The preferential coordination geometry of each Co oxidation state was later confirmed; spectroscopic studies of cobalamin cofactors in aqueous solution established that MeCbl is 6 coordinate octahedral geometry, Co(II)cbl is five coordinate square-pyramidal, and Co(I)cbl is preferentially 4 coordinate square-planar (280,281). Therefore, the redox chemistry carried out by B_{12}-enzymes involves alternating between oxidation states, with each transition involving a change in the number/identity of axial ligands, which have a strong effect on the kinetics and thermodynamics of electron transfers in cobalamins. Indeed, characterization of B_{12}-dependent enzymes in subsequent years demonstrated correlation between oxidation state and enzymatic activity (281,282). Thus, cobalamin cofactors are now known to cycle between oxidization states and coordination of axial ligands to enable each step of catalysis.

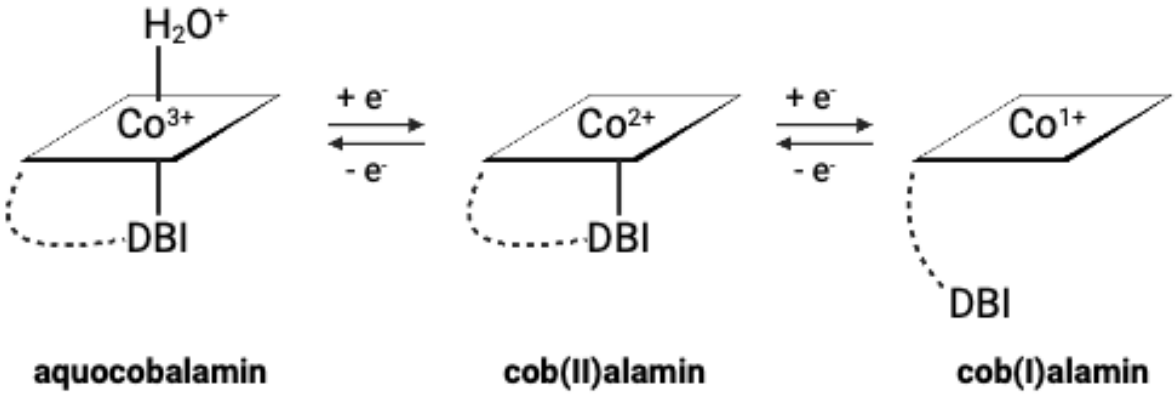


Figure 1.15: Redox-transitions between aquocob(III)alamin, cob(II)alamin, and cob(I)alamin

1.3.1.3 Breaking the bond, the co-c bond

One of the intriguing features of cobalamin cofactors is the formation of an unusual Co-C bond, the breaking (and reformation) of which represents a crucial step in the catalysis of B_{12}-dependent enzymes (262,263). In human cobalamin-dependent enzymes, the organometallic cleavage of the

Co-C bond can occur via homolytic cleavage and heterolytic cleavage (Figure 1.16). A homolytic mechanism involves the cleavage/formation of a single axial bond with the electrons shared equally between the products. This type of reactivity is characteristic of AdoCbl in MCM:

5'-deoxyadenosyl-cob(III)alamin ⇄ cob(II)alamin + 5'deoxyadenosyl radical

Homolytic bond cleavage in Co-C bond in AdoCbl is especially important in its role as a cofactor to MCM (and other AdoCbl-dependent enzymes) as AdoCbl can reversibly function as a radical 'source' (283)[5'-deoxyadenosyl radical] and a radical 'trap' (284)[cobalt radical], the latter of which readily reacts with alkyl radicals (Figure 1.16).

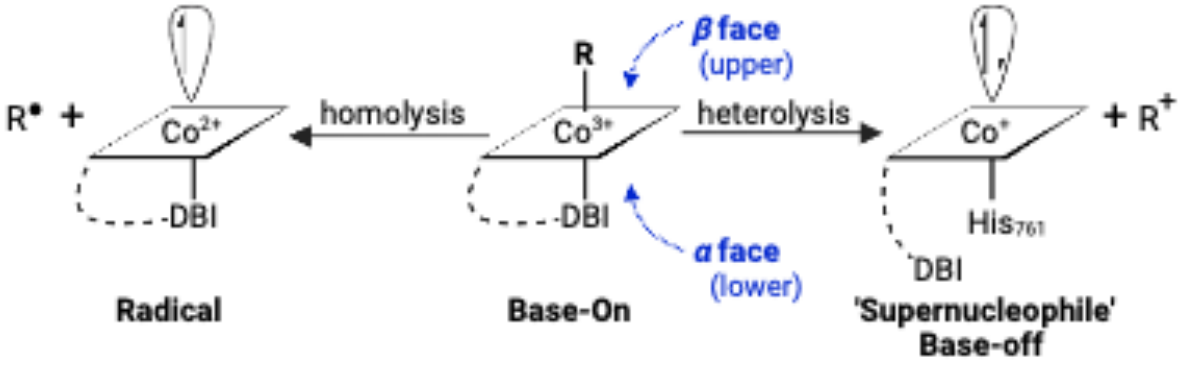


Figure 1.16: Cleavage of the Co-C bond

Both homolytic and heterolytic mechanisms are employed by cobalamin-dependent enzymes.

Alternatively, the nucleophile induced heterolytic mechanism involves the cleavage/formation of two axial bonds in *trans* (Figure 1.16), where both electrons end up on a single molecule. In MS, MeCbl employs this heterolytic mode of reactivity:

methylcob(III)alamin + nucleophile ⇄ cob(I)alamin + methylated product

This mode of reactivity in MeCbl involves the inter-conversion between cob(I)alamin and methylcob(III)alamin, vital in many methyl-transfer reactions carried out by enzymes (Figure 1.16) (262,263,285). Cob(I)alamin is one of the strongest nucleophiles found in nature, deemed a 'supernucleophile'(286). Thus, cob(I)alamin abstracts methyl-groups from challenging donors to form Co-CH_3 [methylcob(III)alamin], subsequently demethylated by nucleophilic species to regenerate the cob(I)alamin supernucleophile. Methyl-transfer catalyzed by Co(I) species can proceed through bimolecular nucleophilic substitution (SN2) where Co(I) acts as a supernucleophile or, under certain conditions, alkylation of Co(I) can progress through a two-step one-electron transfer pathway. In the latter scenario, cob(I)alamin acts as strong one-electron reducing agent forming cob(II)alamin intermediates throughout the reaction (262,287).

Comparison of corrin structure between homolytic and heterolytic mechanisms employed by cobalamins demonstrates a stark difference in structural changes undergone in each mechanism. Notably, homolysis-based reactivity does not dramatically affect the cobalamin structure, consistent with findings that the strength of the Co-C bond in AdoCbl is minimally influenced by the coordination of the lower ligand (262,288). Heterolytic cleavage/formation, on the other hand, requires significant reorganization at both faces (α- and β- face) of the cobalamin. Thus, cobalamins that undergo heterolytic cleavage (ex. MeCbl) are subject to considerable axial ligand effects which tune the reactivity of the cofactor (262,274).

1.3.2 Methionine synthase

MS promotes three different methyl-transfer reactions on Cbl throughout the enzyme's catalytic and reactivation cycles (Figure 1.17). During the MS catalytic cycle, the cofactor uses one of the strongest nucleophiles found in nature, Co(I) (267), to facilitate a thermodynamically challenging

methyl transfer from methyl-tetrahydrofolate (CH_3-H_4folate, MTF) to homocysteine through the formation of a Co-C bond (263,286,289).

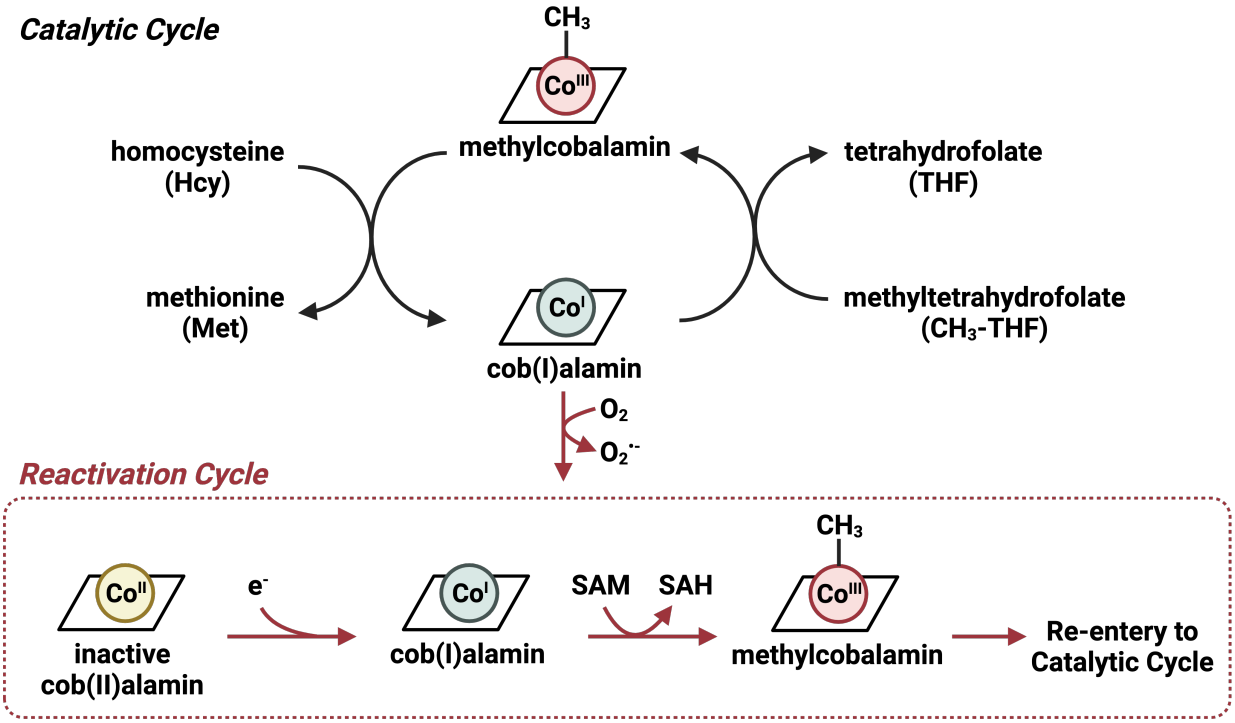


Figure 1.17:Catalytic and reactivation cycles of MS

Catalytic cycle of MS shown in black arrows, with requisite domains highlighted for each reaction. Upon oxidative inactivation, MS undergoes the reactivation cycle (bottom, red arrows) to regenerate the catalytically competent cofactor.

To accomplish this Cbl is involved in two chemically distinct methyl-transfer reactions: the first from CH_3-H_4folate to Cbl(I), and a second from CH_3-Cbl(III) to a homocysteine that has been activated to receive the methyl-group by Zn^{2+} (290–292). Thus, MS reactivity relies on the enzyme continuously cycling between the CH_3-Cbl(III) and Cbl(I) forms of the cofactor (293,294). Once every ~2000 enzyme turnovers, microaerophilic cellular conditions lead to the inactivation of MS, where the Cbl(I) supernucleophile is oxidized to a catalytically inactive Cbl(II) species (Figure 1.17, bottom) (295,296). When this happens, MS undergoes a reactivation cycle to restore catalytic competency. During the reactivation cycle Cbl(II) is enzymatically reduced and

subsequently methylated by S-adenosyl-L-methionine (AdoMet) to create a functional CH_3-Cbl(III) cofactor that can re-enter the catalytic cycle (Figure 1.17) (296–300).

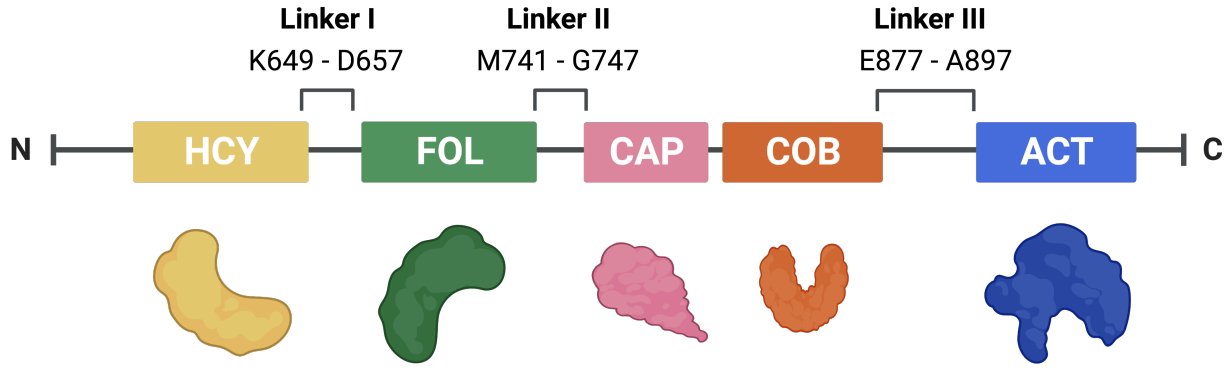


Figure 1.18: Domain architecture of Methionine Synthase

MS is a 5-module protein, where each of the three substrates and the Cbl cofactor bind to a different enzyme module (Figure 1.18). These modules are linearly arranged and connected by intermodular linker regions, like beads on a string (301,302). Previous studies established the general function of each MS module and revealed the dynamic complexity of this system (292,296,303,304). The first two modules [homocysteine-binding acceptor domain (Hcy) and the MTF-binding donor domain (Fol)] work together to catalyze methyl transfers to and from Cbl, which is bound by the fourth (Cob) module (Figure 1.18 and Figure 1.19). Cbl is thought to be protected from unwanted side reactions between catalytic steps by a small 'Cap' module that sits between the Fol and Cob regions (274). The fifth and final MS module is the AdoMet binding reactivation module (Act). Act is required for the reductive reactivation of Cbl upon oxidative inactivation to Cbl(II) (296,305). Each of the methyl transfer reactions MS catalyzes necessitates a different modular arrangement (Figure 1.19, B-E). Therefore, major conformational changes are needed to displace the Cap domain and bring individual module into contact with Cbl for each step of the catalytic and reactivation cycles.

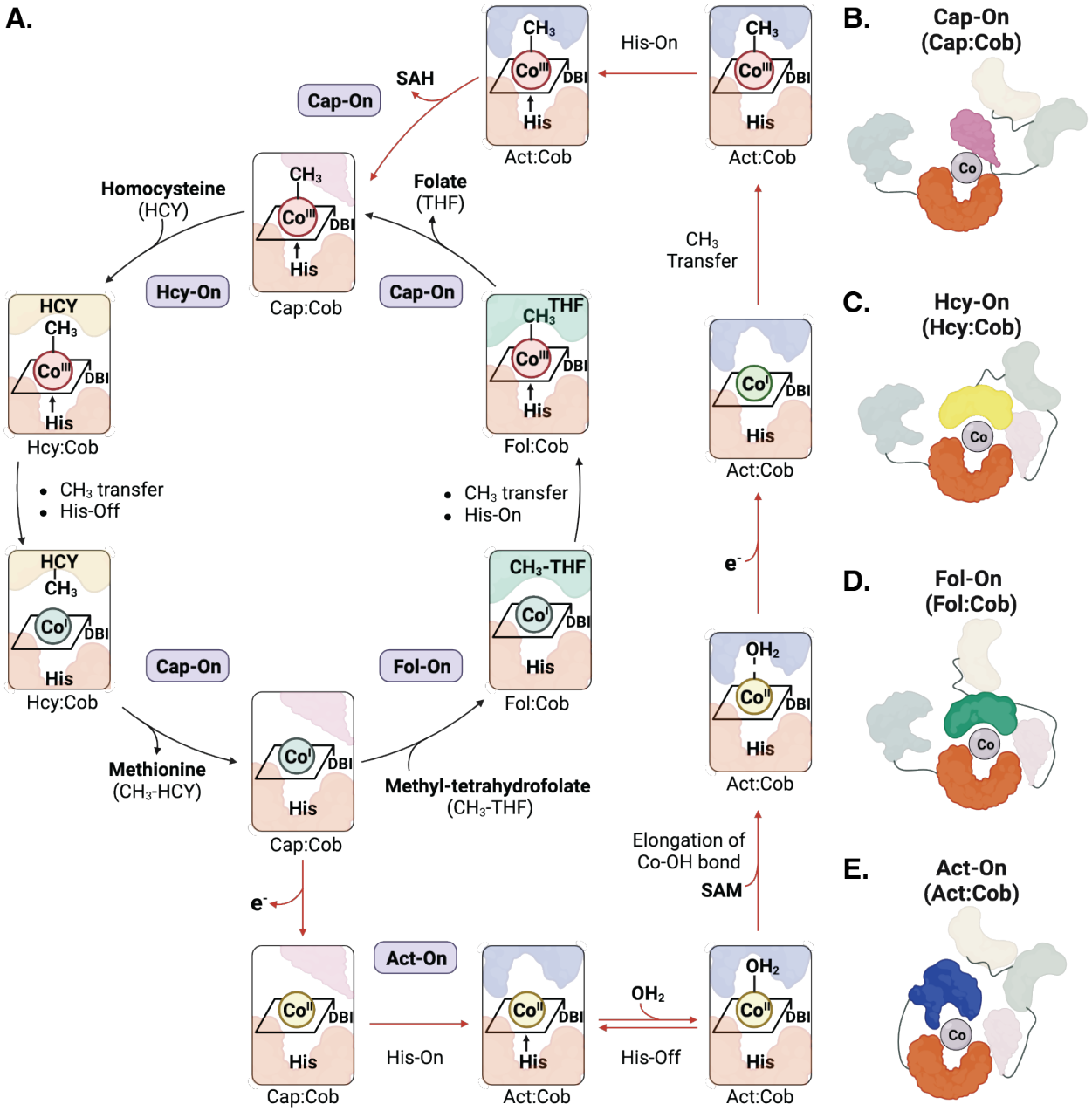


Figure 1.19: Conformational dynamics of MS

(A) The catalytic cycle of MS (black arrows) and the reactivation cycle (red arrows). Conformations indicated in purple boxes. **(B-E)** Conformations adopted by MS throughout the catalytic and reactivation cycles with different substrate binding domains positioned above the cofactor.

MS operates by an alternating domain mechanism, where domains move over large distances to form each catalytic conformation (306–308). Figure 1.19 highlights the four proposed conformations of MS, with the Cob interacting with each of the four other modules. These states

exist in a dynamic equilibrium determined by the oxidation state of the cobalamin cofactor, as well as the concentrations of the substrates or products (302). Substrates and products do not control the distribution of the molecular species through an allosteric mechanism. Instead, the overall distribution is regulated by the relative stabilities of different molecular arrangements (308).

While how the Cbl cofactor facilitates methyl-transfer has been studied for nearly 50 years, a number of key questions about MS have not been answered due to the lack of an MS system amenable to vigorous biochemical manipulation. Significantly, although methyl transfer from MTF to Cbl(I) has been shown to proceed via a S_N2 mechanism, the proton donor for the reaction has not been identified (286,309–311). Furthermore, many of the residues important for activating the homocysteine substrate for catalysis are not known. Although crystal structures implicate several important positions in the enzyme (274,303,306,308), mutations to test the function of these residues have not been possible to either create, express, or purify. Thus, a system more amenable to biochemical manipulation will be of significant importance to the advancement of MS characterization.

1.3.3 Methylmalonyl-CoA Mutase

Methylmalonyl-CoA Mutase (MCM) is a B_{12}-dependent isomerase and indispensable in human metabolism (312). B_{12}-dependent isomerases make up one of the largest subfamilies in bacteria, where they play a vital role in fermentation pathways (312–314). MCM is the exception, however. Found both in both bacteria and *H. sapiens*, MCM catalyzes the reversible isomerization of methylmalonyl-CoA to succinyl-CoA through homolytic cleavage of the Co-C bond (283). In contrast to MS, MCM binds 5'deoxyadenosyl-cobalamin (AdoCbl) and harnesses the reactivity of

the 5'deoxyadenosyl radical for H-abstraction (315) in this important biological transformation (283,316).

MCM is in the subclass of B_{12}-dependent enzymes that binds the cofactor in the 'base-off/his-on' conformation (275,317). The coordinating histidine is found in a conserved DX**H**XXG motif in a loop lining the upper edge of the B_{12}-binding domain. This motif is conserved in both B_{12}-dependent isomerases and methyltransferases (including MS) that bind the cofactor in the 'base-off/his-on' mode (Figure 1.14) (318).

MCM exploits the inherent lability of the Co-C bond to enact radical based arrangements. However, the homolysis products, cob(II)alamin and 5'deoxyadenosyl radical, are not observed in the absence of substrate. Upon substrate addition, the rate of homolytic cleavage is accelerated ~trillionfold compared to the uncatalyzed rate in solution (283,316,319), indicating substrate binding induces labialization of the Co-C bond towards homolysis. This could represent one mechanism utilized by MCM to control the radical to prevent unwanted side reactions and maintain catalytic efficiency. The basis of Co-C bond labialization upon substrate incorporation is proposed to stem from additional strain on the cofactor that is relieved by homolysis of the Co-C bond and tight interactions with the products (320–322).

One key issue regarding MCM catalysis is how the timing and reactivity of the radical is controlled to prevent unwanted side reactions and ensure homolysis of the Co-C bond. This question has been explored in a variety of B_{12}-dependent radical enzymes through kinetic and biochemical techniques. Stopped-flow kinetic analysis has been used to measure the rate of homolysis, and found that it is not rate limiting, and furthermore, the rate of homolysis is sensitive to isotopic substitution (323). Deacceleration of the rate of homolysis has been observed in MCM (323,324) and two other similar B_{12}-dependent radical enzymes, glutamate mutase (325) and

ethanolamine ammonia lyase (326). These findings were rationalized as evidence for kinetic coupling between homolytic bond cleavage and substrate radical generation, where the creation of the stable substrate-centered radical drives the reaction equilibrium towards homolysis (327). Further, this model provides an explanation for the observed isotopic effects: rate of homolysis is dependent on the formation of the substrate radical and as the mechanism involves H-atom abstraction, the rate of homolysis will be sensitive to a deuterium or a hydrogen at that position (327).

The majority of research has been done on MCM from *Propionibacterium shermanii* (MMCM) (328), which was also the first accurate structure of an AdoCbl dependent enzyme (275,329,330) MMCM binds AdoCbl in the 'base-off' orientation, where AdoCbl and MeCbl adopt the 'base-on' orientation in solution. Insights into the cobalamin binding have been inferred from the solution NMR structure of the *apo*B_{12}-binding subunit of glutamate mutase from *C. tetanomorphum* (331,332). These experiments demonstrated the B_{12}-binding subunit is largely a dynamically preformed structure, primed and ready for B_{12}-binding, though a flexible 'nascent' helix becomes ordered upon cofactor binding. One possibility for B_{12} incorporation that has been proposed posits that the 'base-off' form of the cofactor in solution is trapped by the nucleotide tail. The binding of the DBI tail may, in turn, stabilize the protein further, as seen in the nascent helix reordering (331). The mechanism of B_{12} loading, as well as the control of timing and precision of radical trajectories, represent significant questions that remain regarding MCM.

1.4 Thesis overview and conclusions

Proteins are essential biomolecular machines that perform the chemistries that sustain complex life. Given this central role, dysfunction or dysregulation of proteins can have serious biological implications. Like the rest of the natural world, proteins are defined by the Structure-Function relationship, wherein the three-dimensional arrangement of a system is intertwined with – and ultimately enables – its correct function. Accordingly, a large proportion of basic research is devoted to understanding the structural and molecular-level details underlying protein function, as well as the enzymatic basis of dysfunction. These studies not only advance our understanding of the detailed mechanisms that enable life, but also represent a fundamental precursor towards biotechnological or therapeutic development. Thus, in this work, I have leveraged the Structure-Function relationship to explore the basis of improbable chemistries performed by three unique enzymes.

In Chapter 2, I investigate the determinants of substrate selectivity in one of the predominant mRNA modifying enzymes, pseudouridine synthase 7 (Pus7), and ultimately describe the contributions of protein features and substrate structure to Pus7 binding and modification. Pus7 recognizes a consensus sequence around the site of modification, but this 5-mer sequence alone is insufficient to describe substrate selectivity and other elements including substrate structure may contribute to Pus7 recognition. To establish the structural basis of Pus7-RNA interactions, I solved the structure of *Saccharomyces cerevisiae* Pus7 and interrogated the contribution of key protein features, including active site residues and eukaryotic-specific insertions, to Pus7 activity. Additionally, I examined the effect of substrate structure on Pus7 binding and modification using electrophoretic mobility shift assays (EMSA) and tritium release assays, respectively. Surprisingly, Pus7 was able to modify every tRNA, mRNA, and non-native

substrate that contained the consensus sequence. In total, these findings indicate that Pus7 is very promiscuous and that Pus7 selectivity is determined by factors including localization, competition with other RNA-binding proteins, and substrate structure. Pus7 has always been recognized for its multi-substrate specificity, but the basis of selectivity in these interactions was unknown. The work described in Chapter 2 addresses these questions and establishes a foundation to further interrogate Pus7-RNA interactions.

Chapter 3 explores the structural basis of molecular juggling through two novel structures of Cobalamin-dependent Methionine Synthase (MS) from *Thermus thermophilus*. Traditionally, structural characterization of MS has relied on the 'divide and conquer' approach due to biochemical challenges associated with expression and crystallization of MS. Thus, only structures of individual domains or di-domains have been reported. In this work, we identified a MS from *T. thermophilus* that overcomes these longstanding barriers and present the first full-length structure of MS, as well as the first 'catalytic' structure with the folate-binding domain positioned above the B_{12}-binding domain. Together, these structures provide valuable insight into B_{12}-loading and the conformations supporting methyl-transfer in MS. We propose that one of the major determinants of conformational states in MS is the cofactor itself, which acts as the mediator and fosters the primary stabilizing interactions with the domain positioned above, rather than relying on domain:domain interactions. In total, the work presented here addresses significant questions regarding the conformations adopted by MS, provides validation for the 'divide and conquer' approach traditionally used for structural characterization of MS, and offers a solution (in the form of the thermophilic homolog *t*MS) to overcome the longstanding challenges of working with this enzyme.

In Chapter 4, I examine the control of radical trajectories and the mechanism of inhibition in another essential B_{12}-dependent enzyme, Methylmalonyl-CoA Mutase (MCM). MCM uses 5'-deoxyadenosyl-cobalamin (AdoCbl) to catalyze the interconversion of Methylmalonyl-CoA to Succinyl-CoA. For this isomerization, MCM employs a radical mechanism through homolytic cleavage of the Co-C bond, but it is unclear how the radical is controlled to prevent unwanted side reactions. To address this, I solved the structure of *Mycobacterium tuberculosis* MCM with the suicide inhibitor Itaconyl-CoA (I-CoA), a close analog of Succinyl-CoA. In addition to visualizing conformational changes in the heterodimer upon inhibitor binding, subsequent EPR studies performed on the MCM-I-CoA crystals confirmed the presence of an air-stable biradical captured *in crystallo*. Together, this work unambiguously describes the means of suicide inactivation by I-CoA and the mechanisms controlling the radical-based isomerization carried out by MCM.

1.5 References

1. Suki B. Chapter 1 - Introduction to structure-function relationships. In: Suki B, editor. Structure and Function of the Extracellular Matrix [Internet]. Academic Press; 2022 [cited 2022 Oct 27]. p. 1–7. Available from: https://www.sciencedirect.com/science/article/pii/B9780128197165000022

2. Buckminsterfullerine, diamond and graphite illustrate "the very rich chemistry of carbon [Internet]. Dezeen. 2021 [cited 2022 Oct 27]. Available from: https://www.dezeen.com/2021/06/25/carbon-allotropes-buckminsterfullerene-andrei-khlobystov/

3. Belenkov EA, Greshnyakov VA. Classification schemes for carbon phases and nanostructures. New Carbon Materials. 2013 Aug 1;28(4):273–82.

4. Sanvictores T, Farci F. Biochemistry, Primary Protein Structure. In: StatPearls [Internet]. Treasure Island (FL): StatPearls Publishing; 2022 [cited 2022 Oct 27]. Available from: http://www.ncbi.nlm.nih.gov/books/NBK564343/

5. Nourbakhsh M. Analysis of RNA secondary structure. Methods Mol Biol. 2014;1182:35–42.

6. Cook-Deegan R. The science commons in health research: structure, function, and value. J Technol Transf. 2007;32:133–56.

7. Redfern OC, Dessailly B, Orengo CA. Exploring the structure and function paradigm. Curr Opin Struct Biol. 2008 Jun;18(3):394–402.

8. Uversky VN. Protein intrinsic disorder and structure-function continuum. Prog Mol Biol Transl Sci. 2019;166:1–17.

9. Zhang K, Qu G, Liu W, Sun Z. [Structure-function relationships of industrial enzymes]. Sheng Wu Gong Cheng Xue Bao. 2019 Oct 25;35(10):1806–18.

10. Anderson AC. The Process of Structure-Based Drug Design. Chemistry & Biology. 2003 Sep 1;10(9):787–97.

11. Everett JR. Academic drug discovery: current status and prospects. Expert Opin Drug Discov. 2015;10(9):937–44.

12. Takebe T, Imai R, Ono S. The Current Status of Drug Discovery and Development as Originated in United States Academia: The Influence of Industrial and Academic Collaboration on Drug Discovery and Development. Clin Transl Sci. 2018 Nov;11(6):597–606.

13. A B, S G. Analytical Techniques for Structural Characterization of Proteins in Solid Pharmaceutical Forms: An Overview. Pharmaceutics [Internet]. 2021 Apr 11 [cited 2022 Oct 27];13(4). Available from: https://pubmed.ncbi.nlm.nih.gov/33920461/

14. Kwok SC, Mant CT, Hodges RS. Importance of secondary structural specificity determinants in protein folding: Insertion of a native β-sheet sequence into an α-helical coiled-coil. Protein Science. 2002;11(6):1519–31.

15. Bongirwar V, Mokhade AS. Different methods, techniques and their limitations in protein structure prediction: A review. Prog Biophys Mol Biol. 2022 Sep;173:72–82.

16. Deutzmann R. Structural characterization of proteins and peptides. Methods Mol Med. 2004;94:269–97.

17. Papageorgiou AC, Poudel N, Mattsson J. Protein Structure Analysis and Validation with X-Ray Crystallography. Methods Mol Biol. 2021;2178:377–404.

18. Benjin X, Ling L. Developments, applications, and prospects of cryo-electron microscopy. Protein Sci. 2020 Apr;29(4):872–82.

19. Kay LE. NMR studies of protein structure and dynamics. J Magn Reson. 2005 Apr;173(2):193–207.

20. Acharya KR, Lloyd MD. The advantages and limitations of protein crystal structures. Trends Pharmacol Sci. 2005 Jan;26(1):10–4.

21. D'Imprima E, Kühlbrandt W. Current limitations to high-resolution structure determination by single-particle cryoEM. Q Rev Biophys. 2021 Mar 11;54:e4.

22. Turnbull AP, Wu X. Studying RNA-Protein Complexes Using X-Ray Crystallography. Methods Mol Biol. 2021;2263:423–46.

23. Ke A. Crystallization of RNA and RNA?protein complexes. Methods. 2004 Nov;34(3):408–14.

24. Rm G, Rj H. Reaching the information limit in cryo-EM of biological macromolecules: experimental aspects. Biophysical journal [Internet]. 2011 May 18 [cited 2022 Oct 27];100(10). Available from: https://pubmed.ncbi.nlm.nih.gov/21575566/

25. Johnson DT, Jones LM. Hydroxyl radical protein footprinting for analysis of higher order structure. Trends in Biochemical Sciences. 2022 Nov 1;47(11):989–91.

26. Calvaresi V, Redsted A, Norais N, Rand KD. Hydrogen–Deuterium Exchange Mass Spectrometry with Integrated Size-Exclusion Chromatography for Analysis of Complex Protein Samples. Anal Chem. 2021 Aug 24;93(33):11406–14.

27. McKenzie-Coe A, Montes NS, Jones LM. Hydroxyl Radical Protein Footprinting: A Mass Spectrometry-Based Structural Method for Studying the Higher Order Structure of Proteins. Chem Rev. 2022 Apr 27;122(8):7532–61.

28. Tullius TD. Chemical ‘snapshots’ of DNA: using the hydroxyl radical to study the structure of DNA and DNA-protein complexes. Trends in Biochemical Sciences. 1987 Jan 1;12:297–300.

29. Shaytan AK, Xiao H, Armeev GA, Wu C, Landsman D, Panchenko AR. Hydroxyl-radical footprinting combined with molecular modeling identifies unique features of DNA conformation and nucleosome positioning. Nucleic Acids Research. 2017 Sep 19;45(16):9229–43.

30. Wilkinson KA, Merino EJ, Weeks KM. Selective 2’-hydroxyl acylation analyzed by primer extension (SHAPE): quantitative RNA structure analysis at single nucleotide resolution. Nat Protoc. 2006;1(3):1610–6.

31. Deigan KE, Li TW, Mathews DH, Weeks KM. Accurate SHAPE-directed RNA structure determination. Proc Natl Acad Sci USA. 2009 Jan 6;106(1):97–102.

32. Breiling A, Lyko F. Epigenetic regulatory functions of DNA modifications: 5-methylcytosine and beyond. Epigenetics Chromatin. 2015;8:24.

33. Millar AH, Heazlewood JL, Giglione C, Holdsworth MJ, Bachmair A, Schulze WX. The Scope, Functions, and Dynamics of Posttranslational Protein Modifications. Annual Review of Plant Biology. 2019;70(1):119–51.

34. Schaefer M, Kapoor U, Jantsch MF. Understanding RNA modifications: the promises and technological bottlenecks of the ‘epitranscriptome.’ Open Biology. 7(5):170077.

35. Boccaletto P, Stefaniak F, Ray A, Cappannini A, Mukherjee S, Purta E, et al. MODOMICS: a database of RNA modification pathways. 2021 update. Nucleic Acids Research. 2022 Jan 7;50(D1):D231–5.

36. Krutyhołowa R, Zakrzewski K, Glatt S. Charging the code - tRNA modification complexes. Curr Opin Struct Biol. 2019 Apr;55:138–46.

37. Cui L, Ma R, Cai J, Guo C, Chen Z, Yao L, et al. RNA modifications: importance in immune cell biology and related diseases. Sig Transduct Target Ther. 2022 Sep 22;7(1):1–26.

38. Gilbert WV, Bell TA, Schaening C. Messenger RNA modifications: Form, distribution, and function. Science. 2016 Jun 17;352(6292):1408–12.

39. Borchardt EK, Martinez NM, Gilbert WV. Regulation and Function of RNA Pseudouridylation in Human Cells. Annu Rev Genet. 2020 Nov 23;54(1):309–36.

40. Schwartz S, Bernstein DA, Mumbach MR, Jovanovic M, Herbst RH, León-Ricardo BX, et al. Transcriptome-wide mapping reveals widespread dynamic-regulated pseudouridylation of ncRNA and mRNA. Cell. 2014 Sep 25;159(1):148–62.

41. Carlile TM, Rojas-Duran MF, Zinshteyn B, Shin H, Bartoli KM, Gilbert WV. Pseudouridine profiling reveals regulated mRNA pseudouridylation in yeast and human cells. Nature. 2014 Nov 6;515(7525):143–6.

42. Lovejoy AF, Riordan DP, Brown PO. Transcriptome-wide mapping of pseudouridines: pseudouridine synthases modify specific mRNAs in S. cerevisiae. PLoS One. 2014;9(10):e110799.

43. Li X, Zhu P, Ma S, Song J, Bai J, Sun F, et al. Chemical pulldown reveals dynamic pseudouridylation of the mammalian transcriptome. Nat Chem Biol. 2015 Aug;11(8):592–7.

44. Wang S, Lv W, Li T, Zhang S, Wang H, Li X, et al. Dynamic regulation and functions of mRNA m6A modification. Cancer Cell Int. 2022 Jan 29;22:48.

45. Liu J, Huang T, Chen W, Ding C, Zhao T, Zhao X, et al. Developmental mRNA m5C landscape and regulatory innovations of massive m5C modification of maternal mRNAs in animals. Nat Commun. 2022 May 5;13(1):2484.

46. Hsu PJ, Fei Q, Dai Q, Shi H, Dominissini D, Ma L, et al. Single base resolution mapping of 2'-O-methylation sites in human mRNA and in 3' terminal ends of small RNAs. Methods. 2019 Mar 1;156:85–90.

47. Choi J, Indrisiunaite G, DeMirci H, Ieong KW, Wang J, Petrov A, et al. 2′-O-methylation in mRNA disrupts tRNA decoding during translation elongation. Nat Struct Mol Biol. 2018 Mar;25(3):208–16.

48. Harcourt EM, Kietrys AM, Kool ET. Chemical and structural effects of base modifications in messenger RNA. Nature. 2017 Jan 18;541(7637):339–46.

49. Hoernes TP, Hüttenhofer A, Erlacher MD. mRNA modifications: Dynamic regulators of gene expression? RNA Biol. 2016 Jun 28;13(9):760–5.

50. Meyer KD, Jaffrey SR. Rethinking m6A Readers, Writers, and Erasers. Annu Rev Cell Dev Biol. 2017 Oct 6;33:319–42.

51. Nachtergaele S, He C. The emerging biology of RNA post-transcriptional modifications. RNA Biol. 2017 Feb;14(2):156–63.

52. Peer E, Rechavi G, Dominissini D. Epitranscriptomics: regulation of mRNA metabolism through modifications. Curr Opin Chem Biol. 2017 Dec;41:93–8.

53. Roundtree IA, Evans ME, Pan T, He C. Dynamic RNA Modifications in Gene Expression Regulation. Cell. 2017 Jun 15;169(7):1187–200.

54. Shen L, Liang Z, Wong CE, Yu H. Messenger RNA Modifications in Plants. Trends Plant Sci. 2019 Apr;24(4):328–41.

55. Jones JD, Monroe J, Koutmou KS. A molecular-level perspective on the frequency, distribution, and consequences of messenger RNA modifications. WIREs RNA. 2020;11(4):e1586.

56. McCown PJ, Ruszkowska A, Kunkler CN, Breger K, Hulewicz JP, Wang MC, et al. Naturally occurring modified ribonucleosides. WIREs RNA. 2020;11(5):e1595.

57. PubMed [Internet]. PubMed. [cited 2022 Oct 28]. Available from: https://pubmed.ncbi.nlm.nih.gov/

58. van Delft P, Akay A, Huber SM, Bueschl C, Rudolph KLM, Di Domenico T, et al. The Profile and Dynamics of RNA Modifications in Animals. ChemBioChem. 2017;18(11):979–84.

59. Licht K, Jantsch MF. Rapid and dynamic transcriptome regulation by RNA editing and RNA modifications. Journal of Cell Biology. 2016 Apr 4;213(1):15–22.

60. Boo SH, Kim YK. The emerging role of RNA modifications in the regulation of mRNA stability. Experimental & Molecular Medicine. 2020 Mar 1;52(3):400–8.

61. Kumar S, Mohapatra T. Deciphering Epitranscriptome: Modification of mRNA Bases Provides a New Perspective for Post-transcriptional Regulation of Gene Expression. Frontiers in Cell and Developmental Biology [Internet]. 2021 [cited 2022 May 15];9. Available from: https://www.frontiersin.org/article/10.3389/fcell.2021.628415

62. Penning A, Jeschke J, Fuks F. Why novel mRNA modifications are so challenging and what we can do about it. Nat Rev Mol Cell Biol. 2022 Jun;23(6):385–6.

63. Destefanis E, Avşar G, Groza P, Romitelli A, Torrini S, Pir P, et al. A mark of disease: how mRNA modifications shape genetic and acquired pathologies. RNA. 2021 Apr;27(4):367–89.

64. Jonkhout N, Tran J, Smith MA, Schonrock N, Mattick JS, Novoa EM. The RNA modification landscape in human disease. RNA. 2017 Dec;23(12):1754–69.

65. Hsu PJ, Shi H, He C. Epitranscriptomic influences on development and disease. Genome Biology. 2017 Oct 23;18(1):197.

66. Angelova MT, Dimitrova DG, Dinges N, Lence T, Worpenberg L, Carré C, et al. The Emerging Field of Epitranscriptomics in Neurodevelopmental and Neuronal Disorders. Front Bioeng Biotechnol. 2018;6:46.

67. Ben-Haim MS, Moshitch-Moshkovitz S, Rechavi G. FTO: linking m6A demethylation to adipogenesis. Cell Res. 2015 Jan;25(1):3–4.

68. Zhao X, Yang Y, Sun BF, Shi Y, Yang X, Xiao W, et al. FTO-dependent demethylation of N6-methyladenosine regulates mRNA splicing and is required for adipogenesis. Cell Res. 2014 Dec;24(12):1403–19.

69. Wang X, Li Z, Kong B, Song C, Cong J, Hou J, et al. Reduced m6A mRNA methylation is correlated with the progression of human cervical cancer. Oncotarget. 2017 Nov 17;8(58):98918–30.

70. Li H, Chen L, Han Y, Zhang F, Wang Y, Han Y, et al. The Identification of RNA Modification Gene PUS7 as a Potential Biomarker of Ovarian Cancer. Biology (Basel). 2021 Nov 3;10(11):1130.

71. Cui Q, Yin K, Zhang X, Ye P, Chen X, Chao J, et al. Targeting PUS7 suppresses tRNA pseudouridylation and glioblastoma tumorigenesis. Nat Cancer. 2021 Sep;2(9):932–49.

72. Cui Q, Shi H, Ye P, Li L, Qu Q, Sun G, et al. m6A RNA Methylation Regulates the Self-Renewal and Tumorigenesis of Glioblastoma Stem Cells. Cell Rep. 2017 Mar 14;18(11):2622–34.

73. Shaheen R, Tasak M, Maddirevula S, Abdel-Salam GMH, Sayed ISM, Alazami AM, et al. PUS7 mutations impair pseudouridylation in humans and cause intellectual disability and microcephaly. Hum Genet. 2019 Mar;138(3):231–9.

74. Naseer MI, Abdulkareem AA, Jan MM, Chaudhary AG, Alharazy S, AlQahtani MH. Next generation sequencing reveals novel homozygous frameshift in PUS7 and splice acceptor variants in AASS gene leading to intellectual disability, developmental delay, dysmorphic feature and microcephaly. Saudi J Biol Sci. 2020 Nov;27(11):3125–31.

75. Lin TY, Smigiel R, Kuzniewska B, Chmielewska JJ, Kosińska J, Biela M, et al. Destabilization of mutated human PUS3 protein causes intellectual disability. Hum Mutat. 2022 Sep 20;

76. Nøstvik M, Kateta SM, Schönewolf-Greulich B, Afenjar A, Barth M, Boschann F, et al. Clinical and molecular delineation of PUS3-associated neurodevelopmental disorders. Clin Genet. 2021 Nov;100(5):628–33.

77. Du T, Rao S, Wu L, Ye N, Liu Z, Hu H, et al. An association study of the m6A genes with major depressive disorder in Chinese Han population. J Affect Disord. 2015 Sep 1;183:279–86.

78. Charette M, Gray MW. Pseudouridine in RNA: what, where, how, and why. IUBMB Life. 2000 May;49(5):341–51.

79. Blaby IK, Majumder M, Chatterjee K, Jana S, Grosjean H, de Crécy-Lagard V, et al. Pseudouridine formation in archaeal RNAs: The case of Haloferax volcanii. RNA. 2011 Jul;17(7):1367–80.

80. Cohn WE, Volkin E. Nucleoside-5′-Phosphates from Ribonucleic Acid. Nature. 1951 Mar;167(4247):483–4.

81. Davis FF, Allen FW. Ribonucleic acids from yeast which contain a fifth nucleotide. J Biol Chem. 1957 Aug;227(2):907–15.

82. Cohn WE. 5-Ribosyl uracil, a carbon-carbon ribofuranosyl nucleoside in ribonucleic acids. Biochimica et Biophysica Acta. 1959 Jan 1;32:569–71.

83. Cohn WE. Pseudouridine, a Carbon-Carbon Linked Ribonucleoside in Ribonucleic Acids: Isolation, Structure, and Chemical Characteristics. Journal of Biological Chemistry. 1960 May 1;235(5):1488–98.

84. Yu CT, Allen FW. Studies of an isomer of uridine isolated from ribonucleic acids. Biochimica et Biophysica Acta. 1959 Jan 1;32:393–406.

85. Scannell JP, Crestfield AM, Allen FW. Methylation studies on various uracil derivatives and on an isomer of uridine isolated from ribonucleic acids. Biochimica et Biophysica Acta. 1959 Jan 1;32:406–12.

86. Pan T. Modifications and functional genomics of human transfer RNA. Cell Res. 2018 Apr;28(4):395–404.

87. Penzo M, Montanaro L. Turning Uridines around: Role of rRNA Pseudouridylation in Ribosome Biogenesis and Ribosomal Function. Biomolecules. 2018 Jun 5;8(2):38.

88. Karijolich J, Yu YT. Spliceosomal snRNA modifications and their function. RNA Biology. 2010 Mar 1;7(2):192–204.

89. Yu B, Chen X. Analysis of miRNA Modifications. In: Meyers BC, Green PJ, editors. Plant MicroRNAs: Methods and Protocols [Internet]. Totowa, NJ: Humana Press; 2010 [cited 2022 Oct 27]. p. 137–48. (Methods in Molecular Biology). Available from: https://doi.org/10.1007/978-1-60327-005-2_10

90. Kazimierczyk M, Wrzesinski J. Long Non-Coding RNA Epigenetics. Int J Mol Sci. 2021 Jun 7;22(11).

91. Davis DR. Biophysical and Conformational Properties of Modified Nucleosides in RNA (Nuclear Magnetic Resonance Studies). In: Modification and Editing of RNA [Internet]. John Wiley & Sons, Ltd; 1998 [cited 2022 Oct 18]. p. 85–102. Available from: https://onlinelibrary.wiley.com/doi/abs/10.1128/9781555818296.ch5

92. Neumann JM, Bernassau JM, Guéron M, Tran-Dinh S. Comparative conformations of uridine and pseudouridine and their derivatives. Eur J Biochem. 1980 Jul;108(2):457–63.

93. Davis DR, Veltri CA, Nielsen L. An RNA model system for investigation of pseudouridine stabilization of the codon-anticodon interaction in tRNALys, tRNAHis and tRNATyr. J Biomol Struct Dyn. 1998 Jun;15(6):1121–32.

94. Auffinger P, Westhof E. Effects of Pseudouridylation on tRNA Hydration and Dynamics: a Theoretical Approach. In: Modification and Editing of RNA [Internet]. John Wiley & Sons, Ltd; 1998 [cited 2022 Oct 18]. p. 103–12. Available from: https://onlinelibrary.wiley.com/doi/abs/10.1128/9781555818296.ch6

95. Arnez JG, Steitz TA. Crystal structure of unmodified tRNA(Gln) complexed with glutaminyl-tRNA synthetase and ATP suggests a possible role for pseudo-uridines in stabilization of RNA structure. Biochemistry. 1994 Jun 21;33(24):7560–7.

96. Yarian CS, Basti MM, Cain RJ, Ansari G, Guenther RH, Sochacka E, et al. Structural and functional roles of the N1- and N3-protons of psi at tRNA's position 39. Nucleic Acids Res. 1999 Sep 1;27(17):3543–9.

97. Durant PC, Davis DR. Stabilization of the anticodon stem-loop of tRNALys,3 by an A+-C base-pair and by pseudouridine. J Mol Biol. 1999 Jan 8;285(1):115–31.

98. Davis DR. Stabilization of RNA stacking by pseudouridine. Nucleic Acids Res. 1995 Dec 25;23(24):5020–6.

99. Davis DR, Poulter CD. 1H-15N NMR studies of Escherichia coli tRNA(Phe) from hisT mutants: a structural role for pseudouridine. Biochemistry. 1991 Apr 30;30(17):4223–31.

100. Spenkuch F, Motorin Y, Helm M. Pseudouridine: still mysterious, but never a fake (uridine)! RNA Biol. 2014;11(12):1540–54.

101. Lirussi L, Demir Ö, You P, Sarno A, Amaro RE, Nilsen H. RNA Metabolism Guided by RNA Modifications: The Role of SMUG1 in rRNA Quality Control. Biomolecules. 2021 Jan 8;11(1).

102. Tagel M, Ilves H, Leppik M, Jürgenstein K, Remme J, Kivisaar M. Pseudouridines of tRNA Anticodon Stem-Loop Have Unexpected Role in Mutagenesis in Pseudomonas sp. Microorganisms. 2020 Dec 23;9(1).

103. Robertus JD, Ladner JE, Finch JT, Rhodes D, Brown RS, Clark BF, et al. Structure of yeast phenylalanine tRNA at 3 A resolution. Nature. 1974 Aug 16;250(467):546–51.

104. Kim SH, Sussman JL, Suddath FL, Quigley GJ, McPherson A, Wang AH, et al. The general structure of transfer RNA molecules. Proc Natl Acad Sci U S A. 1974 Dec;71(12):4970–4.

105. Grosjean H, Sprinzl M, Steinberg S. Posttranscriptionally modified nucleosides in transfer RNA: their locations and frequencies. Biochimie. 1995;77(1–2):139–41.

106. Hopper AK, Phizicky EM. tRNA transfers to the limelight. Genes Dev. 2003 Jan 15;17(2):162–80.

107. Roovers M, Droogmans L, Grosjean H. Post-Transcriptional Modifications of Conserved Nucleotides in the T-Loop of tRNA: A Tale of Functional Convergent Evolution. Genes. 2021 Feb;12(2):140.

108. Pf A. Decoding the genome: a modified view. Nucleic acids research [Internet]. 2004 Jan 9 [cited 2022 Oct 18];32(1). Available from: https://pubmed.ncbi.nlm.nih.gov/14715921/

109. Agris PF, Narendran A, Sarachan K, Väre VYP, Eruysal E. The Importance of Being Modified: The Role of RNA Modifications in Translational Fidelity. Enzymes. 2017;41:1–50.

110. Ranjan N, Rodnina MV. tRNA wobble modifications and protein homeostasis. Translation. 2016 Jan 2;4(1):e1143076.

111. Khonsari B, Klassen R. Impact of Pus1 Pseudouridine Synthase on Specific Decoding Events in Saccharomyces cerevisiae. Biomolecules. 2020 May 7;10(5).

112. Cochella L, Green R. An active role for tRNA in decoding beyond codon:anticodon pairing. Science. 2005 May 20;308(5725):1178–80.

113. Olejniczak M, Dale T, Fahlman RP, Uhlenbeck OC. Idiosyncratic tuning of tRNAs to achieve uniform ribosome binding. Nat Struct Mol Biol. 2005 Sep;12(9):788–93.

114. Taoka M, Nobe Y, Yamaki Y, Sato K, Ishikawa H, Izumikawa K, et al. Landscape of the complete RNA chemical modifications in the human 80S ribosome. Nucleic Acids Research. 2018 Oct 12;46(18):9289–98.

115. Ofengand J, Fournier MJ. The Pseudouridine Residues of rRNA: Number, Location, Biosynthesis, and Function. In: Modification and Editing of RNA [Internet]. John Wiley & Sons, Ltd; 1998 [cited 2022 Oct 18]. p. 229–53. Available from: https://onlinelibrary.wiley.com/doi/abs/10.1128/9781555818296.ch12

116. Bakin AV, Ofengand J. Mapping of Pseudouridine Residues in RNA to Nucleotide Resolution. In: Martin R, editor. Protein Synthesis: Methods and Protocols [Internet]. Totowa, NJ: Springer New York; 1998 [cited 2022 Oct 18]. p. 297–309. (Methods in Molecular Biology). Available from: https://doi.org/10.1385/0-89603-397-X:297

117. Ofengand J, Malhotra A, Remme J, Gutgsell NS, Del Campo M, Jean-Charles S, et al. Pseudouridines and pseudouridine synthases of the ribosome. Cold Spring Harb Symp Quant Biol. 2001;66:147–59.

118. Szymanski M, Barciszewska MZ, Barciszewski J, Erdmann VA. 5S ribosomal RNA database Y2K. Nucleic Acids Res. 2000 Jan 1;28(1):166–7.

119. Cantara WA, Crain PF, Rozenski J, McCloskey JA, Harris KA, Zhang X, et al. The RNA Modification Database, RNAMDB: 2011 update. Nucleic Acids Res. 2011 Jan;39(Database issue):D195-201.

120. Piekna-Przybylska D, Decatur WA, Fournier MJ. The 3D rRNA modification maps database: with interactive tools for ribosome analysis. Nucleic Acids Res. 2008 Jan;36(Database issue):D178–83.

121. de Crécy-Lagard V, Ross RL, Jaroch M, Marchand V, Eisenhart C, Brégeon D, et al. Survey and Validation of tRNA Modifications and Their Corresponding Genes in Bacillus subtilis sp Subtilis Strain 168. Biomolecules. 2020 Jun 30;10(7).

122. Decatur WA, Fournier MJ. rRNA modifications and ribosome function. Trends Biochem Sci. 2002 Jul;27(7):344–51.

123. Natchiar SK, Myasnikov AG, Kratzat H, Hazemann I, Klaholz BP. Visualization of chemical modifications in the human 80S ribosome structure. Nature. 2017 Nov 23;551(7681):472–7.

124. Natchiar SK, Myasnikov AG, Hazemann I, Klaholz BP. Visualizing the Role of 2'-OH rRNA Methylations in the Human Ribosome Structure. Biomolecules. 2018 Oct 25;8(4):125.

125. Leppik M, Liiv A, Remme J. Random pseuoduridylation in vivo reveals critical region of Escherichia coli 23S rRNA for ribosome assembly. Nucleic Acids Res. 2017 Jun 2;45(10):6098–108.

126. Sumita M, Desaulniers JP, Chang YC, Chui HMP, Clos L, Chow CS. Effects of nucleotide substitution and modification on the stability and structure of helix 69 from 28S rRNA. RNA. 2005 Sep;11(9):1420–9.

127. Barozzi C, Zacchini F, Asghar S, Montanaro L. Ribosomal RNA Pseudouridylation: Will Newly Available Methods Finally Define the Contribution of This Modification to Human Ribosome Plasticity? Frontiers in Genetics [Internet]. 2022 [cited 2022 Oct 18];13. Available from: https://www.frontiersin.org/articles/10.3389/fgene.2022.920987

128. O'Connor M, Leppik M, Remme J. Pseudouridine-Free Escherichia coli Ribosomes. J Bacteriol. 2018 Feb 15;200(4).

129. Zhao Y, Rai J, Yu H, Li H. CryoEM structures of pseudouridine-free ribosome suggest impacts of chemical modifications on ribosome conformations. Structure. 2022 Jul 7;30(7):983-992.e5.

130. Chikne V, Doniger T, Rajan KS, Bartok O, Eliaz D, Cohen-Chalamish S, et al. A pseudouridylation switch in rRNA is implicated in ribosome function during the life cycle of Trypanosoma brucei. Sci Rep. 2016 May 4;6(1):25296.

131. Massenet S, Mougin A, Branlant C. Posttranscriptional Modifications in the U Small Nuclear RNAs. In: Modification and Editing of RNA [Internet]. John Wiley & Sons, Ltd; 1998 [cited 2022 Oct 18]. p. 201–27. Available from: https://onlinelibrary.wiley.com/doi/abs/10.1128/9781555818296.ch11

132. Massenet S, Branlant C. A limited number of pseudouridine residues in the human atac spliceosomal UsnRNAs as compared to human major spliceosomal UsnRNAs. RNA. 1999 Nov;5(11):1495–503.

133. Berglund JA, Rosbash M, Schultz SC. Crystal structure of a model branchpoint-U2 snRNA duplex containing bulged adenosines. RNA. 2001 May;7(5):682–91.

134. Newby MI, Greenbaum NL. Sculpting of the spliceosomal branch site recognition motif by a conserved pseudouridine. Nat Struct Biol. 2002 Dec;9(12):958–65.

135. Newby MI, Greenbaum NL. A conserved pseudouridine modification in eukaryotic U2 snRNA induces a change in branch-site architecture. RNA. 2001 Jun;7(6):833–45.

136. Adachi H, Yu YT. Insight into the mechanisms and functions of spliceosomal snRNA pseudouridylation. World J Biol Chem. 2014 Nov 26;5(4):398–408.

137. Newby MI, Greenbaum NL. Investigation of Overhauser effects between pseudouridine and water protons in RNA helices. Proc Natl Acad Sci U S A. 2002 Oct 1;99(20):12697–702.

138. Kierzek E, Malgowska M, Lisowiec J, Turner DH, Gdaniec Z, Kierzek R. The contribution of pseudouridine to stabilities and structure of RNAs. Nucleic Acids Res. 2014 Mar;42(5):3492–501.

139. Fernández IS, Ng CL, Kelley AC, Wu G, Yu YT, Ramakrishnan V. Unusual base pairing during the decoding of a stop codon by the ribosome. Nature. 2013 Aug 1;500(7460):107–10.

140. Karijolich J, Yu YT. Converting nonsense codons into sense codons by targeted pseudouridylation. Nature. 2011 Jun 15;474(7351):395–8.

141. Svidritskiy E, Madireddy R, Korostelev AA. Structural Basis for Translation Termination on a Pseudouridylated Stop Codon. J Mol Biol. 2016 May 22;428(10 Pt B):2228–36.

142. Karikó K, Muramatsu H, Welsh FA, Ludwig J, Kato H, Akira S, et al. Incorporation of pseudouridine into mRNA yields superior nonimmunogenic vector with increased translational capacity and biological stability. Mol Ther. 2008 Nov;16(11):1833–40.

143. Hoernes TP, Clementi N, Faserl K, Glasner H, Breuker K, Lindner H, et al. Nucleotide modifications within bacterial messenger RNAs regulate their translation and are able to rewire the genetic code. Nucleic Acids Res. 2016 Jan 29;44(2):852–62.

144. Hoernes TP, Heimdörfer D, Köstner D, Faserl K, Nußbaumer F, Plangger R, et al. Eukaryotic Translation Elongation is Modulated by Single Natural Nucleotide Derivatives in the Coding Sequences of mRNAs. Genes (Basel). 2019 Jan 25;10(2):E84.

145. Eyler DE, Franco MK, Batool Z, Wu MZ, Dubuke ML, Dobosz-Bartoszek M, et al. Pseudouridinylation of mRNA coding sequences alters translation. Proceedings of the National Academy of Sciences. 2019 Nov 12;116(46):23068–74.

146. Hamma T, Ferré-D'Amaré AR. Pseudouridine Synthases. Chemistry & Biology. 2006 Nov 1;13(11):1125–35.

147. De Zoysa MD, Yu YT. Posttranscriptional RNA Pseudouridylation. Enzymes. 2017;41:151–67.

148. Rintala-Dempsey AC, Kothe U. Eukaryotic stand-alone pseudouridine synthases – RNA modifying enzymes and emerging regulators of gene expression? null. 2017 Sep 2;14(9):1185–96.

149. Kiss T, Fayet-Lebaron E, Jády BE. Box H/ACA small ribonucleoproteins. Mol Cell. 2010 Mar 12;37(5):597–606.

150. Kiss DJ, Oláh J, Tóth G, Varga M, Stirling A, Menyhárd DK, et al. The Structure-Derived Mechanism of Box H/ACA Pseudouridine Synthase Offers a Plausible Paradigm for Programmable RNA Editing. ACS Catal. 2022 Mar 4;12(5):2756–69.

151. Czekay DP, Kothe U. H/ACA Small Ribonucleoproteins: Structural and Functional Comparison Between Archaea and Eukaryotes. Frontiers in Microbiology [Internet]. 2021 [cited 2022 Oct 27];12. Available from: https://www.frontiersin.org/articles/10.3389/fmicb.2021.654370

152. Mochizuki Y, He J, Kulkarni S, Bessler M, Mason PJ. Mouse dyskerin mutations affect accumulation of telomerase RNA and small nucleolar RNA, telomerase activity, and ribosomal RNA processing. Proc Natl Acad Sci U S A. 2004 Jul 20;101(29):10756–61.

153. Meier UT. The many facets of H/ACA ribonucleoproteins. Chromosoma. 2005 May;114(1):1–14.

154. Garus A, Autexier C. Dyskerin: an essential pseudouridine synthase with multifaceted roles in ribosome biogenesis, splicing, and telomere maintenance. RNA. 2021 Dec;27(12):1441–58.

155. Machado-Pinilla R, Carrillo J, Manguan-Garcia C, Sastre L, Mentzer A, Gu BW, et al. Defects in mTR stability and telomerase activity produced by the Dkc1 A353V mutation in dyskeratosis congenita are rescued by a peptide from the dyskerin TruB domain. Clin Transl Oncol. 2012 Oct;14(10):755–63.

156. Ratnasamy V, Navaneethakrishnan S, Sirisena ND, Grüning NM, Brandau O, Thirunavukarasu K, et al. Dyskeratosis congenita with a novel genetic variant in the DKC1 gene: a case report. BMC Med Genet. 2018 May 25;19(1):85.

157. Thumati NR, Zeng XL, Au HHT, Jang CJ, Jan E, Wong JMY. Severity of X-linked dyskeratosis congenita (DKCX) cellular defects is not directly related to dyskerin (DKC1) activity in ribosomal RNA biogenesis or mRNA translation. Hum Mutat. 2013 Dec;34(12):1698–707.

158. Ley B, Torgerson DG, Oldham JM, Adegunsoye A, Liu S, Li J, et al. Rare Protein-Altering Telomere-related Gene Variants in Patients with Chronic Hypersensitivity Pneumonitis. Am J Respir Crit Care Med. 2019 Nov 1;200(9):1154–63.

159. Elsharawy KA, Mohammed OJ, Aleskandarany MA, Hyder A, El-Gammal HL, Abou-Dobara MI, et al. The nucleolar-related protein Dyskerin pseudouridine synthase 1 (DKC1) predicts poor prognosis in breast cancer. Br J Cancer. 2020 Nov;123(10):1543–52.

160. Watanabe M, Yamamoto G, Fujiyoshi K, Akagi Y, Kakuta M, Nishimura Y, et al. Development of metachronous rectal cancers in a young man with dyskeratosis congenita: a case report. J Med Case Rep. 2019 Apr 27;13(1):117.

161. Ko E, Kim JS, Ju S, Seo HW, Chang Y, Kang JA, et al. Oxidatively Modified Protein-Disulfide Isomerase-Associated 3 Promotes Dyskerin Pseudouridine Synthase 1-Mediated Malignancy and Survival of Hepatocellular Carcinoma Cells. Hepatology. 2018 Nov;68(5):1851–64.

162. Simos G, Tekotte H, Grosjean H, Segref A, Sharma K, Tollervey D, et al. Nuclear pore proteins are involved in the biogenesis of functional tRNA. EMBO J. 1996 May 1;15(9):2270–84.

163. Lahry K, Gopal A, Sahu AK, Marbaniang CN, Shah RA, Mehta A, et al. An Alternative Role of RluD in the Fidelity of Translation Initiation in Escherichia coli. J Mol Biol. 2022 Jun 30;434(12):167588.

164. Tagel M, Tavita K, Hõrak R, Kivisaar M, Ilves H. A novel papillation assay for the identification of genes affecting mutation rate in Pseudomonas putida and other pseudomonads. Mutat Res. 2016 Aug;790:41–55.

165. Jia Z, Meng F, Chen H, Zhu G, Li X, He Y, et al. Human TRUB1 is a highly conserved pseudouridine synthase responsible for the formation of Ψ55 in mitochondrial tRNAAsn, tRNAGln, tRNAGlu and tRNAPro. Nucleic Acids Res. 2022 Aug 26;50(16):9368–81.

166. Becker HF, Motorin Y, Planta RJ, Grosjean H. The yeast gene YNL292w encodes a pseudouridine synthase (Pus4) catalyzing the formation of psi55 in both mitochondrial and cytoplasmic tRNAs. Nucleic Acids Res. 1997 Nov 15;25(22):4493–9.

167. Mukhopadhyay S, Deogharia M, Gupta R. Mammalian nuclear TRUB1, mitochondrial TRUB2, and cytoplasmic PUS10 produce conserved pseudouridine 55 in different sets of tRNA. RNA. 2021 Jan 1;27(1):66–79.

168. Pickerill ES, Kurtz RP, Tharp A, Guerrero Sanz P, Begum M, Bernstein DA. Pseudouridine synthase 7 impacts Candida albicans rRNA processing and morphological plasticity. Yeast. 2019 Nov;36(11):669–77.

169. Kurimoto R, Chiba T, Ito Y, Matsushima T, Yano Y, Miyata K, et al. The tRNA pseudouridine synthase TruB1 regulates the maturation of let-7 miRNA. EMBO J. 2020 Oct 15;39(20):e104708.

170. Song J, Zhuang Y, Zhu C, Meng H, Lu B, Xie B, et al. Differential roles of human PUS10 in miRNA processing and tRNA pseudouridylation. Nat Chem Biol. 2020 Feb;16(2):160–9.

171. Guzzi N, Cieśla M, Ngoc PCT, Lang S, Arora S, Dimitriou M, et al. Pseudouridylation of tRNA-Derived Fragments Steers Translational Control in Stem Cells. Cell. 2018 May 17;173(5):1204-1216.e26.

172. Vicidomini R, Di Giovanni A, Petrizzo A, Iannucci LF, Benvenuto G, Nagel AC, et al. Loss of Drosophila pseudouridine synthase triggers apoptosis-induced proliferation and promotes cell-nonautonomous EMT. Cell Death Dis. 2015 Mar 26;6(3):e1705.

173. Jana S, Hsieh AC, Gupta R. Reciprocal amplification of caspase-3 activity by nuclear export of a putative human RNA-modifying protein, PUS10 during TRAIL-induced apoptosis. Cell Death Dis. 2017 Oct 5;8(10):e3093.

174. Keffer-Wilkes LC, Veerareddygari GR, Kothe U. RNA modification enzyme TruB is a tRNA chaperone. Proc Natl Acad Sci U S A. 2016 Dec 13;113(50):14306–11.

175. Jobert L, Skjeldam HK, Dalhus B, Galashevskaya A, Vågbø CB, Bjørås M, et al. The human base excision repair enzyme SMUG1 directly interacts with DKC1 and contributes to RNA quality control. Mol Cell. 2013 Jan 24;49(2):339–45.

176. Niu Y, Zheng Y, Zhu H, Zhao H, Nie K, Wang X, et al. The Arabidopsis Mitochondrial Pseudouridine Synthase Homolog FCS1 Plays Critical Roles in Plant Development. Plant Cell Physiol. 2022 Jul 14;63(7):955–66.

177. Lu S, Li C, Zhang Y, Zheng Z, Liu D. Functional Disruption of a Chloroplast Pseudouridine Synthase Desensitizes Arabidopsis Plants to Phosphate Starvation. Front Plant Sci. 2017;8:1421.

178. Lai SJ, Tu IF, Wu WL, Yang JT, Luk LYP, Lai MC, et al. Site-specific His/Asp phosphoproteomic analysis of prokaryotes reveals putative targets for drug resistance. BMC Microbiol. 2017 May 25;17(1):123.

179. Mangum JE, Hardee JP, Fix DK, Puppa MJ, Elkes J, Altomare D, et al. Pseudouridine synthase 1 deficient mice, a model for Mitochondrial Myopathy with Sideroblastic Anemia, exhibit muscle morphology and physiology alterations. Sci Rep. 2016 May 20;6:26202.

180. O'Connor M, Gregory ST. Inactivation of the RluD pseudouridine synthase has minimal effects on growth and ribosome function in wild-type Escherichia coli and Salmonella enterica. J Bacteriol. 2011 Jan;193(1):154–62.

181. Gu BW, Ge J, Fan JM, Bessler M, Mason PJ. Slow growth and unstable ribosomal RNA lacking pseudouridine in mouse embryonic fibroblast cells expressing catalytically inactive dyskerin. FEBS Lett. 2013 Jul 11;587(14):2112–7.

182. Song S, Wood TK. Persister cells resuscitate via ribosome modification by 23S rRNA pseudouridine synthase RluD. Environ Microbiol. 2020 Mar;22(3):850–7.

183. Song W, Ressl S, Tracey WD. Loss of Pseudouridine Synthases in the RluA Family Causes Hypersensitive Nociception in Drosophila. G3 (Bethesda). 2020 Dec 3;10(12):4425–38.

184. Tesarova M, Vondrackova A, Stufkova H, Veprekova L, Stranecky V, Berankova K, et al. Sideroblastic anemia associated with multisystem mitochondrial disorders. Pediatr Blood Cancer. 2019 Apr;66(4):e27591.

185. Daniels EG, Alders M, Lezzerini M, McDonald A, Peters M, Kuijpers TW, et al. A uniparental isodisomy event introducing homozygous pathogenic variants drives a multisystem metabolic disorder. Cold Spring Harb Mol Case Stud. 2019 Dec;5(6).

186. Woods J, Cederbaum S. Myopathy, lactic acidosis and sideroblastic anemia 1 (MLASA1): A 25-year follow-up. Mol Genet Metab Rep. 2019 Dec;21:100517.

187. Glioblastoma Is Dependent on tRNA Pseudouridylation. Cancer Discov. 2021 Oct;11(10):2367.

188. Song D, Guo M, Xu S, Song X, Bai B, Li Z, et al. HSP90-dependent PUS7 overexpression facilitates the metastasis of colorectal cancer cells by regulating LASP1 abundance. J Exp Clin Cancer Res. 2021 May 14;40(1):170.

189. Nombela P, Miguel-López B, Blanco S. The role of m6A, m5C and Ψ RNA modifications in cancer: Novel therapeutic opportunities. Molecular Cancer. 2021 Jan 18;20(1):18.

190. Zhang M, Zhao W, Liu S, Liu H, Liu L, Peng Q, et al. H/ACA snoRNP Gene Family as Diagnostic and Prognostic Biomarkers for Hepatocellular Carcinoma. Pharmgenomics Pers Med. 2021;14:1331–45.

191. Stockert JA, Weil R, Yadav KK, Kyprianou N, Tewari AK. Pseudouridine as a novel biomarker in prostate cancer. Urol Oncol. 2021 Jan;39(1):63–71.

192. Hu G, Wang R, Wei B, Wang L, Yang Q, Kong D, et al. Prognostic Markers Identification in Glioma by Gene Expression Profile Analysis. J Comput Biol. 2020 Jan;27(1):81–90.

193. Kan G, Wang Z, Sheng C, Chen G, Yao C, Mao Y, et al. Dual Inhibition of DKC1 and MEK1/2 Synergistically Restrains the Growth of Colorectal Cancer Cells. Adv Sci (Weinh). 2021 May;8(10):2004344.

194. Vieri M, Kirschner M, Tometten M, Abels A, Rolles B, Isfort S, et al. Comparable Effects of the Androgen Derivatives Danazol, Oxymetholone and Nandrolone on Telomerase Activity in Human Primary Hematopoietic Cells from Patients with Dyskeratosis Congenita. Int J Mol Sci. 2020 Sep 29;21(19).

195. Armando RG, Mengual Gómez DL, Juritz EI, Lorenzano Menna P, Gomez DE. Homology Model and Docking-Based Virtual Screening for Ligands of Human Dyskerin as New Inhibitors of Telomerase for Cancer Treatment. Int J Mol Sci. 2018 Oct 18;19(10).

196. Roovers M, Hale C, Tricot C, Terns MP, Terns RM, Grosjean H, et al. Formation of the conserved pseudouridine at position 55 in archaeal tRNA. Nucleic Acids Res. 2006;34(15):4293–301.

197. Fitzek E, Joardar A, Gupta R, Geisler M. Evolution of Eukaryal and Archaeal Pseudouridine Synthase Pus10. J Mol Evol. 2018 Jan;86(1):77–89.

198. Deogharia M, Mukhopadhyay S, Joardar A, Gupta R. The human ortholog of archaeal Pus10 produces pseudouridine 54 in select tRNAs where its recognition sequence contains a modified residue. RNA. 2019 Mar;25(3):336–51.

199. Wrzesinski J, Nurse K, Bakin A, Lane BG, Ofengand J. A dual-specificity pseudouridine synthase: an Escherichia coli synthase purified and cloned on the basis of its specificity for psi 746 in 23S RNA is also specific for psi 32 in tRNA(phe). RNA. 1995 Jun;1(4):437–48.

200. Conrad J, Niu L, Rudd K, Lane BG, Ofengand J. 16S ribosomal RNA pseudouridine synthase RsuA of Escherichia coli: deletion, mutation of the conserved Asp102 residue, and sequence comparison among all other pseudouridine synthases. RNA. 1999 Jun;5(6):751–63.

201. Bakin A, Kowalak JA, McCloskey JA, Ofengand J. The single pseudouridine residue in Escherichia coli 16S RNA is located at position 516. Nucleic Acids Res. 1994 Sep 11;22(18):3681–4.

202. Huang L, Pookanjanatavip M, Gu X, Santi DV. A conserved aspartate of tRNA pseudouridine synthase is essential for activity and a probable nucleophilic catalyst. Biochemistry. 1998 Jan 6;37(1):344–51.

203. Hoang C, Ferré-D'amaré AR. Crystal structure of the highly divergent pseudouridine synthase TruD reveals a circular permutation of a conserved fold. RNA. 2004 Jul 1;10(7):1026–33.

204. Foster PG, Huang L, Santi DV, Stroud RM. The structural basis for tRNA recognition and pseudouridine formation by pseudouridine synthase I. Nat Struct Biol. 2000 Jan;7(1):23–7.

205. Hoang C, Ferré-D'Amaré AR. Cocrystal structure of a tRNA Psi55 pseudouridine synthase: nucleotide flipping by an RNA-modifying enzyme. Cell. 2001 Dec 28;107(7):929–39.

206. Sivaraman J, Iannuzzi P, Cygler M, Matte A. Crystal structure of the RluD pseudouridine synthase catalytic module, an enzyme that modifies 23S rRNA and is essential for normal cell growth of Escherichia coli. J Mol Biol. 2004 Jan 2;335(1):87–101.

207. Hoang C, Chen J, Vizthum CA, Kandel JM, Hamilton CS, Mueller EG, et al. Crystal structure of pseudouridine synthase RluA: indirect sequence readout through protein-induced RNA structure. Mol Cell. 2006 Nov 17;24(4):535–45.

208. Motorin Y, Keith G, Simon C, Foiret D, Simos G, Hurt E, et al. The yeast tRNA:pseudouridine synthase Pus1p displays a multisite substrate specificity. RNA. 1998 Jul;4(7):856–69.

209. Anderson MZ, Brewer J, Singh U, Boothroyd JC. A pseudouridine synthase homologue is critical to cellular differentiation in Toxoplasma gondii. Eukaryot Cell. 2009 Mar;8(3):398–409.

210. Nakamoto MA, Lovejoy AF, Cygan AM, Boothroyd JC. mRNA pseudouridylation affects RNA metabolism in the parasite Toxoplasma gondii. RNA. 2017 Dec;23(12):1834–49.

211. Bykhovskaya Y, Casas K, Mengesha E, Inbal A, Fischel-Ghodsian N. Missense mutation in pseudouridine synthase 1 (PUS1) causes mitochondrial myopathy and sideroblastic anemia (MLASA). Am J Hum Genet. 2004 Jun;74(6):1303–8.

212. Oncul U, Unal-Ince E, Kuloglu Z, Teber-Tiras S, Kaygusuz G, Eminoglu FT. A Novel PUS1 Mutation in 2 Siblings with MLASA Syndrome: A Review of the Literature. J Pediatr Hematol Oncol. 2020 Apr 13;

213. Casas K, Bykhovskaya Y, Mengesha E, Wang D, Yang H, Taylor K, et al. Gene responsible for mitochondrial myopathy and sideroblastic anemia (MSA) maps to chromosome 12q24.33. Am J Med Genet A. 2004 May 15;127A(1):44–9.

214. Patton JR, Bykhovskaya Y, Mengesha E, Bertolotto C, Fischel-Ghodsian N. Mitochondrial myopathy and sideroblastic anemia (MLASA): missense mutation in the pseudouridine synthase 1 (PUS1) gene is associated with the loss of tRNA pseudouridylation. J Biol Chem. 2005 May 20;280(20):19823–8.

215. Zeharia A, Fischel-Ghodsian N, Casas K, Bykhocskaya Y, Tamari H, Lev D, et al. Mitochondrial myopathy, sideroblastic anemia, and lactic acidosis: an autosomal recessive syndrome in Persian Jews caused by a mutation in the PUS1 gene. J Child Neurol. 2005 May;20(5):449–52.

216. Fernandez-Vizarra E, Berardinelli A, Valente L, Tiranti V, Zeviani M. Nonsense mutation in pseudouridylate synthase 1 (PUS1) in two brothers affected by myopathy, lactic acidosis and sideroblastic anaemia (MLASA). BMJ Case Rep. 2009;2009.

217. Gu X, Yu M, Ivanetich KM, Santi DV. Molecular recognition of tRNA by tRNA pseudouridine 55 synthase. Biochemistry. 1998 Jan 6;37(1):339–43.

218. Zhu J, Gopinath K, Murali A, Yi G, Hayward SD, Zhu H, et al. RNA-binding proteins that inhibit RNA virus infection. Proc Natl Acad Sci U S A. 2007 Feb 27;104(9):3129–34.

219. Keffer-Wilkes LC, Soon EF, Kothe U. The methyltransferase TrmA facilitates tRNA folding through interaction with its RNA-binding domain. Nucleic Acids Res. 2020 Aug 20;48(14):7981–90.

220. Chan CM, Huang RH. Enzymatic characterization and mutational studies of TruD – the fifth family of pseudouridine synthases. Archives of Biochemistry and Biophysics. 2009 Sep 1;489(1):15–9.

221. Kaya Y, Ofengand J. A novel unanticipated type of pseudouridine synthase with homologs in bacteria, archaea, and eukarya. RNA. 2003 Jun;9(6):711–21.

222. Behm-Ansmant I, Urban A, Ma X, Yu YT, Motorin Y, Branlant C. The Saccharomyces cerevisiae U2 snRNA:pseudouridine-synthase Pus7p is a novel multisite-multisubstrate RNA:Psi-synthase also acting on tRNAs. RNA. 2003 Nov;9(11):1371–82.

223. Ma X, Zhao X, Yu YT. Pseudouridylation (Psi) of U2 snRNA in S. cerevisiae is catalyzed by an RNA-independent mechanism. EMBO J. 2003 Apr 15;22(8):1889–97.

224. Wu G, Xiao M, Yang C, Yu YT. U2 snRNA is inducibly pseudouridylated at novel sites by Pus7p and snR81 RNP. EMBO J. 2011 Jan 5;30(1):79–89.

225. Carlile TM, Rojas-Duran MF, Gilbert WV. Transcriptome-Wide Identification of Pseudouridine Modifications Using Pseudo-seq. Curr Protoc Mol Biol. 2015 Oct 1;112:4.25.1-4.25.24.

226. Martinez NM, Gilbert WV. Investigating Pseudouridylation Mechanisms by High-Throughput in Vitro RNA Pseudouridylation and Sequencing. Methods Mol Biol. 2021;2298:379–97.

227. Campos CM, Tsai K, Courtney DG, Bogerd HP, Holley CL, Cullen BR. Mapping of pseudouridine residues on cellular and viral transcripts using a novel antibody-based technique. RNA. 2021 Nov 1;27(11):1400–11.

228. Martinez NM, Schaening-Burgos C, Gilbert WV. Pseudouridine site assignment by high-throughput in vitro RNA pseudouridylation and sequencing. Methods Enzymol. 2021;658:277–310.

229. Begik O, Lucas MC, Pryszcz LP, Ramirez JM, Medina R, Milenkovic I, et al. Quantitative profiling of pseudouridylation dynamics in native RNAs with nanopore sequencing. Nat Biotechnol. 2021 Oct;39(10):1278–91.

230. Carlile TM, Martinez NM, Schaening C, Su A, Bell TA, Zinshteyn B, et al. mRNA structure determines modification by pseudouridine synthase 1. Nat Chem Biol. 2019 Oct;15(10):966–74.

231. Pan H, Agarwalla S, Moustakas DT, Finer-Moore J, Stroud RM. Structure of tRNA pseudouridine synthase TruB and its RNA complex: RNA recognition through a combination of rigid docking and induced fit. Proc Natl Acad Sci U S A. 2003 Oct 28;100(22):12648–53.

232. Hoang C, Hamilton CS, Mueller EG, Ferré-D'Amaré AR. Precursor complex structure of pseudouridine synthase TruB suggests coupling of active site perturbations to an RNA-sequestering peripheral protein domain. Protein Sci. 2005 Aug;14(8):2201–6.

233. Kaya Y, Campo MD, Ofengand J, Malhotra A. Crystal Structure of TruD, a Novel Pseudouridine Synthase with a New Protein Fold *. Journal of Biological Chemistry. 2004 Apr 30;279(18):18107–10.

234. Ericsson UB, Nordlund P, Hallberg BM. X-ray structure of tRNA pseudouridine synthase TruD reveals an inserted domain with a novel fold. FEBS Lett. 2004 May 7;565(1–3):59–64.

235. Chaudhuri BN, Chan S, Perry LJ, Yeates TO. Crystal structure of the apo forms of psi 55 tRNA pseudouridine synthase from Mycobacterium tuberculosis: a hinge at the base of the catalytic cleft. J Biol Chem. 2004 Jun 4;279(23):24585–91.

236. Czudnochowski N, Wang AL, Finer-Moore J, Stroud RM. In human pseudouridine synthase 1 (hPus1), a C-terminal helical insert blocks tRNA from binding in the same orientation as in the Pus1 bacterial homologue TruA, consistent with their different target selectivities. J Mol Biol. 2013 Oct 23;425(20):3875–87.

237. Bank RPD. RCSB PDB - 5KKP: Crystal Structure of Human Pseudouridylate Synthase 7 [Internet]. [cited 2021 Oct 21]. Available from: https://www.rcsb.org/structure/5KKP

238. Guegueniat J, Halabelian L, Zeng H, Dong A, Li Y, Wu H, et al. The human pseudouridine synthase PUS7 recognizes RNA with an extended multi-domain binding surface. Nucleic Acids Res. 2021 Oct 28;49(20):11810–22.

239. Hur S, Stroud RM, Finer-Moore J. Substrate recognition by RNA 5-methyluridine methyltransferases and pseudouridine synthases: a structural perspective. J Biol Chem. 2006 Dec 22;281(51):38969–73.

240. Aravind L, Koonin EV. THUMP--a predicted RNA-binding domain shared by 4-thiouridine, pseudouridine synthases and RNA methylases. Trends Biochem Sci. 2001 Apr;26(4):215–7.

241. McCleverty CJ, Hornsby M, Spraggon G, Kreusch A. Crystal structure of human Pus10, a novel pseudouridine synthase. J Mol Biol. 2007 Nov 9;373(5):1243–54.

242. Grünberg S, Doyle LA, Dai N, Corrêa IR, Yigit E, Stoddard BL. The structural basis of mRNA recognition and binding by eukaryotic pseudouridine synthase PUS1 [Internet]. bioRxiv; 2021 [cited 2022 Sep 8]. p. 2021.12.08.471817. Available from: https://www.biorxiv.org/content/10.1101/2021.12.08.471817v1

243. Dong X, Bessho Y, Shibata R, Nishimoto M, Shirouzu M, Kuramitsu S, et al. Crystal structure of tRNA pseudouridine synthase TruA from Thermus thermophilus HB8. RNA Biol. 2006 Jul;3(3):115–22.

244. Hur S, Stroud RM. How U38, 39, and 40 of many tRNAs become the targets for pseudouridylation by TruA. Mol Cell. 2007 Apr 27;26(2):189–203.

245. Grosjean H, Edqvist J, Stråby KB, Giegé R. Enzymatic formation of modified nucleosides in tRNA: dependence on tRNA architecture. J Mol Biol. 1996 Jan 12;255(1):67–85.

246. Becker HF, Motorin Y, Sissler M, Florentz C, Grosjean H. Major identity determinants for enzymatic formation of ribothymidine and pseudouridine in the T psi-loop of yeast tRNAs. J Mol Biol. 1997 Dec 1;274(4):505–18.

247. Safra M, Nir R, Farouq D, Slutzkin IV, Schwartz S. TRUB1 is the predominant pseudouridine synthase acting on mammalian mRNA via a predictable and conserved code. Genome Res. 2017 Mar;27(3):393–406.

248. Tillault AS, Schultz SK, Wieden HJ, Kothe U. Molecular Determinants for 23S rRNA Recognition and Modification by the E. coli Pseudouridine Synthase RluE. J Mol Biol. 2018 Apr 27;430(9):1284–94.

249. Hellmuth K, Grosjean H, Motorin Y, Deinert K, Hurt E, Simos G. Cloning and characterization of the Schizosaccharomyces pombe tRNA:pseudouridine synthase Pus1p. Nucleic Acids Res. 2000 Dec 1;28(23):4604–10.

250. Deryusheva S, Gall JG. Dual nature of pseudouridylation in U2 snRNA: Pus1p-dependent and Pus1p-independent activities in yeasts and higher eukaryotes. RNA. 2017 Jul;23(7):1060–7.

251. Arluison V, Hountondji C, Robert B, Grosjean H. Transfer RNA−Pseudouridine Synthetase Pus1 of Saccharomyces cerevisiae Contains One Atom of Zinc Essential for Its Native Conformation and tRNA Recognition. Biochemistry. 1998 May 1;37(20):7268–76.

252. Arluison V, Buckle M, Grosjean H. Pseudouridine synthetase Pus1 of Saccharomyces cerevisiae: kinetic characterisation, tRNA structural requirement and real-time analysis of its complex with tRNA. J Mol Biol. 1999 Jun 11;289(3):491–502.

253. Sibert BS, Fischel-Ghodsian N, Patton JR. Partial activity is seen with many substitutions of highly conserved active site residues in human Pseudouridine synthase 1. RNA. 2008 Sep;14(9):1895–906.

254. Kräutler B, Ostermann S. Structure, reactions and functions of B12 and B12-proteins. In: Kadish KM, Smith KM, Guilard R, editors. The porphyrin handbook. Oxford: Elsevier Science; 2003.

255. Brown KL. Chemistry and Enzymology of Vitamin B12. Chem Rev. 2005 Jun 1;105(6):2075–150.

256. Lindblom K. Celebrating the History of B Vitamin Research With ACS [Internet]. ACS Axial | ACS Publications. 2016 [cited 2022 Nov 2]. Available from: https://axial.acs.org/2016/12/02/history-b-vitamin-research/

257. Harris, Steven. Vitamin B_{12}. In: Wikipedia [Internet]. 2022 [cited 2022 Nov 2]. Available from: https://en.wikipedia.org/w/index.php?title=Vitamin_B12&oldid=1116830388

258. Rickes EL, Brink NG, Koniuszy FR, Wood TR, Folkers K. Crystalline vitamin B12. Science [Internet]. 1948;107. Available from: https://doi.org/10.1126/science.107.2781.396

259. Smith EL, Parker LFJ. Purification of anti-pernicious anaemia factor. Biochem J. 1948;43.

260. Kräutler B. Organometallic chemistry of b(12) coenzymes. Met Ions Life Sci. 2009;6:1–51.

261. Eschenmoser A. Vitamin-B12 – Experiments concerning the origin of its molecular structure. Angew Chem Int Ed [Internet]. 1988;27. Available from: https://doi.org/10.1002/anie.198800051

262. Kräutler B. B12 coenzymes, the central theme. In: Kräutler B, Arigoni D, Golding BT, editors. Vitamin B12 and B12 proteins [Internet]. Weinheim: Wiley-VCH; 1998. Available from: https://doi.org/10.1002/9783527612192

263. Banerjee R. Chemistry and Biochemistry of B12. John Wiley & Sons; 1999. 960 p.

264. Giedyk M, Goliszewska K, Gryko D. Vitamin B12 catalysed reactions. Chem Soc Rev. 2015 Jun 7;44(11):3391–404.

265. Hodgkin DC, Pickworth J, Robertson JH, Trueblood KN, Prosen RJ, White JG. Crystal structure of the hexacarboxylic acid derived from B12 and the molecular structure of the vitamin. Nature [Internet]. 1955;176. Available from: https://doi.org/10.1038/176325a0

266. Hodgkin DC, Kamper J, Mackay M, Pickworth J, Trueblood KN, White JG. Structure of Vitamin-B12. Nature [Internet]. 1956;178. Available from: https://doi.org/10.1038/178064a0

267. Lenhert PG, Hodgkin DC. Structure of the 5,6-Dimethylbenzimidazolylcobamide Coenzyme. Nature. 1961 Dec;192(4806):937–8.

268. Kratky C, Kräutler B. Molecular structure of B12 cofactors and other B12 derivatives. In: Banerjee R, editor. Chemistry and biochemistry of B12. New York: Wiley; 1999.

269. Randaccio L, Geremia S, Nardin G, Würges J. X-ray structural chemistry of cobalamins. Coord Chem Rev [Internet]. 2006;250. Available from: https://doi.org/10.1016/j.ccr.2005.12.005

270. Ellenbogen L, Cooper BA. Vitamin B12. In: Machlin LJ, editor. Handbook of vitamins, nutritional and clinical aspects Food Science and Technology. New York: Marcel Dekker; 1984.

271. Dolphin, D., Johnson, A. W.: Chem. Commun. 495 (1965).

272. Friedrich W. Vitamins. Berlin: Walter de Gruyter; 1988.

273. Kieninger C, Deery E, Lawrence AD, Podewitz M, Wurst K, Nemoto-Smith E, et al. The Hydrogenobyric Acid Structure Reveals the Corrin Ligand as an Entatic State Module Empowering B12 Cofactors for Catalysis. Angewandte Chemie International Edition. 2019;58(31):10756–60.

274. Drennan CL, Huang S, Drummond JT, Matthews RG, Ludwig ML. How a Protein Binds B12: A 3.0 Å X-Ray Structure of B12-Binding Domains of Methionine Synthase. Science. 1994 Dec 9;266(5191):1669–74.

275. Mancia F, Keep NH, Nakagawa A, Leadlay PF, McSweeney S, Rasmussen B, et al. How coenzyme B12 radicals are generated: the crystal structure of methylmalonyl-coenzyme A mutase at 2 A resolution. Structure. 1996 Mar 15;4(3):339–50.

276. Brown KL, Zou X, Banka RR, Perry CB, Marques HM. Solution structure and thermolysis of Coβ-5′-deoxyadenosylimidazolylcobamide, a coenzyme B12 analogue with an imidazole axial nucleoside. Inorg Chem [Internet]. 2004;43. Available from: https://doi.org/10.1021/ic040079z

277. Kräutler B, Konrat R, Stupperich E, Färber G, Gruber K, Kratky C. Direct evidence for the conformational deformation of the corrin ring by the nucleotide base in vitamin-B12 – synthesis and solution spectroscopic and crystal-structure analysis of Coβ-cyanoimidazolylcobamide. Inorg Chem [Internet]. 1994;33. Available from: https://doi.org/10.1021/ic00096a043

278. Fasching M, Schmidt W, Kräutler B, Stupperich E, Schmidt A, Kratky C. Coα-(1H-imidazolyl)-Coβ-methylcob(III)amide: model for protein-bound corrinoid cofactors. Helv Chim Acta [Internet]. 2000;83. Available from: https://doi.org/3.0.CO;2-T

279. Kräutler B. Biochemistry of B12-cofactors in human metabolism. Subcell Biochem. 2012;56:323–46.

280. Pratt JM. Inorganic chemistry of vitamin B12. New York: Academic; 1972.

281. Lexa D, Savéant JM. The electrochemistry of vitamin-B12. Acc Chem Res [Internet]. 1983;16. Available from: https://doi.org/10.1021/ar00091a001

282. Kräutler B. Electrochemistry and organometallic electrochemical synthesis. In: Banerjee R, editor. Chemistry and biochemistry of B12. New York: Wiley; 1999.

283. Halpern J. Mechanisms of coenzyme B12-dependent rearrangements. Science [Internet]. 1985;227. Available from: https://doi.org/10.1126/science.2857503

284. Endicott JF, Netzel TL. Early events and transient chemistry in the photohomolysis of alkylcobalamins. J Am Chem Soc [Internet]. 1979;101. Available from: https://doi.org/10.1021/ja00508a066

285. Matthews RG. Cobalamin-Dependent Methyltransferases. Acc Chem Res. 2001 Aug 1;34(8):681–9.

286. Schrauzer GN, Deutsch E. Reactions of cobalt(I) supernucleophiles. The alkylation of vitamin B12s, cobaloximes(I), and related compounds. J Am Chem Soc. 1969 Jun 1;91(12):3341–50.

287. Kräutler B, Caderas C. Complementary diastereoselective cobalt methylations of the Vitamin-B12 derivative cobester. Helv Chim Acta [Internet]. 1984;67. Available from: https://doi.org/10.1002/hlca.19840670727

288. Kräutler B. Thermodynamic trans-effects of the nucleotide base in the B12 coenzymes. Helv Chim Acta [Internet]. 1987;70. Available from: https://doi.org/10.1002/hlca.19870700506

289. Schrauzer GN, Deutsch E, Windgassen RJ. The nucleophilicity of vitamin B12. J Am Chem Soc. 1968 Apr 24;90(9):2441–2.

290. Koutmos M, Pejchal R, Bomer TM, Matthews RG, Smith JL, Ludwig ML. Metal active site elasticity linked to activation of homocysteine in methionine synthases. Proceedings of the National Academy of Sciences. 2008 Mar 4;105(9):3286–91.

291. Peariso K, Goulding CW, Huang S, Matthews RG, Penner-Hahn JE. Characterization of the Zinc Binding Site in Methionine Synthase Enzymes of Escherichia coli: The Role of Zinc in the Methylation of Homocysteine. J Am Chem Soc. 1998 Aug 1;120(33):8410–6.

292. Goulding CW, Matthews RG. Cobalamin-Dependent Methionine Synthase from Escherichia coli: Involvement of Zinc in Homocysteine Activation. Biochemistry. 1997 Dec 1;36(50):15749–57.

293. Banerjee RV, Frasca V, Ballou DP, Matthews RG. Participation of Cob(I)alamin in the reaction catalyzed by methionine synthase from Escherichia coli: a steady-state and rapid reaction kinetic analysis. Biochemistry. 1990 Dec 18;29(50):11101–9.

294. Matthews RG, Koutmos M, Datta S. Cobalamin-dependent and cobamide-dependent methyltransferases. Curr Opin Struct Biol. 2008 Dec;18(6):658–66.

295. Matthews RG. Cobalamin-dependent methionine synthase. In: Banerjee R, editor. Chemistry and biochemistry of B12. New York: Wiley; 1999.

296. Drummond JT, Huang S, Blumenthal RM, Matthews RG. Assignment of enzymic function to specific protein regions of cobalamin-dependent methionine synthase from Escherichia coli. Biochemistry. 1993 Sep 1;32(36):9290–5.

297. Fujii K, Huennekens FM. Activation of methionine synthetase by a reduced triphosphopyridine nucleotide-dependent flavoprotein system. J Biol Chem. 1974 Nov 10;249(21):6745–53.

298. Taylor RT, Weissbach H. N5-methyltetrahydrofolate-homocysteine transmethylase. Role of S-adenosylmethionine in vitamin B12-dependent methionine synthesis. J Biol Chem. 1967 Apr 10;242(7):1517–21.

299. Foster MA, Dilworth MJ, Woods DD. COBALAMIN AND THE SYNTHESIS OF METHIONINE BY ESCHERICHIA COLI. Nature. 1964 Jan 4;201:39–42.

300. Banerjee RV, Harder SR, Ragsdale SW, Matthews RG. Mechanism of reductive activation of cobalamin-dependent methionine synthase: an electron paramagnetic resonance spectroelectrochemical study. Biochemistry. 1990 Feb 6;29(5):1129–35.

301. Goulding CW, Postigo D, Matthews RG. Cobalamin-dependent methionine synthase is a modular protein with distinct regions for binding homocysteine, methyltetrahydrofolate, cobalamin, and adenosylmethionine. Biochemistry. 1997 Jul 1;36(26):8082–91.

302. Jarrett JT, Huang S, Matthews RG. Methionine Synthase Exists in Two Distinct Conformations That Differ in Reactivity toward Methyltetrahydrofolate, Adenosylmethionine, and Flavodoxin†. Biochemistry. 1998 Apr 21;37(16):5372–82.

303. Dixon MM, Huang S, Matthews RG, Ludwig M. The structure of the C-terminal domain of methionine synthase: presenting S-adenosylmethionine for reductive methylation of B12. Structure. 1996 Nov 15;4(11):1263–75.

304. Banerjee RV, Johnston NL, Sobeski JK, Datta P, Matthews RG. Cloning and sequence analysis of the Escherichia coli metH gene encoding cobalamin-dependent methionine synthase and isolation of a tryptic fragment containing the cobalamin-binding domain. J Biol Chem. 1989 Aug 15;264(23):13888–95.

305. Drennan CL, Dixon MM, Hoover DM, Jarrett JT, Goulding CW, Matthews RG, et al. Cobalamin-dependent methionine synthase from Escherichia coli: structure and reactivity. In: Kräutler B, Arigoni D, Golding BT, editors. Vitamin B12 and B12-proteins [Internet]. Wiley-VCH: Weinheim; 1998. Available from: https://doi.org/10.1002/9783527612192.ch07

306. Evans JC, Huddler DP, Hilgers MT, Romanchuk G, Matthews RG, Ludwig ML. Structures of the N-terminal modules imply large domain motions during catalysis by methionine synthase. Proc Natl Acad Sci U S A. 2004 Mar 16;101(11):3729–36.

307. Fleischhacker AS, Matthews RG. Ligand Trans Influence Governs Conformation in Cobalamin-Dependent Methionine Synthase. Biochemistry. 2007 Oct 1;46(43):12382–92.

308. Bandarian V, Ludwig ML, Matthews RG. Factors modulating conformational equilibria in large modular proteins: A case study with cobalamin-dependent methionine synthase. Proc Natl Acad Sci U S A. 2003 Jul 8;100(14):8156–63.

309. Garrow TA. Purification, kinetic properties, and cDNA cloning of mammalian betaine-homocysteine methyltransferase. J Biol Chem. 1996 Sep 13;271(37):22831–8.

310. Kallen RG, Jencks WP. The Dissociation Constants of Tetrahydrofolic Acid. Journal of Biological Chemistry. 1966 Dec 25;241(24):5845–50.

311. Hogenkamp HPC, Bratt GT, Sun S. Methyl transfer from methylcobalamin to thiols – a reinvestigation. Biochemistry [Internet]. 1985;24. Available from: https://doi.org/10.1021/bi00344a018

312. Banerjee R. The Yin-Yang of cobalamin biochemistry. Chem Biol [Internet]. 1997;4. Available from: https://doi.org/10.1016/S1074-5521(97)90286-6

313. Banerjee R. Radical peregrinations catalyzed by coenzyme B12-dependent enzymes. Biochemistry. 2001 May 29;40(21):6191–8.

314. Marsh ENG, Drennan CL. Adenosylcobalamin-dependent isomerases: new insights into structure and mechanism. Curr Opin Chem Biol [Internet]. 2001;5. Available from: https://doi.org/10.1016/S1367-5931(00)00238-6

315. Rétey J. Stereospecifity of the coenzyme B12-catalyzed rearrangements and the role of negative catalysis. In: Banerjee R, editor. Chemistry and biochemistry of B12. New York: Wiley; 1999.

316. Finke RG. Coenzyme B12-based chemical precedent for Co–C bond homolysis and other key elementary step. In: Kräutler B, Arigoni D, Golding BT, editors. Vitamin B12 and B12-proteins [Internet]. Weinheim: Wiley-VCH; 1998. Available from: https://doi.org/10.1002/9783527612192.ch25

317. Reitzer R, Gruber K, Jogl G, Wagner UG, Bothe H, Buckel W, et al. Glutamate mutase from Clostridium cochlearium: the structure of a coenzyme B12-dependent enzyme provides new mechanistic insights. Structure [Internet]. 1999;7. Available from: https://doi.org/10.1016/S0969-2126(99)80116-6

318. Marsh EN, Holloway DE. Cloning and sequencing of glutamate mutase component S from Clostridium tetanomorphum. Homologies with other cobalamin-dependent enzymes. FEBS Lett. 1992 Sep 28;310(2):167–70.

319. Hay BP, Finke RG. Thermolysis of the CoóC bond in adenosylcorrins.3. quantification of the axial base effect in adenosylcobalamin by the synthesis and thermolysis of axial base-free adenosylcobinamideóinsights into the energetics of enzyme-assisted cobalt carbon bond homolysis. J Am Chem Soc [Internet]. 1987;109. Available from: https://doi.org/10.1021/ja00260a011

320. Kräutler B, Keller W, Kratky C. Coenzyme B12-chemistry: the crystal and molecular structure of Cob(II)alamin. J Am Chem Soc [Internet]. 1989;111. Available from: https://doi.org/10.1021/ja00206a037

321. Toraya T. Radical catalysis in coenzyme B12-dependent isomerization (eliminating) reactions. Chem Rev [Internet]. 2003;103. Available from: https://doi.org/10.1021/cr020428b

322. Gschösser S, Hannak RB, Konrat R, Gruber K, Mikl C, Kratky C, et al. Homocoenzyme B12 and bishomocoenzyme B12, covalent structural mimics for homolyzed, enzyme-bound coenzyme B12. Chem Eur J [Internet]. 2005;11. Available from: https://doi.org/10.1002/chem.200400701

323. Padmakumar R, Banerjee R. Evidence that cobalt-carbon bond homolysis is coupled to hydrogen atom abstraction from substrate in methylmalonyl-CoA mutase. Biochemistry [Internet]. 1997;36. Available from: https://doi.org/10.1021/bi962503g

324. Chowdhury S, Banerjee R. Evidence for quantum mechanical tunneling in the coupled cobalt-carbon bond homolysis-substrate radical generation reaction catalyzed by

methylmalonyl-CoA mutase. J Am Chem Soc [Internet]. 2000;122. Available from: https://doi.org/10.1021/ja994302g

325. Marsh ENG, Ballou DP. Coupling of cobalt-carbon bond homolysis and hydrogen atom abstraction in adenosylcobalamin-dependent glutamate mutase. Biochemistry [Internet]. 1998;37. Available from: https://doi.org/10.1021/bi980512e

326. Bandarian V, Reed GH. Isotope effects in the transient phases of the reaction catalyzed by ethanolamine ammonia-lyase: determination of the number of exchangeable hydrogens in the enzyme-cofactor complex. Biochemistry. 2000 Oct 3;39(39):12069–75.

327. Banerjee R, Ragsdale SW. The many faces of vitamin B12: catalysis by cobalamin-dependent enzymes. Annu Rev Biochem. 2003;72:209–47.

328. Banerjee R, Chowdhury S. Methylmalonyl-CoA mutase. In: Banerjee R, editor. Chemistry and biochemistry of B12. New York: Wiley; 1999.

329. Evans PR, Mancia F. Insights on the reaction mechanism of methylmalonyl-CoA mutase from the crystal structure. In: Kräutler B, Golding BT, Arigoni D, editors. Vitamin B12 and B12 proteins [Internet]. Weinheim: Wiley-VCH; 1998. Available from: https://doi.org/10.1002/9783527612192.ch13

330. Ludwig ML, Evans PR. X-ray crystallography of B12 enzymes: methylmalonyl-CoA mutase and methionine synthase. In: Banerjee R, editor. Chemistry and biochemistry of B12. New York: Wiley; 1999.

331. Tollinger M, Eichmüller C, Konrat R, Huhta MS, Marsh ENG, Kräutler B. The B12-Binding Subunit of Glutamate Mutase from Clostridium tetanomorphum Traps the Nucleotide Moiety of Coenzyme B12. Journal of Molecular Biology. 2001 Jun 8;309(3):777–91.

332. Tollinger M, Konrat R, Hilbert BH, Marsh EN, Kräutler B. How a protein prepares for B12 binding: structure and dynamics of the B12-binding subunit of glutamate mutase from Clostridium tetanomorphum. Structure. 1998 Aug 15;6(8):1021–33.

CHAPTER 2 Pseudouridine Synthase 7 is an Opportunistic Enzyme that Binds and Modifies Substrates with Diverse Sequences and Structures

Pseudouridine (Ψ) is a ubiquitous RNA modification incorporated by pseudouridine synthase (Pus) enzymes into hundreds of noncoding and protein-coding RNA substrates. Here, we determined the contributions of substrate structure and protein sequence to binding and catalysis by pseudouridine synthase 7 (Pus7), one of the principal messenger RNA (mRNA) modifying enzymes. Pus7 is distinct among the eukaryotic Pus proteins because it modifies a wider variety of substrates and shares limited homology with other Pus family members. We solved the crystal structure of *Saccharomyces cerevisiae* Pus7, detailing the architecture of the eukaryotic-specific insertions thought to be responsible for the expanded substrate scope of Pus7. Additionally, we identified an insertion domain in the protein that fine-tunes Pus7 activity both *in vitro* and in cells. These data demonstrate that Pus7 preferentially binds substrates possessing the previously identified UGUAR (R = purine) consensus sequence and that RNA secondary structure is not a

---

This chapter was adapted with permission from Purchal, M. K.; Eyler, D. E.; Tardu, M.; Franco, M. K.; Korn, M. M.; Khan, T.; McNassor, R.; Giles, R.; Lev, K.; Sharma, H.; Monroe, J.; Mallik, L.; Koutmos, M.; Koutmou, K. S. Pseudouridine Synthase 7 Is an Opportunistic Enzyme That Binds and Modifies Substrates with Diverse Sequences and Structures. *Proc Natl Acad Sci U S A* **2022**, *119* (4), e2109708119.

Author contributions: M.P. performed structural characterization of Pus7 (Figures 2.1, 2.2A, 2.5 and SI Figures 2.2-2.4, SI Table 2.1), EMSA binding assays (Table 2.1 and Figures 2.3, 2.4 and SI Figure 2.8, SI Tables 2.2, and 2.3), CLAP assays (2.6B, C and SI Figure 2.12, 2.13). D.E. performed tritium-release activity assays (Figures 2.2B, 2.4, 2.5, 2.6A and Table 2.1 and SI Table 2.4). M.T. performed phylogenetic analysis (SI Figure 2.1), ribosome foot printing (SI Figure 2.7), and cell growth/*in vivo* experiments (SI Figures 2.5, 2.6, 2.14, 2.15). M.K.F., J.M., T.K., R.M., M.M.K., R.G., K.L., H.S., and L.M. contributed to this work, specifically stopped-flow, thermal-stability, and *in vivo*. Data Analysis and writing: M.P., D.E., M.K., and K.S.K.

strong requirement for Pus7-binding. In contrast, the rate constants and extent of Ψ incorporation are more influenced by RNA structure, with Pus7 modifying UGUAR sequences in less-structured contexts more efficiently both in vitro and in cells. Although less-structured substrates were preferred, Pus7 fully modified every transfer RNA, mRNA, and nonnatural RNA containing the consensus recognition sequence that we tested. Our findings suggest that Pus7 is a promiscuous enzyme and lead us to propose that factors beyond inherent enzyme properties (e.g., enzyme localization, RNA structure, and competition with other RNA-binding proteins) largely dictate Pus7 substrate selection.

2.1 Introduction

Posttranscriptional modifications to the four standard RNA nucleosides increase the structural and functional complexity of RNAs. The C5-glycosidic isomer of uridine, pseudouridine (Ψ), is incorporated into multiple RNA species including transfer RNAs (tRNAs), ribosomal RNAs (rRNAs), and eukaryotic messenger RNAs (mRNAs) (Figure 2.1A). While Ψ has been studied in the context of noncoding RNAs for decades, the significance of Ψ in mRNAs is less well understood. Like all mRNA modifications, Ψ has the potential to affect every step in the life cycle of an mRNA (1). In line with this idea, mRNA pseudouridylation has been implicated as a regulator of human alternative splicing, yeast protein synthesis, and toxoplasma mRNA metabolism (2–4). Nonetheless, despite its prevalence, how cells select mRNAs for Ψ modification and the impacts of individual mRNA Ψ-sites on biological processes remain to be established.

Ψ is installed into RNAs in all organisms by pseudouridine synthases (Pus). This large class of enzymes is categorized into six families: TruA, TruB, RluA, RsuA, TruD, and Pus10 (SI Figure 2.1). Despite the observation that many Pus enzymes are not required for cellular viability in non-stressed conditions, wild-type cells outcompete Pus-deficient cells, suggesting that these enzymes confer a fitness advantage (5–7). Consistent with this, Pus proteins markedly enhance cellular fitness under heat shock (8–11). Furthermore, mutations in the human TruA, TruB, and TruD family members (Pus1, Pus4, and Pus7, respectively) are linked to inherited diseases such as mitochondrial myopathy and sideroblastic anemia and intellectual disabilities (12–15). In addition to their enzymatic roles, there is emerging evidence that some Pus proteins can have alternative functions in the cell—acting as a tRNA folding chaperones or prions (16–18). One of the next horizons for the Pus field will be deconvoluting the contributions of each of the varied Pus activities (noncoding RNA– and mRNA-modifying or nonenzymatic) to gene expression.

Bacterial Pus family members have been studied for decades, and their structures, chemical mechanisms, and modes of RNA target selection are well characterized (19,20). However, it is unclear why some eukaryotic Pus enzymes exhibit an increased substrate scope, modifying hundreds of mRNAs in addition to their traditional noncoding targets (21,22). The enzymes responsible for catalyzing the bulk of Ψ incorporation into eukaryotic mRNAs are Pus1, Pus4, and Pus7 (23,24). Recent transcriptome-wide Ψ-mapping and reporter studies indicate that mRNA secondary structure and sequence contribute to substrate selection by human Pus1 and Pus4 (25,26). Pus7 is distinct from Pus1 and 4, as it shares ∼10% sequence identity with these enzymes and incorporates Ψ into a particularly diverse set of RNAs (tRNA, tRNA fragments, small nuclear RNA [snRNA], rRNA, and mRNA) (23,24,27–29). Pus7 has been shown to preferentially incorporate Ψ into UGUAR (U = U converted to Ψ, R= A or G) consensus sequences (23,24,28). However, this sequence motif does not explain how Pus7 selects its substrates, because only a limited set (<5%) of the UGUAR sequences present in RNAs are Pus7 targets.

To establish the determinants of Pus7 substrate selection, we first solved a crystal structure of *Saccharomyces cerevisiae* Pus7. Comparing the yeast and human Pus7 structures with their prokaryotic TruD counterparts, we noticed a large eukaryotic-specific insertion (insertion domain I, ID-I) sitting atop the active site in two distinct conformations. Both of the observed conformations for this domain are well positioned to potentially interact with incoming RNA substrates. Our biochemical and cell-based studies of Pus7 enzymes lacking ID-I ($Pus7_{\Delta IDI}$) support this supposition. We also evaluated the impact of RNA sequence and structure on Pus7 substrate binding and modification. Our results indicate that Pus7 preferentially associates with RNAs containing a UGUAR consensus sequence but only minimally discriminates between binding different RNA structures. The enzyme has a greater degree of selectivity during catalysis

than binding, with the rate constant and extent of pseudouridylation increasing on RNAs with less predicted-structure both *in vitro* and in cells. Nonetheless, despite these preferences, Pus7 modifies all of the UGUAR containing substrates that we presented it *in vitro*. Our observations lead us to propose an opportunistic model for Pus7 substrate selection. In this model, Pus7 rapidly samples RNAs, binding and modifying any UGUAR sequence that it can access for long enough. This model is supported by observations that UGUAR sequences in structured contexts are modified more rapidly and efficiently at elevated temperatures, when RNA structures are destabilized. These findings extend the current framework for thinking about Pus enzymes by proposing that protein compartmentalization, local RNA structure, and RNA–protein interactions are among the most-substantive determinants for Pus7 substrate discrimination.

2.2 Results

2.2.1 *S. cerevisiae* Pus7 structure reveals flexibility in the architecture of the eukaryotic insertion domains

Pus7, like other TruD family members, shares little sequence identity with the four other Pus families (TruA, TruB, RluA, and RsuA) (30). Available structures of TruD from *Escherichia coli* (Protein Data Bank [PDB]: 1SI7, 1ZSW, and 1SB7), *Methanosarcina mazei* (PDB: 1Z2Z), and Pus7 from humans (PDB: 5KKP) show that the core of the enzyme has a V-shape formed by two conserved domains: the catalytic PUS domain and the TRUD domain (31–33). The structure of the PUS domain is highly conserved across all pseudouridine synthases (34), with the invariant catalytic aspartate falling on a β-strand that lies across the center of this domain. The TRUD domain is a mixed αβ fold conserved across all TruD homologs (30–33). This domain is oriented adjacent to the PUS domain, and together they form a catalytic cleft (V) lined with positively charged residues. The TRUD and PUS domains are connected by an extensive loop-rich region at the base of the cleft where the active site sits. Structural comparison of the five biological assemblies present in the three *E. coli* TruD structures demonstrates the ability of these core domains to flex toward each other by 18° (32). This motion could be facilitated by the hinge-like, loop-rich region connecting the domains and potentially help the enzyme to "clamp" down on RNA substrates.

Archaeal and eukaryotic TruD homologs contain large insertions at specific sites in the TRUD domain (33). Eukaryotic homologs possess two additional unique insertions decorating the catalytic domain (31–33). The largest of these eukaryotic-specific insertions, ID-I, is a small domain first visualized in the structure of human Pus7 (PDB: 5KKP). This domain is conserved in yeast Pus7 (SI Figure 2.2) (30). Eukaryotic members of the TruD family target a much-larger range

of RNAs than their bacterial counterparts, which only modify U13 on $tRNA^{Glu}$ (30). Given their conservation and the expanded substrate selection observed, it stands to reason that structural features (insertions) unique to the eukaryotic enzymes might play a crucial role in the recognition and modification of their additional substrates.

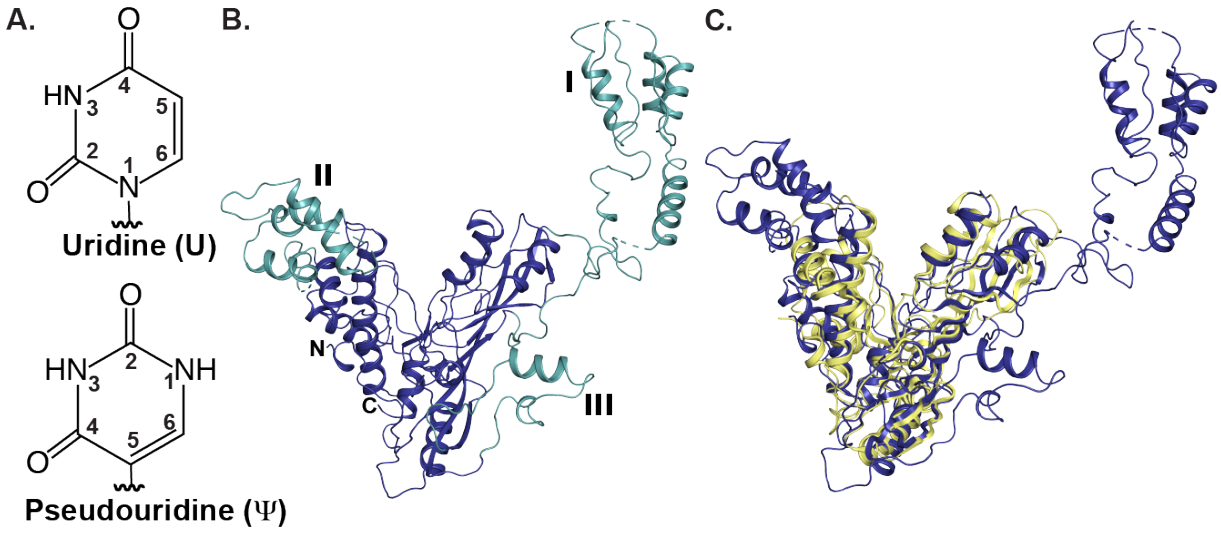


Figure 2.1: *S. cerevisiae* Pus7 structure.

(A) Uridine and pseudouridine. **(B)** X-ray structure of Pus7 at 3.2-Å resolution (PDB: 7MZV). The structurally conserved, V-shaped enzyme core housing the PUS and TRUD domains (blue). The three eukaryotic-specific insertions (green) are numbered I through III. **(C)** Superimposition of the *S. cerevisiae* Pus7 (blue) and *E. coli* TruD (yellow, PDB: 1SB7) structures demonstrating the structural conservation of the enzyme's catalytic core.

To that end, we sought to develop a structural basis for understanding how eukaryotic insertions can contribute to Pus7 substrate selection and modification. We used a combination of molecular replacement (using human Pus7 as a model) and single-wavelength anomalous dispersion phasing to obtain a crystal structure of *S. cerevisiae* Pus7 to 3.2 Å resolution (Figure 2.1B, SI Figure 2.2, and SI Table 2.1) (PDB: 7MZV). The core of the Pus7 structure strongly resembles that of TruD (RMSD = 3.74 A˚ based on 144 atoms), adopting the characteristic V-shape conformation of the PUS and TRUD domains (Figure 2.1C and SI Figure 2.2). The active site sits near the bottom of the cleft adjacent to the hinge region where the two domains interface. The universally conserved catalytic aspartate (D256) resides on a loop between β3 and β4, is

surrounded by residues conserved within the TruD family (K61, F67, E71, N305, and F307) (SI Figure 2.3) and sits ~4 Å closer to the conserved active site residues in Pus7 than in TruD (SI Figure 2.2). Such positioning is consistent with the location of the catalytic Asp in the cocrystal structure of TruB with an RNA (PDB: 1K8W) (32). The most-notable structural difference between Pus7 and TruD is the presence of the three eukaryotic-specific insertions that decorate the enzyme's V-shaped core (ID-I [aa 75 to 215], II [Ins-II, aa 365 to 443], and III [Ins-III, aa 594 to 634]) (Figure 2.1B). These insertions are connected to the catalytic domain by flexible linkers.

A structural alignment of yeast and human Pus7 reveals that the three insertions adopt similar folds and are present at equivalent locations in both structures (SI Figure 2.2). ID-I folds into an extensive domain that is anchored to the top exterior (side opposite the active site) of the PUS domain by flexible linkers. ID-I is observed in two distinct orientations above the active site (SI Figure 2.2). Ins-III, the smallest insertion, is located on the exterior of the catalytic domain near the hinge region. Ins-II is a helical bundle perched atop the TRUD domain, and together, ID-I and Ins-II elongate the cleft between the PUS and TRUD domains. These insertions are ideally positioned to potentially interact with RNA substrates. Comparison of human and yeast Pus7 structures demonstrates that ID-I and Ins-II can adopt different orientations relative to the core of the protein. In yeast Pus7, ID-I extends laterally away from the protein, while in human Pus7, it is slightly rotated to be positioned directly above the active site. This suggests that ID-I can move as a rigid body, both swiveling laterally (akin to a flag on a pole) over the top of or away from the active site. Ins-II is also found in slightly different orientations in each structure, hinting to its rotational and translational freedom of motion. Ins-II in yeast Pus7 is pivoted away from the active site but angled in toward the catalytic cleft in human Pus7. Though the physiological relevance of these orientations is not immediately apparent, the freedom of motion and inherent flexibility could

enable Ins-II or ID-I to play a part in recognizing regions of the substrate distal to the site of modification.

To determine whether ID-I shares homology with known RNA-binding motifs, we performed a protein structure comparison search with the DALI server (35). Using ID-I from both the yeast and human Pus7 structures, our results indicate that ID-I shares strong structural homology with the RNA-binding R3H domains. The strongest match was to a Poly(A)specific ribonuclease domain (PDB: 2A1S, Z = 6.3, %ID =21) that utilizes an R3H motif to bind single-stranded nucleic acids. This motif contains an invariant arginine separated from a conserved histidine by three residues. Sequence alignments demonstrate that these residues are present and conserved in ID-I (SI Figure 2.3). Additionally, the structure of the R3H motif is characterized by two α-helices packed against a three-strand β-sheet, which aligns well with the structures of ID-I in yeast and human Pus7 (SI Figure 2.4). ID-I's predicted ability to interact with RNAs and its stark positional variance relative to the active site suggest that ID-I may serve as an attenuator of RNA-binding.

2.2.2 Pus7 enhances *S. cerevisiae* viability under translational stress

Pus7 enhances cellular fitness under temperature stress and modifies a wider variety of mRNA targets under heat shock (24). Given the recently discovered links between Pus7 mutations and neurological defects, we hypothesized that Pus7 is important for cellular health under additional stress conditions (12). To test this supposition, we compared the growth of wild-type and *pus7Δ S. cerevisiae* cells under 15 different conditions: multiple temperatures (22 °C, 30 °C, and 37 °C), elevated salt concentrations (NaCl and $MgCl_2$), varied pH (pH 4.5 and 8.5), carbon sources (glucose, sucrose, and galactose), proteosome stress (MG132), and translational stress (puromycin, paromomycin, cycloheximide, and hygromycin). Cell growth was assessed by spot-plating on

solid media and growth curves in liquid media (SI Figures 2.5 and 2.6). *pus7Δ* cells do not exhibit a growth defect relative to wild-type cells in YPD (yeast extract peptone dextrose) media at 22 °C or 30°C but are sensitive to increased temperature (37 °C) as previously reported (11). We did not observe any carbon source–dependent growth changes between the wild-type and the knockout cells. The *pus7Δ* strain has a slight sensitivity to high concentrations of NaCl but not $MgCl_2$. This is consistent with high-throughput studies that identified *pus7Δ* as one of ~300 yeast knockouts that are more sensitive than wild-type cells to hyperosmotic (1 M NaCl) stress (36).

Translation inhibitors had the largest impact on *pus7Δ* growth relative to wild-type cells. *pus7Δ* cells are more sensitive to puromycin, cycloheximide, and hygromycin and exhibit a decreased sensitivity to paromomycin (SI Figures 2.5 and 2.6). This makes sense as Pus7 modifies both tRNAs and mRNAs, which could impact translation. To test hypothesis, we analyzed available ribosome-profiling data sets for *pus7Δ* (37,38). We observe that ribosome occupancy is increased on Pus7 targeted mRNA codons in *pus7Δ* cells (SI Figure 2.7), consistent with recent reports that mRNA pseudouridylation slows translation elongation (3,39). Together, our results indicate that Pus7 is likely to be particularly important when cellular translation is under stress.

2.2.3 Conserved Pus7 active site residues enhance RNA modification

Our crystal structure revealed that the Pus7 active site is similar to that of TruD and suggests that ID-I might be positioned to contribute to enzyme function. We next wanted to test whether conserved Pus7 active site residues and ID-I enhance the ability of the enzyme to modify a reported mRNA substrate, CDC8 (24). To accomplish this, we measured the single-turnover rate constants (k_{obs}) for Ψ incorporation into a 61 nt–long region of CDC8 by wild-type and mutant enzymes (D256A, K61A, E71A, F67A, H161A, N305A, F307A, F307Y, and $Pus7_{\Delta ID\text{-}I}$) (Figure 2.2 and Table 2.1).

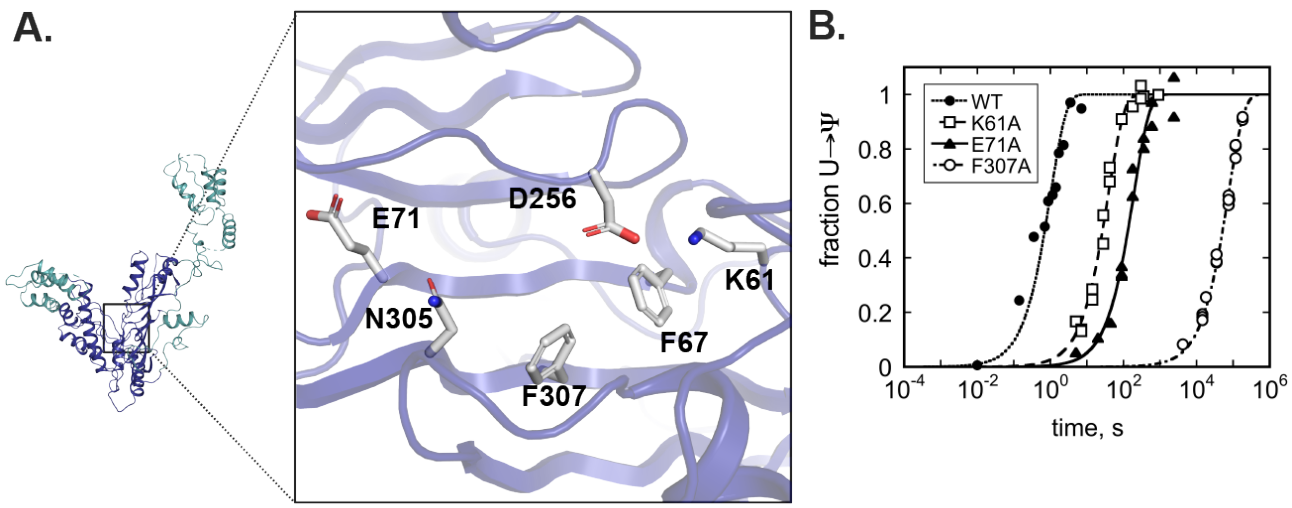


Figure 2.2: Pus7 active site residues enhance catalysis

(A) Conserved Pus7 active site residues investigated in this study**. (B)** Time courses for Ψ incorporation into a CDC8 mRNA by saturating concentrations of (●) wild-type Pus7 and Pus7 active site mutants ([□] K61A, [▲] E71A, and [○] F307A). The single-turnover rate constants (k_{obs}) for alanine substitutions of all of the residues displayed in A are reported in Table 2.1.

In these experiments, the Pus7 enzymes (2 to 10 μM) were each incubated with ^{3}H-labeled CDC8 RNA (<100 nM) and ^{3}H release upon conversion of U to Ψ was monitored at discrete time points (40). As expected, mutation of the catalytic D256 residue to alanine abolishes Pus7 activity, with no Ψ formation observed after 16 h (41). Alanine substitution of the nearby active site residues K61, F67, and E71 reduced the k_{obs} for CDC8 modification by 40- to 200-fold, consistent with the proposed role of these residues in substrate positioning during catalysis in TruD. Mutations to the conserved active site NxF motif had larger impacts on Pus7 activity, with N305A and F307A mutants decreasing k_{obs} by up to 50,000-fold relative to the wild-type enzyme. The F307A defect was partially rescued by a F307Y mutation (400-fold reduction in k_{obs}), suggesting that the F307 base stacks with the target uridine to enhance CDC8 modification. In contrast to the active site mutants, k_{obs} is unchanged by the ID-I point mutation H161A. However, removal of ID-I (Pus7$_{\Delta ID\text{-}I}$) reduced k_{obs} by twofold (SI Table 2.2). These data reveal that conserved active site residues are

important for Ψ modification by Pus7, while ID-I does not have a large influence on the rate determining step for Pus7 under saturating enzyme concentrations.

Table 2.1: Wild-type and mutant Pus7 single turnover rate constants and dissociation constants for CDC8

PUS7	k_{obs} (s^{-1})*	$K_{D,app1}$ (nM)†
WT	$9.9 \pm 1.0 \times 10^{-1}$	75 ± 15
K61A	$2.6 \pm 0.1 \times 10^{-2}$	180 ± 40
F67A	$4.6 \pm 0.2 \times 10^{-3}$	480 ± 50
E71A	$5.2 \pm 0.3 \times 10^{-3}$	210 ± 50
H161A	$6.9 \pm 0.9 \times 10^{-1}$‡	170 ± 40
D256A	N.A.§	60 ± 15
N305A	$4.0 \pm 0.1 \times 10^{-4}$	230 ± 60
F307A	$1.3 \pm 0.1 \times 10^{-5}$	340 ± 170
F307Y	$2.6 \pm 0.1 \times 10^{-3}$	N.D.¶
ΔID1	$3.8 \pm 0.6 \times 10^{-1}$‡	160 ± 40

*k_{obs} values were determined by tritium release assays using 2 μM PUS7 and <100 nM CDC8 substrate. At least three replicate curves were collected for each mutant.
†$K_{D,app1}$ values were determined by EMSAs ($n \geq 2$) using Pus7 containing the catalytically inactive D256A mutation in addition to the mutation indicated in the first column.
‡k_{obs} value determined using 10 μM Pus7 and <100 nM CDC8 substrate.
§N.A., no activity.
¶N.D., not determined.

2.2.4 Multiple Pus7 enzymes bind to unmodified and modified CDC8

To assess the contributions of active site residues and ID-I to RNA substrate-binding, we performed electrophoretic mobility shift assays (EMSAs) with a 5'-fluorescein labeled CDC8 and a series of catalytically inactive Pus7 mutants (D256A mutant background). We observed a single band shift at low enzyme concentrations (<50 nM), indicating the formation of a Pus7-CDC8 (ES) complex (Figure 2.3). For all of the mutants evaluated the singly bound species is supershifted when enzyme concentrations are increased (Figure 2.3 and SI Figure 2.8). These additional bands indicate the formation of complexes that have multiple Pus7 proteins associated with CDC8 (ESE_n

complexes). We speculate that these lower-affinity binding events reflect nonspecific interactions between Pus7 and the RNA phosphodiester backbone. This may be a general property of Pus enzymes, as we also observe a similar binding behavior for the Pus family member Pus1 interacting with CDC8 (SI Figure 2.9). For both enzymes, the transition from the 1:1 complex to the n+1:1 complex occurs over a very narrow concentration range, reflecting the large number of nonspecific binding sites available on each RNA. Based on these observations, we considered several binding models (SI Figure 2.10) and ultimately fit our data to a simplified mechanism in which Pus7 independently binds a unique site tightly ($K_{D,app1}$: low nM) and multiple additional sites with reduced affinities ($K_{D,app2}$: high nM–μM). Such a model is supported by stopped-flow studies with 5'-fluorescein–labeled CDC8 and unlabeled D256A Pus7. At low enzyme concentrations, we observed a single exponential phase, whose rate constant (k_{obs1}) is linearly dependent on enzyme concentration (SI Figure 2.11). As we increase enzyme concentration, a second phase emerged (k_{obs2}) consistent with our EMSAs indicating that multiple proteins bind to Pus7 RNA targets.

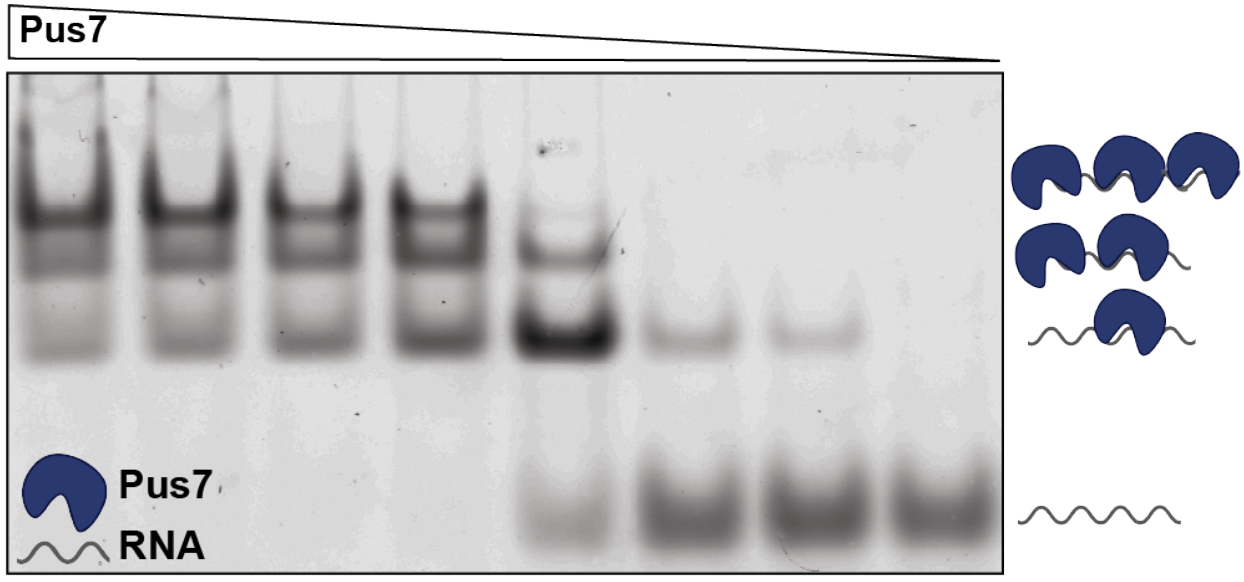


Figure 2.3: Multiple Pus7 proteins bind to CDC8 RNA

The association of increasing concentrations of catalytically inactive D256A Pus7 with limiting amounts of 50-fluorescein-labeled CDC8 visualized on a nondenaturing agarose gel. Increased concentrations of D256A resulted in multiple binding events.

We applied our binding model to obtain $K_{D,app1}$ values for D256A, D256A/K61A, D256A/F67A, D256A/E71A, D256A/ H161A, D256A/N305A, D256A/F307A, and D256A/Pus7$_{\Delta ID-I}$ binding to CDC8 (Table 2.1). D256A binds CDC8 tightly with $K_{D,app1}$ = 60 ± 15 nM, consistent with the K_D estimated for D256A from our stopped-flow assays ($k_{on,app}$ = 4.3 × 10^8 $M^{-1}s^{-1}$, $k_{off,app}$ = 35 s^{-1}, and $k_{off,app}/k_{on,app}$ = K_D ~85 nM) (SI Figure 2.10). Additional active site point mutations increased the $K_{D,app1}$ for Pus7 binding CDC8 by two- to eightfold relative to D256A (Table 2.1 and SI Figure 2.8). Similarly, removal of ID-I (D256A/Pus7$_{\Delta ID-I}$) increased the $K_{Dapp,1}$ for CDC8 by twofold. Many enzymes bind their products with different affinities than their substrates. To test whether Pus7 discriminates between unmodified and Ψ-modified transcripts, we measured the $K_{Dapp,1}$ values for Pus7 and Pus7$_{\Delta ID-I}$ binding to modified CDC8. We found that wild-type Pus7 bound Ψ-modified CDC8 sequences with an affinity similar to its catalytically inactive counterpart (D256A) for an unmodified CDC8 substrate (Table 2.1). In contrast, Pus7$_{\Delta ID-I}$ has an eightfold weaker affinity for a Ψ-modified CDC8 than D256A/Pus7$_{\Delta ID-I}$. This indicates that Pus7 does not distinguish between substrate or product during binding and that ID-I promotes this lack of discrimination.

2.2.5 ID-1 influences the extent of pseudouridine incorporation in full-length mRNA *in vitro* and in cells

Our binding studies suggest that ID-I has the potential to impact substrate selection. We tested this hypothesis by evaluating Ψ incorporation under conditions in which Pus7 must distinguish between all cellular RNAs. Pus7 and Pus7$_{\Delta ID-I}$ proteins (50 μM) were incubated for 10 min at 30 °C with 150 μg of total RNA purified from *pus7Δ* cells. The extent of Ψ incorporation on previously reported Pus7 modified substrates (ARG5,6, BET2, TEF5, RTC3$_{(U77)}$, RTC3$_{(U288)}$, TEF2$_{(U555)}$, and TEF2 $_{(U1104)}$) was measured using N-cyclohexyl-N0-β-(4-

methylmorpholinium)ethylcarbodiimide–reverse transcription and ligation-assisted PCR analysis of Ψ modification (CLAP) (42). The RNAs treated with Pus7 and Pus7$_{\Delta ID-I}$ exhibited a similar range of stoichiometries (6 to 40% versus 7 to 60%) (SI Figure 2.12). While the range of Ψ incorporation levels is similar, the presence of ID-I impacted the extent of Ψ addition on four of the seven substrates that we examined—with Pus7$_{\Delta ID-I}$ incorporating Ψ at lower levels than Pus7 in ARG5,6, BET2, and TEF2$_{(U555)}$ and adding more Ψ in RTC3$_{(U288)}$ (SI Figure 2.13). We verified that ID-I influences how frequently Pus7 incorporates Ψ in cells by using CLAP to measure Ψ-levels on BET2, RTC3$_{(U77)}$, RTC3$_{(U288)}$, TEF2$_{(U555)}$, and TEF2 $_{(U1104)}$ mRNAs purified from cells containing or lacking ID-I (SI Figure 2.13). Both *in vitro* and in cells, ID-I does not have a uniform effect on the substrates that we examined. Together, these data suggest that ID-I may act as a rheostat to fine-tune how Pus7 interacts with individual sequences. The idea that ID-I makes subtle contributions to Pus7 function is supported by our observation that cells expressing Pus7$_{\Delta ID-I}$ do not have a growth defect (SI Figures 2.14 and 2.15).

2.2.6 Pus7 tightly binds RNAs with an array of sequences and secondary structures

The sequences targeted by Pus7 in cells are common in RNAs, yet only a small subset of potential target uridines are converted to Ψ. Pus1 and Pus4 are reported to use RNA secondary structure to recognize their substrates, and we wondered whether RNA structure similarly dictates which uridines Pus7 modifies (25,26). To begin asking this question, we predicted the secondary structure context of *S. cerevisiae* mRNA sequences modified by Pus7 (23). For each sequence, we modeled a 100-nucleotide region surrounding the site of Ψ incorporation. Secondary structure models were obtained using two different folding algorithms in the RNAstructure software package (free energy minimization and maximum expected accuracy) (23,43). These predictions indicate that in cells Pus7 modifies Us in a wide variety of structural contexts including unstructured regions, loops,

bulges, and helices (SI Figure 2.16). We noted that multiple Pus7 consensus sequences are often present within a single targeted mRNA. Therefore, we also examined the structural context of nonmodified sequences within Pus7 targeted mRNAs. Comparison of our structural models suggest that targeted uridines more commonly exist in less-structured contexts than nontargeted uridines present on the same mRNAs (SI Figure 2.16).

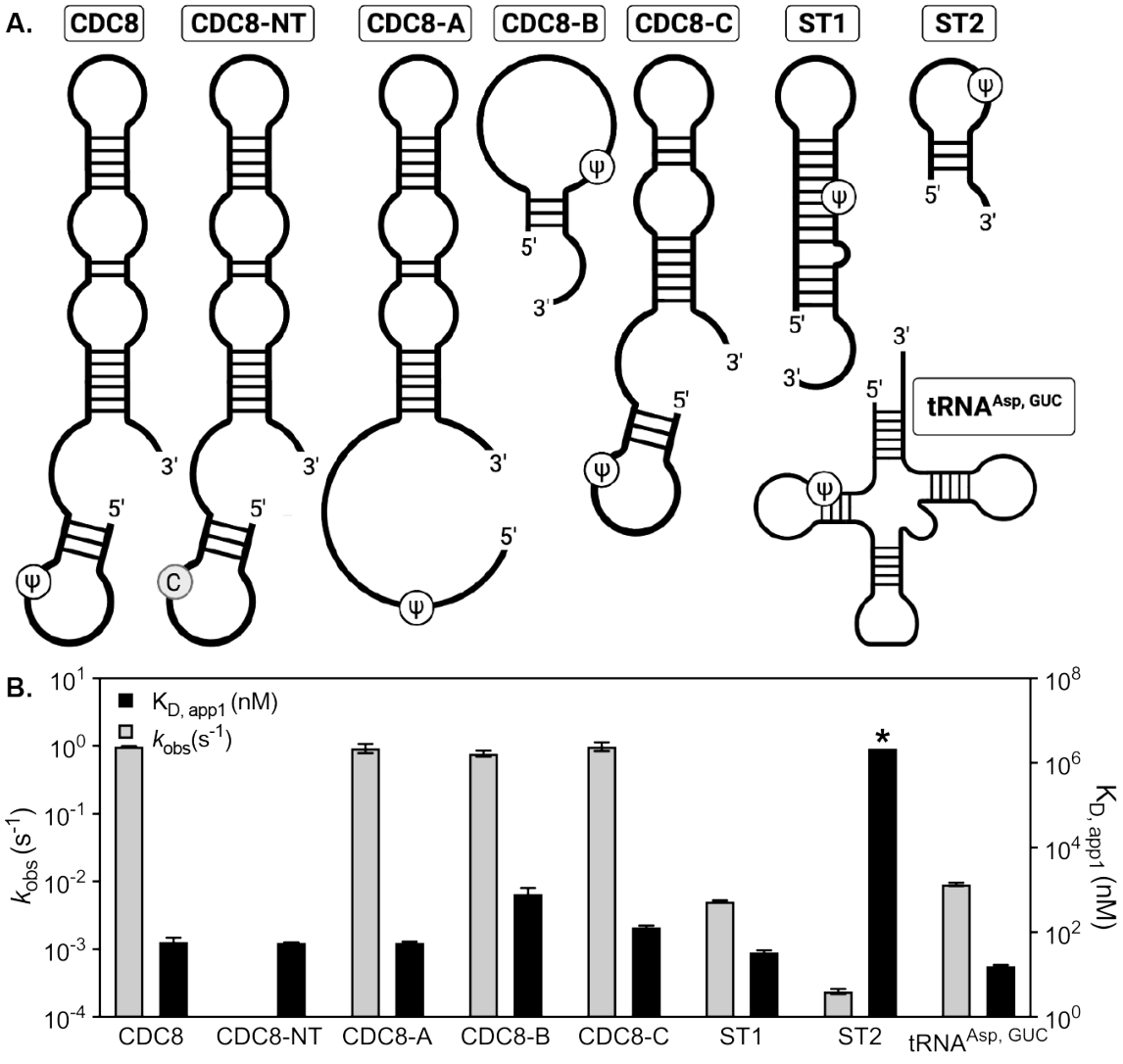


Figure 2.4: Pus7 can bind and modify a variety of RNA substrates

(A) Secondary structures of the RNAs investigated in this study. The substrate sequences are available in SI Table 2.5. **(B)** $K_{D,app1}$ (right y-axis, black bars) and k_{obs} (left y-axis, gray bars) values for Pus7 binding and modifying the substrates displayed in A. The $K_{D,app1}$ displayed for ST2 (*) is a lower limit for this value (SI Figure 2.9).

To experimentally evaluate the ability of Pus7 to interact with RNAs in a variety of structural contexts, we measured the binding of a catalytically inactive D256A Pus7 mutant to a series of 5'-fluorescein-labeled RNAs (Figure 2.4A). These RNAs differ in both their sequence and secondary structures and include a natural tRNA target ($tRNA^{Asp}$), three truncated CDC8 mRNAs (CDC8-A, CDC8-B, and CDC8-C), and two nonnatural substrates that place the target U in different structural contexts (ST1 and ST2). D256A bound to nearly all of the RNA substrates that we tested with similar affinities (16 to 130 nM) (SI Table 2.3). Only the short (19 to 25 nt) CDC8-B and ST2 substrates significantly increased the $K_{D,app1}$ for D256A, though the enzyme still bound the CDC8-B with a sub-μM dissociation constant ($K_{D,app1}$ = 800 ± 320 nM) (Figure 2.4 and SI Table 2.3). These findings indicate that Pus7 has a substantial affinity for RNAs in general and, together with our stopped-flow binding data, lead us to propose that the enzyme rapidly searches for consensus sequences amid many nonspecific binding sites (SI Figure 2.17).

2.2.7 Pus7 can rapidly incorporate pseudouridine on a diverse set of RNAs

Our results indicate that PUS7 substrate-binding is largely independent of the RNA secondary structures and sequences that we investigated. We wondered whether RNA structural properties play a bigger role in determining the ability of Pus7 to modify substrates. To test this, we compared the single-turnover rate constants for Ψ incorporation into $tRNA^{Asp}$, CDC8, CDC8A/B/C, ST1, and ST2 (Figures 2.4 and 2.5). We find that wild-type Pus7 modified the CDC8 mRNA substrates ~10-fold faster than $tRNA^{Asp}$ ($k_{obs,tRNA}$ = 0.009 s^{-1} ± 0.0005 and $k_{obs,cdc8}$ = 0.99 s^{-1} ± 0.1). The $k_{obs,tRNA}$ value that we measured is slower than previous studies of *E. coli* TruA, TruB, and RluA Pus enzymes, which have rate constants between 0.1 and 0.7 s^{-1} on their noncoding targets (44–46). As expected, no Ψ is incorporated into substrates when the target uridine is mutated to a cytidine ($tRNA^{Asp}_{NT}$ and $CDC8_{NT}$). While the truncation of CDC8 (CDC8-A/B/C) does not alter

the rate constant for pseudouridylation, we find that Pus7 incorporates Ψ 200- to 4,000-fold more slowly into the shorter nonnatural UGUAG-containing RNA sequences (ST1 and ST2). We anticipate that the reduced k_{obs} for ST2 reflects its weak binding of Pus7 (Figure 2.4 and SI Figure 2.8). Notably, the removal of ID-I (Pus7$_{\Delta ID\text{-}I}$) partially recovered enzyme activity on ST1 (increasing $k_{obs,ST1}$ by two-fold), suggesting that this domain may help to serve as a gatekeeper for substrate selection (Figure 2.5). Together, our data demonstrate that Pus7 is capable of fully modifying any substrate containing a UGUAG target sequence, regardless of context. However, it does convert uridines to Ψ more quickly when they are present in regions of RNAs >25 nt in length predicted to be flexible single stranded.

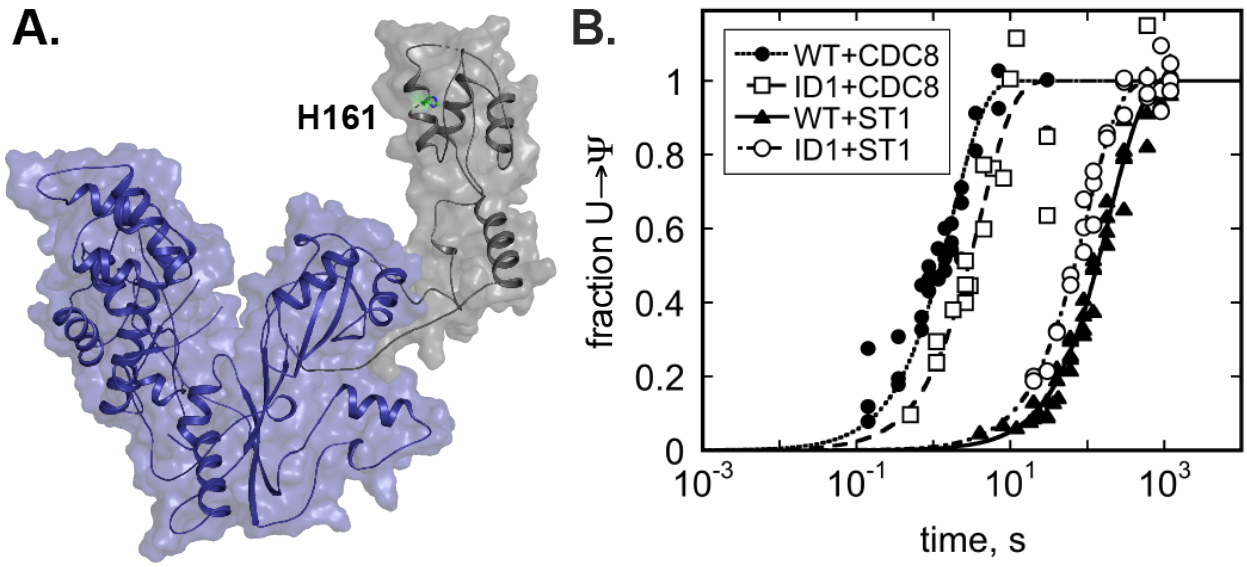


Figure 2.5: ID-I enhances Pus7 selectivity for CDC8 over ST1

(A) Pus7 (blue, PDB: 7MZV) with insertion domain I (ID-I) shown in gray. The Pus7$_{\Delta ID1}$ protein (blue) lacks ID-I (gray). **(B)** Time-courses of Ψ incorporation into CDC8 and ST1 by wild-type Pus7 (●: CDC8; ▲:ST1) and Pus7$_{\Delta ID1}$ (□: CDC8; ○: ST1). The single-turnover k_{obs} values for these reactions are reported in SI Table 2.4. ID-I enhances the ability of Pus to discriminate between CDC8 and ST1; wild-type $k_{obs,CDC8}/k_{obs,ST1}$ = 178, and Pus7$_{\Delta ID1}$ $k_{obs,CDC8}/k_{obs,ST1}$ = 40.

2.2.8 Pus7 activity towards Us predicted to be in structured regions is enhanced at increased temperatures

We further explored our observation that Pus7 appears to be more active on less structured UGUAR sequences by measuring the rate constants for Ψ incorporation into the ST1 substrate at elevated temperatures (37 °C, 42 °C, and 50 °C) where the structural stability of the RNA is reduced and the molecule should be more dynamic. If, as we hypothesize, base-pairing limits Pus7 activity toward substrates, then we anticipated that heating samples should increase pseudouridylation. Indeed, we find that $k_{obs,ST1}$ is increased by 18-fold between 30 °C and 42 °C (Figure 2.6A and SI Figure 2.19). There is still an enhancement in $k_{obs,ST1}$ at 50 °C, albeit less than that at 42 °C, despite being closer to the T_m of ST1. We modeled the thermostability of Pus7, and our model indicates that Pus7 is not stable at 50 °C (SI Figure 2.20), likely accounting for the decrease in activity observed at this temperature.

In addition to our *in vitro* observations, we modeled the structures of 20 randomly selected mRNAs reported to be modified by Pus7 under heat shock at 30 °C and 45 °C (SI Figure 2.20). Comparison of these models reveals that most of the targeted Us that we modeled are in different, often less-structured contexts at 45 °C than at 30 °C (SI Figure 2.13). We tested if the predicted decrease in substrate structure correlates with increased modification levels by measuring the degree of Ψ incorporation on nine full-length Pus7 substrates (ARG5,6, BET2, TEF5, $RTC3_{(U77)}$, $RTC3_{(U288)}$, $TEF2_{(U555)}$, $TEF2_{(U1104)}$, $U2snRNA_{(U35)}$, and $U2snRNA_{(U56)}$) at different temperatures by CLAP (42). In these assays, we reacted 150 μg of total RNA purified from *pus7Δ* cells with 50 μM of purified Pus7 at 30 °C and 37 °C for 10 min. The extent of pseudouridylation increased at 37 °C on all but one of the "heat shock" targets with more predicted secondary structure (TEF5, $RTC3_{(U77)}$, $RTC3_{(U288)}$, $TEF2_{(U555)}$, and $TEF2_{(U1104)}$) (Figure 2.6B). In contrast, the level of pseudouridylation on targets previously observed in Ψ-mapping studies at 30 °C (ARG5,6, BET2,

and U2snRNA$_{(U35)}$) or under nutrient starvation (U2snRNA$_{(U56)}$) were either unaffected or decreased (Figure 2.6B).

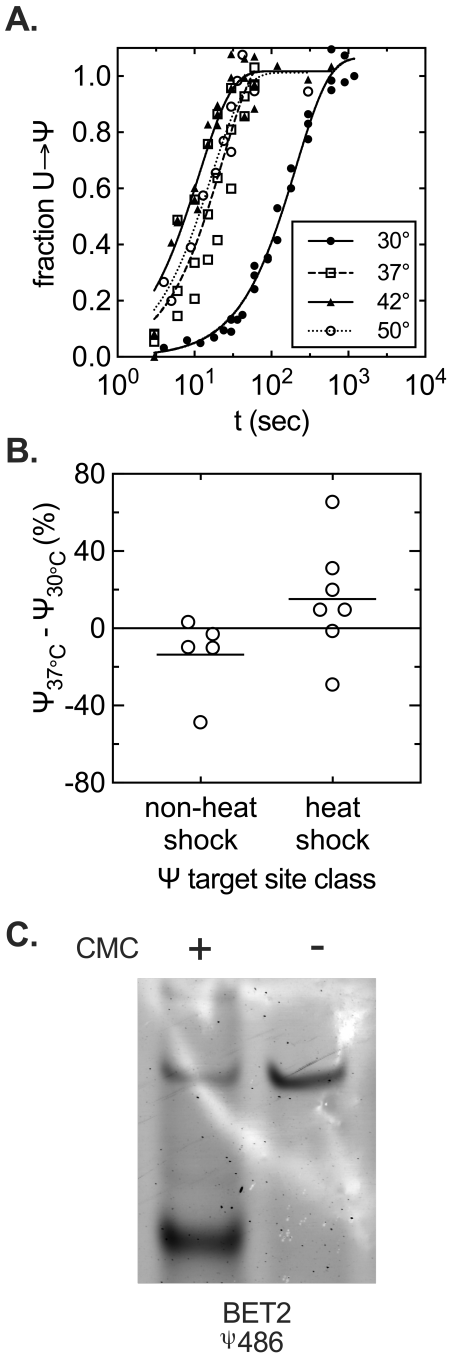


Figure 2.6: Pus7 is more active at elevated temperatures on substrates with UGUAR sequences predicted to be in secondary structures.

(A) Time courses of Ψ incorporation into ST1 by wild-type Pus7 at varying temperatures (●:30°C; □:37°C; ▲:42°C; and ○:50°C). The single-turnover k_{obs} values for these reactions are reported in SI Figure 2.19. **(B)** Differences in the

stoichiometry of Ψ incorporation at 30 °C and 37 °C in full-length RNA substrates measured by CLAP. The level of Ψ addition is generally enhanced at sites that were only detectable under heat shock in Schwartz et al. (24). **(C)** Representative CLAP gel of BET2 pseudouridylated target site from total RNA purified from BY4741 *yeast pus7Δ::kanMX*. Black arrow denotes the truncated, pseudouridylated product. The upper band is the unmodified, full-length product.

2.3 Discussion

Our fundamental understanding of the structure and mechanism of pseudouridine synthases is largely built on foundational studies of bacterial enzymes that exclusively modify noncoding RNAs. The discovery of pseudouridine in eukaryotic mRNAs, coupled with the identification of heritable diseases caused by mutations to human Pus enzymes, have ignited a renewed interest in eukaryotic pseudouridine synthases. While all Pus enzymes share a structurally conserved catalytic core, the eukaryotic enzymes possess additional insertions with unidentified functions. Pus7, a homolog of the bacterial TruD enzyme, is among the Pus enzymes that modify the largest number of mRNAs in eukaryotes (23,24). We solved the crystal structure of *S. cerevisiae* Pus7, revealing the inherent flexibility in the form and position of Pus7 eukaryotic-specific insertions (Figure 2.1B). The largest insertion (ID-I) is a positively charged domain connected to the Pus7 core by long flexible linkers. Normal mode analyses of our structure with elNémo and DynOmics Portal 1.0, along with the differing positions of ID-I observed in the two Pus7 structures, suggest that this domain is very flexible. ID-I appears capable of swinging away the core of the molecule, which may allow it to form contacts with RNA substrates distal from their modification sites. Our findings that the removal of ID-I ($Pus7_{\Delta IDI}$) reduces RNA-binding and enhances substrate discrimination are consistent with such a model (Figure 2.5 and Table 2.1). Nonetheless, in contrast to the large effect of active site mutations on substrate catalysis (38- to 74,000 fold), the impacts of ID-I deletion both in vivo and in vitro are modest (two- to fourfold reduction in Ψ incorporation) and indicate that the role of ID-I is more likely to fine-tune RNA substrate selection (Figure 2.5 and SI Figure 2.13).

Pus7 is distinguished from other pseudouridine synthases by the apparent variety of substrates that it has been reported to modify in cells. Our in vitro biochemical assays support this idea, as we find that Pus7 can bind and pseudouridylate UGUAR motifs in diverse sequence and structural contexts (Figure 2.4). Our kinetic and CLAP data indicate that although Pus7 can modify uridines predicted to be in strong secondary structures, it is most active on Us in regions with less-predicted structure (Figures 2.4 and 2.6, and SI Figure 2.19). The relatively slow k_{obs} value that we measured for $tRNA^{Asp}$, in which the targeted uridine is base paired, exemplifies this trend. These biochemical observations are in line with the predicted structural contexts of uridines targeted by Pus7 in yeast cells. The models we obtained using RNAstructure reveal a trend for Pus7 substrate selection in which modified UGUAR sequences are in less-structured regions than unmodified UGUAR sequences (SI Figure 2.16). These observations can help to partially rationalize which UGUAR sites Pus7 modifies in cells.

Under heat shock, Pus7 modifies 15-fold more mRNAs (24). Consistent with this, we found that the ST1 RNA, predicted to contain a base-paired UGUAG sequence, is modified more efficiently at elevated temperatures when the stability of the base-paired region of the molecule is significantly reduced (Figure 2.6A and SI Figure 2.19). Additionally, we measured the stoichiometry of Ψ incorporation in full-length RNAs previously reported to be modified under either unstressed or heat-shock conditions (23,24). We saw that heat shock–induced sites are more efficiently modified at higher temperatures than Ψ sites detected in unstressed cells (Figure 2.6B). This is notable because Pus7 should be less stable (SI Figure 2.20) and presumably less active at elevated temperatures. Our data suggest that increased RNA dynamics are more important than having optimal enzyme activity on these substrates.

Although our data indicate that Pus7 more quickly modifies structurally unconstrained Us, we find that if left for long enough (2 to 10 min) Pus7 converts 100% of Us to Ψs in all of the model UGUAG-containing sequences that we present it (Figure 2.4). Similarly, we also observed that Pus1 is able to modify sequences not predicted to contain its preferred secondary structure if allowed to react for 30 min (SI Figure 2.9). Since a significant portion of Us identified as Pus7 targets in sequencing studies are predicted to be in structured regions, a simplistic model in which Pus7 only interacts with single-stranded uridines does not satisfactorily explain either our *in vitro* studies or the breadth of targets identified by Ψ-mapping in cells. Our findings suggest that instead of identifying motifs that Pus7 can modify, we need to address why this promiscuous enzyme does not modify every UGUAR sequence in cells. We propose that Pus7 rapidly samples RNA sequences and opportunistically selects substrates that contain an accessible (even if only transiently) UGUAR sequence motif (SI Figure 2.15). Such a mechanism is reminiscent of DNA glycosylases that use facilitated diffusion to quickly scan nonspecific sites in their search for damaged bases (47,48).

The work that we present here suggests that Pus7 is a promiscuous enzyme and that factors beyond inherent enzyme properties (e.g., enzyme localization, RNA structure, and competition with other RNA-binding proteins) significantly contribute to shaping Pus7 substrate selection (49,50). This idea is exemplified by the observed relocalization of Pus7 to the cytoplasm and subsequent increased substrate scope under heat shock (24). The importance of protein localization and cellular conditions to Pus target selection is unlikely to be unique to Pus7. Pus4 was recently reported relocalize to the cytoplasm and have increased activity toward its mRNA substrates when it is in a prion conformation (18). Collectively, these findings indicate that the environment of

RNA substrates, which remodels in response to changing cellular conditions, plays a previously unrecognized role in determining the Ψ modification landscape.

2.4 Methods

2.4.1 Expression and purification of wild-type, mutant and truncated Pus7 proteins

Saccharomyces cerevisiae wild-type and truncated ($Pus7_{\Delta N34C9}$, $Pus7_{\Delta IDI}$) Pus7 protein encoding DNA-sequences were ordered from GeneArt. Ligation independent cloning was used to incorporate these sequences into a pMCSG7 vector containing an N-terminal His_6-tag and TEV cleavage site. Single and double mutants were incorporated into the Pus7 sequence by QuikChange® site-directed mutagenesis (Stratagene) using appropriate primers (IDT). Sequences were confirmed by Sanger DNA sequencing (UMich sequencing core). All proteins were expressed in BL21(DE3)-P-LysS *E. coli* cells grown in 1L Terrific Broth, 100 μg/mL ampicillin at 37°C and 250 RPM. Protein expression was induced by the addition of isopropyl β-D-1-thiogalactopyranoside (IPTG) to a final concentration of 0.15 mM when cells reached OD_{600} of ~0.6. Following induction, cells were grown for 18 hours at 20°C and harvested by centrifugation at 5,000 RPM for 30 minutes. Pus7 proteins were purified on a Ni^{2+} Hi-Trap column (GE healthcare), the His-tag was removed by TEV protease treatment followed by a second Ni^{2+} Hi-Trap column. The protein was further purified by anion exchange chromatography on a 5 mL Resource™ Q column (GE Healthcare), and size exclusion chromatography using a Superdex 200 column (GE Healthcare). Purified protein was either concentrated and stored at -80° C or used immediately for crystallization.

2.4.2 Selenomethionine Expression

$Pus7_{\Delta N34C9}$ was expressed in BL21(DE3) cells grown in Terrific Broth media (4% glycerol), 100 μg/mL ampicillin at 37°C overnight. The cells were pelleted resuspended in 1.1L of selenomethionine minimal media, supplemented with 50 μg/mL L-selenomethionine, and 100 mL

of freshly prepared, and sterile filtered nutrient solution 20% (w/v) glucose, 0.3% (w/v) $MgSO_4$, 0.1mg/mL $Fe(II)(SO_4)_3$, 0.1 mg/mL Thiamine, adjust to pH 7.4, sterile. The cells were then grown at 37°C and 250 RPM until OD_{600} of 0.6. The cells were induced with IPTG to a final concentration of 0.2 mM and grown for 18 hours at 20°C before harvesting by centrifugation.

2.4.3 Crystallization

Unlabeled and SeMet derivatized $Pus7_{\Delta N34C9}$ was concentrated 10 mg/mL in 50 mM TRIS, pH 7.5, 50 mM NaCl, 1 mM tris(2-carboxyethyl)phosphine (TCEP). Crystals of Pus7 were obtained by the sitting-drop vapor-diffusion method at 20°C by mixing 0.5uL of protein solution (10mg/mL) with 0.5uL of the reservoir solution which contained 2 M ammonium sulfate, 10 mM nickel (II) chloride, 100 mM TRIS pH 8.5. The crystals were then cryoprotected in a solution of 15% glycerol, 1.7 M ammonium sulfate, 0.85 mM nickel (II) chloride, 85 mM TRIS pH 8.5 before being flash cooled in liquid-nitrogen.

2.4.4 Crystal data processing

Diffraction data were collected at 100 K and at the Se edge on LS-CAT 21-ID-D at Advanced Photon Source, Argonne National Laboratory using a DECTRIS EIGER 9M. Three data sets were collected from two crystals, and all were separately processed with XDS to 3.2 Å resolution were Friedel pairs were treated as equal. Reflections from a total of 1500 selected frames (first 500 from 2 datasets and first 400 from the third) were merged and scaled with Aimless (51) and the resulting reflection file was used for subsequent refinements of our Pus7 model. The data were indexed to space group C222 (unit-cell parameters a = 117.9, b = 171.8, c = 105.3 Å) with 1 molecule in the asymmetric unit (Matthew's coefficient V_M = 3.46 $Å^3$ Da^{-1}, 64.5% solvent content). 500 frames

from a single data set were processed anomalous (Friedel pairs were not treated as equal) with XDS to 3.2 Å and the resulting reflection file was used for the SAD phasing.

2.4.5 Crystal structure solution

Initial structure solutions were obtained by molecular replacement using the human Pus7 (PDB:5KKP) as a search model and initial phases were calculated using Phaser (52). However, we were unable to obtain a structure solution for insertion domain one, which necessitated the growth of Se-Met Pus7 crystals. AutoSol (53) was used to identify selenium sites and calculate density-modified 3.3 Å experimental maps based on a single-wavelength anomalous dispersion (SAD) data set from SeMet Pus7 (the experimentally determined SeMet f' and f'' values that were used were -7.4 and 5.0 respectively). Specifically, 16 selenium sites were located and used for SAD phasing, using phenix.hyss. Subsequently, Phaser was used to calculate the experimental phases, followed by density modification with RESOLVE (figure of merit 0.36 before and 0.78 after density modification). The experimental density map showed clear features of the protein backbone and well-defined side chains. RESOLVE traced and automatically built 389 residues and their side chains in the experimental electron density. The final experimental model was in really good agreement with our original MR derived model but also provided us with a partial model of ID-1. The partial model of ID-1 included residues 129 to 148, a region of ID-1 that packs against the core of an adjacent monomer and includes the only SeMet present in ID1. The electron density corresponding to the insertion domain is overall poor and of rather low resolution, as also reflected in the very high average B-factors (165.02) as compared to the average B-factors (117.44) for the rest of the protein (SI Figure 2.2E). Ultimately, using SAD phasing, in combination with our MR model, we were able to obtain a structure solution for the insertion domain, completing our structure model. An overlay of the final Pus7 model with all 16 experimentally determined

selenium heavy atoms is shown in SI Figure 2.2. The structural model was refined with REFMAC5 as part of the CCP4I2 package (51) using isotropic individual *B*-factors with maximum-likelihood targets where the Babinet model for bulk-solvent scaling was utilized. Refinement was followed by model building and modification with Coot (54). We performed several iterative rounds of refinement followed by model building and modification. All crystallographic information as well as refinement statistics are provided in Table 2.1. The geometric quality of the model and its agreement with the structure factors were assessed with MolProbity (55). Figures displaying crystal structures were generated by PyMOL (56).

2.4.6 Preparation of 5'-fluorescein labeled RNA substrates.

RNA was prepared via *in vitro* transcription from DNA oligonucleotide templates ordered from Integrated DNA Technologies (IDT) and transcribed by recombinant T7 RNA polymerase (57). Transcription reactions were carried out in 50 mM Tris-HCl pH 8.0, 4 mM $MgCl_2$, 1 mM spermidine, 5 mM DTT, 4 mM ATP, 4 mM CTP, 4 mM UTP, 1 mM GTP, 4 mM guanosine-5'-O-monophosphorothiolate (GMPS), 350 µg/mL purified T7 RNA polymerase, 12.5 µM purified DNA template containing T7 promoter and 4 U/µl SUPERaseIn. After stopping the transcription by the addition of 50 mM EDTA and 500 mM NaCl, the RNA was washed with degassed TE pH 7.2 three times using Amicon spin column (10 kDa MWCO). The washed RNA (~250 µl) was incubated with 20 µl 45 mM fluorescein overnight at 37°C to label the 5'end. All following steps were carried out in the dark. The reaction was stopped by addition of an equal volume of 2X loading dye (0.05% Bromophenol Blue, 0.05% Xylene Cyanol dye, 50% m/v urea, 0.1 M EDTA) and run on a 12% urea-polyacrylamide gel. The RNA was eluted via crush-and-soak method into buffer (TE, 0.1% SDS, and 0.5 M NaCl) overnight at 4°C. The elution products were subsequently filtered, washed, and concentrated using degassed TE and an Amicon spin column (10 kDa

MWCO). The RNA was then ethanol precipitated at -20°C for 12 hours. The resulting pellet was resuspended in 20 µl of RNase free H_2O. The concentration of the total and labeled RNA were measured photometrically using A_{260} and A_{494} respectively, using a Nanodrop spectrophotometer. Select Fl-labeled substrates were also purchased from Dharmacon.

2.4.7 Electrophoretic mobility shift assays (EMSAs)

For gel-shift assays, serially diluted protein (0-2000 µM) was incubated with 10 nM of 5'-fluorescein labeled RNA in 10 µL reaction volumes for ≥ 5 min at 25°C in a binding buffer containing 100 mM NH_4OAc, 100 mM Tris, pH 8.0, 5 mM $MgCl_2$, 2 mM DTT, and 6% (w/v) sucrose. An aliquot of each reaction (5 µL) was loaded on a preequilibrated, native 6% polyacrylamide (37.5:1) gel in 1xTBE. Gels were electrophoresed at 30V for ~4h at 4°C. When fluorescently labeled RNA substrates were used, electrophoresis was performed in the dark. Gels were then rinsed in 1xTBE and imaged on an Amersham Typhoon Biomolecular Imager (GE Healthcare). If unlabeled RNA was used, the gel was stained with SYBR® Gold Nucleic Acid Gel Stain (Invitrogen) in 1xTBE for ≥ 30 min in the dark before imaging on the Typhoon. Band intensities were quantified using ImageQuant (Cytiva) and the percentage of RNA bound calculated using Equation 2.1:

$$RNA_{bound}\,(\%) = 100 \times \frac{{E_T}^{n_h}}{{K_{D,app}}^{n_h} + {E_T}^{n_h}} \qquad \text{Equation 2.1}$$

Binding data were fit using equations derived from the binding models shown in SI Figure 2.10. In general, simpler models were tried first, and if systematic errors remained in the fit, more complex models were used to fit the data. The simplest model used was a Hill curve, Equation 2.2:

$$RNA_{bound}\,(\%) = 100 \times \frac{{E_T}^{n_h}}{{K_{D,app}}^{n_h} + {E_T}^{n_h}} \qquad \text{Equation 2.2}$$

In this model, $K_{D,app}$ is the apparent K_D for binding of Pus7 to one of the many sites on a given RNA; $K_{D,app}{}^{n}{}_{h}$ is the concentration of Pus7 at which 50% of available sites are bound. When systematic errors remained in the fit, a more complex model was used in which Pus7 bound first to a single specific site on the RNA, followed by the binding of multiple Pus7 moieties to multiple nonspecific sites on the same RNA (SI Figure 2.10B). These data were analyzed using Equation 2.3:

$$RNA_{bound}\,(\%) = 100 \times \frac{\left({E_T}/{K_{D,app1}}\right) \times \left(1 + \left({E_T}/{K_{D,app2}}\right)^{n_h}\right)}{1 + \left({E_T}/{K_{D,app1}}\right) \times \left(1 + \left({E_T}/{K_{D,app2}}\right)^{n_h}\right)} \qquad \text{Equation 2.3}$$

Neither of these models are theoretically correct, in particular because there is no evidence for cooperative binding of Pus7 to RNA. A theoretically correct binding model would need to account for random binding of Pus7 to all of the possible binding sites on a given RNA. Each RNA has many binding sites, which are not all equivalent because of differences in sequence and structure, and the binding sites can interact with one another negatively (via steric occlusion, for example) and positively (e.g., binding of Pus7 at one site changes structure at a second site, increasing binding affinity). Our experimental methods do not provide enough information to develop such a model. The simplified models we use to analyze the data are therefore the best available tool and allow for quantitative comparison of differences in binding that are identifiable via visual inspection of EMSA gel images.

2.4.8 Single-turnover pseudouridylation assays

RNA substrates containing 5,6-[^{3}H]-uridine were prepared by *in vitro* transcription (57) and denaturing gel purification. Reaction buffer was as described for the EMSA experiments. RNA substrates were folded in 1X reaction buffer by heating to 60°C for 5 minutes, followed by a 30 minute incubation at 30°C (44). Indicated concentrations of protein were mixed with the smallest detectable amount of substrate (~3,000 cpm per uridine in each timepoint, which allows reliable detection of tritium release above 5% turnover). At each timepoint an aliquot of reaction mix (containing ~3,000 cpm/U) was quenched in 1,250 µL 0.1 M HCl (final) containing 250 µg Norit-A. Quenched timepoints were mixed, centrifuged at 21,000 x *g* for 5 minutes, and 1000 µL of supernatant was transferred to a new tube containing 250 µL of 0.1 M HCl with 250 µg Norit-A. Mixing and centrifugation were repeated, and 1000 µL of the supernatant was filtered through glass wool in a 1 mL pipet tip to remove residual charcoal. Aliquots of the filtrate (500 µL) were removed for liquid scintillation counting in a Beckman LSC-6500. For each reaction mix, input controls were prepared by passing an aliquot of reaction through the same process using 0.1 M HCl without the Norit-A. Counts observed in the input sample are used to calculate cpm/uridine, allowing calculation of the amount of Ψ produced at each timepoint. Background counts were determined by processing an RNA only reaction aliquot through the sample pipeline; these counts were routinely equivalent to background in our instrument (~30 cpm). Fraction of target U converted to Ψ data were fitted using Equation 2.4:

$$U > \Psi\ (fraction) = 1 - \ e^{-k_{obs} \times t} \qquad \text{Equation 2.4}$$

2.4.9 Stop Flow Assays: Pus7/D256A binding with fluorescently labeled mRNA

D256A Pus7 and 5'-fluorescein labeled CDC8 were generated and purified as described as above. Kinetic binding experiments were performed using the Kintek SF-300x stop-flow apparatus.

Fluorescently labeled mRNA (5 nM final concentration) was mixed with D256A at varied concentrations (20 nM – 750 nM final). Binding experiments were performed at room temperature in same buffer used in the EMSA experiments over the span of 1-1.5 seconds. Lower concentrations of Pus7/D256A (0-100 nM) displayed monophasic behavior and were fit with a single exponential equation (Equation 2.5) obtain a k_{obs1}.

$$A_1 e^{-k1t} + c \qquad \text{Equation 2.5}$$

Higher concentrations displayed biphasic behavior and therefore were fit with a double exponential equation (Equation 2.6) to obtain k_{obs1} and k_{obs2}.

$$A_1 e^{-k1t} + A_2 e^{-k2t} + c \qquad \text{Equation 2.6}$$

The k_{obs1} values from both fits were then plotted against the concertation of D256A PUS7 mutant, displaying a linear relationship. The y-intercept gave a k_{off} of approximately 35 s^{-1} and the slope gave a k_{on} of ~4.3 x 10^8 $M^{-1}s^{-1}$. The K_D For D256A binding CDC8 was obtained using Equation 2.7:

$$K_D = {}^{k_{off}}/_{k_{on}} \qquad \text{Equation 2.7}$$

2.4.10 Wild-type and *pus7Δ* growth assessment

Wild-type and *pus7*Δ yeast cells were inoculated into 3 mL YPD media and grown overnight. Then, they were diluted to OD_{600}=1 as a starting point, and 7 ml of 10-fold serial dilutions were spotted on fresh YPD agar plates supplemented with 0.75-1.0 M NaCl, 250 mM $MgSO_4$, 200 mM puromycin, 100 ng/mL cycloheximide, 25-50 mg/mL hygromycin B, 50 mM MG132 and 1.5-3 mg/mL paromomycin. Growth of the cells were also tested in the presence of different carbon sources including 2% glucose, 2% sucrose and 2% galactose in YEP agar media (1% yeast extract and 2% peptone). The plates were incubated for 2-3 days at 30°C unless otherwise indicated.

2.4.11 Phylogenetic tree generation

Annotated TruD/Pus7 sequences (>400 total sequences) from GenBank (NCBI) were aligned using ClustalW (58). Then, a representative 44 amino acid sequence was used for further analysis. Evolutionary analyses were conducted using MEGAX tool (59). The phylogenetic tree was generated using the Maximum Likelihood method (60). The bootstrap consensus tree inferred from 100 replicates is taken to represent the evolutionary history of the taxa analyzed (61). The percentage of replicate trees in which the associated taxa clustered together in the bootstrap test (100 replicates) are shown next to the branches. Initial tree(s) for the heuristic search were obtained automatically by applying Neighbor-Join and BioNJ algorithms to a matrix of pairwise distances estimated using the JTT model, and then selecting the topology with superior log likelihood value. A discrete Gamma distribution was used to model evolutionary rate differences among sites (5 categories (+G, parameter = 1.3722)). The rate variation model allowed for some sites to be evolutionarily invariable ([+I], 0.89% sites). There were a total of 1344 positions in the final dataset.

2.4.12 Ribosome profiling data analysis

Raw ribosome profiling sequencing data from two studies (37,38) were downloaded and processed using the procedures described below. Briefly, adapter contaminations and low-quality reads were filtered out from the raw reads using the Cutadapt tool (62) like as previously described (63). Subsequently rRNA and tRNA contaminations were removed by aligning reads to the non-coding RNA (ncRNA) sequences of *S. cerevisiae* using Bowtie2 (64). Next, the remaining unaligned sequences were aligned against the transcriptome (coding RNA) of *S. cerevisiae* (R64-1-1 genome built) using TopHat2 (65). After that perfect match alignments were extracted from the TopHat

output. For further downstream analysis, 3'- and 5'-end P-site offset values were determined using riboWaltz (66). These P-site offset values are required to identify where ribosomes are located on each ribosome protected footprints (RPFs). After P-site offset calculation, actively translating ribosomes that represent trinucleotide periodicity were identified. Then the number of mapped RPFs was counted for each codon position within a gene using Samtools (67).

2.4.13 Modeling of thermal stability of PUS7

Using the established relationship between a protein's stability and its heat capacity (ΔC_p), stability (ΔG) chain length can be reasonably modeled as a function of chain length (N) and temperature (T) (68–71). Pus 7's stability curve was modeled as a function of N and T using previously published model seen in Equation 2.8 (69,71–73).

$$\Delta G(N,T) = \Delta H_R + \Delta C_p(T - T_R) - T\Delta S_R - T\Delta C_p \; ln \; ln \left(\frac{T}{T_R}\right) \qquad \text{Equation 2.8}$$

Where enthalpy (ΔH_R) and entropy (ΔS_R) are calculated at a reference temperature and (ΔC_p) is the heat capacity of a protein, T_R is the reference temperature of 373 K for both ΔH_R and ΔS_R. These previous studies took advantage of the correlation between a protein's thermodynamic parameter and chain length to derive linear equations from experimental measurements collected into databases. The linear equations can be expressed as Equations 2.9, 2.10, and 2.11 (69,71).

$$\Delta H_R = m_h \cdot N + b_h \qquad \text{Equation 2.9}$$

$$\Delta S_R = m_s \cdot N + b_s \qquad \text{Equation 2.10}$$

$$\Delta C_p = m_c \cdot N + b_c \qquad \text{Equation 2.11}$$

Where m_h and b_h are the slope and intercept of ΔH_R, m_s and b_s are the slope and intercept of ΔS_R and m_c and b_c are the slope and intercept of ΔC_p when these parameters are plotted as a function

of *N*. Equations 2.9, 2.10, and 2.11 can be inserted into Equation 2.8, in order to get stability as a function of N and T as seen in Equation 2.12.

$$\Delta G(N, T) =$$

$$(m_h N + b_h) + (m_c N + b_c)(T - T_R) - T(m_s N + b_s) - T(m_c N + b_c)\ ln\left(\frac{T}{T_R}\right)$$

Equation 2.12

2.4.14 Detection and quantification of pseudouridylation: CLAP Assay

The CLAP assay was adapted from Zhang, 2019 (42).

2.4.14.1 Pseudouridylation of total RNA or in vitro transcribed CDC8

Briefly, 150 µg of total RNA purified from BY4741 yeast *Δpus7::kanMX* was mixed with 50 µM Pus7-WT or Pus7-ΔID1 and incubated for 10 minutes at 30°C or 37°C in 1X pseudouridinylation buffer (100 mM TRIS-HCl pH 8.0, 100 mM NH_4OAc, 2 mM DTT, 5 mM $MgCl_2$) to modify the RNA. The reaction was stopped by adding $1/10^{th}$ volume of 3 M NaOAc pH 5.2, followed by two phenol:chloroform (1:1) extractions with saturated acid phenol, and a final chloroform extraction to isolate the RNA. The RNA was then precipitated by adding an equal volume of 100% EtOH and 1 µL of GlycoBlue (Thermo Fisher, AM9515) and incubated at -20°C for 3 hours.

2.4.14.2 CMC treatment

RNA was resuspended in 41.5 µL of BEU buffer (50 mM Bicine pH 8.3, 4 mM EDTA, 7 M Urea). For CMC treated samples, 8.5 µL of freshly prepared 1 M CMC dissolved in BEU buffer was added, for a final concentration of 170 mM CMC. For CMC non-treated samples, 8.5 µL of BEU buffer was added, for a final reaction volume of 50 µL. Samples were incubated at 37°C for 20 minutes. The reaction was stopped by adding 100 µL of Stop Buffer (300 mM NaOAc pH 5.2, 0.1

mM EDTA) for a final volume of 150 µL. Excess CMC was removed by two sequential ethanol precipitations. Briefly, 700 µL 100% EtOH, and 1 µL GlycoBlue were added to the reaction before incubating 3 hours at -20°C. Sample was spun down for 30 min, 15kRPM, at 4°C before removing the supernatant, and washing the pellet by adding 500 uL of 70% EtOH, and spin for 5 min at 15kRPM. Remove supernatant and allow pellet to dry. Resuspend the RNA pellet in 100 µL of Stop Buffer and repeat ethanol precipitation and wash.

2.4.14.3 Alkali Treatment

Resuspend the pellet in 40 µL of 50 mM Na_2CO_3 pH 10.4 (pH taken at 37°C, temperature of incubation) and incubate for 3 hours at 37°C. Precipitate RNA via ethanol precipitation, as described above, with an additional 70% ethanol wash. Let pellet air dry. Resuspend the pellet in 20 µL sterile water and determine concentration by nano-drop.

2.4.14.4 RNA 5' Phosphorylation

To 6 µg RNA in 6.5 µL, add 1 µL 10X T4 PNK reaction buffer (NEB B0201S), 1 µL of 1 mM ATP, 0.5 µL 20 U/µL SUPERase•In RNase Inhibitor (Thermo Fisher AM2694), and 1 µL 10 U/µL T4 PNK (NEB M0201L) for a final volume of 10 µL. Incubate at 37°C for 30 minutes.

2.4.14.5 Blocker Ligation

To the reaction above, add 1 µL 10X T4 RNA Ligase reaction buffer (NEB B0216L), 1 µL of 100 µM 5' RNA blocker oligo (IDT /5AmMC6/rArCrCrCrA), 1 µL of 1 mM ATP, 1 µL 20 U/µL SUPERase•In RNase Inhibitor (Thermo Fisher AM2694), 3 µL DMSO, 2 µL sterile water and 1 µL 10 U/µL T4 RNA Ligase I (NEB M0204L) for a final volume of 20 µL. Incubate at 16°C for 16 h. Stop ligation reaction by adding 1.2 µL 200 mM EDTA.

2.4.14.6 Reverse transcription

For reverse transcription, the RT primer was first annealed by taking 3 μL of ligation mixture, adding 1 μL of 10 X annealing buffer (250 mM Tris-HCl, 480 mM KCl, pH 7.4) and 1 μL of 0.5 μM target specific reverse transcription primer (IDT). Samples were heated to 95°C for 3 minutes and slowly cooled to 37°C at a rate of -0.01°C/s (~15 min). To annealed sample, add 5 μL of 2 X AMV reverse transcription reaction mixture (1.2 U/μL AMV RT (NEB M0277L), 2X AMV RT buffer, and 1 mM of each dNTP) for a final concentration of 0.6 U/μL AMV RT, 1X AMV RT buffer, and 0.5 mM of each dNTP. Incubate at 42°C for one hour. Inactivate AMV RT by heating to 85 °C for 5 min before placing on ice. To digest RNA, add 1 μL of 5 U/μL RNaseH and incubate at 37°C for 20 minutes. Inactivate RNaseH by heating to 85 °C for 5 min and before placing reaction on ice. Add 1 μL of splint/adaptor oligo mix (1.5 μM adaptor oligo, 1.5 μM splint oligo) and incubate mixture at 75°C for 3 minutes followed 3 minutes at room temperature to anneal the splint/adaptor. Add 4 μL of 4x ligation mixture (40 U/μL T4 DNA ligase, 4X T4 DNA ligase buffer, and 50% DMSO) for a final concentration of (10 U/μL T4 DNA ligase, 1X T4 DNA ligase buffer, and 12.5% DMSO). Incubate at 16°C for 16 h. Heat reaction to 65°C for 10 min to deactivate T4 DNA ligase, place immediately on ice.

2.4.14.7 PCR

Use 2 μL of reaction above, mix with 3.5 μL of 5 μM forward primer and 3.5 μL of 5 μM reverse primer (or reverse transcription primer). Add components for Q5 DNA polymerase reaction to a final volume of 35 μL and final concentration of 1X Q5 reaction buffer, 1X Q5 GC enhancer, 200 μM of each dNTP, 0.5 μM of forward and reverse primers, and 0.2 U/μL Q5 high fidelity DNA polymerase (NEB M0491L). Perform 35 cycles of PCR at requisite annealing temperatures for

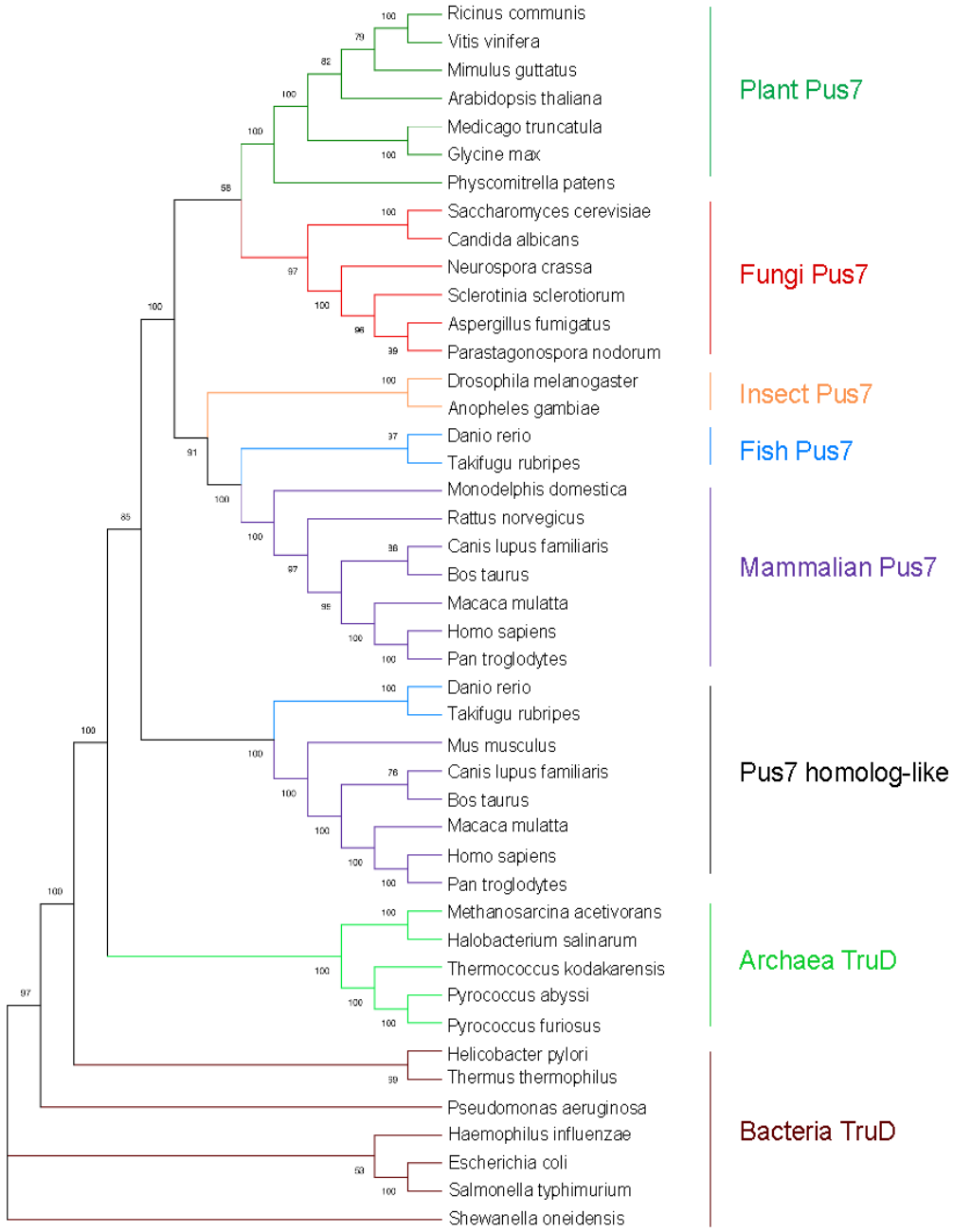
each site. 5 μL of PCR reaction was mixed with 1 μL of 6X TriTrack DNA loading dye and loaded on to a native 10% acrylamide (29:1) gel in 1X TBE pre-run at 10V/cm for 1 hour. Gel ran 3 hours at 10V/cm before being stained in 1X SYBR gold nucleic acid gel stain in 1X TBE for ~10 minutes. Gels were imaged on Amersham Typhoon imager and quantified using ImageQuant.

2.4.14.8 CLAP Primers

ARG5,6_RT	CCCATAGCAAGATTAATATTT
ARG5,6_FWD	TAGTTATTGGTGGTTTCA
ARG5,6_REV	TGCAGACATTGAGTAGC
ARG5,6_ADAPT	pCCATGTGAAACCACCAATAACTA
ARG5,6_SPLINT	TTTCACATGGAGTTGTTTGC/3SpC3/
BET2_RT	GCTTGAGCTGCATGGGATTCA
BET2_FWD	ACTATCAATTTTGGGTGAATTAA
BET2_REV	GCATTAGGACATAATCCAAAG
BET2_ADAPT	pCCATGTTAATTCACCCAAAATTGATAGT
BET2_SPLINT	ATTAACATGGAGACTTTGTA/3SpC3/
U2snRNA_RT	TATTATTTTGGGTGCCAAAAA
U2snRNA_56_FWD	CCTTTTGGCTTAGATCAA
U2snRNA_REV	ATGTGTATTGTAACAAATTAAAAGG
U2snRNA_56_ADAPT	pCCATGTTGATCTAAGCCAAAAGG
U2snRNA_56_SPLINT	ATCAACATGGAACAACTGAA/3SpC3/
U2snRNA_35_FWD	ACGAATCTCTTTGCCTTT
U2snRNA_35_ADAPT	pCCATGAAAGGCAAAGAGATTCGT
U2snRNA_35_SPLINT	CCTTTCATGGAGTATCTGTT/3SpC3/
CDC8_RT	ATATGCGTACTCAAAACAGGC
CDC8_FWD	GCTATTGGATAAAGAGATAAGGA
CDC8_REV	TCAACGATTTGCCAAATAAGC
CDC8_ADAPT	pCCATGTCCTTATCTCTTTATCCAATAGC
CDC8_SPLINT	AAGGACATGGAGACGTTACT/3SpC3/
EFB1/TEF5_81_RT	GTTGAACCATCTGGAGAATTC
EFB1/TEF5_81_FWD	GAAACAATTAAACGCTTCTTT
EFB1/TEF5_81_REV	TGGGTAAGCAGATTGGAAA

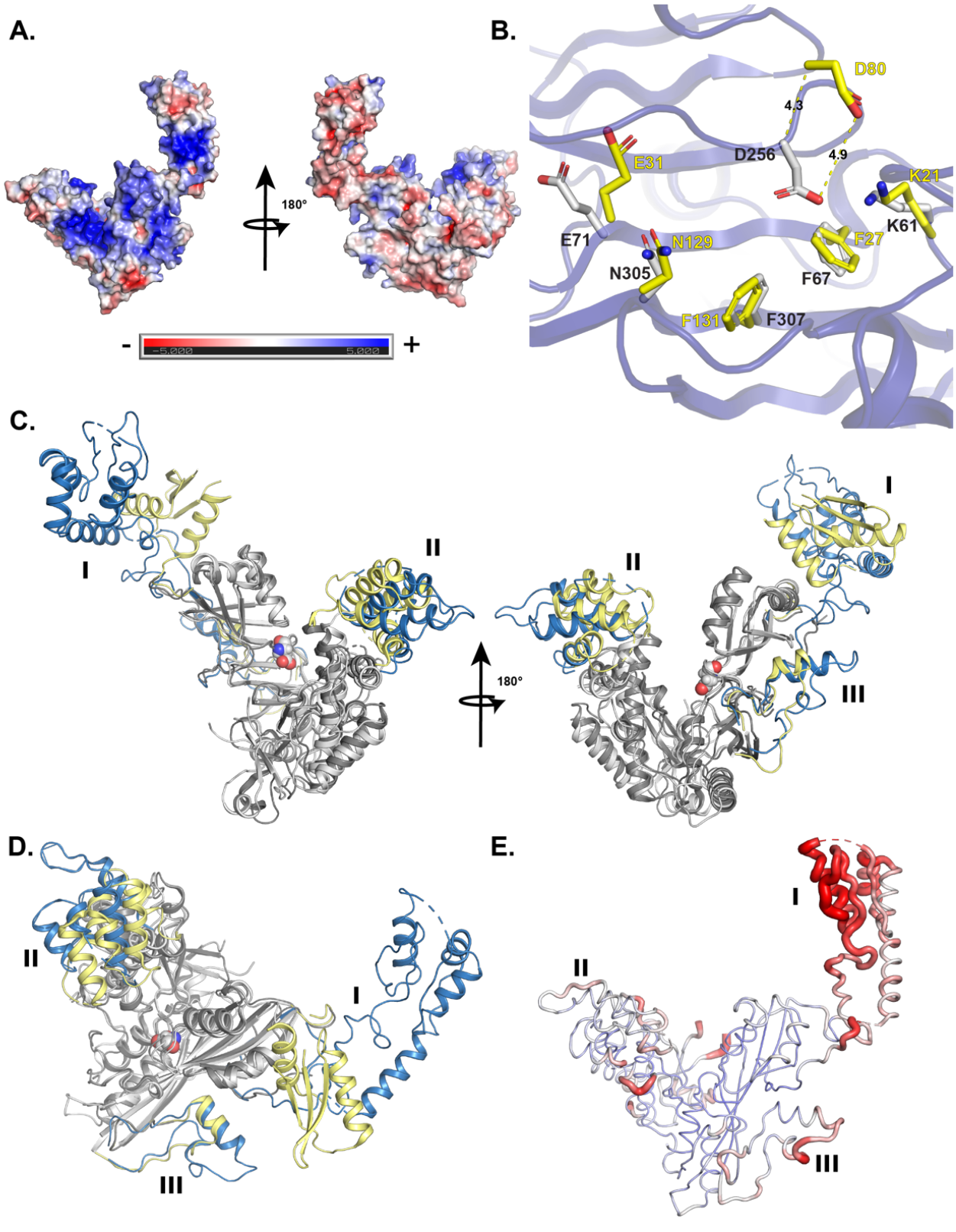
EFB1/TEF5_81_ADAPT	pCCATGAAAGAAGCGTTTAATTGTTTC
EFB1/TEF5_81_SPLINT	TCTTTCATGGACTGCTGTTT/3SpC3/
RTC3_77_RT	TCCTGAGGAGTGAAAACTTCG
RTC3_77_FWD	GGTGAAAATACAGATTTGATTG
RTC3_77_REV	AAGAGTTCGACAACTTCAGAT
RTC3_77_ADAPT	pCCATGCAATCAAATCTGTATTTTCACC
RTC3_77_SPLINT	GATTGCATGGAGACGAATAT/3SpC3/
RTC3_288_RT/REV	TCAATTGTAGGCTTTGGTTC
RTC3_288_FWD	GTTATCGATTTGATATTGAGAAA
RTC3_288_ADAPT	pCCATGTTTCTCAATATCAAATCGATAAC
RTC3_288_SPLINT	AGAAACCATGGAGTCTCAAAA/3SpC3/
TEF2_555_RT	GGACTTCAAGAACTTTGGATG
TEF2_555_FWD	GAAACCTCCAACTTTATCAA
TEF2_555_REV	GGTGGTAGCTTCAATCATGTT
TEF2_555_ADAPT	pCCATGTTGATAAAGTTGGAGGTTTC
TEF2_555_SPLINT	ATCAACATGGGTTCCATTCG/3SpC3/
TEF2_1104_RT	ACCCTTGTACCATGGAGCGTT
TEF2_1104_FWD	TTACTCTCCAGTTTTGGA
TEF2_1104_REV	GTCTTCCAACTTCTTACCAGA
TEF2_1104_ADAPT	pCCATGTCCAAAACTGGAGAGTAA
TEF2_1104_SPLINT	TTGGACATGGAGATTCGACG/3SpC3/

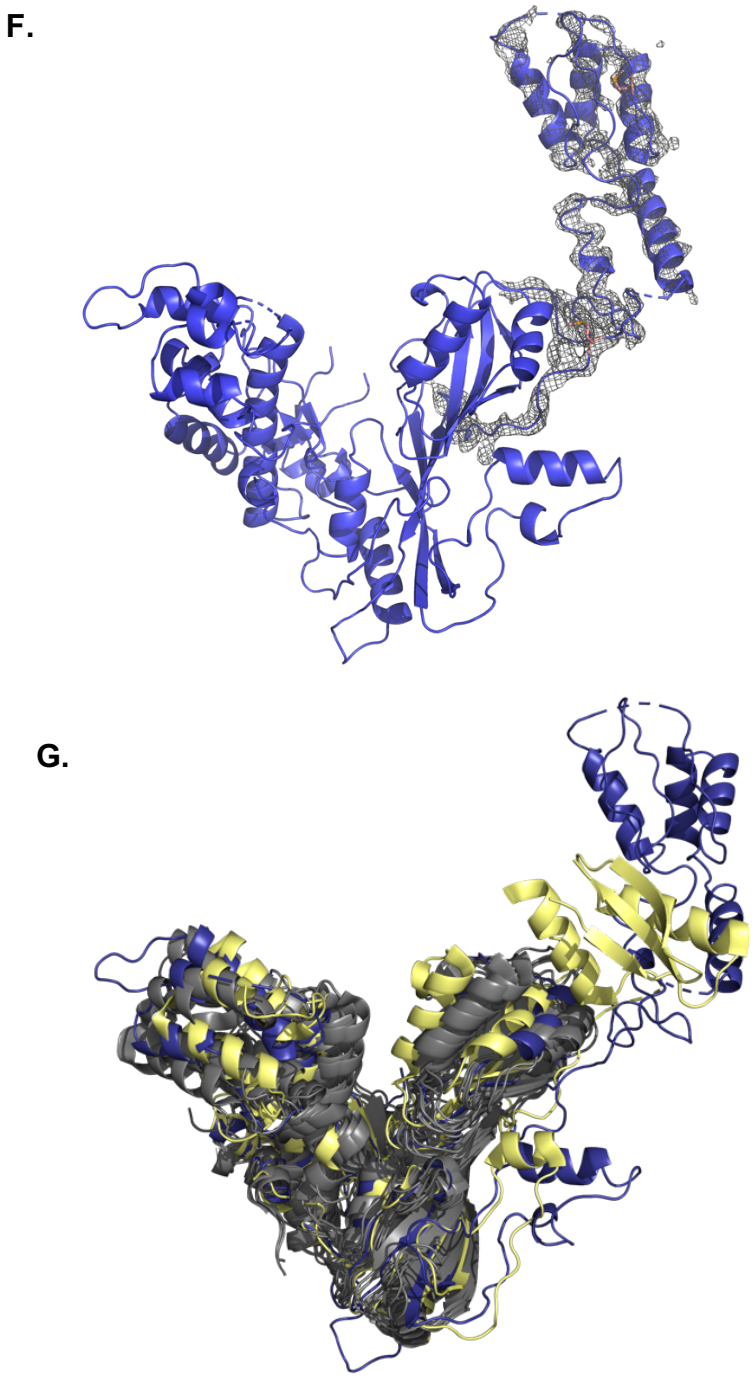
2.5 Supplementary Information



SI Figure 2.1: Phylogenetic relations in TruD and Pus7 family

This tree shows the relation of Pus7 family proteins in different species. It also represents the relation between Pus7 family with TruD family proteins

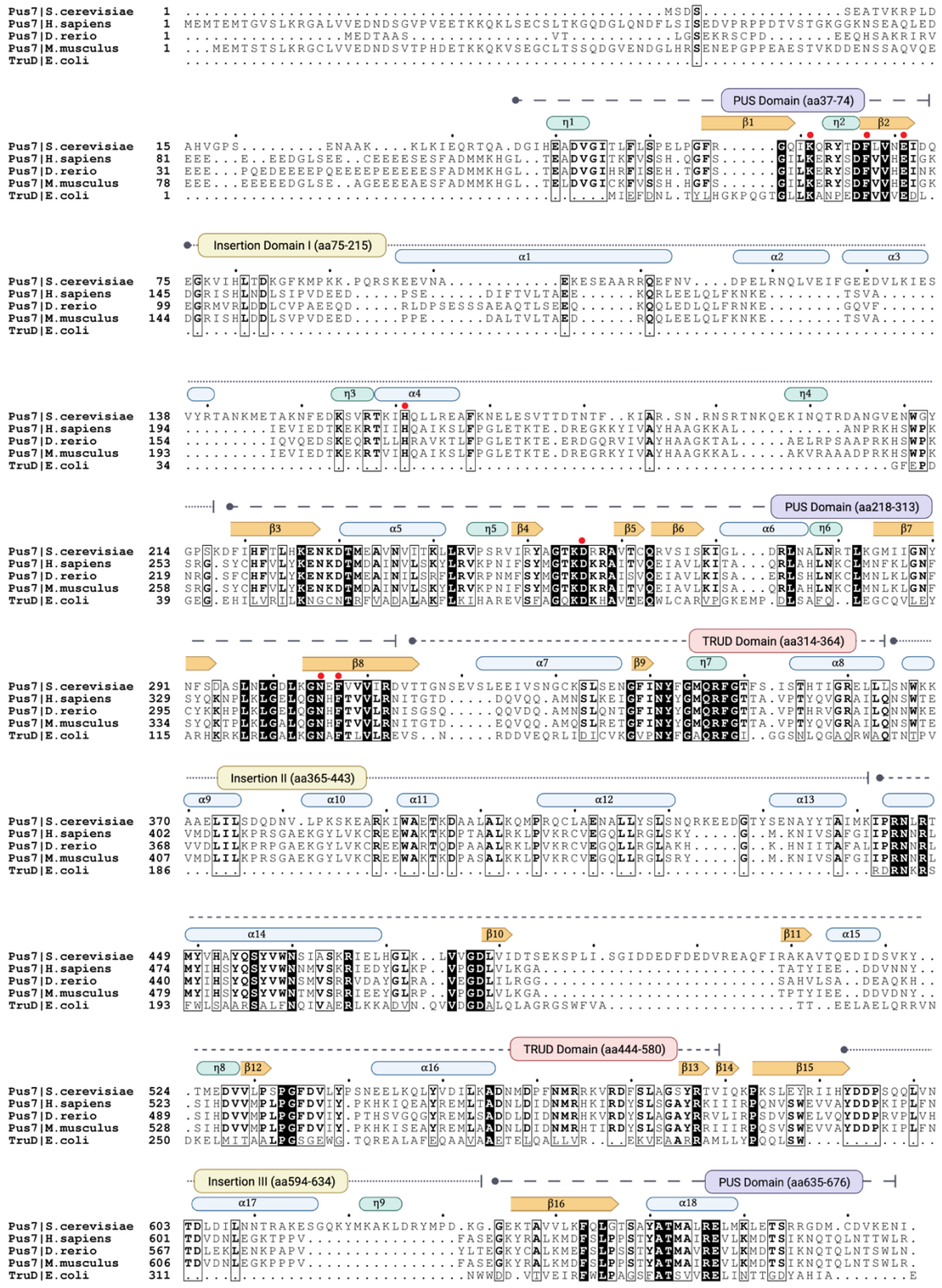

A.
180°
-
+
-5.000
5.000
B.
D80
4.3
E31
D256
4.9
K21
K61
E71
N129
F27
N305
F67
F131
F307
C.
I
II
180°
I
II
III
D.
II
I
III
E.
I
II
III



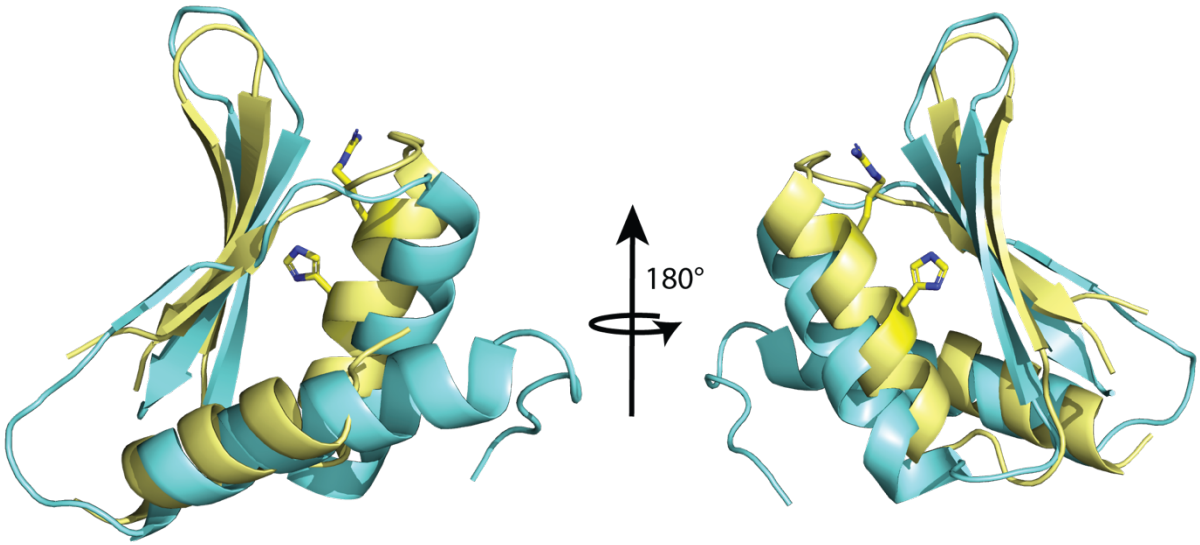
SI Figure 2.2: Comparison of Pus7 structures

(A) Rendering of the electrostatic surface potential of *S. cerevisiae* Pus7 generated with ABPS Electrostatics (74). Negatively charged regions are shown in red, and positively charged regions are shown in blue. **(B)** Catalytic aspartate residue (D256, *S. cerevisiae* Pus7 numbering) is shifted ~4Å relative to the same residue in *E. coli* TruD (D80). Figure shows alignment of the *S. cerevisiae* Pus7 active site (purple, PDB: 7MZV) and residues (gray sticks) with the equivalent residues in TruD (yellow sticks, PDB:1SB7) (32). Pus7 numbering in black, TruD numbering in yellow. Distances measured both from C_{Alpha} position and from the carboxyl on D256/D80. **(C)** Superposition of *S. cerevisiae*

Pus7 (light gray, blue) and *H. sapiens* Pus7 (dark gray, yellow, PDB: 5KKP) (75), (C_{alpha} RMSD = 3.743 Å for 144 atoms) and rotated 180 degrees to show the difference in position of the insertions (I, II, and III) in *S. cerevisiae* (blue) and *H. sapiens* (yellow) Pus7. The catalytic residue D256, *S. cerevisiae* numbering, is shown in the active site (light gray spheres). **(D)** Top-down view of *S. cerevisiae* and *H. sapiens* Pus7 superposition, looking down into the active site. **(E)** Putty representation of *S. cerevisiae* Pus7 colored according to *B* factors. Residues with the lowest *B* factors in blue (min = 20Å) and maximum in red (max=200Å). **(F)** 2Fo-Fc maps showing experimental electron density (gray mesh) around *S. cerevisiae* Pus7 ID-I (blue) contoured at 1.5σ. Methionine residues (M88 and M145, orange sticks), used for SAD phasing. **(G)** Superposition (using 136 – 336 C_{alpha} atoms of the TRUD and PUS domains, RMSD: ~2.52 Å for these domains) of TruD homologs, including: each molecule in the asymmetric unit of each *E.coli* TruD structures (gray, PDB: 1SB7, 1SI7, 1SZW) , both TruD molecules in the asymmetric unit of the *Methanosarcina mazei* structure (gray, PDB: 1Z2Z), the single Pus7 molecule in the *H. sapiens* structure (yellow, PDB: 5KKP), and the single molecule in the *S. cerevisiae* Pus7 structure reported here (blue, PDB: 7MZV).

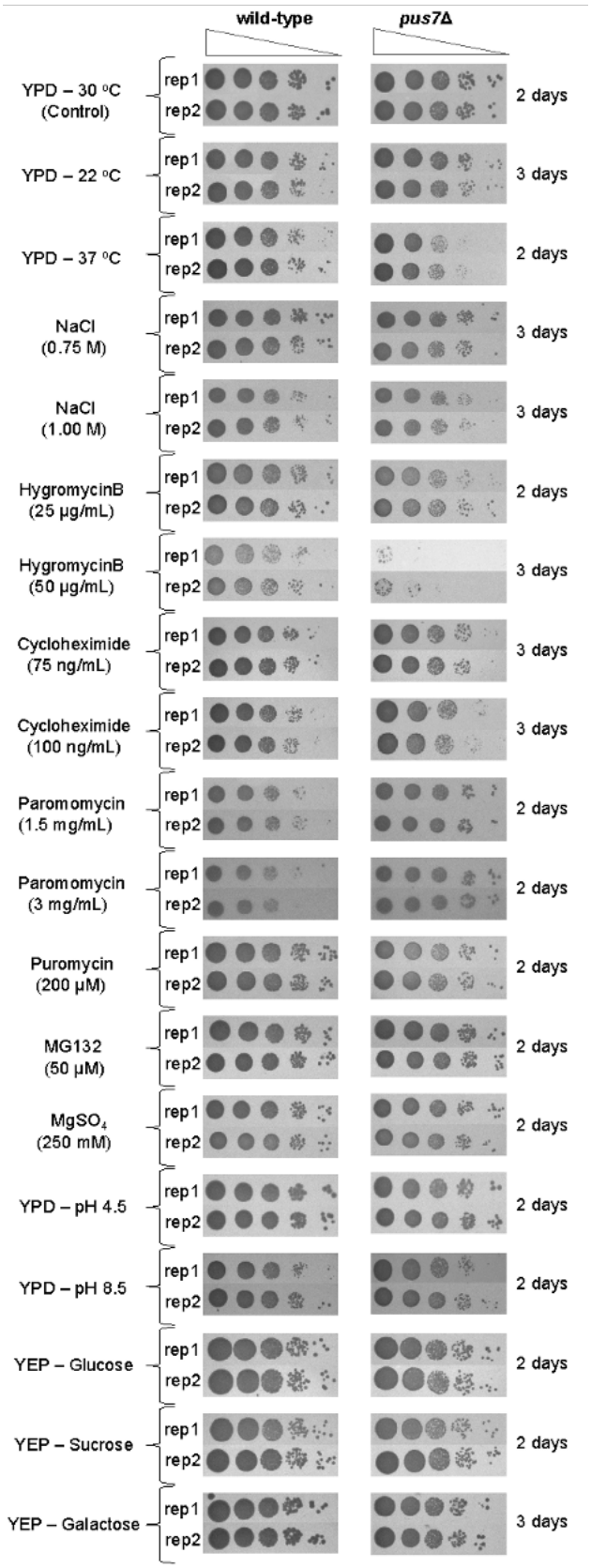


SI Figure 2.3: Sequence alignment of representative TruD family members

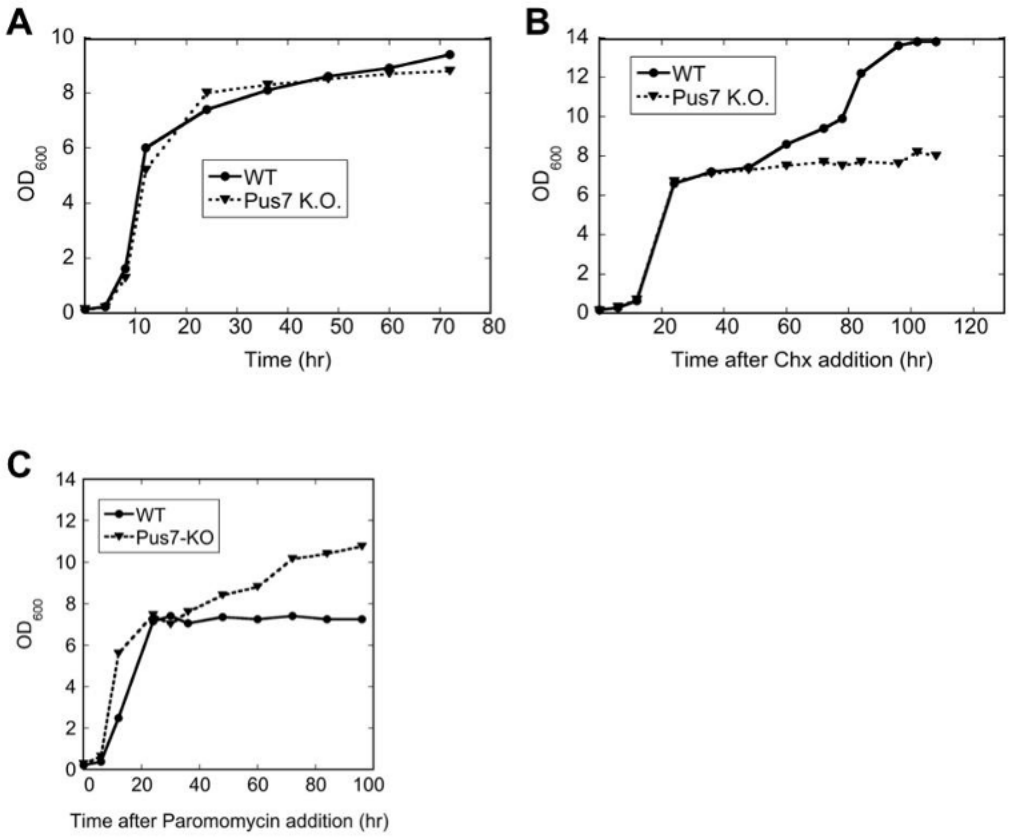


SI Figure 2.4: ID-I contains a single strand nucleic acid binding R3H domain

DALI (76) alignment of the R3H motifs (based on 56 atoms, RMSD: 2.788 Å) from *H. sapiens* Pus7 ID-I (yellow, PDB: 5KKP) and *H. sapiens* PARN nuclease domain (blue, PDB: 2A1S) (35,75,77).

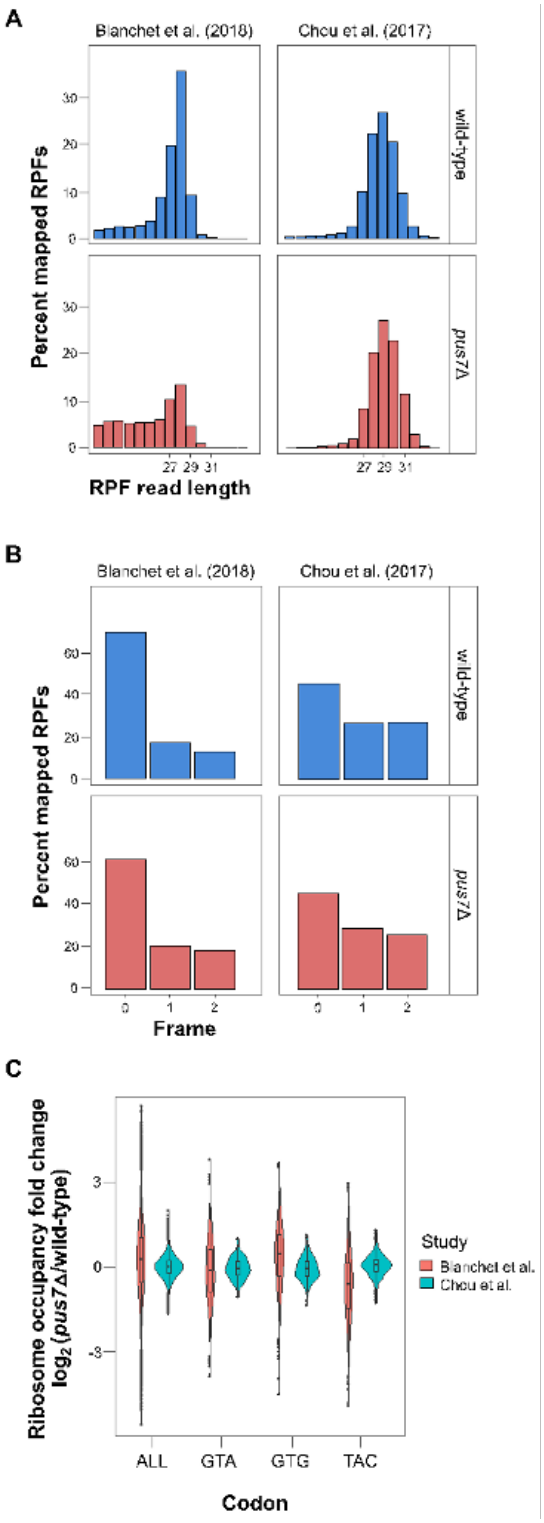


SI Figure 2.5: *S. cerevisiae* cell growth under different conditions



SI Figure 2.6: Growth curves for wild-type and *pus7*Δ cells

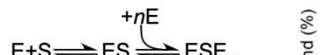
Wild-type and *pus7*Δ cells grown in YPD at 30°C after the addition of **(A)** nothing, **(B)** cycloheximide, and **(C)** paromomycin.



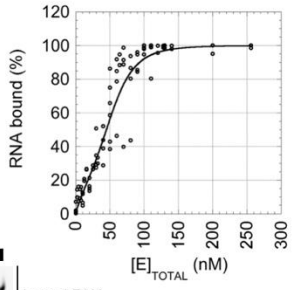
SI Figure 2.7: Ribosome occupancies are affected in *pus7*Δ cells

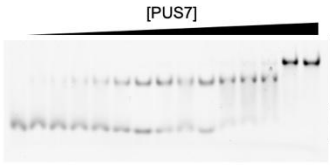
(A) Ribosome protected footprint (RPF) read length distribution. Distribution of ribosome protected fragments (RPFs) length show that most of the RPFs are between 27-30 nucleotide length. **(B)** ~50-60 % of these RPFs are in-frame (frame 0). **(C)** Ribosome occupancies are altered in *pus7*Δ compared to wild-type cells. Fold change in the ribosome codon occupancies was simply calculated by dividing the number of mapped RPFs in the P-site of *pus7*Δ to wild-type..

cdc8-FL PUS7 WT

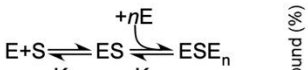


$K_{D,s}$ = 75±15 nM

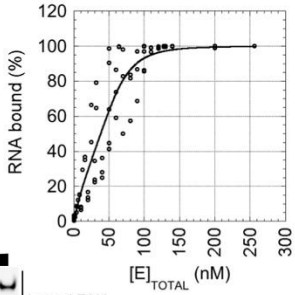


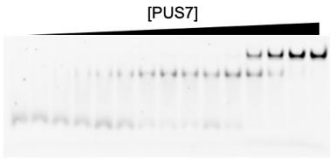


cdc8-FL PUS7 D256A

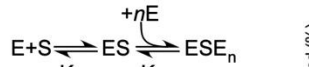


$K_{D,s}$ = 60±15 nM

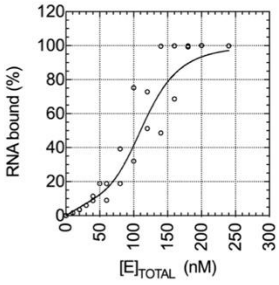


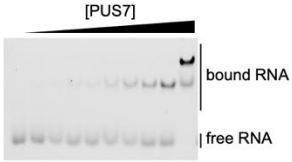


cdc8-FL K61A D256A

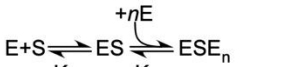


$K_{D,s}$ = 180±40 nM

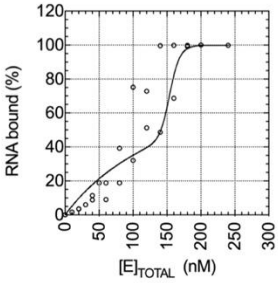


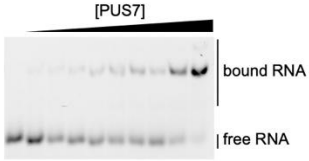


cdc8-FL F67A D256A

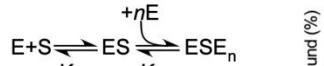


$K_{D,s}$ = 480±50 nM

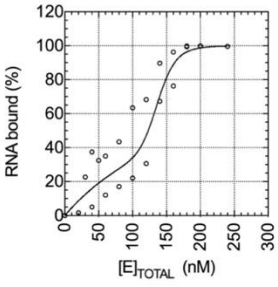


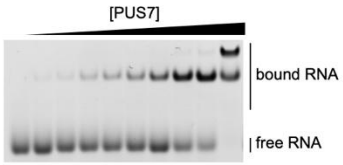


cdc8-FL E71A D256A

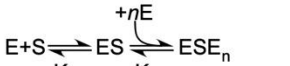


$K_{D,s}$ = 210±50 nM

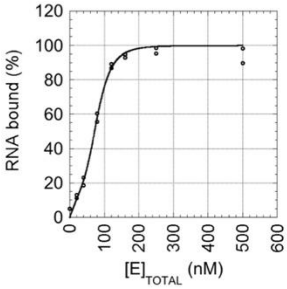


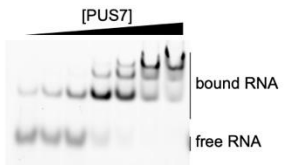


cdc8-FL H161A D256A

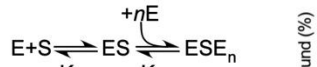


$K_{D,s}$ = 170±40 nM

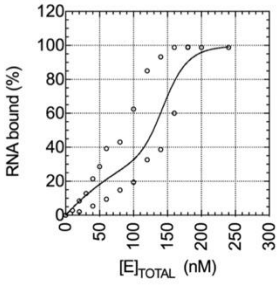


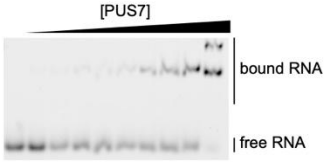


cdc8-FL N305A D256A

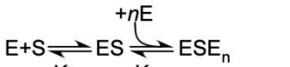


$K_{D,s}$ = 230±60 nM

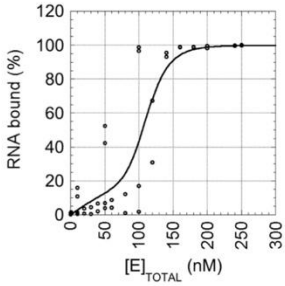


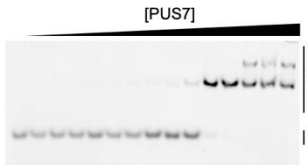


cdc8-FL F307A D256A

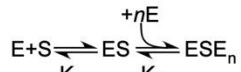


$K_{D,s}$ = 340±170 nM

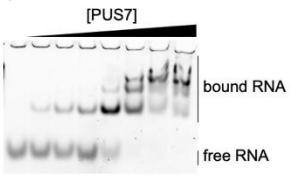


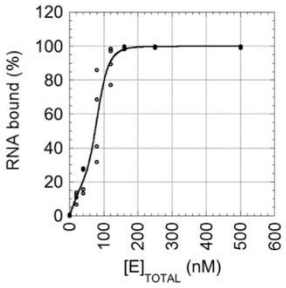


cdc8-FL ΔID1 D256A

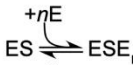


$K_{D,s}$ = 160±40 nM

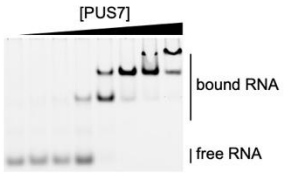


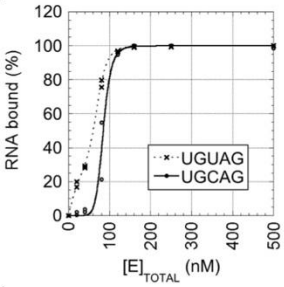


cdc8-FL nontarget D256A

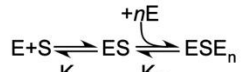


nontarget: UG**C**AG
target: UG**U**AG

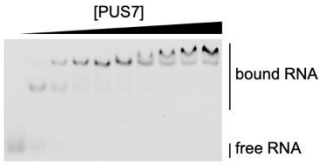


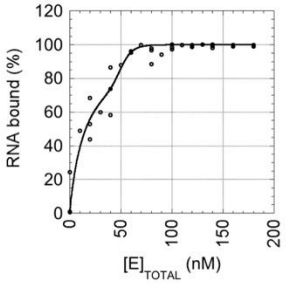


cdc8-A PUS7 WT

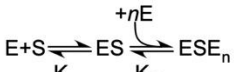


$K_{D,s}$ = 15±2 nM

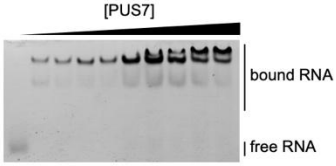


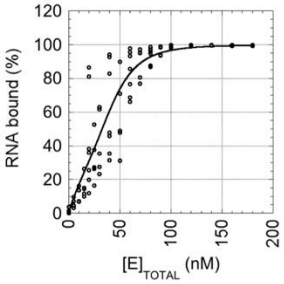


cdc8-A PUS7 D256A

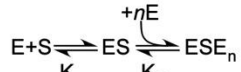


$K_{D,s}$ = 57±4 nM

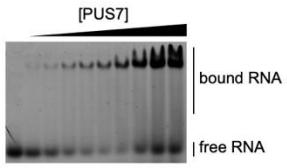


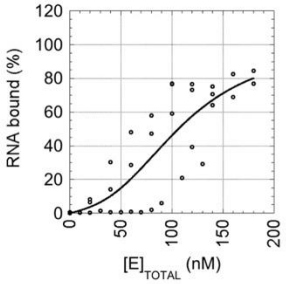


cdc8-B PUS7 WT

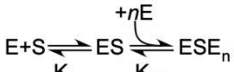


$K_{D,s}$ = 630±1500 nM

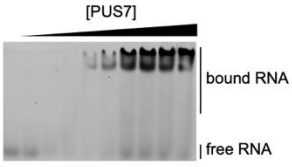


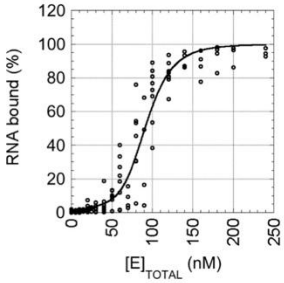


cdc8-B PUS7 D256A

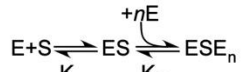


$K_{D,s}$ = 800±320 nM

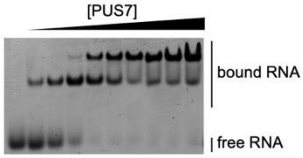


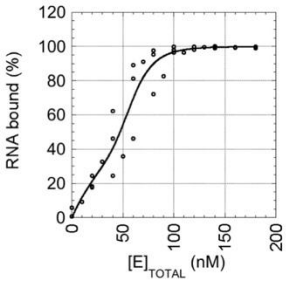


cdc8-C PUS7 WT

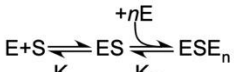


$K_{D,s}$ = 74±18 nM

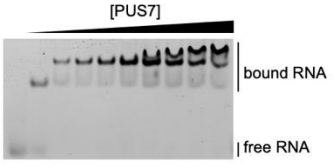


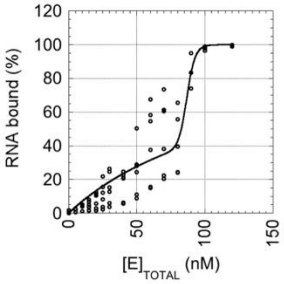


cdc8-C PUS7 D256A

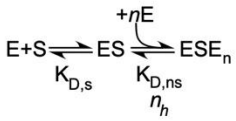


$K_{D,s}$ = 130±13 nM

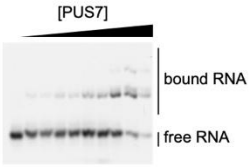


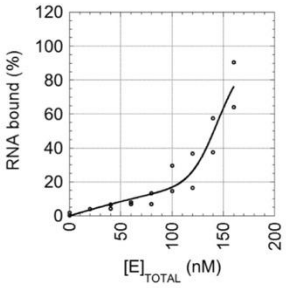


tArg PUS7-WT

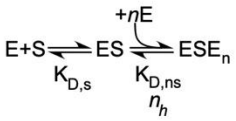


$K_{D,s}$ = 550±160 nM

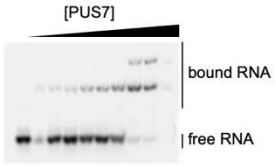


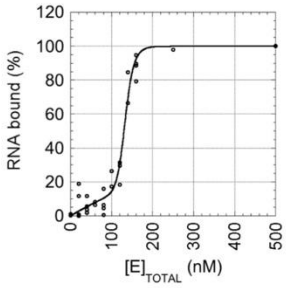


tArg PUS7-D256A

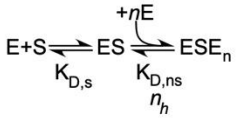


$K_{D,s}$ = 700±170 nM

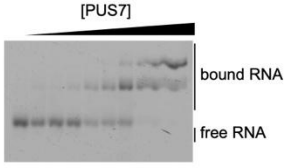


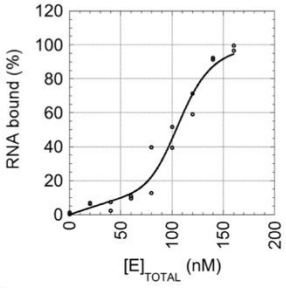


tArg-C13U PUS7-WT

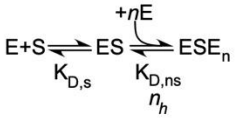


$K_{D,s}$ = 470±180 nM

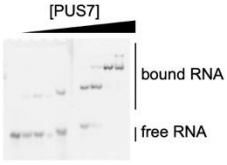


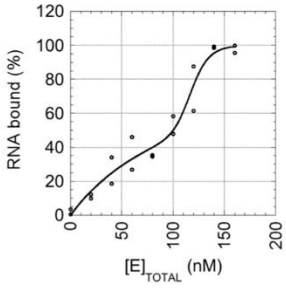


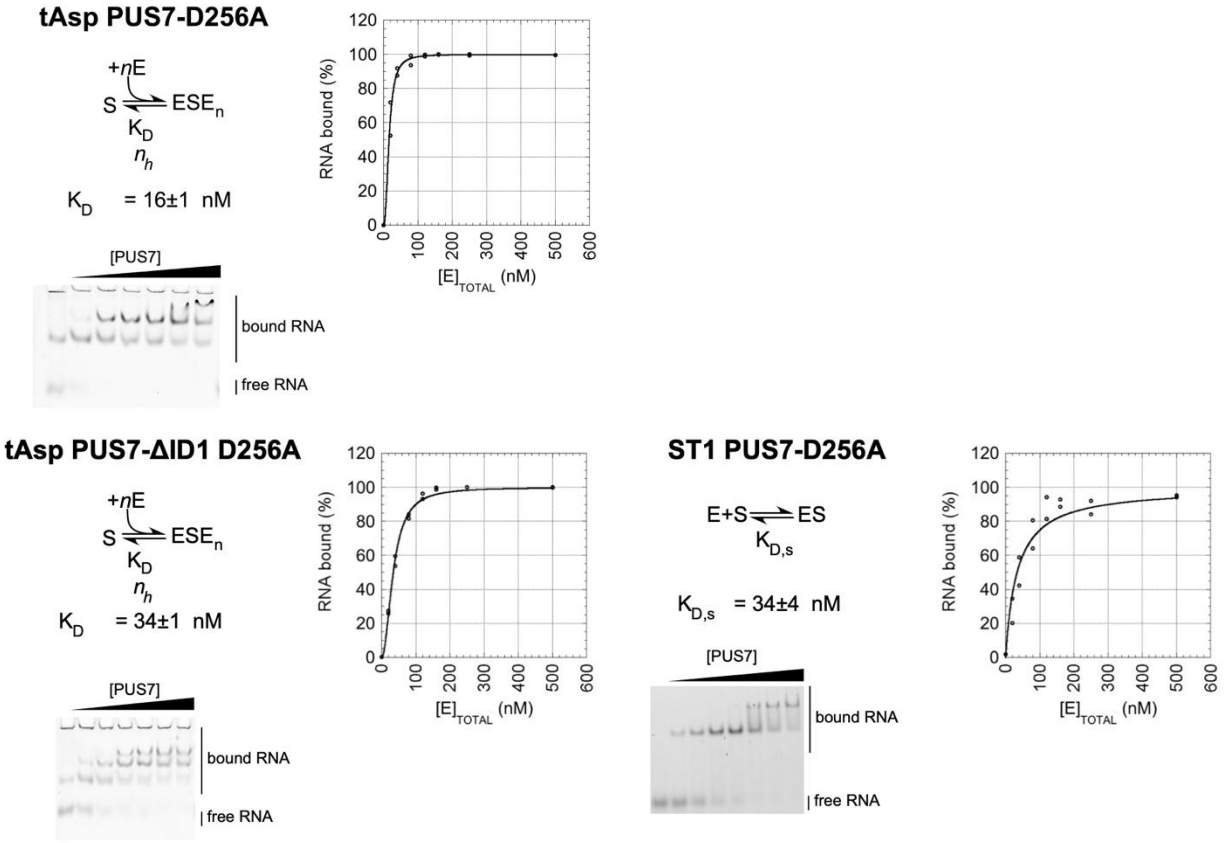
tArg-C13U PUS7-D256A



$K_{D,s}$ = 120±20 nM

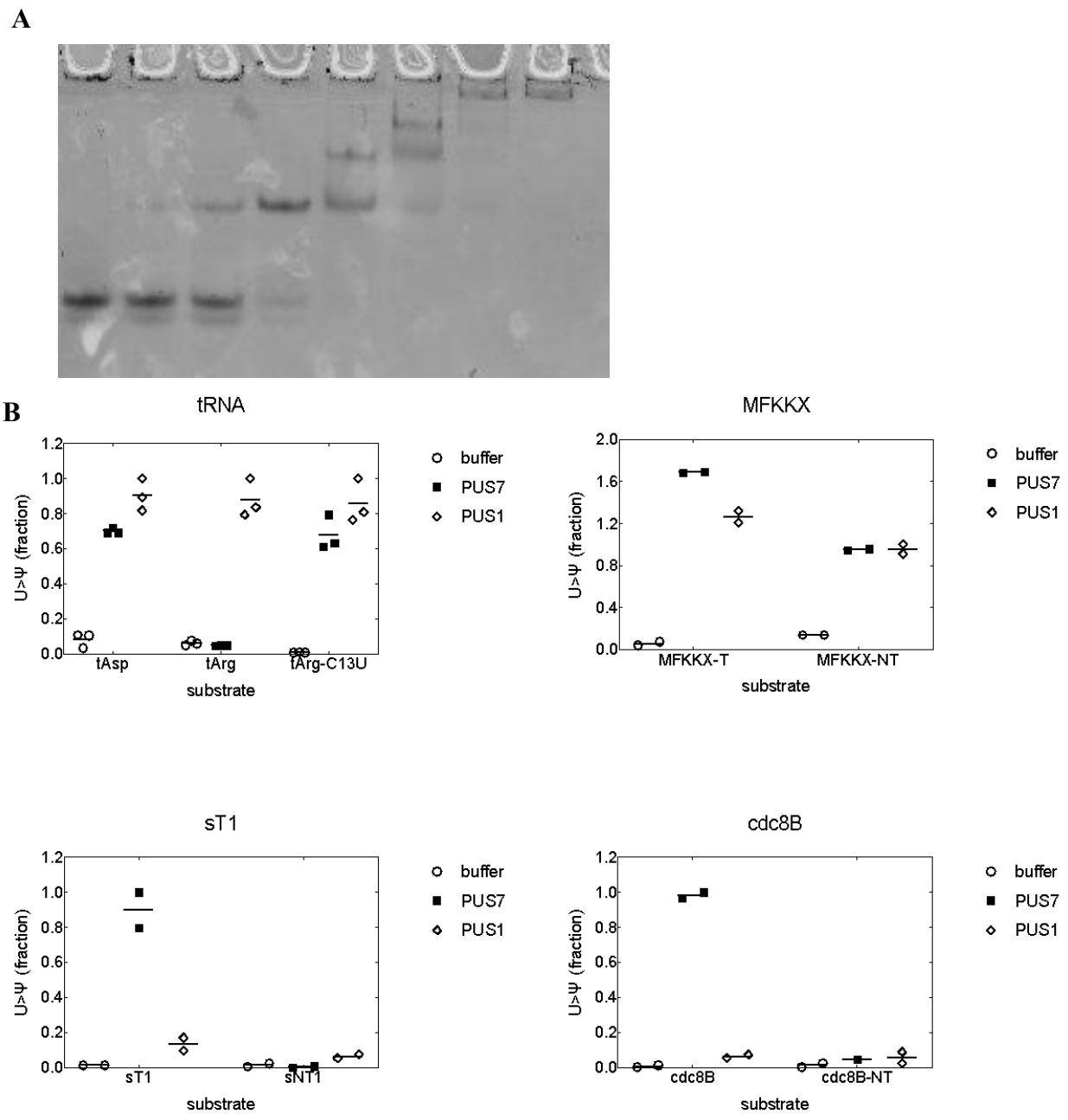






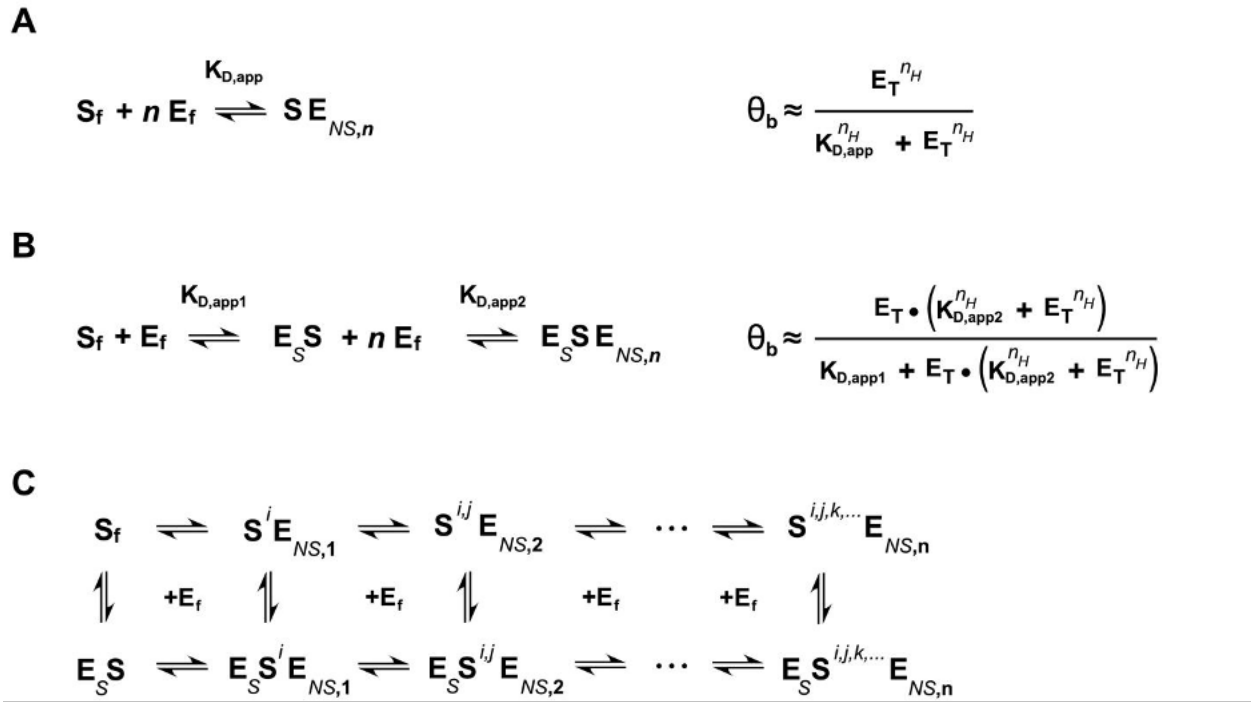
SI Figure 2.8: Raw EMSA data

Each panel is titled "SUBSTRATE PROTEIN" in bold text. Panels are grouped by substrate and then by protein mutations. Each panel shows the binding model used for curve fitting, one gel image, and a single curve fitted to all replicate data sets. The dissociation constant for the specific binding step of the model is noted along with the error of the fitted parameter.



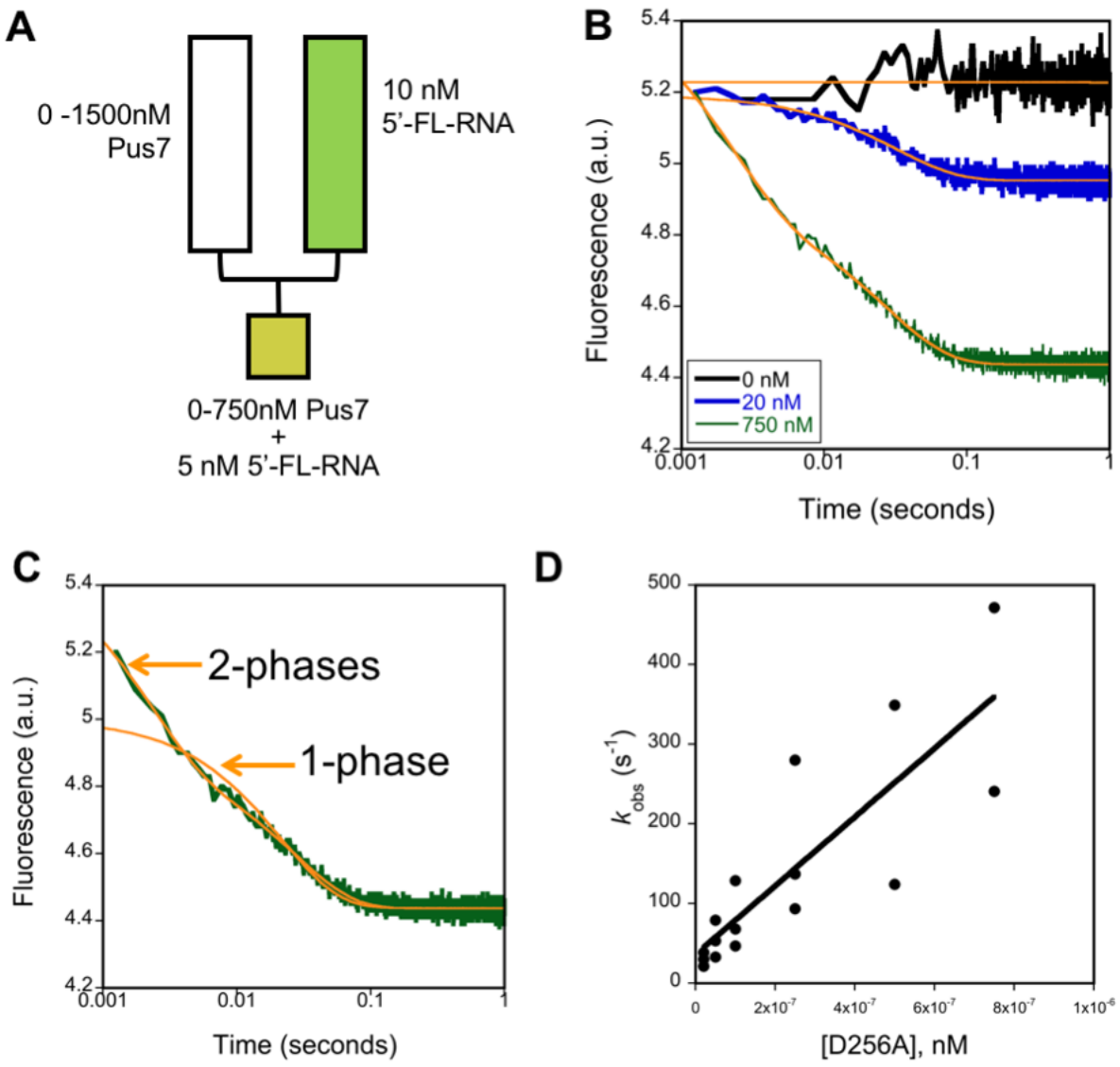
SI Figure 2.9: *S. cerevisiae* PUS1 nonspecifically binds RNA and catalyzes Ψ formation outside its consensus sequence

(A) EMSA using PUS1 and its GLK1 target RNA showing specific and nonspecific binding events. **(B)** Measurement of pseudouridine synthase activity on a variety of PUS7 and PUS1 substrate RNAs. The tRNA substrates are positive controls and show the expected pattern of activity. The MFKKX substrate contains two UGUAG motifs and mutation of one of them eliminates pseudouridylation at that site by both *S. cerevisiae* PUS1 and *S. cerevisiae* PUS7.



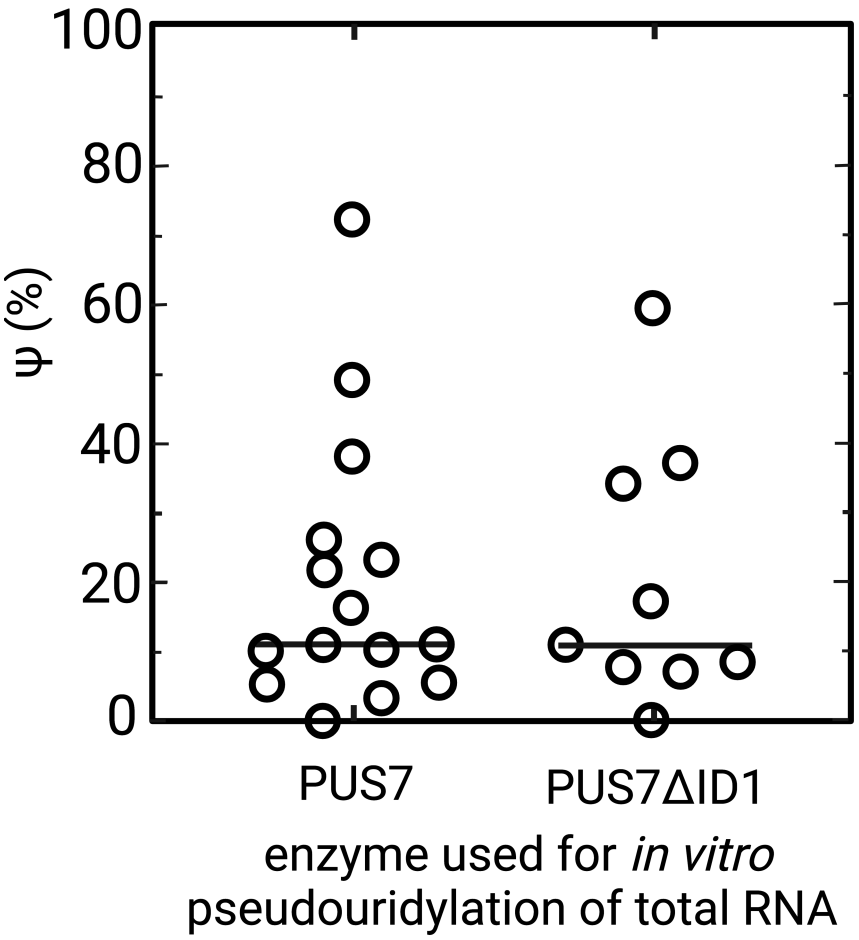
SI Figure 2.10: Models utilized in analysis of EMSA data

(A) Simultaneous binding of *n* PUS7 enzymes to a single RNA substrate, otherwise known as a Hill binding curve. This model was used when no specific binding was apparent. Binding curves were fit to the equation shown. The EMSA assay allowed direct estimation of free enzyme concentrations, so we fit the data using both free and total enzyme concentrations. The differences between these fits was much smaller than the difference in fits of data from independent replicates. Since using total enzyme as the independent variable was not the limiting factor in the precision of our measurements, we used total enzyme as the independent variable for simplicity. **(B)** Model for binding of one PUS7 enzyme to a single specific site on the substrate RNA, followed by simultaneous binding of *n* PUS7 enzymes to *n* non-specific sites on the same RNA. This model was used to fit data when the Hill equation underestimated the fraction bound at lower concentrations of enzyme, reflecting the existence of a unique site with a lower K_D for PUS7. **(C)** A realistic model for binding of one or more PUS7 enzymes, in arbitrary order, to a single specific site and one or more nonspecific sites on a single RNA. Occupancy of nonspecific sites is indicated by superscripts *i*, *j*, *k*, … on the S. Nonspecific sites can be bound in any order (e.g., *k,l,i,j*) but are depicted in alphabetical order for convenience.



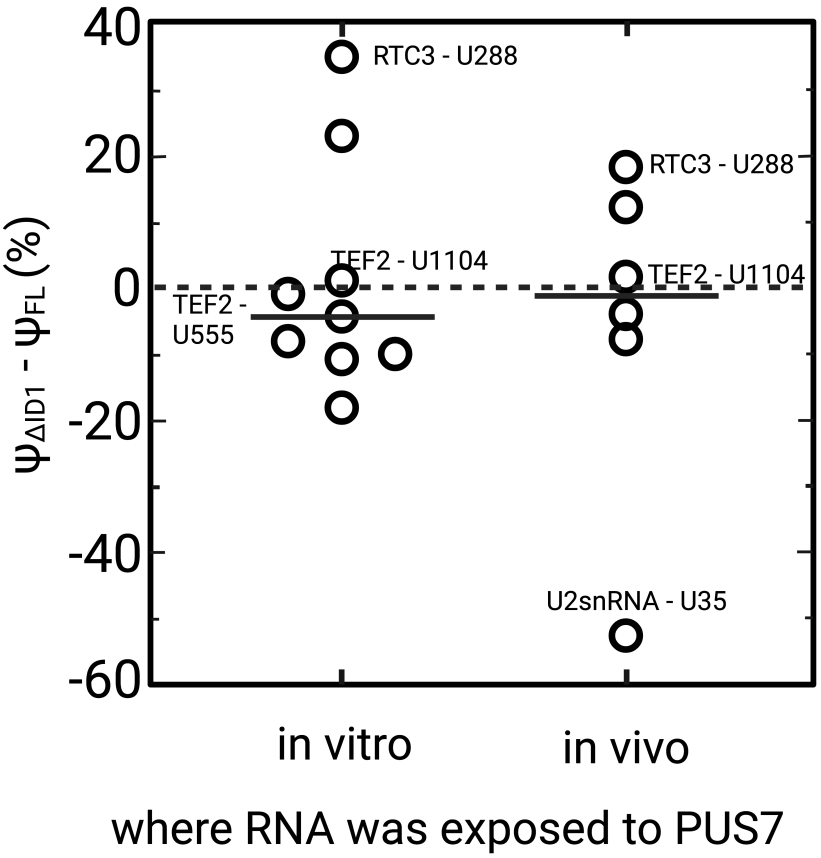
SI Figure 2.11: Stopped flow assessment of binding kinetics

(A) Experimental set-up, as described in the corresponding methods**. (B)** Stopped-flow traces of Fl-CDC8 rapidly mixed with 0, 20 and 750 nM of D256A Pus7 protein. **(C)** Traces at higher D256A concentrations were biphasic. This shows a 750 nM trace fit with one or two phases. **(D)** All of the $k_{obs,1}$ values measured are plotted as a function of D256A Pus7 concentration.



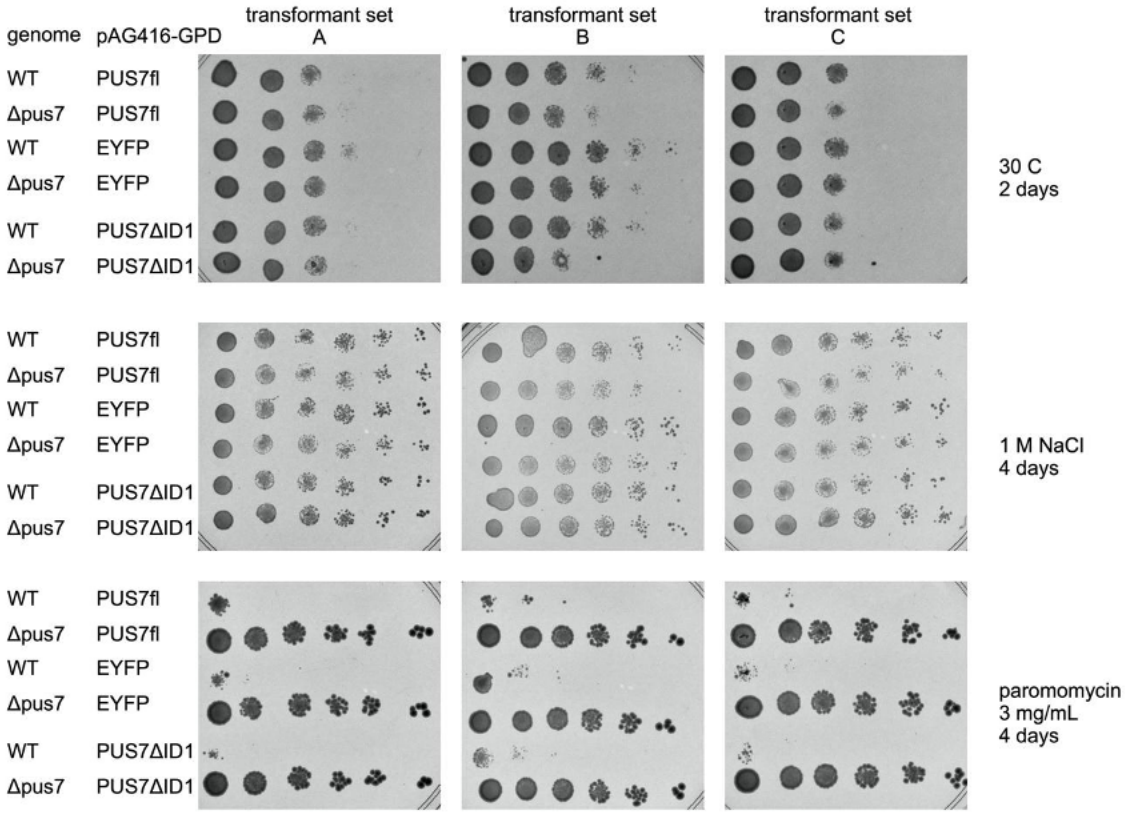
SI Figure 2.12: Deletion of ID1 does not broadly affect pseudouridylation of total RNA *in vitro*

Total cellular RNA extracted from Δ*pus7::kanMX* was pseudouridylated *in vitro* using PUS7FL or PUS7ΔID1. Pseudouridylation of known sites was assayed using CLAP (42).



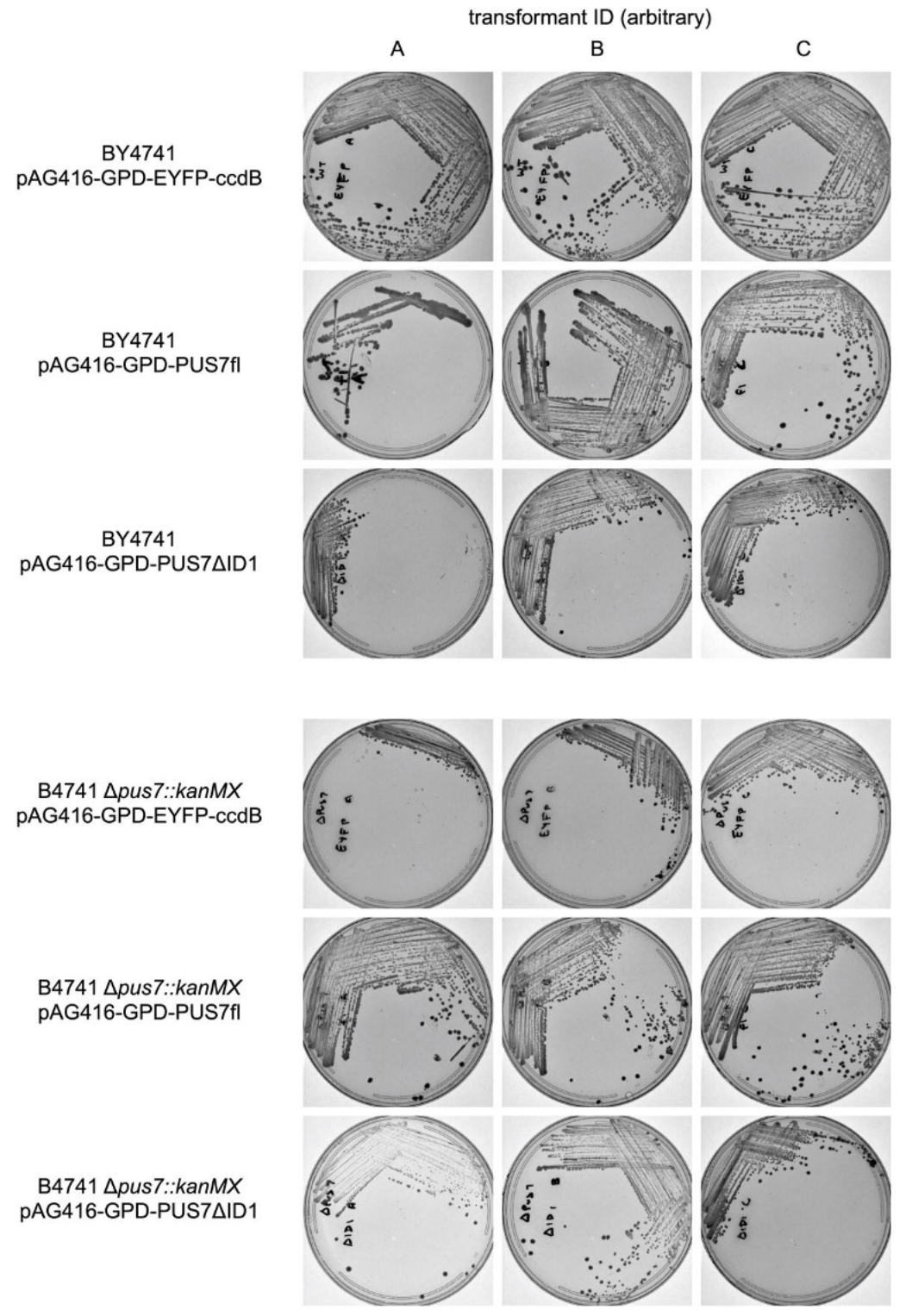
SI Figure 2.13: Deletion of ID-I influences pseudouridylation efficiency in a target dependent manner

Total cellular RNA was extracted from *pus7::kanMX* yeast and pseudouridylated *in vitro* with PUS7FL or PUS7ΔID1 (left side) or extracted from *pus7::kanMX* yeast expressing PUS7FL or PUS7ΔID1 (right side). Pseudouridylation was assayed at specific sites using CLAP (42). The difference between mean pseudouridylation level at sites in RNA exposed to PUS7ΔID1 and RNA exposed to PUS7FL is shown on the y axis.



SI Figure 2.14: Expression of PUS7ΔID1 confers no obvious phenotypic defects relative to PUS7FL

PUS7 was expressed from a CEN plasmid under the control of a GPD promoter in WT and *Δpus7::kanMX* yeast and assayed by spot plating under the indicated conditions. Three independent transformants were assayed for each plasmid.



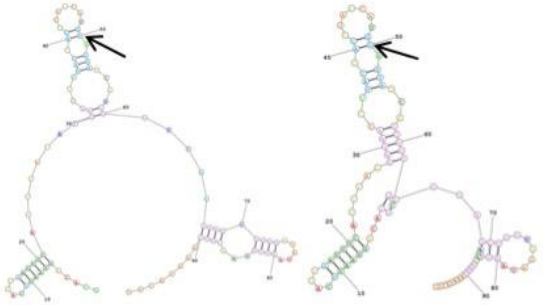
SI Figure 2.15: Isolation of PUS7FL and PUS7ΔID1 expressing clones

Three independent transformants were isolated for each strain/plasmid combination.

YKL103C

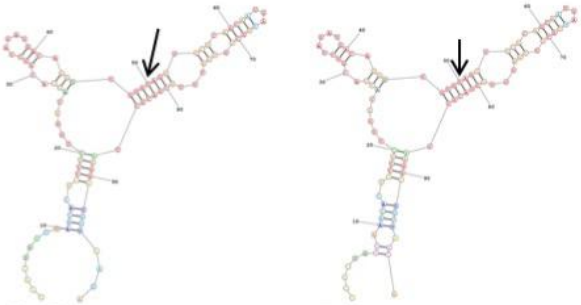
Target UGUAG

GCACAGGAAUUCAUUGAUUUCAUUUACAAGAACC
CUACCACUUACCA**UGUAG**UAUCAUUUUUCGCGG
AGCUGUUAGAUAAGCAUAACUUCAAAUACUU



Non-target UGUAG

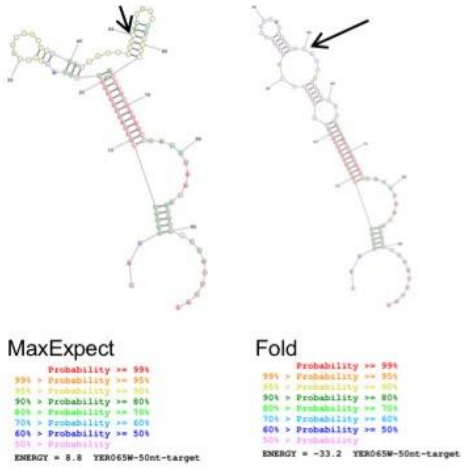
GUAUUGGUAAACACUUCCUUUUUGCACCACGUCU
AGAUGACAGGUUG**UGUAG**UUUCGCAGCAAUGAUU
GCUUUGAUUUGCUACGCUAAGGAUGUUAAUA



YER069W

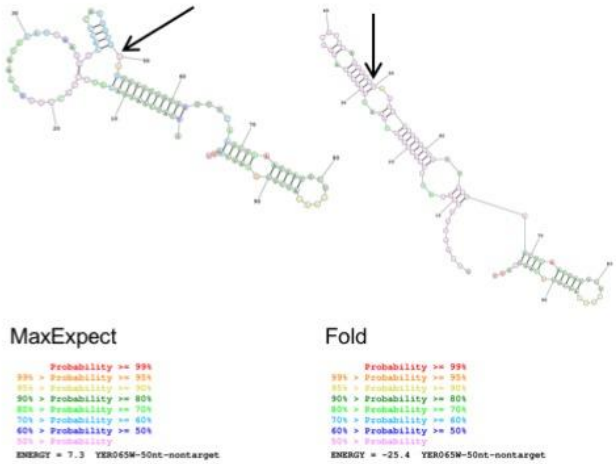
Target UGUAG

CAUGGUGUAGUUAUUGGUGGUUUCAAGCUGAAU
GAUGCUGAAGAUCG**UGUAG**UUGUUUGCGCAACC
AUCGAUAACUUACUUAAAGGCGCCGCUACUCAA



Non-target UGUAG

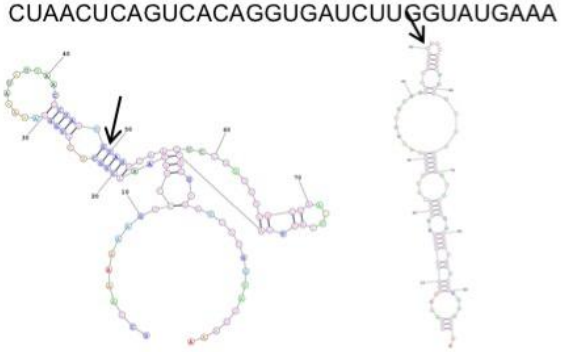
GUCAUCGAUGAUAUCCCACUGGUUAAAGAUAUUG
AGGGCACCCAUGG**UGUAG**UUAUUGGUGGUUUCAA
GCUGAAUGAUGCUGAAGAUCGUGUAGUUGUU



YMR062C

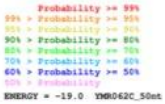
Target UGUAG

UCGAAAAAAGUCCUUGAAACUGCUCGUGGUAAAA
AUAUUAACGCUAU**UGUAG**UCAAUUCCGGUUGUG
CUAACUCAGUCACAGGUGAUCUUGGUAUGAAA



MaxExpect

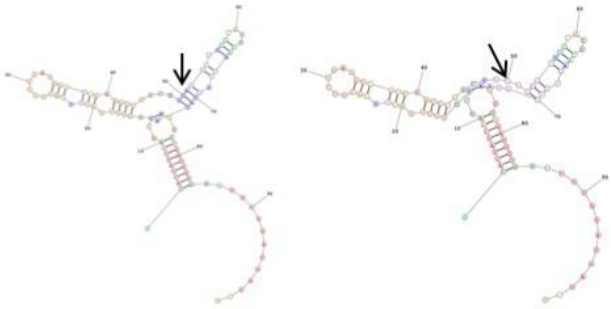
Fold



YPR176C

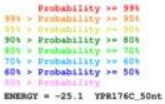
Target UGUAG

GCACUAUCAAUUUUGGGUGAAUUAACGUCUGAA
GUUGUUGACCCUGC**UGUAG**ACUUUGUACUCAAG
UGUUAUAAUUUUGAUGGUGGCUUUGGAUUAUGU



MaxExpect

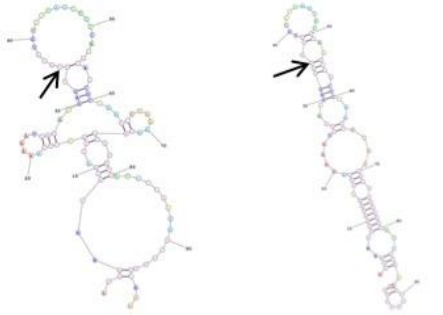
Fold



YNL336W

Target UGUAG

AUGAAAGAGAAUGAACUUAAAAAUGAGAAGAG**UG**
UAGAUGUAUUAUCCUUCAAACAGCUCGAAUCCC
AAAAGAUUGUUCUACCUCAAGAUCUUUUCAGA



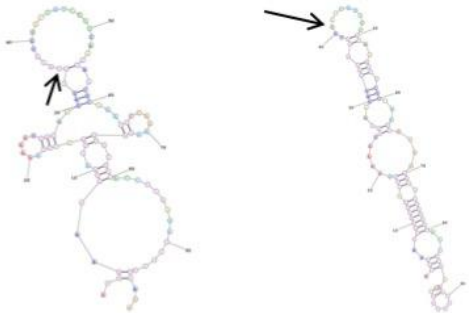
MaxExpect

Fold

YBR032C

Target UGUAG

AUGAAAGAGAAUGAACUUAAAAAUGAGAAGAG**UG**
UAGAUGUAUUAUCCUUCAAACAGCUCGAAUCCC
AAAAGAUUGUUCUACCUCAAGAUCUUUUCAGA



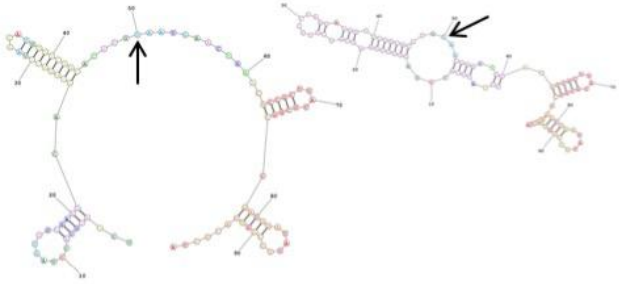
MaxExpect

Fold

YDR380W

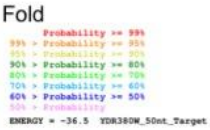
Target UGUAG

UCUAUUCACCUAGUGUUGAAUCAGCUGGCCUAA
GAUGGGUCGGCACG**UGUAA**UGAACUGAACGCCG
CUUAUGCGGCCGACGGAUAUUCCCGUUACUCUA



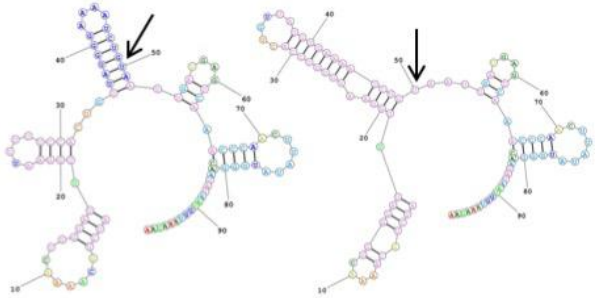
MaxExpect

Fold



Non-Target UGUAA-1

AUCUGCAAAACUGGGAUUUGGAAUUUUUCCACU
GUUAUGGGAAAAUC**UGUAA**UUGAUGAGUCAAAC
CCAACUUAUAUGGGUCAAUAUAAUGGUAAAGAA

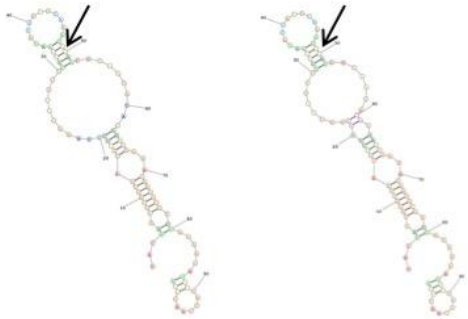


MaxExpect

Fold

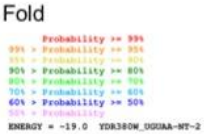
Non-Target UGUAA-2

UACAAGCGCAUUGACGUUUCUAAACUUUCUUUG
CAAUAUGAUUCAAA**UGUAA**CUCAAUAUACGAACG
AAACAAUGCGGUUAGAAGAUCCUACCAAUGGA



MaxExpect

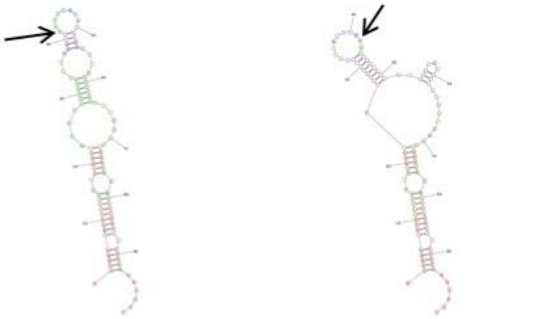
Fold



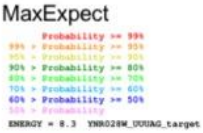
YNR028W

Target UUUAG

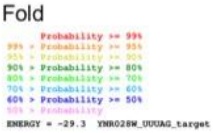
AUGAAGAGCUUUUUUCUUUAUCUUUAUGUGGCA
UUCAUG**UUUAG**UUGCAUAACGGCUCUGCCAUUA
CCUGUGGAUAACAAAAGGGCUUCUUCAGACUCC



MaxExpect



Fold



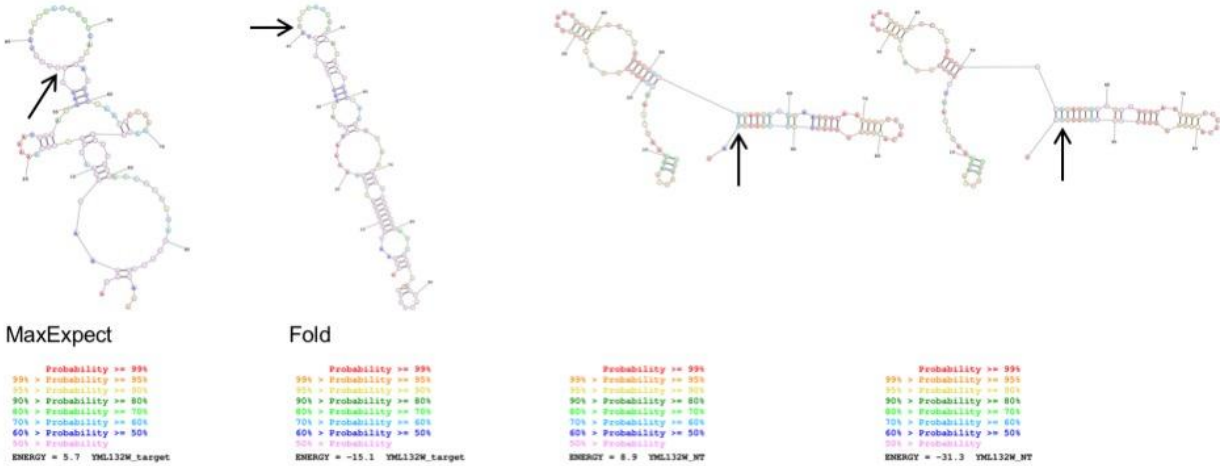
YML132W

Target UGUAG

AUGAAAGAGAAUGAACUUAAAAAUGAGAAGAG**UG**
UAGAUGUAUUAUCCUUCAAACAGCUCGAAUCCC
AAAAGAUUGUUCUACCUCAAGAUCUUUUCAGA

Non-Target UGUAG

UCUGCGAAGAAAUCUAUGUGGCCUUUACCAUUG
AAUGUGGAACUAUGGCCAUACAUUAAAGAAGCG
CAAUUAUCCCGCAGUGAGGUGCUCUUAG**UGUAG**



YJR161C

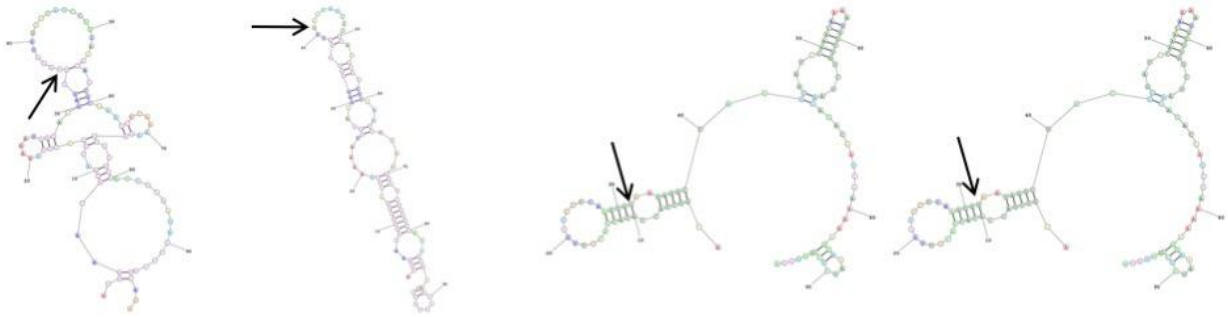
Target UGUAG

AUGAAAGAGAAUGAACUUAAAAAUGAGAAGAG**UG**
UAGAUGUAUUAUCCUUCAAACAGCUCGAAUCCC
AAAAGAUUGUUCUACCUCAAGAUCUUUUCAGA

YBR036C

Target UGUAA

AUGUCUACCACACUACUUUGGUUUUCAAG**UGUA**
AUAGGCUACGUGAUUCAAACAAAAUGUUUGUCU
AACAUACAAUCUAAAAAGGAAAUCUCCGUGGGG



YPR124W

Target AUUAU

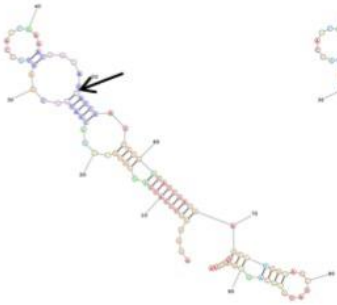
CAGUAGCUCUUCAGGGAUGGAUAUGGACAUGAG
UAUGGGAAUGAACU**AUUAU**CUGACUCCCACAUA
UAAAAACUAUCCAGUUUUGUUUCACCAUUUGCA

MaxExpect

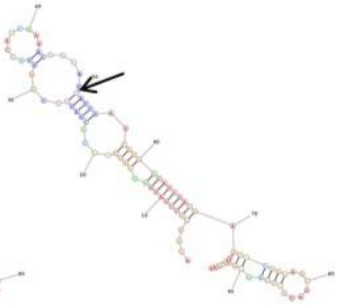
Fold

Non-Target AUUAU - 1

AUGGCGUCGAUGAGCAUGGAUGCGAUGUCUAGU
GCCAGCAAAACGAU**AUUAU**CGAGCAUGUCAUCG
AUGAGCAUGGAAGCGAUGUCCAGUGCGAGCAAA



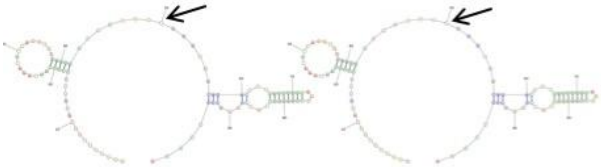
MaxExpect



Fold

Non-Target AUUAU - 2

UAAAUAUUCCACUUUACCUUCAGCAAACUCCAAA
GACGAAGGAAAAC**AUUAU**GACACAGAGAAUAAUU
UUGAAAUUCAAGGUUUACCUAAGCUGCCAAA

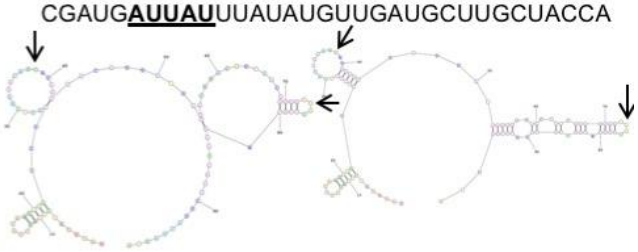


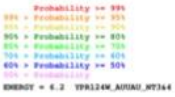
MaxExpect

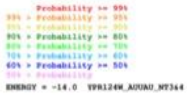
Fold

Non-Target AUUAU – 3 & 4

UAUUUGUUCCAUCUUUAAUGGAUCUCUUUCAUG
AC**AUUAU**AAGGGCGUUCUUAGUAUUUACCUCUA
CGAUG**AUUAU**UUAUAUGUUGAUGCUUGCUACCA



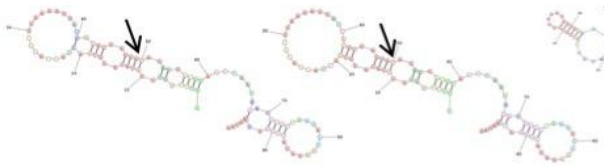




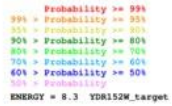
YDR152W

Target UGUAG

GGCGUGGUGGGCACAGGCGACGAGUAUAUGGAA
GAGGAUGACGCGAG**UGUAG**AUGACGUAGCCAAG
GGACUUGCCAAGACCGAAAUAGCAAAUCAAUAA



MaxExpect

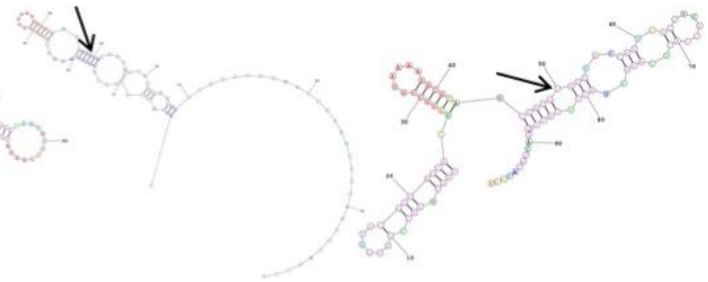


Fold

YLR258W

Target UGUAG

UGGUUGGCCGGUGUUGCGUUACCAUUAUGCCG
UAAAAGGCGUAUCGA**UGUAG**UUACCAUUtgtUUCA
CCACUCAUGCUACUUUAUUGGGACGGUAUUUA



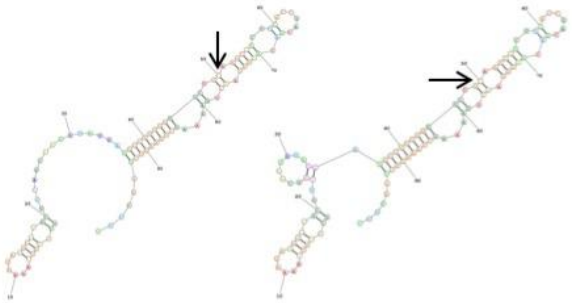
MaxExpect

Fold

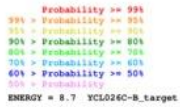
YCL026B-C

Target UGUAA

GCCGGUCAAAAGAGGCCUGCUUCAGCAAGGGAU
GAGGCCUUUGGUUC**UGUAA**UCUUCUUCACCGAC
GACAAGGUAACUGAAAAGCUAAAGGCUGACUUC



MaxExpect

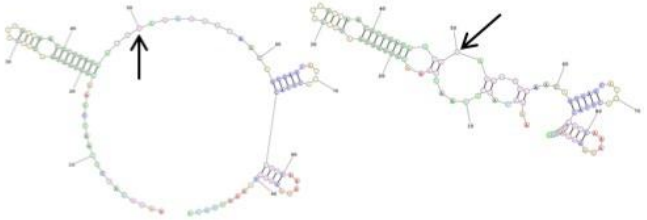


Fold

YDR497C

Target UGUAG

AAGACAUCGCAAUCAAAUGUUGGUGAUGCCGUU
GGCAACGCUGAUAG**UGUAG**AGUUCAACAGUGAG
CAUGACUCACCUUCAAAGAGAGGUAAAAUUACA



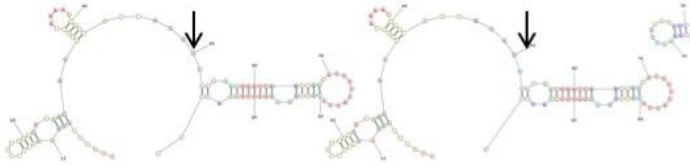
MaxExpect

Fold

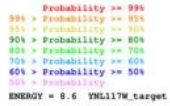
YNL117W

Target UGUAG

AUUAUCGUGAGACCCCGUGGUUGGCACAUGGUG
GAAAAGCACCUUUA**UGUAG**AUGAUGAACCAAUCA
GCGCUUCCAUCUUUGAUUUUGGUUUAUAUUUC



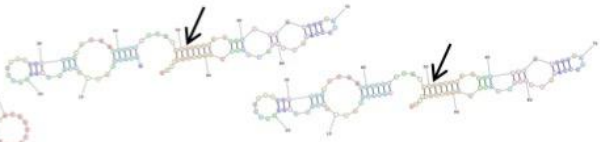
MaxExpect



Fold

Non-Target UGUAG

AAGAUGGAGCACCACUUGGAAGCUAAACUAUGG
AACGACGUCUUCUG**UGUAG**CUCAAGAUUACAUU
GGGAUCCCAAGGGGUACAAUCAGAGCUACUGUG



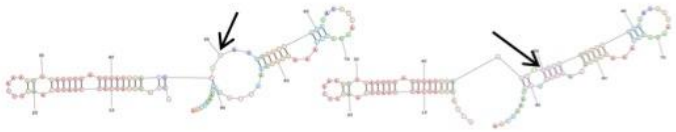
MaxExpect

Fold

YBR122C

Target UGUAG

GCUAUUUAUCACCAAUUCAAUGUCAAAAUGGAGU
UGAGUGAUGGAAG**UGUAG**UUAUCCGGAGAUCCC
AAUAUCCAAAGGGUGAAAUUAGAUUAAUUCAA



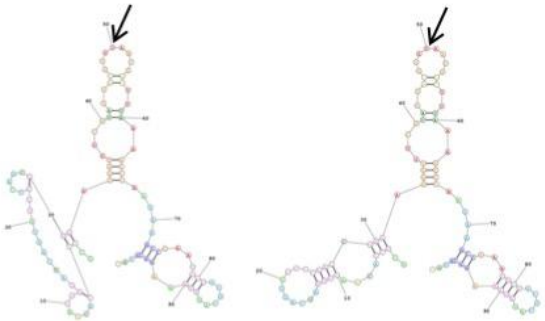
MaxExpect

Fold

YNR036C

Target UGUAA

GCUCCACAACUGGAUCAGUGUCCCCAGCGGAAG
GGUGUCGUAUUACG**UGUAA**UGGUUUUAAAGCCC
AAGAAGCCGAAUUCUGCGCAGAGAAAGGCGUGC



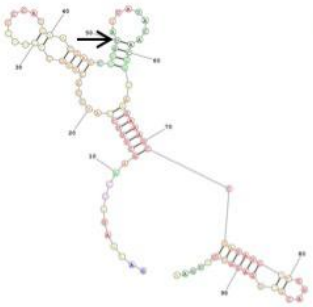
MaxExpect

Fold

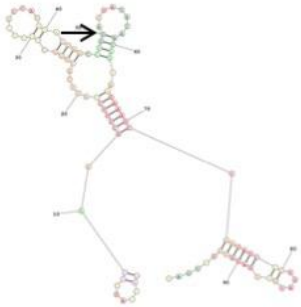
YCL043C

Target UGUAG

GAUGAUUUCAAGCUUUCUAUUUACUUGCCCUCC
GCCAUGGACGAGCC**UGUAG**UAUACAACGGUAAG
AAAGCCGAUAUCGCUGACGCUGAUGUUUUUGAA



MaxExpect

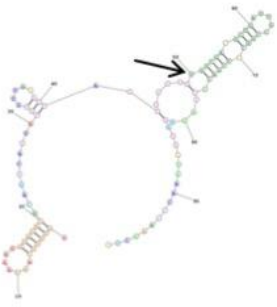


Fold

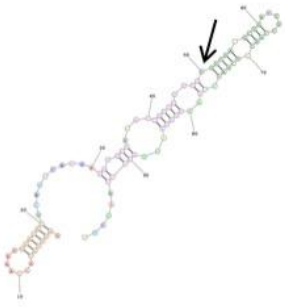
YOL038W

Target UGUGU

GGUGUGCAACAAAGGUACACGCAGUCAGGAGGU
GUUAGACCAUUUGG**UGUGU**CGACGCUGAUUGCC
GGCUUCGACCCGAGAGAUGAUGAACCCAAGCUU



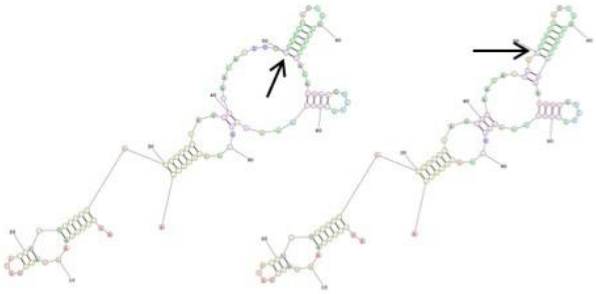
MaxExpect



Fold

Non-target UGUGU -1

AGGCGGUAAAGAGGGGUACCUGUGCUGUAGGUG
UCAAGGGUAAGAAU**UGUGU**AGUAUUGGGCUGCG
AAAGAAGGUCUACUUUGAAGCUGCAAGACACUA

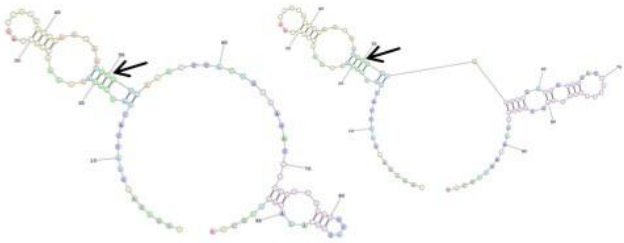


MaxExpect

Fold

Non-target UGUGU - 2

UGGAGAAGAAUUACGAUCGCAAAGAACCACCAGC
CACAGUGGAAGAA**UGUGU**CAAACUUACUGUAAG
AUCUCUGUUGGAGGUAGUUCAAACAGGUGCAA



MaxExpect

Fold

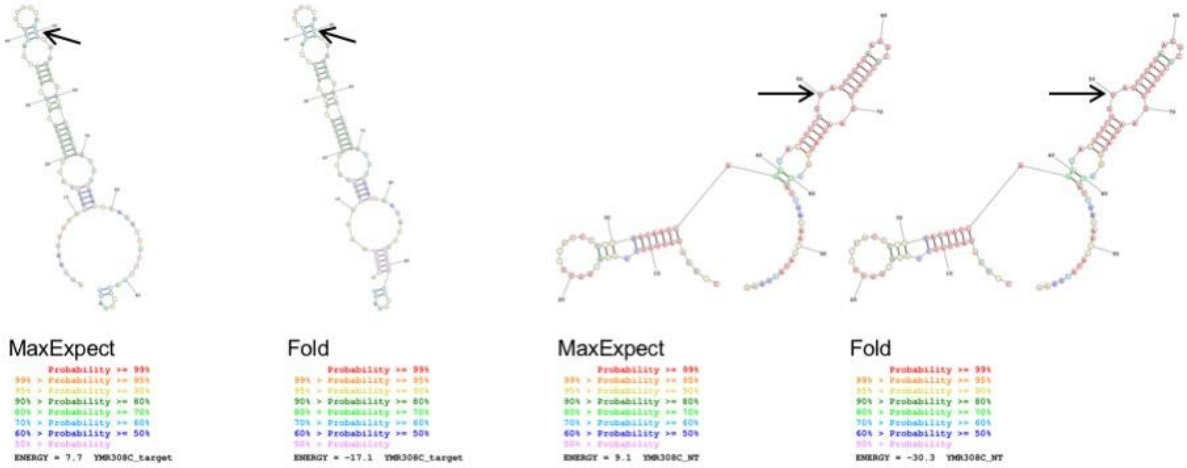
YMR308C

Target UGUAG

AUUGAAAAUAAUUCGCCAAUUGUGUGUGCUCAA
UCUAAUAUCUCCGC**UGUAG**UUGAUUCAGUCAUA
CAAGCCUUGAAUGAGAGAAGUUUGACCGAAAGG

Non-Target UGUAG

CUUCAGAUUGUCCAGGCGUUUGCUUCCCCUGAC
AAUCAAAUACGUUC**UGUAG**CUGAGAAGGCUCUU
AGUGAAGAAUGGAUUACCGAAAACAAUAUUGAG



YHR141C

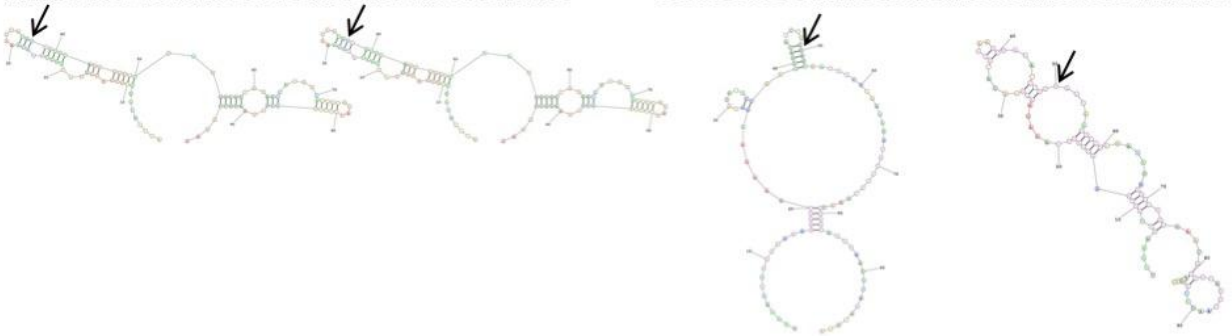
Target UGUUA

AUGGGUAUGUGGACGAUUAGGAAUAGACAAACC
A**UGUUA**UUUAUCUCCAUUAGGGCGUGAGAGUGU
AAUUAGUACACAGGUACUACUAGAAUGCUAAAG

YDR322C-A

Target UGUAG

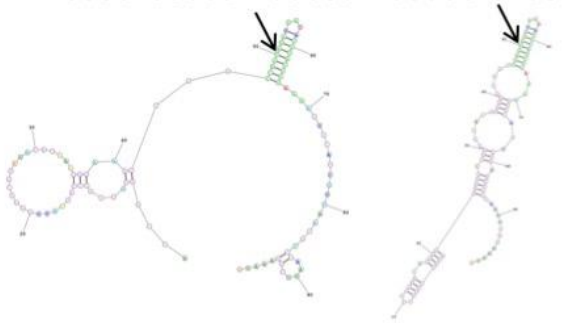
UUGAAGCUGGUAGAGGAGGCAAAGAAGGAAUAC
GCCAAGCUACACCC**UGUAG**UAACUCCUAAAGAU
GUGCCUGCGAACGCCUCAUUUAAUUUGGAAGAU



YOR305W

Target AGUAG

ACCGGUGGCGCCCAUGAUGGUGGCGUAGACAUA
AAGGGAAGCUGGCC**AGUAG**AUGAUAUUUAUUGG
AAAAUUUCAUCGUUAAUGCCCAAUUUGGAAAUG

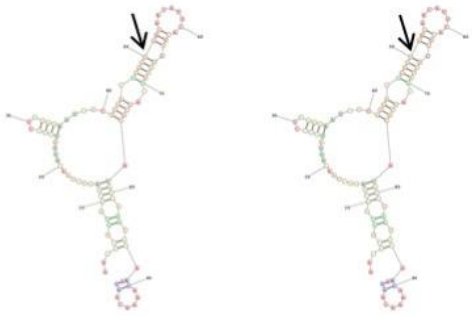


MaxExpect

Fold

Non-Target AGUAG

AAGGCUUUCACUAAAUCAAAAUUAUCUCCUAGAG
AAUUCCGUGAGUU**AGUAG**GAACAUUCACCUCAC
UAGUAUCACAUAGUCAACGCAACAAAACAGUA



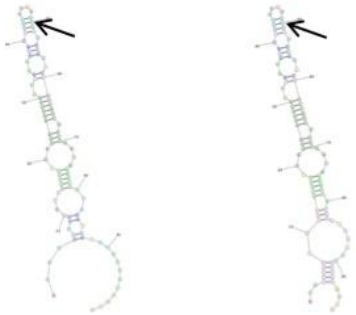
MaxExpect

Fold

YPR198W

Target UGUAG

AGAGUCGUUGCUGGGUUUGGAGGAAGUGGAAUU
GAAUCACUUGCUUU**UGUAG**UUGGAACAUCCAUU
GUCCGAGAAAACCAUAGAGGAAUUAUGAUAACG

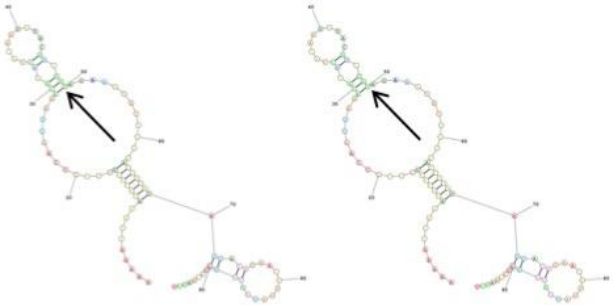


MaxExpect

Fold

Non-Target UGUAG

AAAAACCAGAUGCGCUCUUCCCAAUUAAACAUAC
CACAAUUUACGUC**UGUAG**AAACACUUUUAGCGU
AUAGCACAGAACAUUAUGAUGGCCCCCAAUCU



MaxExpect

Fold

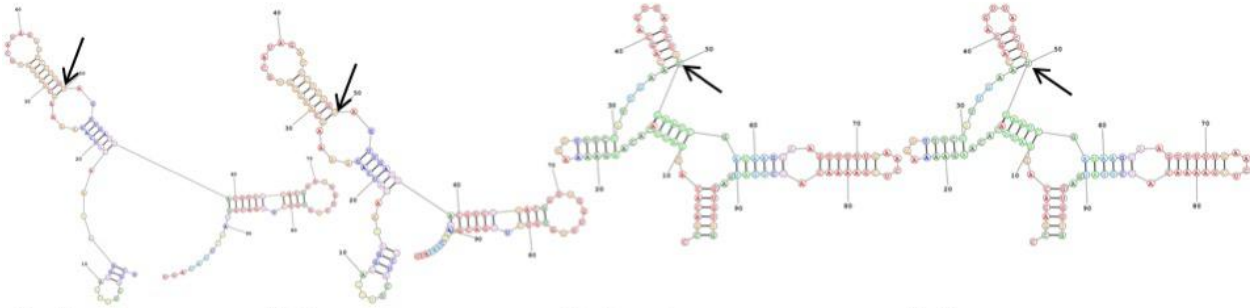
YGL115W

Target UGUAG

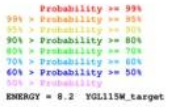
UCUACUAUUAUGGAUAACAUCAGAAAAGCAAGGG
UGCAUAGAUUCUU**UGUAG**UUGAUGACGUCGGAC
GGUUGGUUGGUGUCUUGACGUUAAGCGAUAUU

Non-Target UGUAG

CCGACACAGGAUUCACAAGAAAAGGUUUCUAUU
GAACAGCAGUUAGC**UGUAG**AAUCGAUAAGGAAG
UUUUUGAACUCGAAAACAUCUUAUGACGUGUUG



MaxExpect



Fold

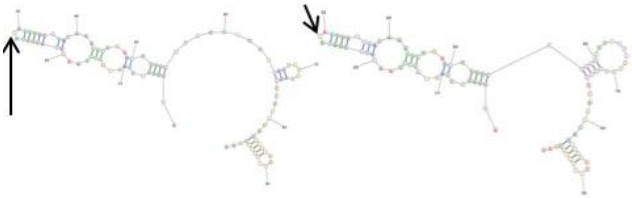
MaxExpect

Fold

YPR009W

Target UGUAA

AUGAAGCCGAACAAUCGAACUUGUGA**UGUAA**UU
ACCAACAAAGAUGAAUCUCUUCCUGCACUUUUG
CUGCCUGCACUGAACAGUUACACCUGUGACGAU



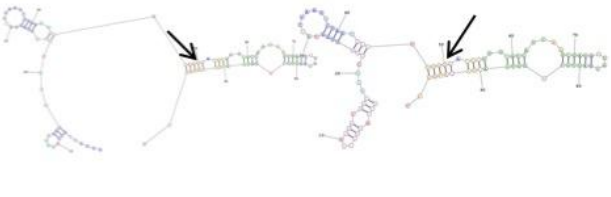
MaxExpect

Fold

YER150W

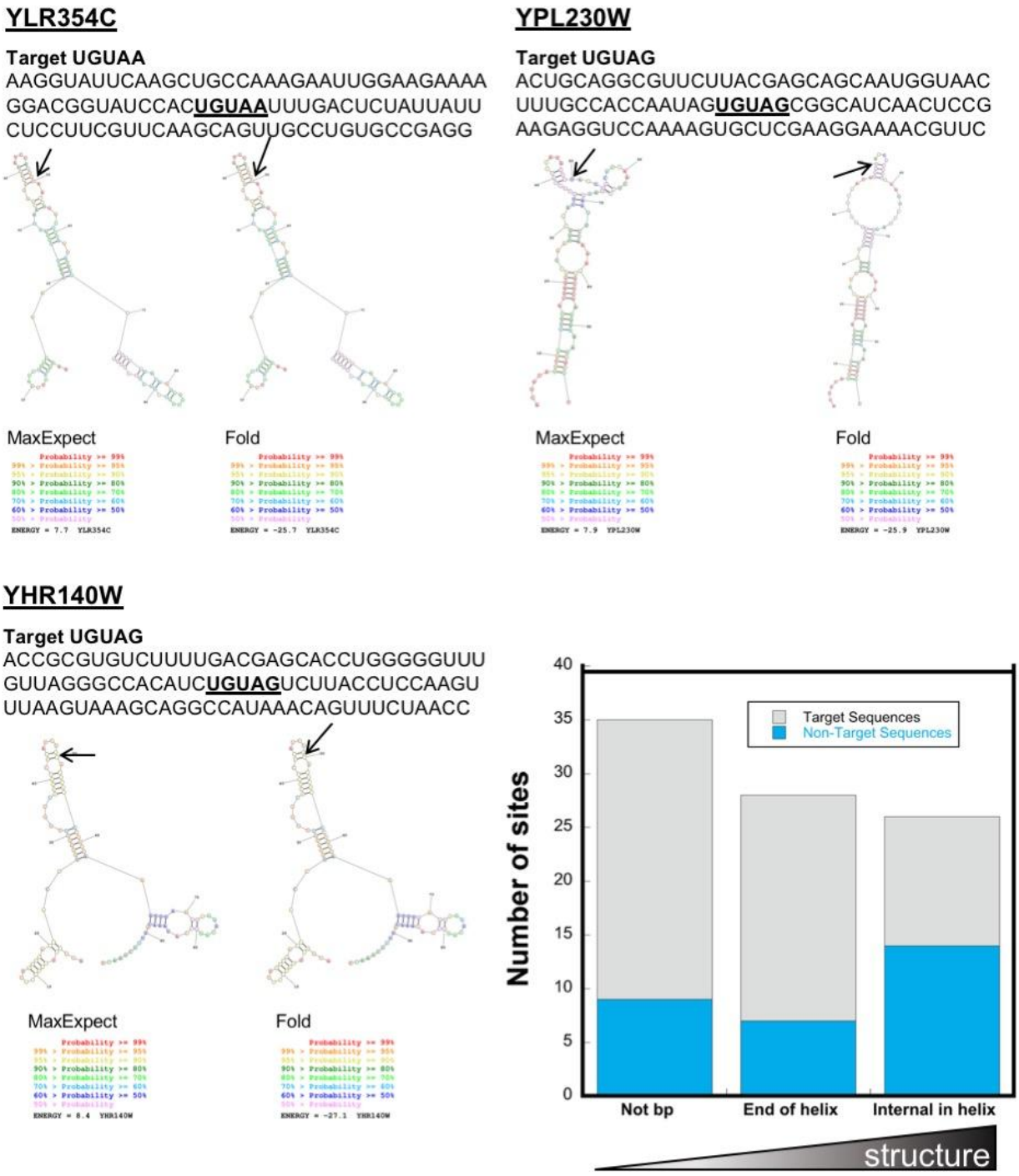
Target UGUAA

GCCAUGGCCUCUACGGCUCUCGGAUUGGUAUCU
AAUUCUAGUUCCUC**UGUAA**UCGUGGUACCAUCA
AGCGAUGCUACUAUUGCCGGUAACGAUACAGCC



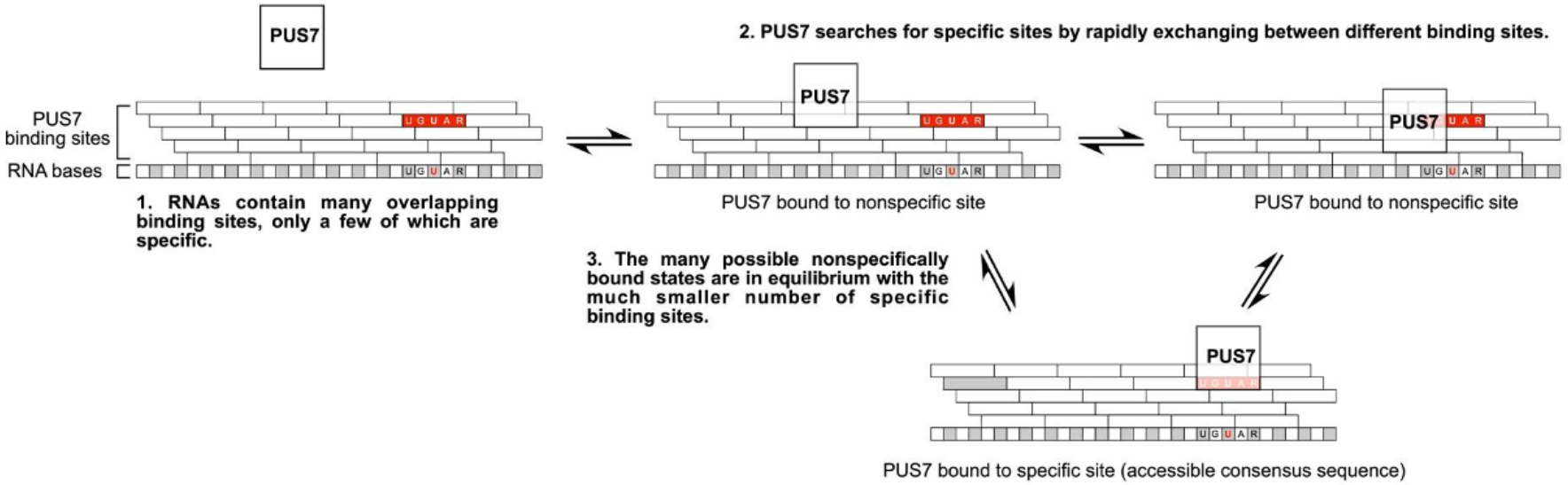
MaxExpect

Fold



SI Figure 2.16: Secondary structure prediction of Pus7 modified sites in mRNA coding regions

Sites as reported in Carlile, et al. Nature (2014) (23).



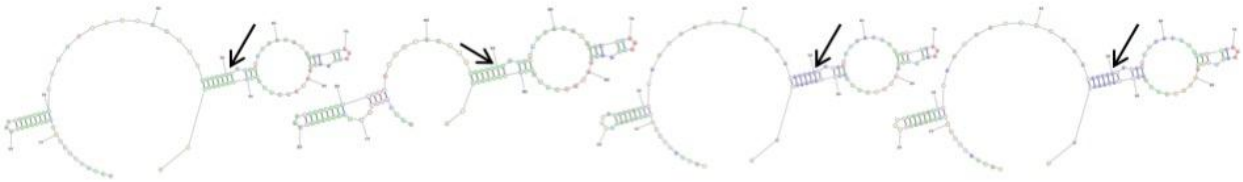
SI Figure 2.17: Model - Pus7 rapidly samples RNAs for specific modifiable sequences

RNAs contain multiple (often overlapping) potential Pus7 binding sites. These sites have varying degrees of accessibility to Pus7 due to their secondary/tertiary structures or occlusion by RNA-binding proteins. Pus7 rapidly samples all accessible sites on a given RNA, forming nonspecific interactions with most sequences. When Pus7 interacts with a modifiable (e.g. UGUAR) sequence, it forms a tighter, 'specific' interaction that results in Ψ installation. Only a handful of the potential Pus7 sites are modifiable and 'specific'.

YBL030C

Target UGUAG – 30°C **45°C**

UACAAUGCUCUAAUGAAAAUAUUUUUCAUUGUAU
CUACCGCUUACAU**UGUAG**UGCUAUUACAAGGGU
CUAAAAGAACCAACACCAUUGCGUAUAAUGAA

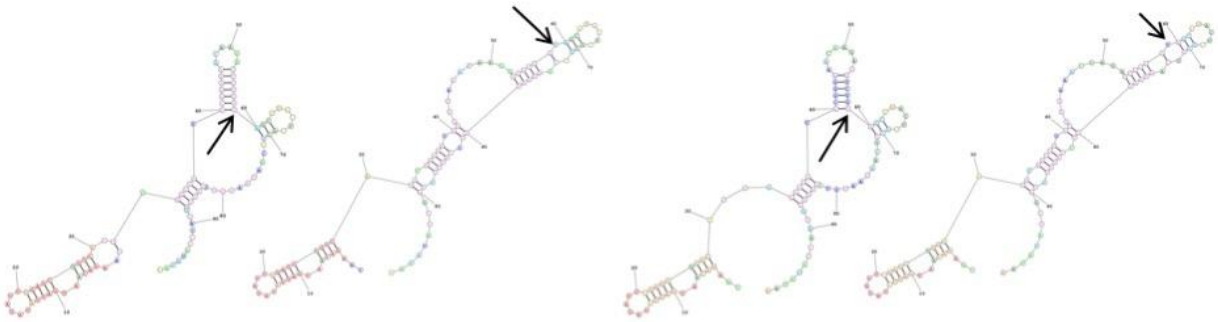


MaxExpect Fold MaxExpect Fold

YBR103W

Target UGUAG – 30°C **45°C**

CAAGACAAUGAUUAUAUAUUUGAUCUGUCAUGGA
ACUGUGCGGGAAACAAGAUUUC**UGUAG**CAUACU
CGCUUCAAGAAGGUUCAGUUGUAGCCAUAUAA

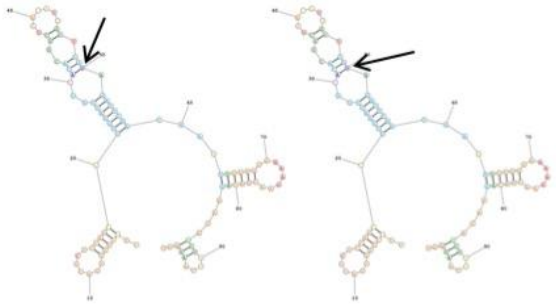


MaxExpect Fold MaxExpect Fold

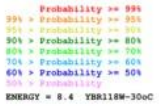
YBR118W

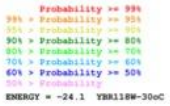
Target UGUAG – 30°C **45°C**

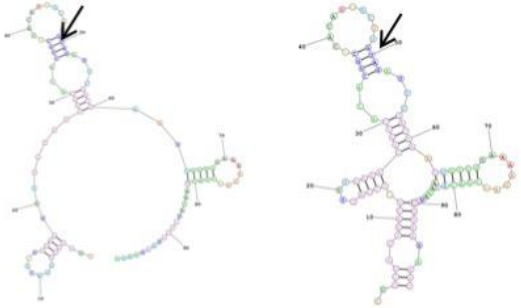
CUGCUGGUUACUCUCCAGUUUUGGAUUGUCACA
CUGCUCACAUUGCU**UGUAG**AUUCGACGAAUUGU
UGGAAAAGAACGACAGAAGAUCUGGUAAGAAGU

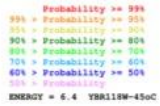


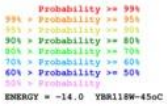
MaxExpect Fold MaxExpect Fold







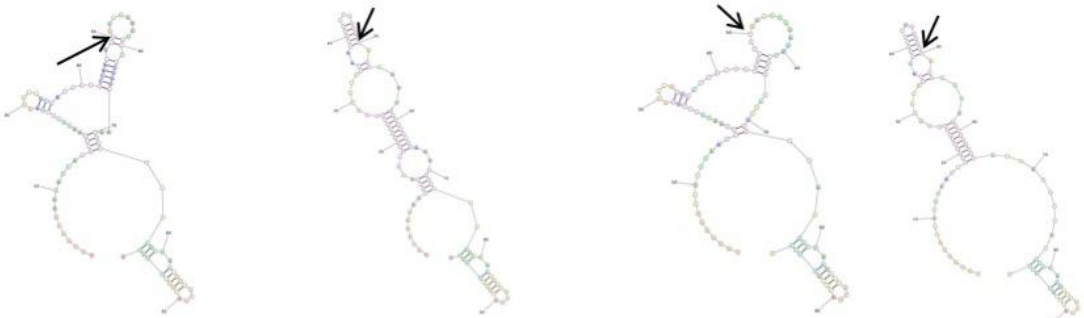




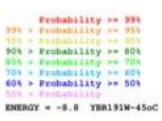
YBR191W

Target UGUAG – 30°C **45°C**

UAGAAAAAAGAUUAAACUUAAGAGUUGAACACAU
CAAGCACUCCAAG**UGUAG**ACAAGAAUUUUUGGAA
AGAGUUAAGGCCAAUGCUGCUAAGCGUGCUG



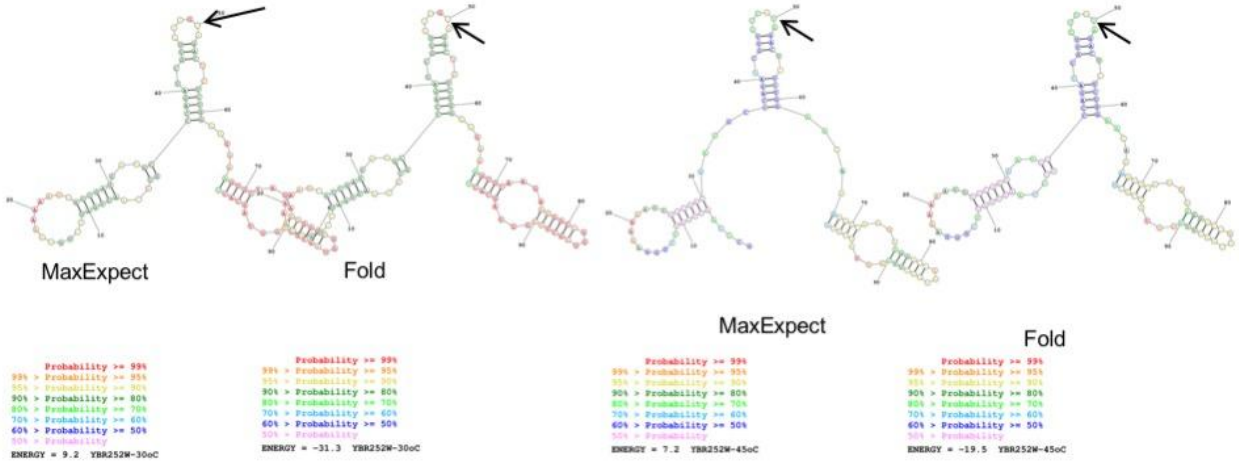
MaxExpect Fold MaxExpect Fold



YBR252W

Target UGUAG – 30°C **45°C**

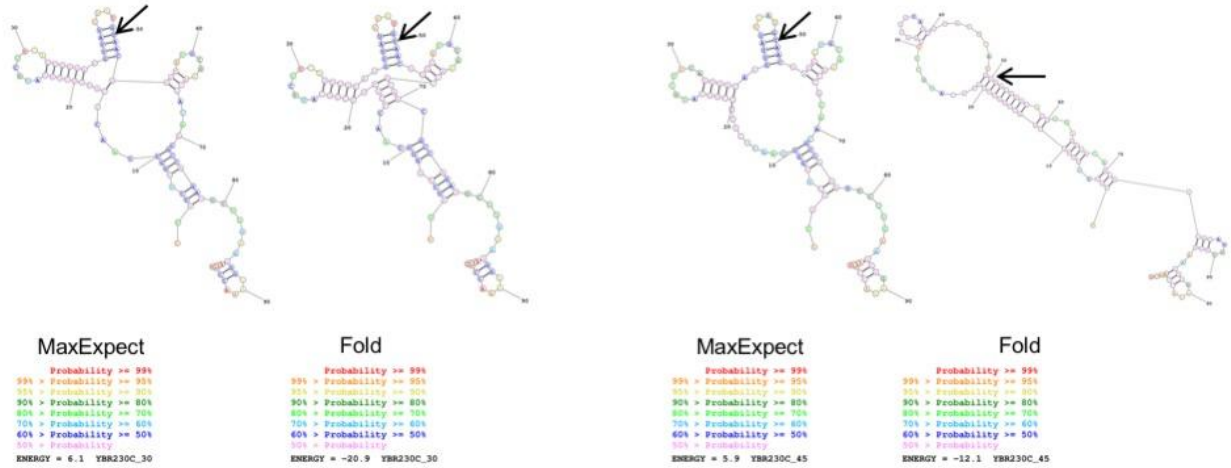
GCCCAAUUGAUUCUGGAAAAAAUUGUCGAUGAU
GCCCAGAUCGUUGU**UGUAG**ACUCUCUGGAAGAA
AGUGCAAGAGGGGCCGGUGGCUUUGGUAGCACU



YBR230C

Target UGUAA – 30°C **45°C**

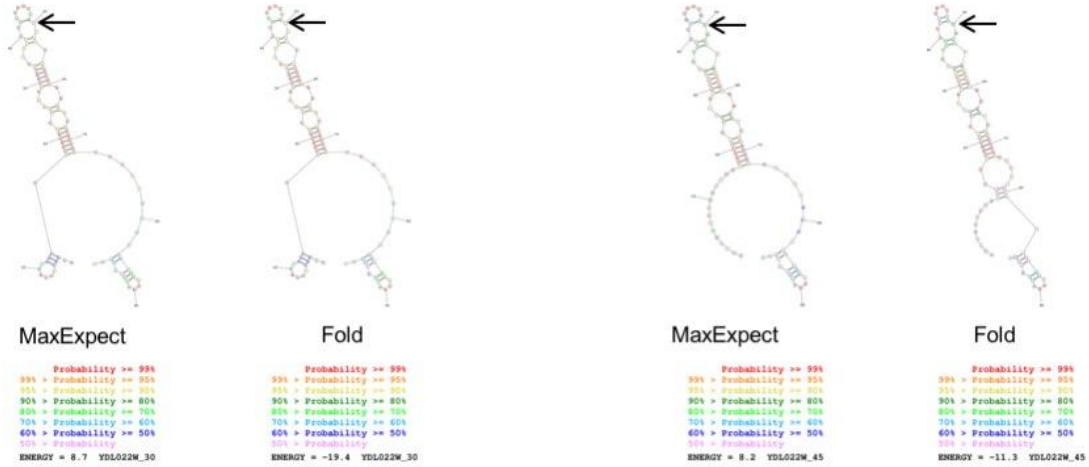
UUGUCUUGGUCAACUUGUUGCUGCAUUCUUCAG
UGGUAUGUUAUUUA**UGUAA**CGGGUAUGCGAACC
ACAACGCCAGAUUCUUGAAGGGGAAACCUAACU



YDL022W – U403

Target UGUAG – 30°C **45°C**

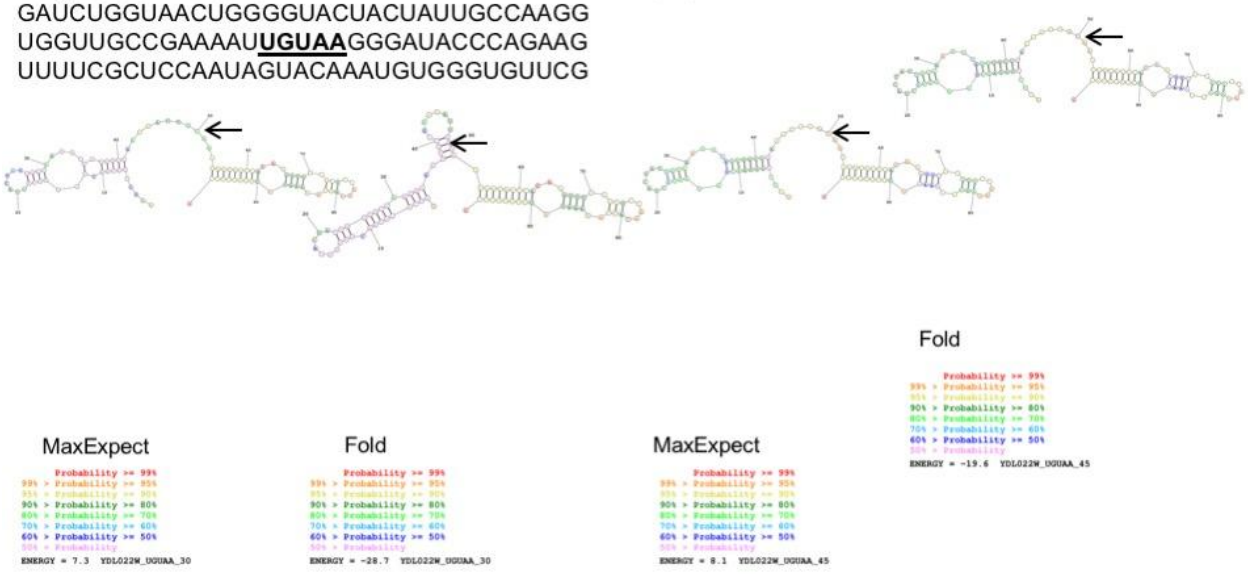
UCGACAUCAUCGUUUUCAACAUUCCACAUCAAUU
UUUGCCCCGUAUC**UGUAG**CCAAUUGAAAGGUCA
UGUUGAUUCACACGUCAGAGCUAUCUCCUGUC



YDL022W – U172

Target UGUAA – 30°C **45°C**

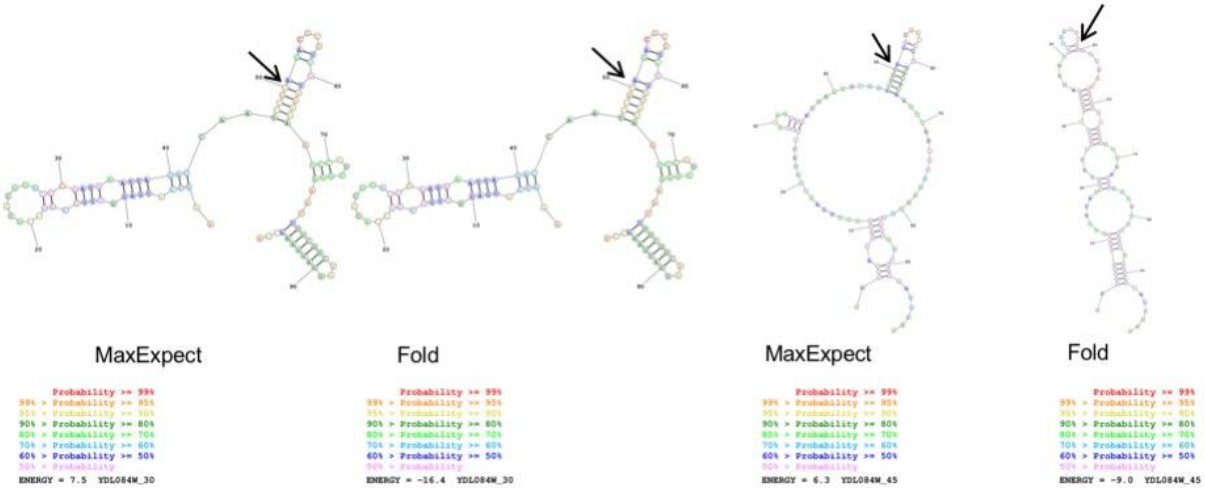
GAUCUGGUAACUGGGGUACUACUAUUGCCAAGG
UGGUUGCCGAAAAU**UGUAA**GGGAUACCCAGAAG
UUUUCGCUCCAAUAGUACAAAUGUGGGUGUUCG



YDL084W

Target UGUAA – 30°C **45°C**

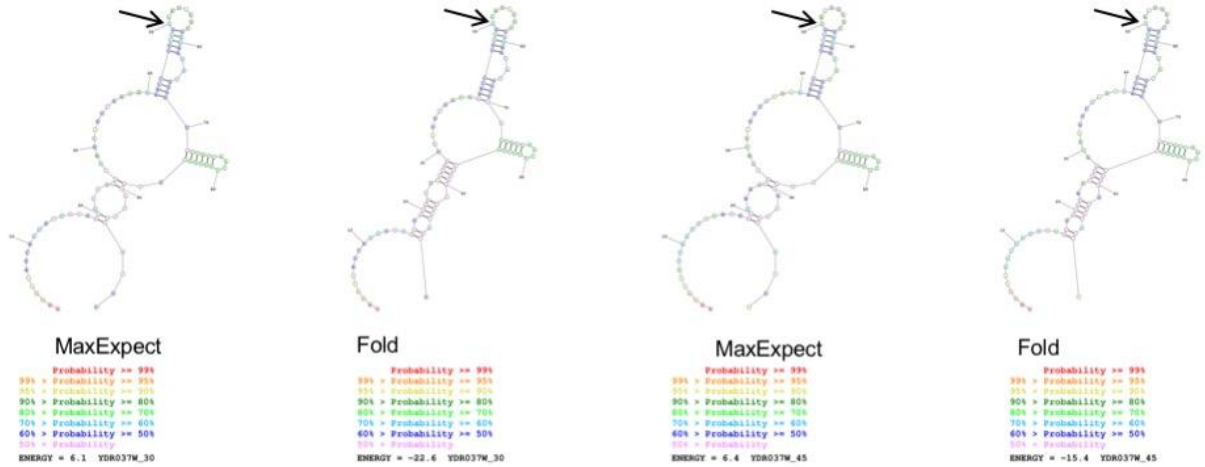
AAGUCAUGAUGUUUUCAGCCACACUUUCUCAAGA
AAUUAGACCAAUU**UGUAG**ACGCUUCUUACAGAAU
CCAUUGGAAAUUUUCGUCGAUGAUGAAGCUA



YDR037W

Target UGUAA – 30°C **45°C**

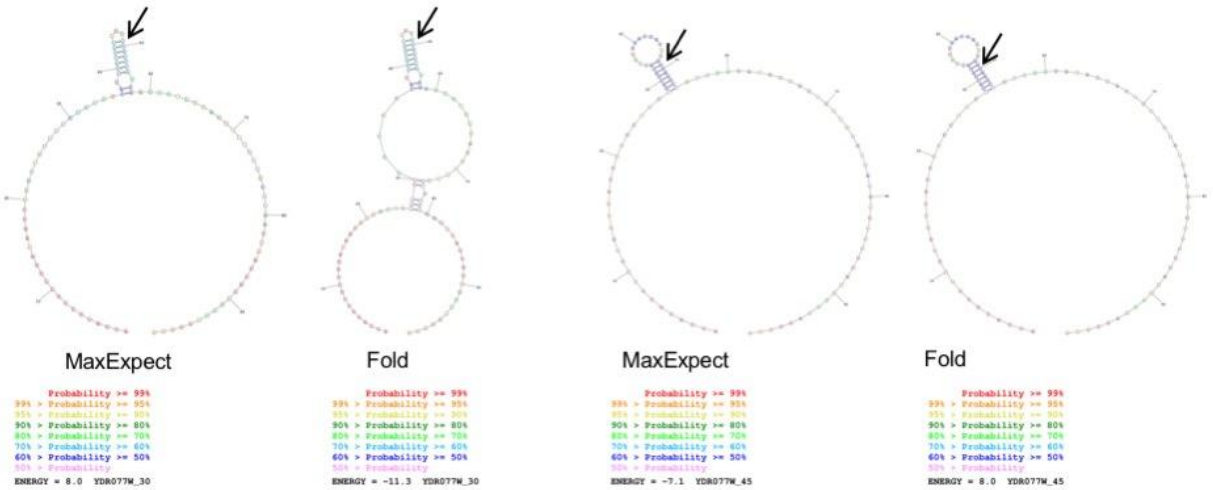
AAAAGGAUCAAGGUGAUGACGAAGCUCAAUUAGU
CGAUGAAACCUUC**UGUAA**UGCUCUAGAAUACGG
UUUACCACCAACUGGUGGUUGGGGUUGUGGUA



YDR077W

Target UGUAG – 30°C **45°C**

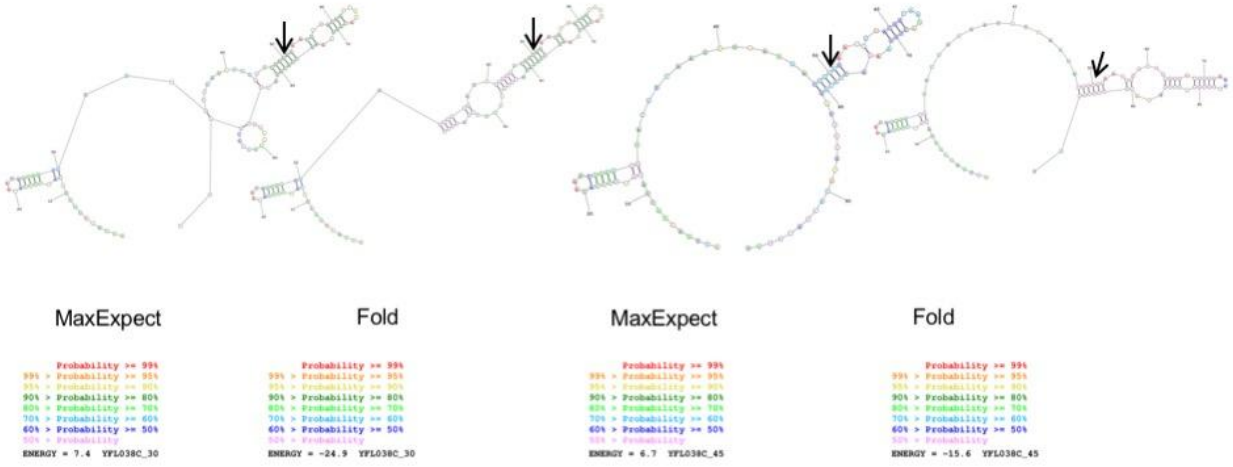
ACCACUCCUCCUUACAACCCAUCUACUGACUACA
CCACUGACUACAC**UGUAG**UCACUGAAUAUACUAC
UUACUGUCCAGAACCAACCACUUUCACCACA



YFL038C

Target UGUAG – 30°C **45°C**

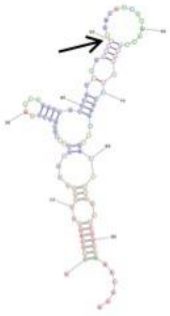
ACCAACGACUACAUCUCCACAAUUGGAGUGGACU
UCAAGAUUAAGAC**UGUAG**AACUGGACGGCAAGA
CUGUAAAGCUACAGAUUUGGGACACUGCAGGU



YGL116W

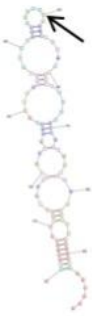
Target UGUAG – 30°C **45°C**

AUUGCAGAUCUGGACGUUCUUUUGGGCCUAAAU
GGUCGCUCGUCGGA**UGUAG**ACAUGACAACCACA
UUGCCGAGUUUGAAGCCACCUCUGCAAAACGGA



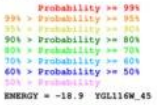
MaxExpect

Fold



MaxExpect

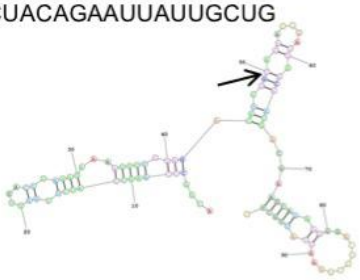
Fold

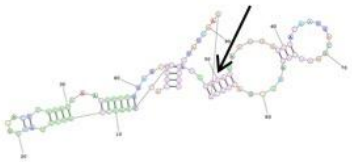


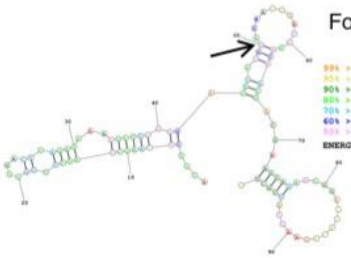
YIL076W

Target UGUAG – 30°C **45°C**

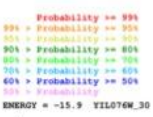
ACGGCUCAAGCUAUCUUGGGUGACUUAGAUAAA
AGUUUGGAGACAUG**UGUAG**AAGGGAUUGACAAU
GACGAAGCAGAAGGGACUACAGAAUUAUUGCUG

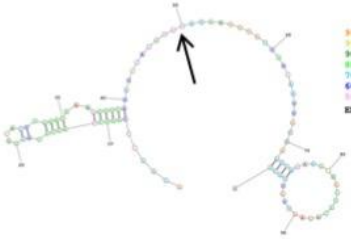




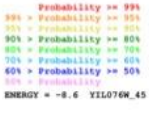


Fold





Fold



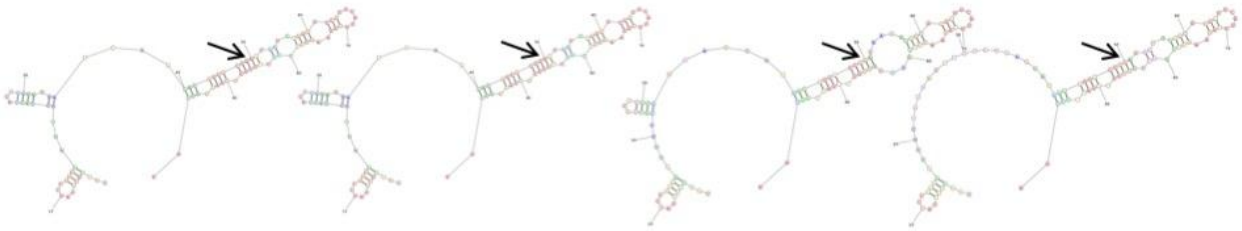
MaxExpect

MaxExpect

YJL158C

Target UGUAG – 30°C **45°C**

CAAUGUCUAUCCGGCAAUUUCUACAACUUGUAU
GAUCAAAACGUCGCCGAACAA**UGUAG**UGCCAUU
CAUUUGGAAGCUGUUUCUUUGGUCGACUGUUAA

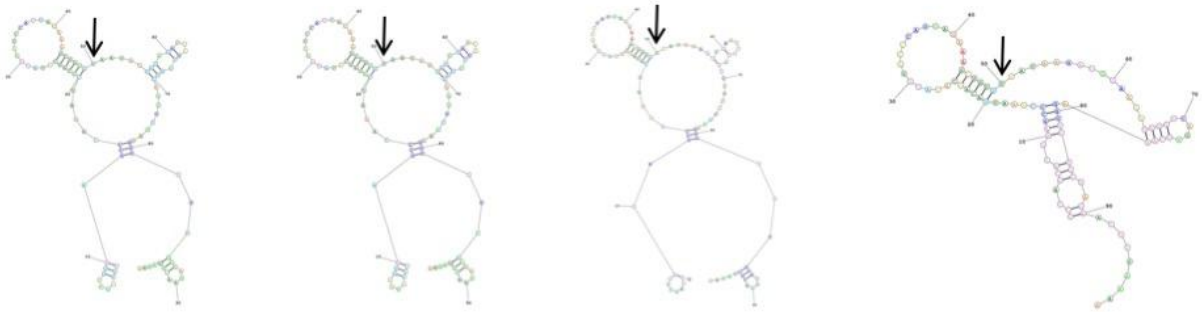


MaxExpect Fold MaxExpect Fold

YLR168C

Target UGUAG – 30°C **45°C**

UCCAGCAGGAAGCCCAAAUUACAGCAUAUGGAUC
CAUUAGAAAGCUG**UGUAA**UAAGAUGGAAGAUUG
GUCAGUUCAAAGGUUUUGCGAGAACGCUAAAA

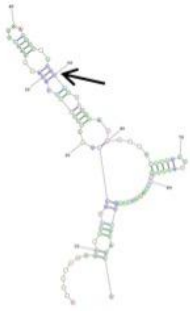


MaxExpect Fold MaxExpect Fold

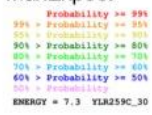
YLR259C

Target UGUAG – 30°C **45°C**

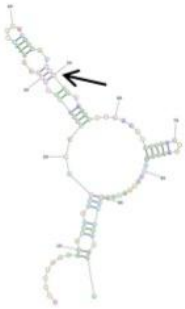
CCUUUAUUGCGUCGUGCUUACUCCUCUCAUAAA
GAAUUGAAAUUCGG**UGUAG**AAGGAAGAGCCUCC
CUUCUUAAGGGUGUCGAAACUUUAGCUGAAGCG



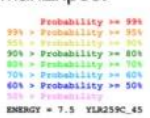
MaxExpect

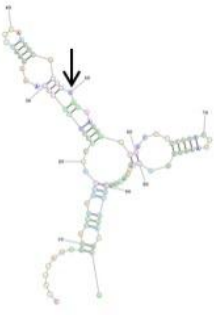


Fold



MaxExpect



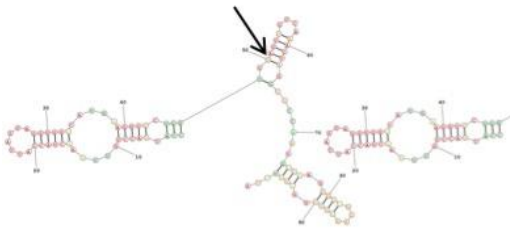


Fold

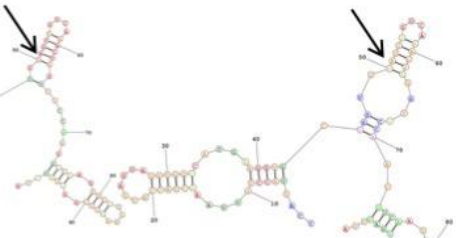
YMR083C

Target UGUAG – 30°C **45°C**

CCAUUACCUGUUAAACUACCAUUAGUAGGUGGU
CAUGAAGGUGCUGG**UGUAG**UUGUCAAACUAGGU
UCCAAUGUCAAGGGCUGGAAAGUCGGUGAUUUA

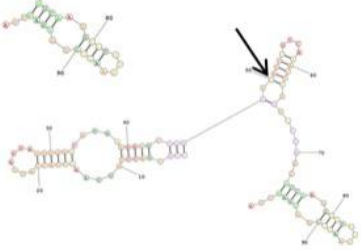


MaxExpect

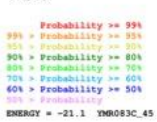


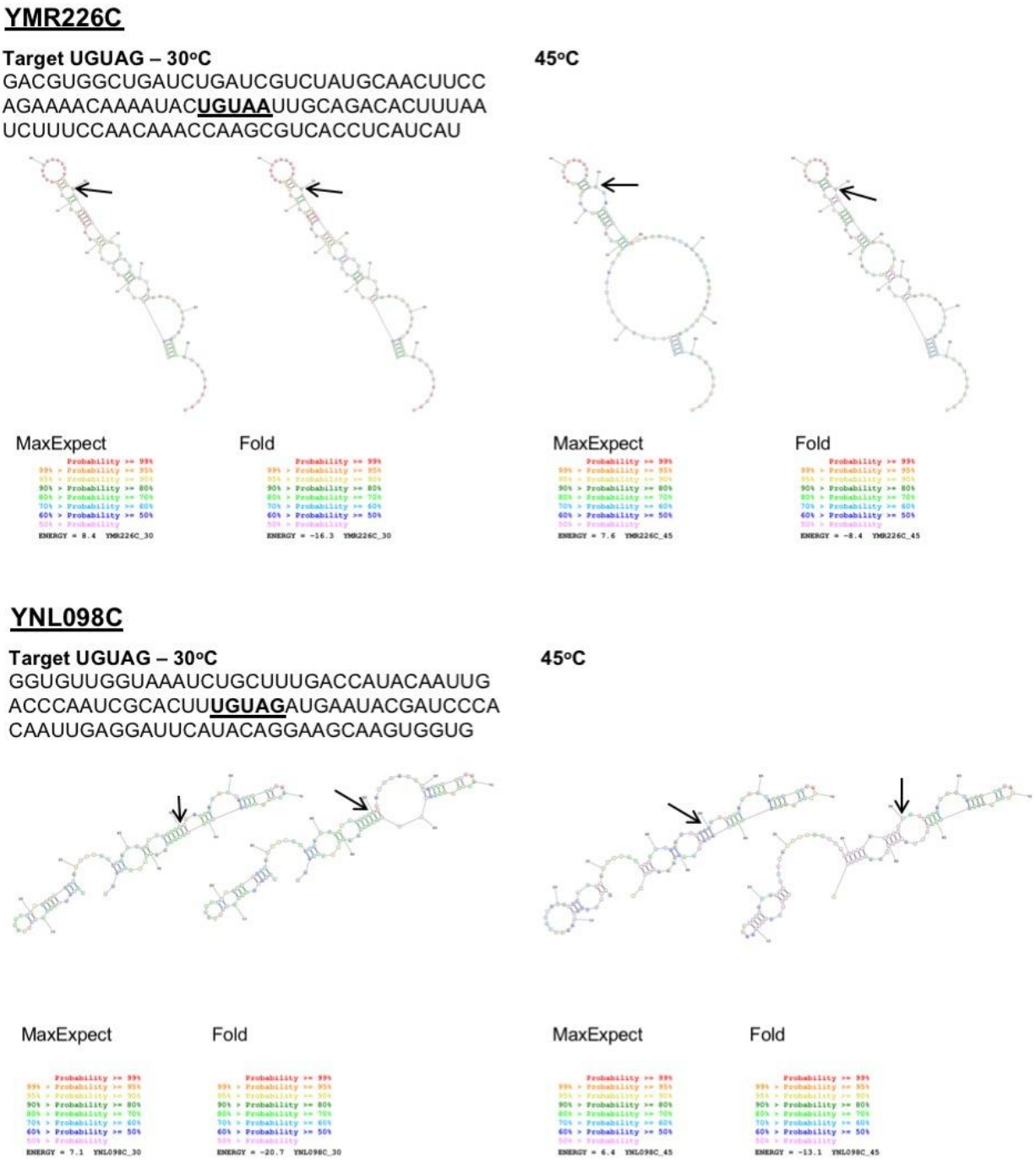
Fold

MaxExpect



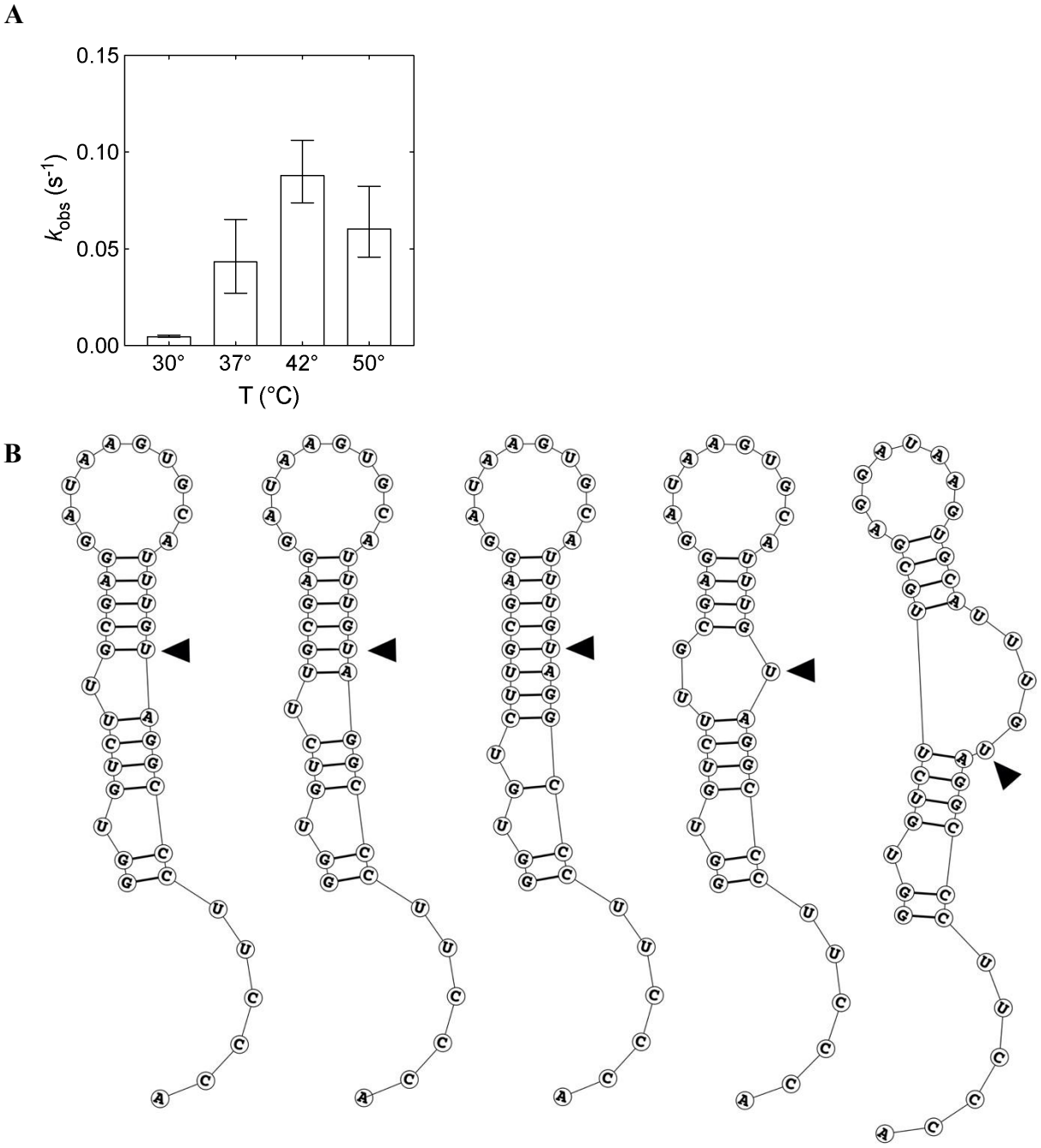
Fold





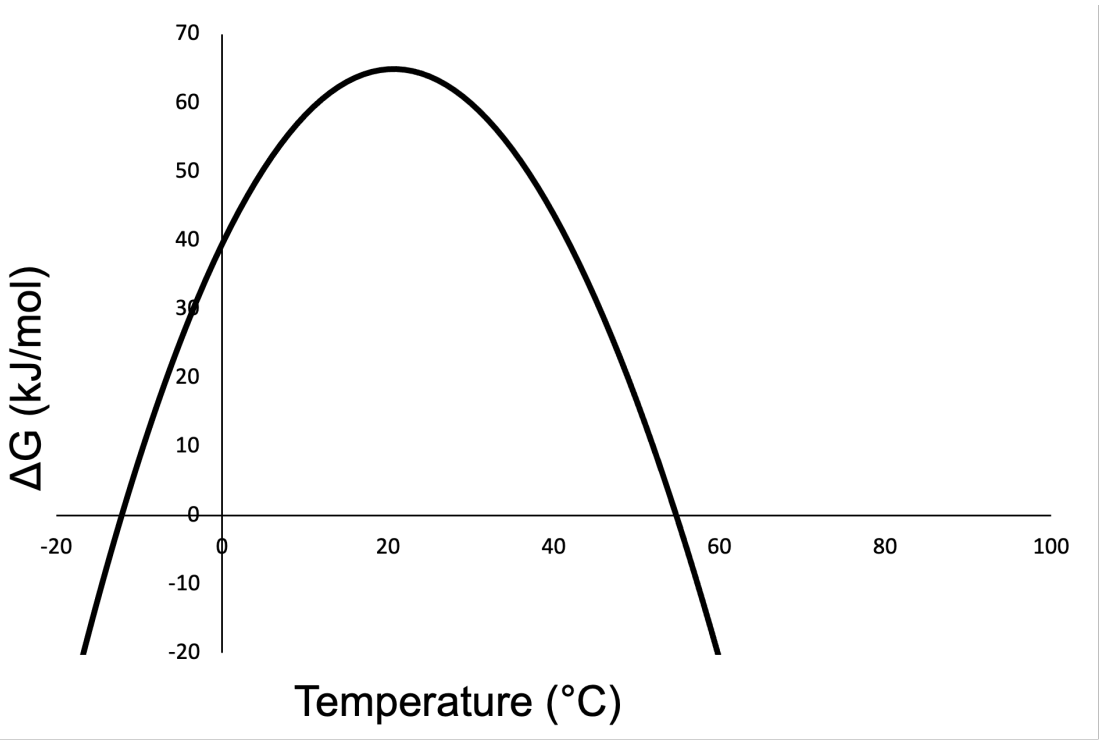
SI Figure 2.18: Secondary structure predictions at 30°C and 45°C of randomly selected Pus7 heat shock targets

Pus7 heat shock targets reported in Schwartz, et al. Cell (2014) (24).



SI Figure 2.19: The observed rate constant for pseudouridylation on short target 1 (ST1) is increased ~10-fold at elevated temperature

(A) Observed rate constants for pseudouridylation increase more than 10-fold as temperature increases, suggesting that increased conformational flexibility of the RNA structure allows more rapid access of PUS7. **(B)** A set of stochastic structure predictions (78) demonstrating possible temperature-dependent changes in the structural environment of the target U in substrate sT1.



SI Figure 2.20: Modeled thermal stability of Pus7

Using the chain length (N) of Pus7, its stability was modeled as a function temperature range to find its maximum stability. Its maximum stability of about 65 kJ/mol is at approximately 22° C.

SI Table 2.1: X-Ray Data Collection and Refinement Statistics for Pus7

	Pus7
Data collection	
Beamline	APS, LSCAT 21-IDD
Wavelength (Å)	0.979
Temperature (K)	100
Resolution (Å)	48.27-3.20 (3.42-3.20)
Space group	*C*222
Cell dimensions (Å)	a = 117.9, b = 171.8, c = 105.3
Cell dimensions (°)	α = β = γ = 90
Observed reflections	184,895 (31,045)
Unique reflections	18,019 (3,207)
R_{meas} (%)	17.8 (141.7)
R_{merge} (%)	17.8 (132.6)
<I/σ>	9.5 (2.0)
CC(1/2)	0.996 (0.802)
Multiplicity	10.3 (9.7)
Completeness (%)	99.9 (100)
Overall *B* (Å^2) (Wilson plot)	121.9
Refinement	
Resolution range	46.32 - 3.20
Number of reflections (work/test set)	18017/881
R_{work}/R_{free} (%)	22.4/27.6
No. of non-H atoms	
Protein	9394
Water	14
Ligand	15
B-factors (Å^2)	
Protein	130.1
Water	88.7
Ligand	164.3
Rmsd deviations	
Bond lengths (Å)	0.0025
Bond angles (°)	1.21
Estimated coordinate error (Å); maximum likelihood based	0.4200
Cruickshank's DPI[1] (Å)	0.4688
Ramachandran plot	
Favored/allowed/outliers	87.7/12.1/0.2
MolProbity Score	1.63 (100th percentile)
PDB	7MZV

SI Table 2.2: Impact of Pus7 mutations on CDC8 binding and modification

variant	k_{obs} (s^{-1}) [a,b]			k_{obs} defect (fold) [c]	$K_{D, app1}$ (nM) [d,e]		
WT[e]	9.9×10^{-1}	±	1.0×10^{-1}	1	76	±	15
D256A[e]	no reaction				60	±	16
K61A	2.6×10^{-2}	±	0.1×10^{-2}	38 ± 6	400	±	200
F67A	4.6×10^{-3}	±	0.2×10^{-3}	210 ± 30	180	±	40
E71A	5.2×10^{-3}	±	0.3×10^{-3}	190 ± 30	210	±	50
F307Y	2.6×10^{-3}	±	0.1×10^{-3}	390 ± 60	378	±	102
N305A	4.0×10^{-4}	±	$\leq 1 \times 10^{-5}$	2,400 ± 300	230	±	60
F307A	1.3×10^{-5}	±	$\leq 1 \times 10^{-6}$	74,000 ± 9,000	344	±	170
WT[f]	8.4×10^{-1}	±	0.5×10^{-1}	1	n.d.[g]		
H161A[f]	6.9×10^{-1}	±	0.9×10^{-1}	1.2 ± 0.2	170	±	40
ΔID1[f]	3.8×10^{-1}	±	0.6×10^{-1}	2.2 ± 0.5	160	±	40

[a] errors are standard error of the fit

[b] k_{obs} determined by fitting a curve of the form $y = 1 - e^{-k_{obs} \times t}$

[c] relative to WT at the same concentration on full-length cdc8 substrate

[d] K_D determined by curve fitting as described in Methods

[e] All K_D values determined using D256A-double mutants, except for WT (no mutation), and the F307Y and D256A single mutants.

[f] Concentration of [Pus7] used to determine k_{obs} = 10 μM. All other reactions carried out with [Pus7] = 2 μM

[g] n.d., not determined

SI Table 2.3: Dissociation constants for Pus7 binding to various substrates

substrate	variant	$K_{D,app1}$ (nM) [a,b]		
cdc8-A	WT	16	±	2
	D256A	57	±	4
cdc8-B	D256A	802	±	320
cdc8-C	WT	74	±	19
	D256A	131	±	13
sT1	D256A	34	±	4
	ΔID1-D256A	69	±	13
sT2	D256A	not analyzed – very weak		
	ΔID1-D256A	not analyzed – very, very weak		
$tRNA^{Asp}_{,GUC}$	D256A	16	±	1
	ΔID1-D256A	34	±	1

[a] errors are standard error of the fit
[b] K_D determined by curve fitting as described in the Methods

SI Table 2.4: Observed rate constants for pseudouridylation on different substrates

substrate	variant	[PUS7] (μM)	k_{obs} (s^{-1}) [a,b]		
cdc8-A	WT	1	4.9×10^{-1}	±	0.5×10^{-1}
		5	9.3×10^{-1}	±	1.4×10^{-1}
		10	7.7×10^{-1}	±	0.8×10^{-1}
cdc8-B	WT	2	6.4×10^{-1}	±	1.0×10^{-1}
		10	7.8×10^{-1}	±	0.8×10^{-1}
cdc8-C	WT	2	8.9×10^{-1}	±	2.1×10^{-1}
		10	9.9×10^{-1}	±	1.4×10^{-1}
sT1	WT	10	5.1×10^{-3}	±	0.2×10^{-3}
	ΔID1	10	9.4×10^{-3}	±	0.7×10^{-3}
sT2	WT	10	2.4×10^{-4}	±	0.2×10^{-4}
tRNA ^{Asp,GUC}	WT	10	9.1×10^{-3}	±	0.5×10^{-3}

[a] errors are standard error of the fit

[b] k_{obs} determined by fitting a curve of the form $fraction\ U \rightarrow = 1 - e^{-k_{obs} \times t}$

SI Table 2.5: RNAs used for biochemical assays

substrate name	RNA sequence
cdc8-FL	GUCAAUCACGAUUGU**U**AGACGUUACUAAUAAGGGCAUUCAGGAAGU UGAAGCGCUUAUUUGG
CDC8-FL-NT	GUCAAUCACGAUUGCAGACGUUACUAAUAAGGGCAUUCAGGAAGU UGAAGCGCUUAUUUGG
CDC 8-A	GAUUGU**U**AGACGUUACUAAUAAGGGCAUUCAGGAAGUUGAAGCGCU UAUUUGG
CDC 8-A-NT	GAUUGCAGACGUUACUAAUAAGGGCAUUCAGGAAGUUGAAGCGCU UAUUUGG
CDC 8-B	GUCAAUCACGAUUGU**U**AGACGUUACU
CDC 8-B-NT	GUCAAUCACGAUUGCAGACGUUACU
CDC 8-C	GUCAAUCACGAUUGU**U**AGACGUUACUAAUAAGGGCGGAAGUGCGCU UAUUUGG
CDC 8-C-NT	GUCAAUCACGAUUGCAGACGUUACUAAUAAGGGCGGAAGUGCGCU UAUUUGG
ST1	GGUGUCUUGCGAGGAUAAGUGCAUUUGU**U**AGGCCCUUCCCA
SNT1	GGUGUCUUGCGAGGAUAAGUGCAUUUGCAGGCCCUUCCCA
ST2	GGGAUCUGU**U**AGCCCACCAA
SNT2	GGGAUCUGCAGCCCACCAA
tRNA^{Asp,GUC}	GCCGUGAUAG**U**UUAAUGGUCAGAAUGGGCGCUUGUCGCGUGCCA GAUCGGGGUUCAAUUCCCCGUCGCGGCGCCA
tRNA^{Asp,GUC}-NT	GCCGUGAUAGCUUAAUGGUCAGAAUGGGGCUUGUCGCGUGCCAG AUCGGGGUUCAAUUCCCCGUCGCGGCGCCA
CLAP-CDC8	GGCUAUUGGAUAAAGAGAUAAGGAAAGGCGAUGAGUCAAUCACGA UUGUAGACGUUACUAAUAAGGGCAUUCAGGAAGUUGAAGCGCUUA UUUGGCAAAUCGUUGAGCCUGUUUUGAGUACGCAUAU

2.6 References

1. Jones JD, Monroe J, Koutmou KS. A molecular-level perspective on the frequency, distribution, and consequences of messenger RNA modifications. WIREs RNA. 2020;11(4):e1586.

2. Martinez NM, Su A, Burns MC, Nussbacher JK, Schaening C, Sathe S, et al. Pseudouridine synthases modify human pre-mRNA co-transcriptionally and affect pre-mRNA processing. Molecular Cell. 2022 Feb 3;82(3):645-659.e9.

3. Levi O, Arava YS. Pseudouridine-mediated translation control of mRNA by methionine aminoacyl tRNA synthetase. Nucleic Acids Res. 2021 Jan 11;49(1):432–43.

4. Nakamoto MA, Lovejoy AF, Cygan AM, Boothroyd JC. mRNA pseudouridylation affects RNA metabolism in the parasite Toxoplasma gondii. RNA. 2017 Dec;23(12):1834–49.

5. Gutgsell N, Englund N, Niu L, Kaya Y, Lane BG, Ofengand J. Deletion of the Escherichia coli pseudouridine synthase gene truB blocks formation of pseudouridine 55 in tRNA in vivo, does not affect exponential growth, but confers a strong selective disadvantage in competition with wild-type cells. RNA. 2000 Dec;6(12):1870–81.

6. Del Campo M, Kaya Y, Ofengand J. Identification and site of action of the remaining four putative pseudouridine synthases in Escherichia coli. RNA. 2001 Nov;7(11):1603–15.

7. Breslow DK, Cameron DM, Collins SR, Schuldiner M, Stewart-Ornstein J, Newman HW, et al. A comprehensive strategy enabling high-resolution functional analysis of the yeast genome. Nat Methods. 2008 Aug;5(8):711–8.

8. Urbonavicius J, Durand JMB, Björk GR. Three modifications in the D and T arms of tRNA influence translation in Escherichia coli and expression of virulence genes in Shigella flexneri. J Bacteriol. 2002 Oct;184(19):5348–57.

9. Ishida K, Kunibayashi T, Tomikawa C, Ochi A, Kanai T, Hirata A, et al. Pseudouridine at position 55 in tRNA controls the contents of other modified nucleotides for low-temperature adaptation in the extreme-thermophilic eubacterium Thermus thermophilus. Nucleic Acids Res. 2011 Mar;39(6):2304–18.

10. Kinghorn SM, O'Byrne CP, Booth IR, Stansfield I. Physiological analysis of the role of truB in Escherichia coli: a role for tRNA modification in extreme temperature resistance. Microbiology (Reading). 2002 Nov;148(Pt 11):3511–20.

11. Sinha H, David L, Pascon RC, Clauder-Münster S, Krishnakumar S, Nguyen M, et al. Sequential elimination of major-effect contributors identifies additional quantitative trait loci conditioning high-temperature growth in yeast. Genetics. 2008 Nov;180(3):1661–70.

12. de Brouwer APM, Abou Jamra R, Körtel N, Soyris C, Polla DL, Safra M, et al. Variants in PUS7 Cause Intellectual Disability with Speech Delay, Microcephaly, Short Stature, and Aggressive Behavior. Am J Hum Genet. 2018 Dec 6;103(6):1045–52.

13. Zeharia A, Fischel-Ghodsian N, Casas K, Bykhocskaya Y, Tamari H, Lev D, et al. Mitochondrial myopathy, sideroblastic anemia, and lactic acidosis: an autosomal recessive syndrome in Persian Jews caused by a mutation in the PUS1 gene. J Child Neurol. 2005 May;20(5):449–52.

14. Bykhovskaya Y, Casas K, Mengesha E, Inbal A, Fischel-Ghodsian N. Missense mutation in pseudouridine synthase 1 (PUS1) causes mitochondrial myopathy and sideroblastic anemia (MLASA). Am J Hum Genet. 2004 Jun;74(6):1303–8.

15. Shaheen R, Han L, Faqeih E, Ewida N, Alobeid E, Phizicky EM, et al. A homozygous truncating mutation in PUS3 expands the role of tRNA modification in normal cognition. Hum Genet. 2016 Jul;135(7):707–13.

16. Keffer-Wilkes LC, Veerareddygari GR, Kothe U. RNA modification enzyme TruB is a tRNA chaperone. Proc Natl Acad Sci U S A. 2016 Dec 13;113(50):14306–11.

17. Keffer-Wilkes LC, Soon EF, Kothe U. The methyltransferase TrmA facilitates tRNA folding through interaction with its RNA-binding domain. Nucleic Acids Res. 2020 Aug 20;48(14):7981–90.

18. Garcia DM, Campbell EA, Jakobson CM, Tsuchiya M, Shaw EA, DiNardo AL, et al. A prion accelerates proliferation at the expense of lifespan. Dang W, Tyler JK, editors. eLife. 2021 Sep 21;10:e60917.

19. Penzo M, Guerrieri AN, Zacchini F, Treré D, Montanaro L. RNA Pseudouridylation in Physiology and Medicine: For Better and for Worse. Genes (Basel). 2017 Nov 1;8(11):301.

20. Hamma T, Ferré-D'Amaré AR. Pseudouridine Synthases. Chemistry & Biology. 2006 Nov 1;13(11):1125–35.

21. Rintala-Dempsey AC, Kothe U. Eukaryotic stand-alone pseudouridine synthases – RNA modifying enzymes and emerging regulators of gene expression? null. 2017 Sep 2;14(9):1185–96.

22. Boccaletto P, Stefaniak F, Ray A, Cappannini A, Mukherjee S, Purta E, et al. MODOMICS: a database of RNA modification pathways. 2021 update. Nucleic Acids Research. 2022 Jan 7;50(D1):D231–5.

23. Carlile TM, Rojas-Duran MF, Zinshteyn B, Shin H, Bartoli KM, Gilbert WV. Pseudouridine profiling reveals regulated mRNA pseudouridylation in yeast and human cells. Nature. 2014 Nov 6;515(7525):143–6.

24. Schwartz S, Bernstein DA, Mumbach MR, Jovanovic M, Herbst RH, León-Ricardo BX, et al. Transcriptome-wide mapping reveals widespread dynamic-regulated pseudouridylation of ncRNA and mRNA. Cell. 2014 Sep 25;159(1):148–62.

25. Safra M, Nir R, Farouq D, Slutzkin IV, Schwartz S. TRUB1 is the predominant pseudouridine synthase acting on mammalian mRNA via a predictable and conserved code. Genome Res. 2017 Mar;27(3):393–406.

26. Carlile TM, Martinez NM, Schaening C, Su A, Bell TA, Zinshteyn B, et al. mRNA structure determines modification by pseudouridine synthase 1. Nat Chem Biol. 2019 Oct;15(10):966–74.

27. Guzzi N, Cieśla M, Ngoc PCT, Lang S, Arora S, Dimitriou M, et al. Pseudouridylation of tRNA-Derived Fragments Steers Translational Control in Stem Cells. Cell. 2018 May 17;173(5):1204-1216.e26.

28. Behm-Ansmant I, Urban A, Ma X, Yu YT, Motorin Y, Branlant C. The Saccharomyces cerevisiae U2 snRNA:pseudouridine-synthase Pus7p is a novel multisite-multisubstrate RNA:Psi-synthase also acting on tRNAs. RNA. 2003 Nov;9(11):1371–82.

29. Ma X, Yang C, Alexandrov A, Grayhack EJ, Behm-Ansmant I, Yu YT. Pseudouridylation of yeast U2 snRNA is catalyzed by either an RNA-guided or RNA-independent mechanism. EMBO J. 2005 Jul 6;24(13):2403–13.

30. Kaya Y, Ofengand J. A novel unanticipated type of pseudouridine synthase with homologs in bacteria, archaea, and eukarya. RNA. 2003 Jun;9(6):711–21.

31. Ericsson UB, Nordlund P, Hallberg BM. X-ray structure of tRNA pseudouridine synthase TruD reveals an inserted domain with a novel fold. FEBS Lett. 2004 May 7;565(1–3):59–64.

32. Hoang C, Ferré-D'amaré AR. Crystal structure of the highly divergent pseudouridine synthase TruD reveals a circular permutation of a conserved fold. RNA. 2004 Jul 1;10(7):1026–33.

33. Kaya Y, Campo MD, Ofengand J, Malhotra A. Crystal Structure of TruD, a Novel Pseudouridine Synthase with a New Protein Fold *. Journal of Biological Chemistry. 2004 Apr 30;279(18):18107–10.

34. Mueller EG. Chips off the old block. Nat Struct Mol Biol. 2002 May;9(5):320–2.

35. Holm L. Using Dali for Protein Structure Comparison. Methods Mol Biol. 2020;2112:29–42.

36. Yoshikawa K, Tanaka T, Furusawa C, Nagahisa K, Hirasawa T, Shimizu H. Comprehensive phenotypic analysis for identification of genes affecting growth under ethanol stress in Saccharomyces cerevisiae. FEMS Yeast Research. 2009 Feb 1;9(1):32–44.

37. Blanchet S, Cornu D, Hatin I, Grosjean H, Bertin P, Namy O. Deciphering the reading of the genetic code by near-cognate tRNA. Proceedings of the National Academy of Sciences. 2018 Mar 20;115(12):3018–23.

38. Chou HJ, Donnard E, Gustafsson HT, Garber M, Rando OJ. Transcriptome-wide Analysis of Roles for tRNA Modifications in Translational Regulation. Mol Cell. 2017 Dec 7;68(5):978-992.e4.

39. Eyler DE, Franco MK, Batool Z, Wu MZ, Dubuke ML, Dobosz-Bartoszek M, et al. Pseudouridinylation of mRNA coding sequences alters translation. Proceedings of the National Academy of Sciences. 2019 Nov 12;116(46):23068–74.

40. Czekay DP, Schultz SK, Kothe U. Assaying the Molecular Determinants and Kinetics of RNA Pseudouridylation by H/ACA snoRNPs and Stand-Alone Pseudouridine Synthases. Methods Mol Biol. 2021;2298:357–78.

41. Veerareddygari GR, Singh SK, Mueller EG. The Pseudouridine Synthases Proceed through a Glycal Intermediate. J Am Chem Soc. 2016 Jun 29;138(25):7852–5.

42. Zhang W, Eckwahl MJ, Zhou KI, Pan T. Sensitive and quantitative probing of pseudouridine modification in mRNA and long noncoding RNA. RNA. 2019 Sep;25(9):1218–25.

43. Bellaousov S, Reuter JS, Seetin MG, Mathews DH. RNAstructure: Web servers for RNA secondary structure prediction and analysis. Nucleic Acids Res. 2013 Jul;41(Web Server issue):W471-474.

44. Wright JR, Keffer-Wilkes LC, Dobing SR, Kothe U. Pre-steady-state kinetic analysis of the three Escherichia coli pseudouridine synthases TruB, TruA, and RluA reveals uniformly slow catalysis. RNA. 2011 Dec;17(12):2074–84.

45. Huang L, Pookanjanatavip M, Gu X, Santi DV. A conserved aspartate of tRNA pseudouridine synthase is essential for activity and a probable nucleophilic catalyst. Biochemistry. 1998 Jan 6;37(1):344–51.

46. Ramamurthy V, Swann SL, Spedaliere CJ, Mueller EG. Role of cysteine residues in pseudouridine synthases of different families. Biochemistry. 1999 Oct 5;38(40):13106–11.

47. Taylor EL, Kesavan PM, Wolfe AE, O'Brien PJ. Distinguishing Specific and Nonspecific Complexes of Alkyladenine DNA Glycosylase. Biochemistry. 2018 Jul 31;57(30):4440–54.

48. Esadze A, Stivers JT. Facilitated Diffusion Mechanisms in DNA Base Excision Repair and Transcriptional Activation. Chem Rev. 2018 Dec 12;118(23):11298–323.

49. Vaidyanathan PP, AlSadhan I, Merriman DK, Al-Hashimi HM, Herschlag D. Pseudouridine and N6-methyladenosine modifications weaken PUF protein/RNA interactions. RNA. 2017 May;23(5):611–8.

50. Gerber AP, Herschlag D, Brown PO. Extensive Association of Functionally and Cytotopically Related mRNAs with Puf Family RNA-Binding Proteins in Yeast. PLOS Biology. 2004 Mar 16;2(3):e79.

51. Potterton L, Agirre J, Ballard C, Cowtan K, Dodson E, Evans PR, et al. CCP4i2: the new graphical user interface to the CCP4 program suite. Acta Crystallogr D Struct Biol. 2018 Feb 1;74(Pt 2):68–84.

52. McCoy AJ, Grosse-Kunstleve RW, Adams PD, Winn MD, Storoni LC, Read RJ. Phaser crystallographic software. J Appl Cryst. 2007 Aug 1;40(4):658–74.

53. Terwilliger TC, Adams PD, Read RJ, McCoy AJ, Moriarty NW, Grosse-Kunstleve RW, et al. Decision-making in structure solution using Bayesian estimates of map quality: the PHENIX AutoSol wizard. Acta Crystallogr D Biol Crystallogr. 2009 Jun;65(Pt 6):582–601.

54. Emsley P, Lohkamp B, Scott WG, Cowtan K. Features and development of Coot. Acta Crystallogr D Biol Crystallogr. 2010 Apr;66(Pt 4):486–501.

55. Williams CJ, Headd JJ, Moriarty NW, Prisant MG, Videau LL, Deis LN, et al. MolProbity: More and better reference data for improved all-atom structure validation. Protein Sci. 2018 Jan;27(1):293–315.

56. Schrödinger, LLC. The PyMOL Molecular Graphics System, Version 1.8. 2015.

57. Milligan JF, Groebe DR, Witherell GW, Uhlenbeck OC. Oligoribonucleotide synthesis using T7 RNA polymerase and synthetic DNA templates. Nucleic Acids Res. 1987 Nov 11;15(21):8783–98.

58. Thompson JD, Higgins DG, Gibson TJ. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. Nucleic Acids Res. 1994 Nov 11;22(22):4673–80.

59. Kumar S, Stecher G, Li M, Knyaz C, Tamura K. MEGA X: Molecular Evolutionary Genetics Analysis across Computing Platforms. Mol Biol Evol. 2018 Jun 1;35(6):1547–9.

60. Le SQ, Gascuel O. An Improved General Amino Acid Replacement Matrix. Molecular Biology and Evolution. 2008 Jul 1;25(7):1307–20.

61. Felsenstein J. CONFIDENCE LIMITS ON PHYLOGENIES: AN APPROACH USING THE BOOTSTRAP. Evolution. 1985 Jul;39(4):783–91.

62. Martin M. Cutadapt removes adapter sequences from high-throughput sequencing reads. EMBnet.journal. 2011 May 2;17(1):10–2.

63. Tardu M, Jones JD, Kennedy RT, Lin Q, Koutmou KS. Identification and Quantification of Modified Nucleosides in Saccharomyces cerevisiae mRNAs. ACS Chem Biol. 2019 Jul 19;14(7):1403–9.

64. Langmead B, Salzberg SL. Fast gapped-read alignment with Bowtie 2. Nat Methods. 2012 Apr;9(4):357–9.

65. Kim D, Pertea G, Trapnell C, Pimentel H, Kelley R, Salzberg SL. TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. Genome Biology. 2013 Apr 25;14(4):R36.

66. Lauria F, Tebaldi T, Bernabò P, Groen EJN, Gillingwater TH, Viero G. riboWaltz: Optimization of ribosome P-site positioning in ribosome profiling data. PLOS Computational Biology. 2018 Aug 13;14(8):e1006169.

67. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, et al. The Sequence Alignment/Map format and SAMtools. Bioinformatics. 2009 Aug 15;25(16):2078–9.

68. Myers JK, Pace CN, Scholtz JM. Denaturant m values and heat capacity changes: relation to changes in accessible surface areas of protein unfolding. Protein Sci. 1995 Oct;4(10):2138–48.

69. Ghosh K, Dill KA. Computing protein stabilities from their chain lengths. Proceedings of the National Academy of Sciences. 2009 Jun 30;106(26):10649–54.

70. Sawle L, Ghosh K. How Do Thermophilic Proteins and Proteomes Withstand High Temperature? Biophys J. 2011 Jul 6;101(1):217–27.

71. Watson MD, Monroe J, Raleigh DP. Size-Dependent Relationships between Protein Stability and Thermal Unfolding Temperature Have Important Implications for Analysis of Protein Energetics and High-Throughput Assays of Protein-Ligand Interactions. J Phys Chem B. 2018 May 31;122(21):5278–85.

72. Rees DC, Robertson AD. Some thermodynamic implications for the thermostability of proteins. Protein Sci. 2001 Jun;10(6):1187–94.

73. Robertson AD, Murphy KP. Protein Structure and the Energetics of Protein Stability. Chem Rev. 1997 Aug 1;97(5):1251–68.

74. Jurrus E, Engel D, Star K, Monson K, Brandi J, Felberg LE, et al. Improvements to the APBS biomolecular solvation software suite. Protein Sci. 2018 Jan;27(1):112–28.

75. Zeng H, Dong A, Walker JR, Bountra C, Arrowsmith CH, Edwards AM, et al. Crystal Structure of Human Pseudouridylate Synthase 7 (PDB: 5KKP). Protein Data Bank (PDB). 2016;

76. Holm L, Rosenström P. Dali server: conservation mapping in 3D. Nucleic Acids Res. 2010 Jul;38(Web Server issue):W545-549.

77. Wu M, Reuter M, Lilie H, Liu Y, Wahle E, Song H. Structural insight into poly(A) binding and catalytic mechanism of human PARN. The EMBO Journal. 2005 Dec 7;24(23):4082–93.

78. Harmanci AO, Sharma G, Mathews DH. Stochastic sampling of the RNA structural alignment space. Nucleic Acids Res. 2009 Jul;37(12):4063–75.

CHAPTER 3 Establishing the Structural Basis of Methyl-Transfer in Cobalamin Dependent Methionine Synthase

3.1 Introduction

B_{12}-cofactors and their cobalamin (Cbl) derivatives catalyze a range of challenging biological processes [methyl-transfer (1–5), carbon-skeleton rearrangement (3,6,7), dehalogenation (8), etc.] essential for complex life, including the biosynthesis of amino acids and CO_2 conversion (1,9). The power to mediate intricate chemistries derives from the reactivity of the central Cobalt ion (10,11) and enzymes dependent on this prosthetic group have developed strategies to control its unique organometallic properties (Figure 3.1) (2,12). Cobalamin-dependent methionine synthase (MS) is an essential enzyme that leverages the reactivity of Cbl to catalyze three distinct methyl-transfer reactions (Figure 3.2) (1). MS is central to one-carbon metabolism, serving as the enzymatic link between the folate and methionine metabolic cycles (1,13,14), and defects in MS result in severe pathologies including megaloblastic anemia, substantial birth abnormalities, and neural tube defects (15–17).

This chapter was adapted with permission from Purchal, M.; Yamada, K.; Mendoza, J.; Castillo, C.; Conway, P.; Koutmos, M. Methionine Synthase. *Manuscript in preparation.*

Author contributions: M.P. performed data analysis (Figures 3.1 – 3.8 and SI Figures 3.5-3.13 and SI Table 3.1) and manuscript writing. K.Y. performed crystallization (Figure 3.6, 3.7 and Table 3.1, 3.2) and biochemical assays (Figure 3.5 and SI Figures 3.3, 3.4). M. K. and J. M. performed model refinement.

MS catalyzes the formation of H_4folate (tetrahydrofolate, THF) and methionine from CH_3-H_4folate (methyltetrahydrofolate, MTF) and homocysteine (HCY), respectively (Figure 3.2). To achieve these thermodynamically challenging reactions, MS uses one of the strongest nucleophiles

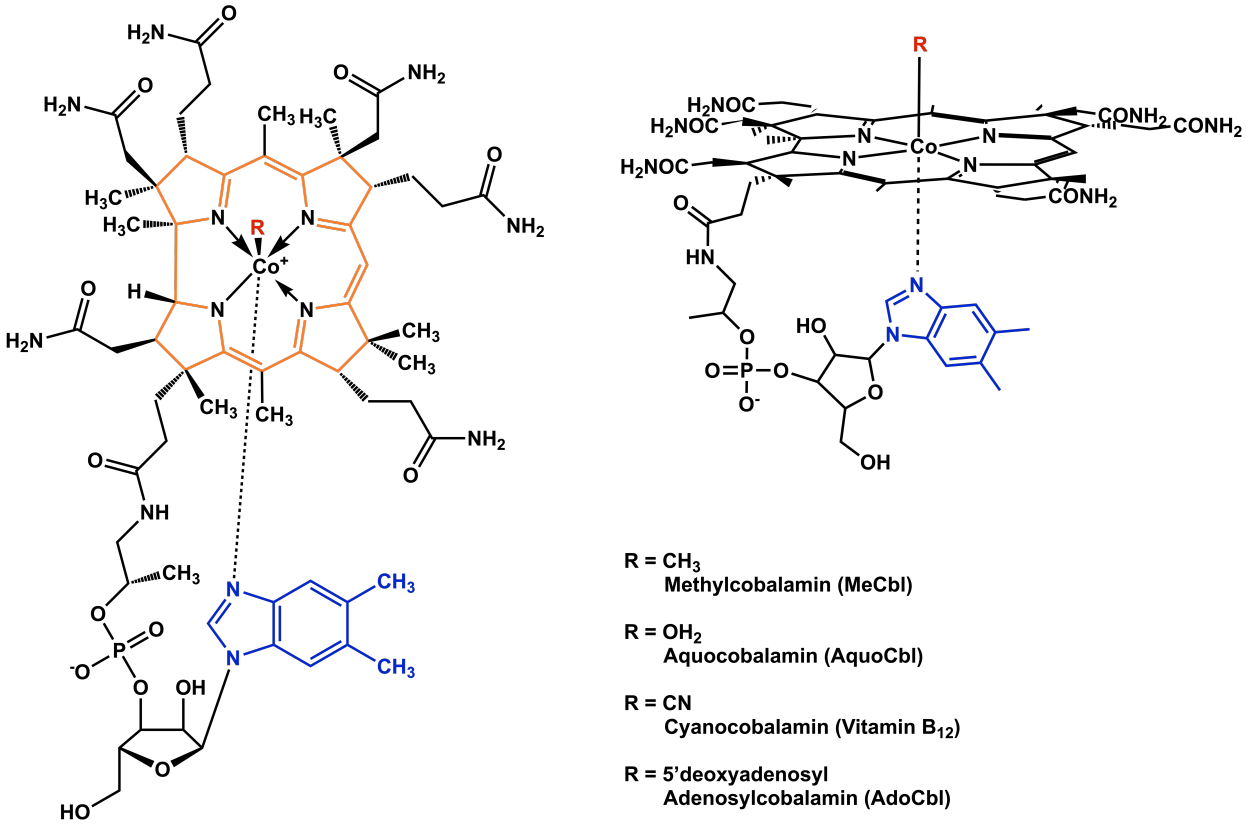


Figure 3.1: B_{12}-cofactor and cobalamin derivatives

(left) Basic structure of cobalamins shown in 'base-on' form that is adopted when free in solution, where the DBI tail (blue) is coordinated to Co in the lower axial position. Upper axial position denoted by R (red). R=CH_3, OH_2, 5'deoxyadenosyl, and CN. **(right)** Side view of cofactor.

found in nature, Co(I) (18,19), to abstract a methyl group from MTF, yielding H_4-folate (THF), a poor leaving group, and methylcob(III)alamin [MeCbl, CH_3-Co(III)] (18,20,21). The methyl group is then transferred from MeCbl to Zn^{2+} activated HCY, forming methionine and regenerating the Co(I) cofactor (Figure 3.2) (22–24**).** Thus, MS reactivity relies on the enzyme continuously cycling between the CH_3-Co(III) and Co(I) forms of the cofactor (Figure 3.2 and Figure 3.3A) (1,25). Once every ~2000 turnovers, microaerophilic cellular conditions lead to the inactivation of MS, where

the Co(I) supernucleophile is oxidized to a catalytically inactive Co(II) species (13,26). To restore catalytic competency, MS undergoes a reactivation cycle where Co(II) is enzymatically reduced and subsequently methylated by S-adenosyl-L-methionine (AdoMet, SAM) to regenerate functional CH_3-Co(III) cofactor that can re-enter the catalytic cycle (Figure 3.2) (26–30).

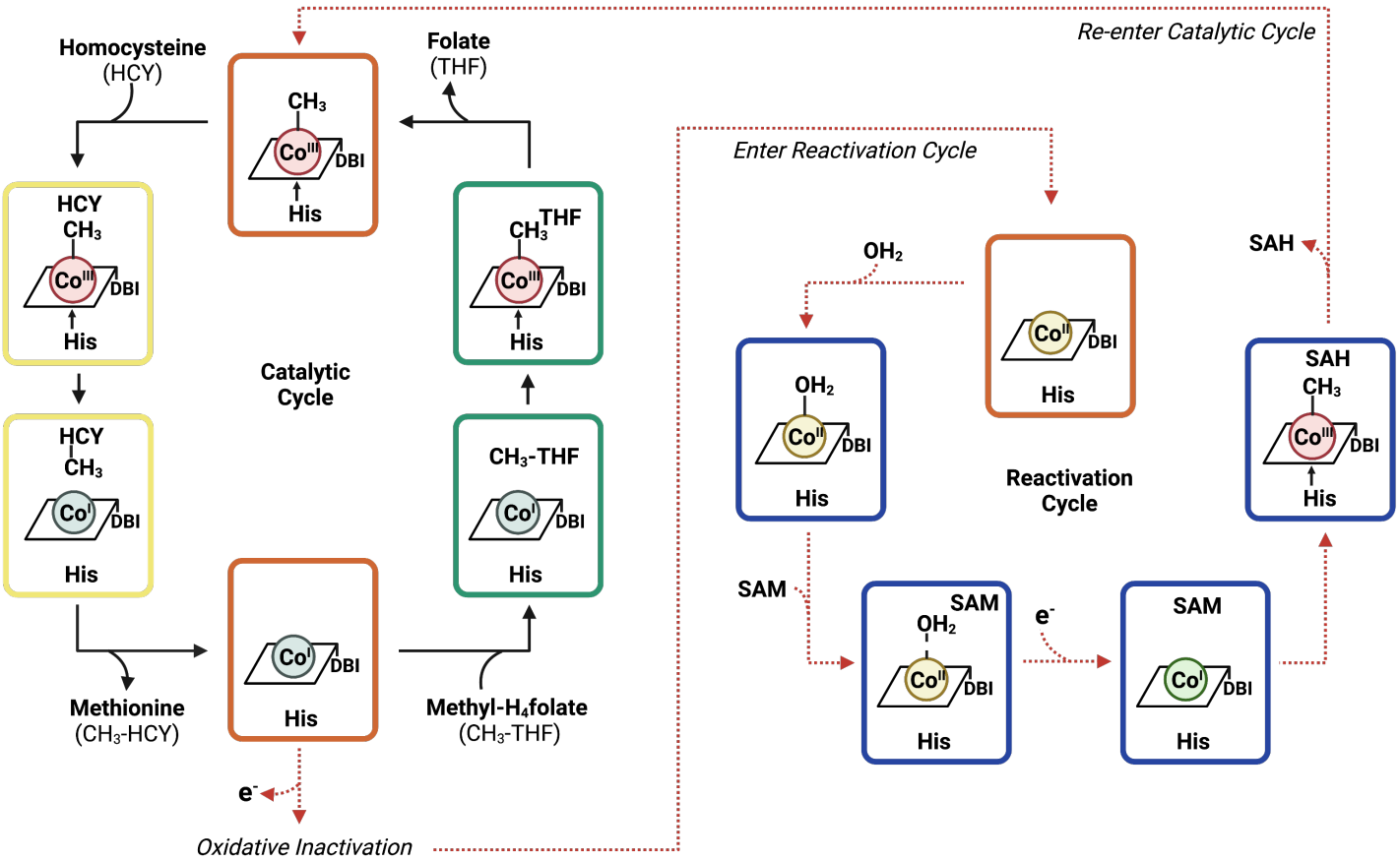


Figure 3.2: Catalytic and reactivation cycles of Methionine Synthase

During the catalytic cycle (black arrows) MS catalyzes two methyl-transfers, generating methionine and H_4-folate. Upon inactivation, MS enters the reactivation cycle (red dashed arrows) to restore catalytic competency to the cofactor. Box color refers to the MS conformation that supports the reaction, each with different substrate binding domains positioned above the cofactor. Red boxes indicate the 'cap-on' or 'resting state' that is adopted between catalytic steps. Yellow indicates the Hcy domain is positioned above the Cob-domain, and green corresponds to Fol-domain on top. Finally, the reactivation cycle is supported by the Act-domain positioned above Cob-domain, shown in blue.

In addition to cycling between different oxidation states, the coordination environment of the cofactor is tightly controlled by association and dissociation of different axial ligands to achieve each reaction (Figure 3.3) (31–34). Cobalamins consist of a cobalt (Co) center equatorially

suspended in a tetrapyrrole corrin ring with a lower nucleotide loop holding the 5,6-dimethylbenzimidazole (DBI) base (Figure 3.1 and Figure 3.3B). In the 'base-on' form, the DBI tail is ligated to Co in the lower axial position (α face) (Figure 3.3B). MS is part of a subclass of Cbl-dependent enzymes that bind Cbl in the unique 'base-off/His-on' form where the DBI tail dissociates and is replaced by a conserved histidine that coordinates Co from the lower α face (Figure 3.3B). Throughout the catalytic cycle, His alternates between association with the Co (His-on) and dissociation (His-off), and this, coupled with the identity of the upper axial ligand (β face), effectively tunes the reactivity of the Co center to achieve each successive methyl-transfer (33–35).

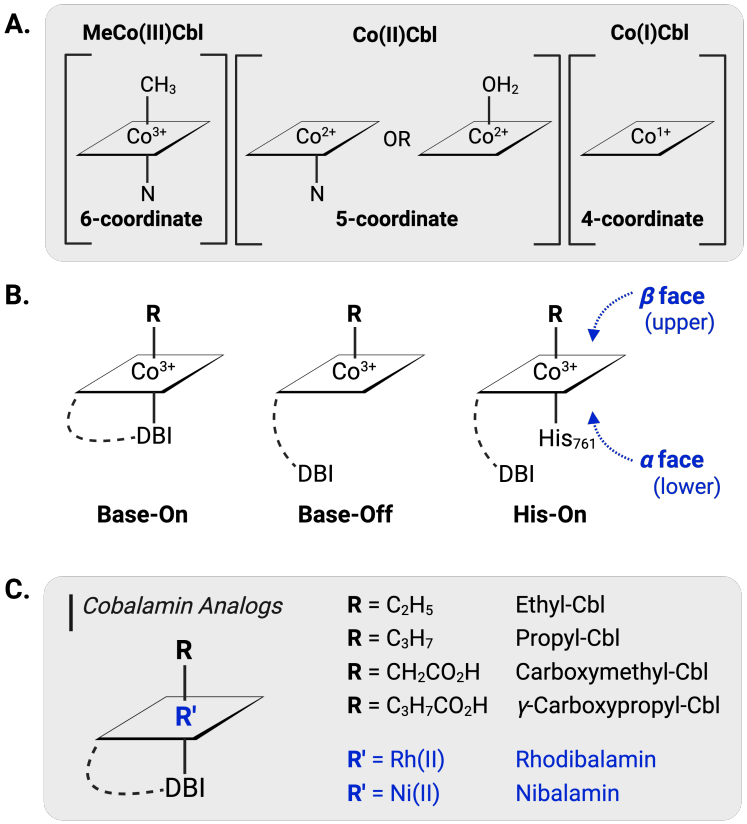


Figure 3.3: Cobalamin coordination environment

(A) Coordination environment of cobalamin in different cobalt oxidation states. **(B)** Base-on/Base-off. In solution, the DBI tail is coordinated in the lower axial position. Upon binding, the DBI tail is replaced with His to achieve the His-on state. **(C)** Naturally occurring and synthetic analogs of cobalamin cofactors.

To support these varied reactions, MS is a large multi-modular enzyme with specialized domains that bind each requisite substrate (HCY, MTF, and SAM) and the Cbl cofactor (Figure 3.4) (35,36). Previous studies established the overall function of each module and revealed the dynamic complexity of this system (24,26,37,38). The conserved Cobalamin-binding domain (Cob, red) carries Cbl, and the adjacent Capping-domain (Cap, pink) protects the reactive cofactor from unwanted side-reactions between catalytic steps (Figure 3.4C) (39). The Homocysteine-binding domain (Hcy, yellow) activates HCY for methylation to methionine and the closely associated Folate-binding domain (Fol, green) fosters the methyl-transfer from MTF to Co(I), forming THF and completing the catalytic cycle. The reactivation cycle is supported by the Activation-domain (Act, blue), which binds SAM to regenerate active MeCbl cofactor after inactivation (Figure 3.4) (26,40).

Given the complexity of this enzyme, understanding the dynamic molecular motions to protect, activate, and catalyze each reaction is a significant challenge. Crystal structures have offered some structural insights but are largely limited to excised single domains or di-domain constructs (SI Figure 3.1) (22,37,39,41–46). This 'divide-and-conquer' strategy has been required to structurally characterize MS due to significant challenges in obtaining and crystallizing the entire protein (39,40). Thus, the full-length enzyme has never been captured, nor has MS been captured in a catalytic conformation, with the Fol- or Hcy- domain positioned above the Cob domain (SI Figure 3.1). The structure of the B_{12}-binding fragment from *E. coli* MS (PDB: 1BMT) depicts a 'resting' state of MS (Figure 3.4C, Cap-on) (39). The cofactor is bound by the Cob-domain in the 'Base-off/His-on' conformation with the DBI tail inserted into the nucleotide binding cleft. The upper β face of MeCbl is shielded by the helical Cap domain, suggesting that removal of the Cap, or 'uncapping', is a precursor to any desired chemistry at the B_{12}-cofactor

(39,47). Given the reactivity of Cbl, this structure likely represents an intermediate/transition state of the enzyme that is adopted between each catalytic and reactivation step to protect the fidelity of the cofactor (Figure 3.2 and Figure 3.4C). Using tri-domain constructs (Cap+Cob+Act), MS was captured in the 'reactivation' conformation (Act:Cob, PDB: 3BUL (41), 1K98 (42), 3IVA (43)) where the Act domain is positioned above the Cob-domain and the Cap is displaced from its protective position (41–43). *Apo*- and *holo*- structures of the Hcy and Fol domains (PDB: 1Q7M (45), 3BOL (22), 5VON (46), 5VOP (46)) inform on substrate binding and activation but are insufficient to interrogate the mechanistic details of MS-catalyzed methyl transfer (22,45,46). Structures of MS in catalytically relevant conformations (i.e., with a substrate-binding domain positioned above the cofactor) are lacking but essential to address persistent questions about the orientations and dynamic molecular juggling that support catalysis.

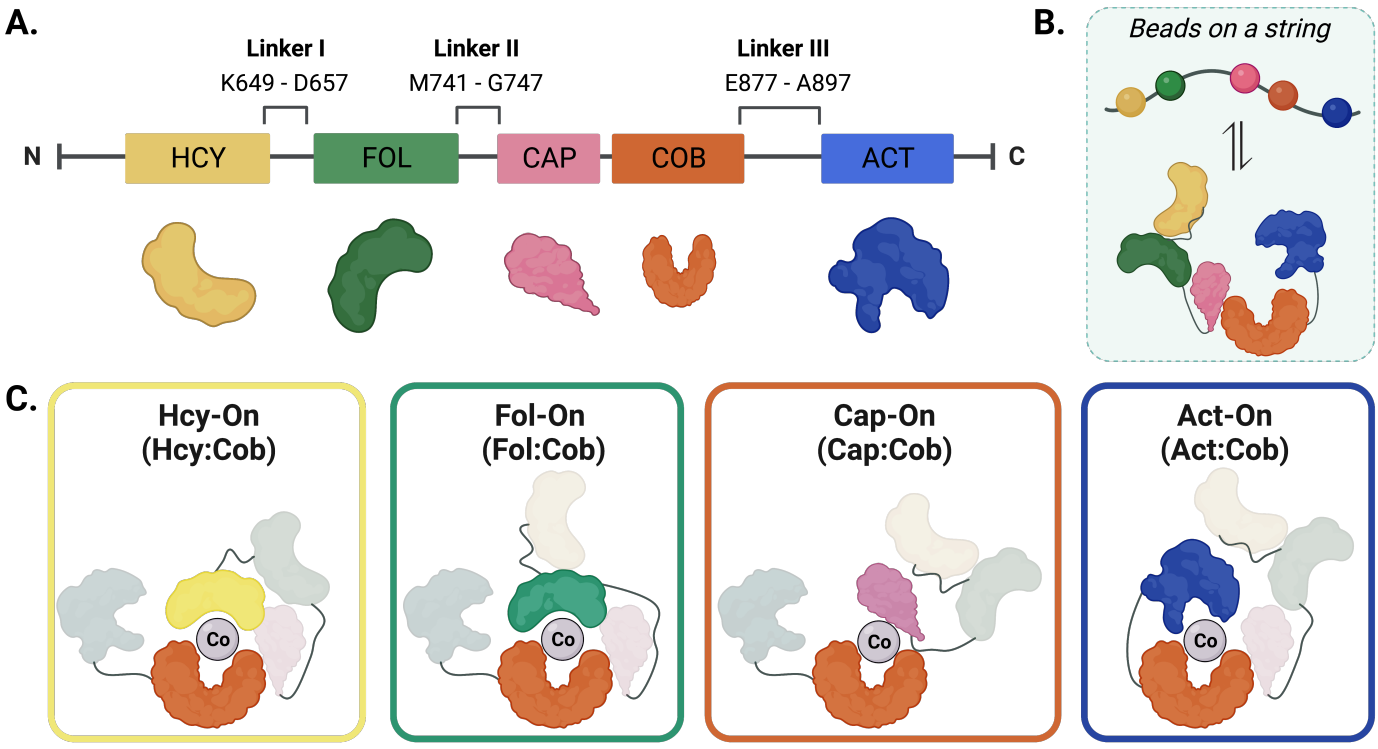


Figure 3.4: Modular organization and dynamic conformations of MS

(A) Domain organization of modules in MS, including the homocysteine binding domain (HCY, yellow), folate binding domain (FOL, green), cap (pink), cobalamin binding domain (red) and the reactivation domain (blue). Cartoon

representations of each domain shown below. **(B)** The MS domains are connected by long linkers (indicated in A) akin to beads on a string. **(C)** MS conformations that must be adopted for catalysis and reactivation.

Not for lack of effort, the dearth of structural insight into MS catalysis and dynamics is largely due to substantial biochemical challenges in working with MS. Traditional sources of MS include homologs from *H. sapiens*, *E. coli*, and *Thermotoga maritima*. These homologs are proven to be highly susceptible to proteolytic degradation (39,40) and require expression in specialized media supplemented with B_{12} to ensure proper folding and cofactor incorporation. This has prevented characterization of the *apo*-form of the protein, as well as loading of any natural or synthetic cobalamin analogs (Figure 3.3C). Mutations and biochemical manipulations (ex. chemical labelling) are challenging and often unsuccessful. In total, the difficulties associated with expression and handling have posed a significant barrier to structural and enzymatic characterization of MS.

To alleviate these challenges, we sought out alternative sources of MS that would allow us to capture the catalytic conformations of MS, as well as the full-length structure to facilitate robust characterization of MS dynamics and catalysis. To that end, we selected cobalamin-dependent MS from *Thermus thermophilus* (*t*MS) as a model system as thermophiles have long been appreciated for their enhanced stability which can facilitate the crystallization of challenging proteins (48–52). Using this model, we successfully captured the first full-length structure of MS and the first catalytic structure with the folate domain positioned above the cofactor.

3.2 Results and Discussion

3.2.1 MS without the BS

The *Thermus thermophilus* HB8 genome codes for a five-module cobalamin-dependent MS that shares the same domain architecture as *H. sapiens* and *E. coli* MS (SI Figure 3.1 and SI Figure 3.2). Despite relatively high levels of sequence conservation between these homologs, the systematic challenges typically associated with overexpression and purification of MS are entirely absent in *T. thermophilus* MS. *t*MS easily expresses in standard media (LB/TB), does not require supplementation with B_{12}, and is generally resistant to proteolytic degradation. Remarkably, for the first time we can robustly express and purify the apoenzyme and incorporate exogenous Cbl in a simple post-purification heat step to obtain active enzyme (SI Figure 3.3). We have successfully incorporated 8 different Cbl analogues including forms that exist naturally but are not cofactors for MS (e.g., Adenosyl- and CN-Cbl), and multiple synthetic Cbl-analogues (e.g., γ-carboxypropyl-Cbl) (Figure 3.3C). Furthermore, *t*MS is highly amenable to protein engineering and biochemical manipulation - we have been able to create, express, and purify all of the mutants and engineered constructs that we have attempted and each of the *t*MS modules can be expressed and purified individually and as combined tetra-, tri-, and di-modules.

Intrigued by the apparent versatility of this system, we sought out to determine if the catalytic properties of *t*MS are comparable to that of human or *E. coli* MS. We first evaluated the thermostability of *apo-t*MS, which remained in the soluble fraction up to 60°C (SI Figure 3.3). Upon the addition of MeCbl, *t*MS remained soluble at 70°C, highlighting the stabilizing effect of cofactor incorporation. Heat-tolerance is expected from enzymes derived from thermophilic organism and poses an added benefit to ease purification by adding a heat-step to denature host-enzymes, leaving just the thermophilic target.

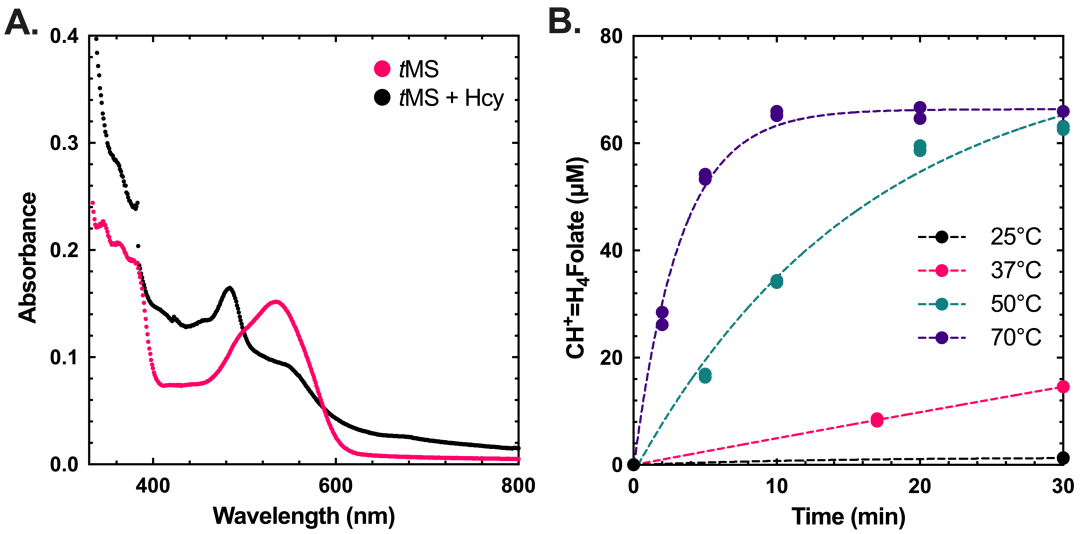


Figure 3.5: Biochemical characterization of *t*MS

(A) Spectra of methylated and de-methylated of CH_3-cobalamin bound to MS. CH_3-cobalamin-bound tMS was prepared using CH_3-cobalamin and apo-tMS (pink line). The holoenzyme was incubated with excess homocysteine at 25 ˚C in the presence of dithiothreitol and tris(2-carboxyethyl)phosphine (black line). **(B)** Effect of temperature on catalytic turnover. A fixed amount of tMS was used, and the reaction mixture was incubated for up to 30 min. The product, H_4folate, was converted to methenyltetrahydrofolate, which was measured at each time point.

We confirmed *t*MS was active by measuring the absorbance change of MeCbl-*t*MS upon the addition of substrate HCY (Figure 3.5A). To probe the effect of temperature on *t*MS activity and substrate affinity, *t*MS was incubated at various temperatures with the requisite substrates for turnover (MTF and HCY) and reactivation (SAM); product formation was monitored by UV-VIS at 350 nm (Figure 3.5B and SI Figure 3.4). *t*MS was moderately active at 37°C and 50°C, and most active at 70°C, with K_M values for MTF and HCY of 18 μM and 9.3 μM, respectively. Overall, *t*MS activity and affinity is comparable to that of *H. sapiens* and *E. coli* MS (SI Table 3.1). Taken together, this suggests that *t*MS may be an exceptional model to interrogate the mechanisms and molecular motions that support each methyl-transfer by MS.

3.2.2 First look at full-length methionine synthase

Given the significant stability of this model, we sought to establish the structural context of MS catalysis and visualize the enzyme in its entirety. To that end, we crystallized full-length 132 kDa

*t*MS and solved the structure to 2.7Å resolution (Figure 3.6 and Table 3.1). *t*MS adopts a compact, wreath-like form and each domain aligns well with previously determined structures of equivalent excised domain or di-domain constructs. This represents the first full-length structure of MS and not only proves the validity of the 'divide-and-conquer' approach that, up to this point, has been exclusively used to structurally characterize MS, but also provides the previously unknown relative orientation between the N- and C-terminal halves of MS and gives insights into possible mechanisms of B_{12} loading.

The active site of the helmet-shaped Act-domain is positioned above the Cob-domain, where the cofactor sits, similar to the position of these domains in the reactivation conformation (Figure 3.6A)(41–43). Notably, *t*MS crystallized in the apo-form; despite the absence of Cbl, the Cob-domain retains the same Rossmann-fold demonstrated by superposition with the Cbl-bound Cob-domain of *E. coli* MS (SI Figure 3.5). The helical Cap is displaced ~25Å from the resting state and is virtually identical in form and position to previously captured structures of MS in the reactivation conformation (SI Figure 3.6).

The tandem TIM-barrel Hcy- and Fol- domains are poised adjacent and outside of the Act:Cob domains and align well with previous structures (SI Figure 3.7). The Fol-domain is characterized by a unique TIM-barrel (46). In contrast to the canonical $\beta_8\alpha_8$-barrel, the Fol-domain is a $\beta_8\alpha_7$-motif where the last antiparallel helix-turn-helix motif is replaced by a helical bundle that interacts with the exterior of the domain. This fold is highly conserved and essentially identical in the excised Fol-domain and full-length structures (SI Figure 3.7). The helical bundle is contiguous with the linker connecting the Fol-domain to the Cap and may allow the restructuring of the linker to position the Fol-domain above the cofactor for catalysis.

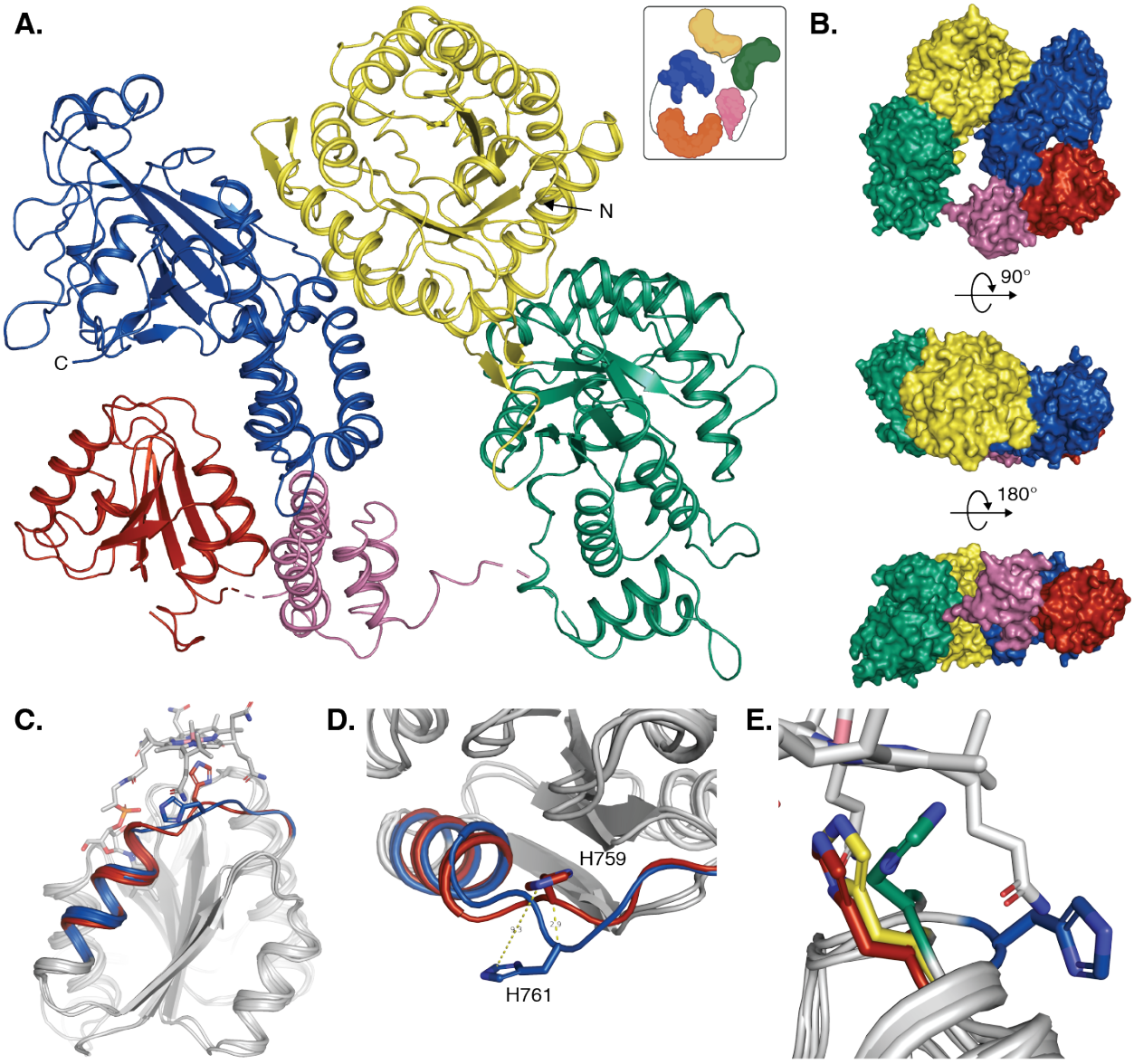


Figure 3.6: Full-length methionine synthase

(A) Structure of Full-length cobalamin-dependent methionine synthase from Thermus thermophilus. Homocysteine binding domain (Hcy) shown in yellow, Folate binding domain (Fol) shown in green, Cap is shown in pink, Cobalamin binding domain (Cob) shown in red, and the Activation domain (Act) in blue. **(B)** Surface representations of Full-length MS with same domain color as A. **(C)** Superposition of the Apo Cob-domain from tMS (gray, blue helix) with the "Resting state" Cob-domain from E-coli MS (PDB: 1BMT, gray, red helix) bound to MeCbl (gray) showing restructuring of nacent alpha helix 1 upon cobalamin binding. **(D)** Top-down view of nascent alpha helix one. Conserved Histidine761 from tMS shown in blue sticks, and His759 from E. coli MS (1BMT) shown in red sticks. **(E)** Alignment of Cob-domains with conserved Histidine shown in sticks from Full-length tMS (blue), E.Coli MS in the reactivation conformation [Cob(II), His-off] (PDB: 3BUL, green), E. Coli MS in reactivation conformation [AquoCbl, His-on] (PDB: 3IV9, yellow), and E.coli MS in the resting state [MeCbl, His-on] (PDB: 1BMT, yellow).

The Hcy- and Fol-domains move together as a rigid body, where an inter-domain linker (Linker I, Figure 3.4A) serves to 'zip' the two domains together, effectively securing them in place.

The edge of the Hcy-domain rests against the distal side of the Act-domain (Figure 3.6a), forming extensive interdomain contacts through >25 hydrogen bonds and salt bridges resulting in a buried surface area of ~900Å (SI Table 3.2). PISA analysis of this interface, however, gives a complex significance score (CSS) of 0, indicating that this interaction is not strong and not essential for complex formation. Three linker regions – Lys649-Asp657 (Linker I), Met741-Gly747 (Linker II), and Glu877-Ala897 (Linker III)– could not be modeled due to lack of sufficient electron density, emblematic of the dynamic nature of these linkers. Overall, this represents the first full-length structure of methionine synthase from any organism and consolidates years of research on the individual domains: finally visualizing the entire ensemble.

MS binds Cbl in the Base-off/His-on state where the DBI tail inserts into the conserved nucleotide binding cleft of the Cob-domain and is replaced by His761 in the lower axial position. His761 is part of the highly conserved regulatory triad (DxHxxG) on helix *α1* (aa747-761) that forms the upper edge of the cofactor binding site. Without Cbl bound, helix *α1* is partially unstructured and His761 is flipped out, serving to open the nucleotide binding cleft to facilitate DBI insertion (Figure 3.6C and D). Upon Cbl binding, helix *α1* becomes more ordered and His761 moves into position to interact with the cobalt (Figure 3.6E). This finding is consistent with NMR studies of the homologous MutS, the Cbl-binding domain from Glutamate Mutase that also binds Cbl in the Base-off/His-on state (53,54). They observed the stabilization of the nascent helix *α1* upon Cbl incorporation, and found that, despite this structural change, the loop with the coordinating His remains relatively dynamic.

Furthermore, B_{12} binding confers overall stability to the Cob-domain, which is reported to have a high degree of flexibility (47). The average B-factors of *apo*-Cob are 1.6X higher than the average for the full-length structure (SI Figure 3.8 and 3.9). In comparison, the average B-factors

Table 3.1: X-Ray Data Collection and Refinement Statistics for Full-length *t*MS$_{\Delta N35}$

	***t*MS$_{\Delta N35}$**
Data collection	
Beamline	APS, GMCA 23-IDB
Wavelength (Å)	1.033
Temperature (K)	100
Resolution (Å)	106.67-2.67 (2.74-2.67)
Space group	*P41212*
Cell dimensions (Å)	a = 134.84, b = 134.84, c = 174.37
Cell dimensions (°)	$\alpha = \beta = \gamma = 90$
Observed reflections	712,125 (19,504)
Unique reflections	46,133 (2,128)
R_{meas} (%)	19.43 (242.8)
R_{merge} (%)	20.09 (257.2)
$\langle I/\sigma \rangle$	10.96 (0.73)
CC(1/2)	0.998 (0.617)
Multiplicity	15.4 (9.2)
Completeness (%)	99.65 (93.21)
Overall *B* (Å^2) (Wilson plot)	62.57
Refinement	
Resolution range	52.77 - 2.87
Number of reflections (work/test set)	46,140/2,239
R_{work}/R_{free} (%)	22.5/25.9
No. of non-H atoms	
Protein	8741
Water	140
Ligand	11
B-factors (Å^2)	
Protein	115.67
Water	73.66
Ligand	153.54
Rmsd deviations	
Bond lengths (Å)	0.032
Bond angles (°)	2.54
Estimated coordinate error (Å); maximum likelihood based	0.372
Cruickshank's DPI[1] (Å)	0.391
Ramachandran plot	
Favored/allowed/outliers	94.9/4.7/0.5
MolProbity Score	1.15 (100th percentile)
PDB	----

for holo-Cob in the resting and reactivation structures are comparable the same as (0.96X and 1.06X, respectively) the total, indicating significant stabilization upon cofactor incorporation. Taken together, these findings suggest that (1) the Cob-domain is a dynamically performed structure in the absence of the cofactor, requiring minimal restructuring to facilitate B_{12} incorporation, and (2) Cbl binding likely involves a 'lock and key' type fit rather than an induced fit model. Due to the high degree of conservation, the performed nature of the Cob-domain is likely a general characteristic of this class of B_{12}-dependent enzymes.

Cobalamin-dependent MS is not known to release or exchange their cofactor after initial loading, so the full-length MS structure likely represents a 'primed-to-be-loaded' state that precedes Cbl incorporation. A significant question remains regarding how Cbl is loaded as the Act-domain blocks entry to the Cob-binding cleft. To address this, we analyzed the interacting surfaces between each domain and compared the CSS of each interface between the full-length and reactivation structures (SI Table 3.2). In the full-length structure, the most extensive interaction interface is between Act and Hcy. Though Act also forms a handful of H-bonds with the other domains (Cap/Cob/Fol), the CSS score for each interface is 0, indicating overall weak interactions. This, however, can be rationalized as displacement of Act is required for cofactor loading and strong interdomain interactions would pose a significant barrier. In contrast, the interfaces between Act, Cob, and Cap in the reactivation structure are stronger and more extensive (CSS = 0.225-1) (SI Table 3.2). Reactivation requires Act to position itself over Cob and stabilizing interdomain interactions could help maintain that position through the reactivation cycle. The comparatively weaker interfaces in the FL structure provide the requisite flexibility to easily break these interactions to enable cofactor loading.

3.2.3 Visualizing (barely) Methyl-transfer in *t*MS

As the rare stability of *t*MS enabled the successful characterization of the elusive full-length apo structure, we reasoned that *t*MS' unique cofactor tolerance could be leveraged to capture a catalytically competent methyltransferase complex to address long standing questions surrounding MS catalysis. We sought out a cofactor with a bulky group in the upper axial position that would destabilize the 'Cap-on' conformation and give substrate binding domains access to the cofactor. To that end, we crystallized the Fol-Cap-Cob *t*MS tri-module with γ-carboxypropyl Cbl (Figure 3.3C), a synthetic Cbl that does not support catalysis, and solved the structure to 2.8 Å (Figure 3.7 and Table 3.2).

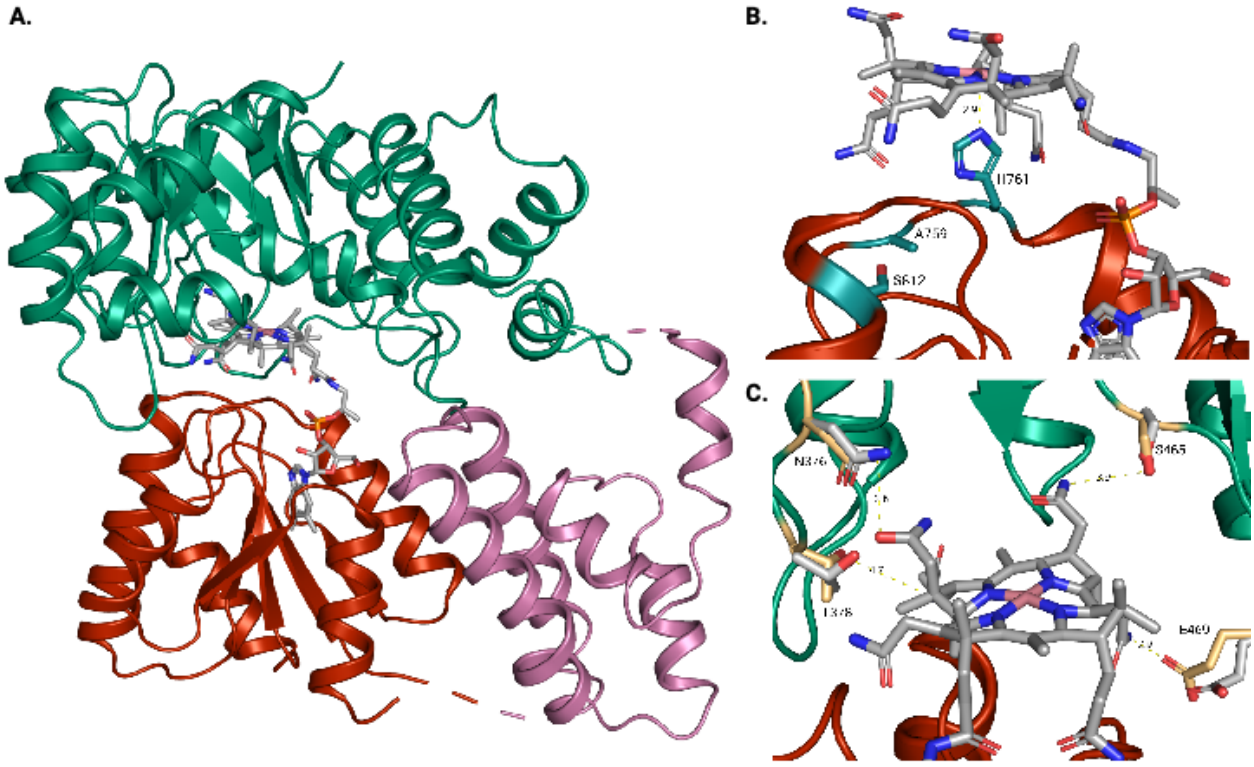


Figure 3.7: MS captured in Fol-on conformation

(A) We captured the Fol-on conformation of MS where the active site of the Folate binding domain (green) is positioned above the B_{12} cofactor (gray) bound by the cobalamin-binding domain (red). The cap (pink) is displaced from its resting position on top of the cofactor. **(B)** The catalytic triad (teal) includes His761, Ser812, and Asp759 (mutated to alanine in this construct) of the cobalamin binding domain (red). His 761 coordinates the cobalamin (gray) in the lower axial position. **(C)** Interactions between the cofactor (gray) and residues of the folate binding domain (domain in green, residues in tan), compared with structure of excised Folate domain bound to MTF (gray residues, PDB: 5VOP).

Table 3.2: X-Ray Data Collection and Refinement Statistics for Fol-on (Fol:Cob) *t*MS$_{Fol:Cob}$

	***t*MS$_{Fol:Cob}$**
Data collection	
Beamline	APS, GMCA 23-IDB
Wavelength (Å)	1.033
Temperature (K)	100
Resolution (Å)	61.21-2.87 (3.03-2.87)
Space group	*P6122*
Cell dimensions (Å)	a = 95.15, b = 95.15, c = 274.29
Cell dimensions (°)	α = β = 90, γ = 120
Observed reflections	243,752 (36,533)
Unique reflections	17,746 (2,503)
R_{meas} (%)	14.3 (247.9)
R_{merge} (%)	13.8 (239.3)
<I/σ>	11.0 (1.0)
CC(1/2)	0.999 (0.893)
Multiplicity	13.7 (14.6)
Completeness (%)	100.0 (100)
Overall *B* (Å^2) (Wilson plot)	126.55
Refinement	
Resolution range	52.77 - 2.87
Number of reflections (work/test set)	17668/905
R_{work}/R_{free} (%)	22.4/27.3
No. of non-H atoms	
Protein	7857
Water	35
Ligand	206
B-factors (Å^2)	
Protein	126.5
Water	88.7
Ligand	124.16
Rmsd deviations	
Bond lengths (Å)	0.017
Bond angles (°)	2.40
Estimated coordinate error (Å); maximum likelihood based	0.372
Cruickshank's DPI[1] (Å)	0.391
Ramachandran plot	
Favored/allowed/outliers	95.9/3.9/0.2
MolProbity Score	1.51 (100th percentile)
PDB	----

In this conformation, the Fol-domain is positioned over the Cob-domain, placing the MTF-binding site directly above Cbl-cofactor (Figure 3.7). The cofactor is bound in the 'his-on' state where His761 is coordinated through its imidazole side chain 2.8 Å from Co (Figure 3.7B). In addition to His761, two other residues (Asp759 and Ser812) make up the conserved catalytic triad that forms a hydrogen bonding network that facilitates the transition from base-off Co(I) to base-on methylcobalamin (55) (Figure 3.7B). There is only minimal electron density in the upper axial position where the carboxypropyl sits, indicating it was likely photoreduced during data collection and the cofactor has been modeled into the structure as cob(II)alamin. The Cap, destabilized from its protective position, lies to the periphery of the Cob-domain in a comparable position to the displaced Cap in the reactivation conformation (Act:Cob) (SI Figure 3.10 and 3.11). This suggests that the position of the displaced Cap is programmed, rather than random. In total, this structure represents the Fol-on (Fol:Cob) state, adopted for the methyl-transfer from MTF to Cob(I). As the only catalytic structure visualized for any methionine synthase, this structure gives unique insight into the required orientation of the Fol and Cob domains and informs on how methyl transfer proceeds.

MTF binds in the shallow, solvent exposed active site at the entrance to the TIM-barrel, comprised of several highly conserved residues involved in substrate binding (Asn464, Asn573, and Thr534) and activation (Asp411, Asp443, Asp531). The active site is largely performed, requiring only minimal restructuring to facilitate MTF binding (46). In the Apo- and MTF-bound structures, ASP443 is positioned to hydrogen bond with N8 of the MTF pterin ring. In the Fol-on structure, ASP443 is repositioned towards the cofactor, in proximity of one of the arms of the corrin ring (SI Figure 3.12). MTF binding precedes adoption of the Fol:Cob conformation and, therefore, association with the cofactor (13). Because ASP443 lies directly between the MTF

binding site and the cofactor and because no substrate is present in our structure, the repositioning of ASP443 could serve to block the cofactor from moving closer to the active site, acting as a signal that the Fol-domain is not ready for catalysis to prevent MS adopting a non-productive conformation.

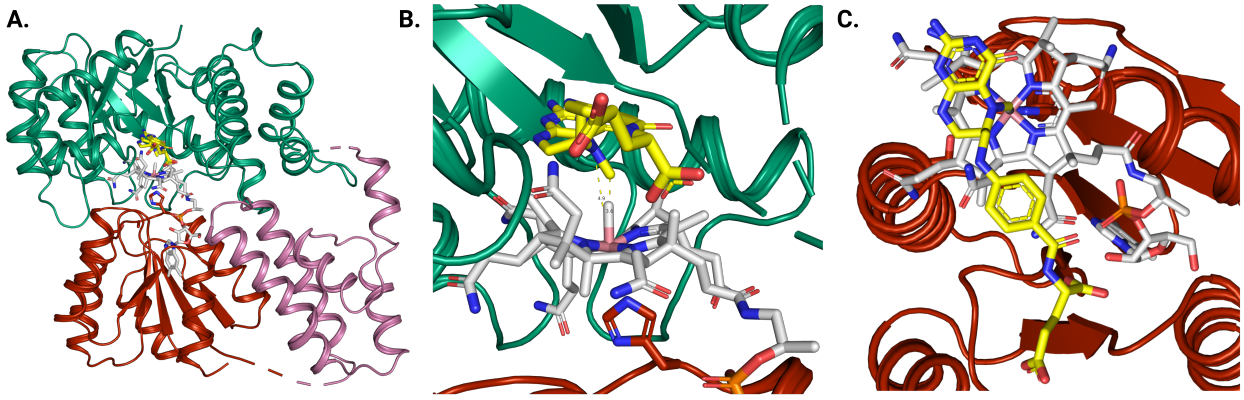


Figure 3.8: MTF modeled into folate active site

(A) Folate-on structure with MTF (yellow, PDB:5VOP) modeled into the active site of the Folate domain (green). MTF is positioned above the cobalamin cofactor (gray) bound by the cobalamin binding domain (red), adjacent to the cap (pink). **(B)** The cobalt ion is 3.6Å from the carbon of the methyl group and 4.9Å from N5 of the pterin ring. **(C)** Top-down view of MTF and cobalamin where the methyl group is positioned in-line with cobalt.

Previous structures of MTF bound MeTr/CFeSp was captured in an 'en route' conformation as the distance between the cofactor and MTF was too large to support catalysis (~8 Å) and further motions are required to bring the Co closer to the methyl group (47). SN2 nucleophilic displacement, favored as the mechanism of methyl transfer in Cbl-dependent enzymes, requires an expected reaction distance of 3-4 Å between the cobalt center and the MTF-methyl (56). To determine if the orientation captured in the Fol-on structure could support catalysis, we modeled MTF into the active site and found that the N5-methyl of MTF is positioned directly above and in-line with the cofactor, ~3.6-4.9 Å from Co-center (Figure 3.8). Both the arrangement and proximity of the reacting centers, this likely represents either a catalytically-competent structure, or a 'right-before-catalysis' state where only minimal, local rearrangements

are required to bring the cobalt into reaction distance. Regardless, this structure is the closest to a catalytic structure captured for any corrinoid protein.

3.2.4 Molecular basis of conformational dynamics/molecular juggling

The distribution of MetH conformations is tightly controlled and largely dependent on the identity of bound ligands and the oxidation state of the cofactor. Questions remain, however, in determining the factors that directly stabilize each individual domain over the cofactor at different points in the catalytic cycle. To address this, we analyzed the protein:protein and protein:cofactor interactions/interfaces in the Fol-on structure and compared how they change from resting, and reactivation structures. From above, Cobalamin is secured/positioned directly under the Fol- active site through four H-bonds between conserved residues (Asn376, Thr378, Ser465, and Glu469) lining the outer edge of the MTF-binding site (PISA: CSS 1) (Figure 3.7C and SI Table 3.2). Compared to the MTF-bound Fol-domain, Glu469 has repositioned to move closer to the cofactor by shifting ~1.9Å and rotating towards the cofactor (SI Figure 3.13).

From below, Cobalamin is held by the Cob-binding domain. In the resting structure, 11 hydrogen bonds secure the cofactor to the cobalamin binding domain. In the fol-on structure, the Cob-domain forms four hydrogen bonds with cobalamin: three H bonds are through the same residues (Ser806, Ala860, Val760) at the same position on the cofactor and a novel H bond made to O8R from conserved Ala 860 not formed in the resting state structure (SI Table 3.2). Notably, the hydrogen bonds are slightly longer in the Fol-on structure - for the three H-bonds shared between the two structures, the average H-bond length in the Fol-on structure is 3.44Å vs. 2.98Å in the resting structure. The same trend holds for the resting vs. reactivation structures. The Cob-domain in the reactivation structure maintains 6 of the 11 hydrogen bonds found in the resting state

and the average H-bond length is longer in the reactivation conformation (3.23Å vs 3.01Å, shared H-bonds) (SI Table 3.2).

Rather than relying on protein:protein domain interactions, cobalamin appears to be the primary mediator of interactions with whichever domain is coming on top. In the Fol-on conformation, there are minimal inter-domain interactions formed and this trend is carried over in the reactivation structure. Fol moving on top allows for the redistribution of Cbl interactions, where Cbl weakens the interactions between the Cob-domain in favor of forming new, stabilizing contacts with the Fol-domain. The result of severing Cob-Cbl interactions allows for the requisite movement of the cofactor towards MTF, as was observed in the CFeSP/MeTr structure. Mediation of interactions through the cofactor allows positioning of the right domain on top at the right time, preventing the protein from adopting a non-productive conformation if the right form of the cofactor or substrate is not present.

3.3 Methods

3.3.1 Expression of *Thermus thermophilus* MS

Construction of expression vector - The original clone of *Thermus thermophillus* methionine synthase (NCBI Gene Locus tag: TTHA0618) was obtained from Riken BioResource Center (57). The gene, originally cloned in pET-11a, was subcloned into pMCSG7 vector using ligation independent cloning technique (58). As a result, tMS was expressed as an N-terminal His-tagged form, which was removed by recombinant *Tobacco Etch Virus* Protease treatment. The expression vector was designated pMCSG7(tMS^{wt}). For crystallization purposes, 35 N-terminal amino acid residues were removed. To express the truncated tMS, pMCSG7($tMS^{\Delta N35}$) was constructed.

Protein Expression - To produce the protein, *BL21star(DE3)* was used as a host. The transformed *E. coli* with pMCSG7(tMS^{wt}) was propagated at 37 ˚C in Lennox Broth containing 50μg/ml ampicillin, and Isopropyl β-D-1-thiogalactopyranoside (final concentration 0.1 mM) was added to produce the protein. Cells were harvested four hours after induction and stored at -80 ˚C.

3.3.2 Purification of *Thermus thermophilus* MS

For holoenzyme purification, CH_3-cobalamin (Sigma) was added into the crude cell lysate to form holo-tMS. The crude extract with CH_3-cobalamin was incubated at 70 ˚C for 15 min, then centrifuged. The supernatant was pooled and applied to a Ni-affinity column (His-trap Chelating HP, GE). Protein was eluted with buffer consisting of 50mM KPB, 0.3M sodium chloride, and 75mM imidazole. Colored fractions were collected and dialyzed at 4 ˚C overnight in the presence of *Tobacco Etch Virus* Protease to remove the Histidine-tag on the N-terminus. The dialyzed sample was incubated at 70 ˚C for 15 min to quench the protease function, then centrifuged. Concentrated, purified tMS was stored at 4˚C. The highly purified enzyme was stable for at least

a month. Purification of apo-tMS was done in the absence of cobalamin. The procedure was the same as that of holo-tMS preparation, except heat-treatment was omitted. The N-terminal truncation did not affect expression efficiency and protein properties.

3.3.3 Biochemical characterization of *Thermus thermophilus* MS

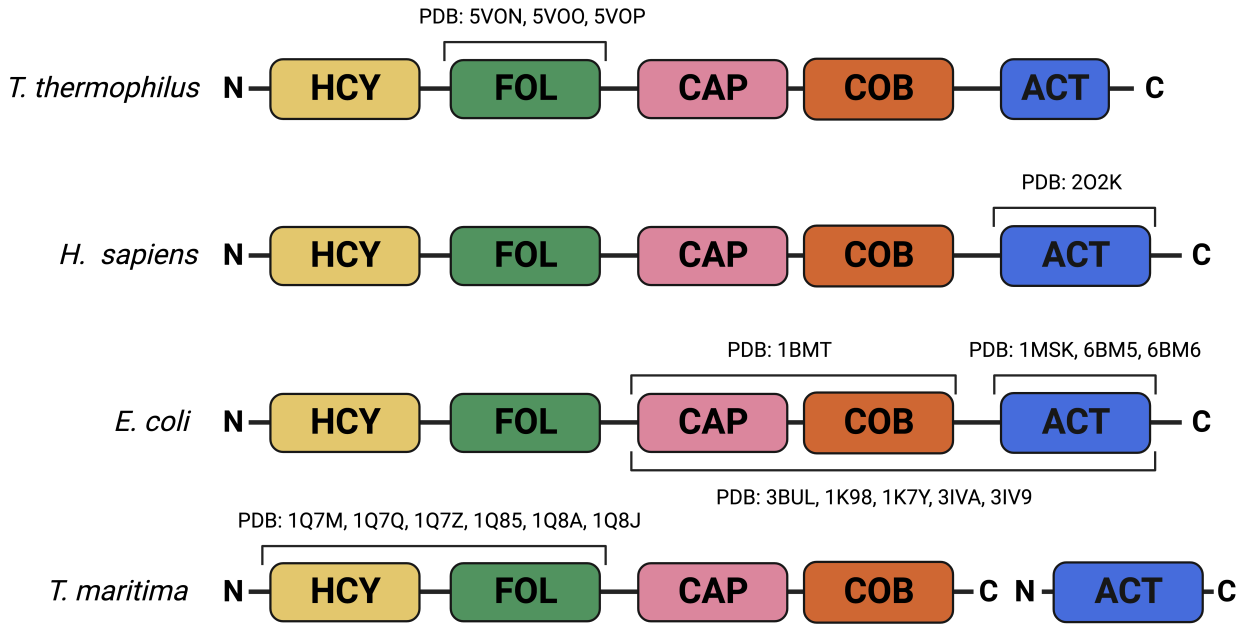
Non-radioactive assays were employed to measure tMS activities with minor modification (59). Briefly, the reaction mixture contained 0.5 M DTT, 0.5 mM aquacobalamin (Sigma), 0.1 M homocysteine, 0.25 mM *6S*-CH_3-H_4folate (Merck Eprova AG), and 3.8 mM AdoMet (Sigma) in 100 mM KPB (pH 7.2). Preincubation was started with addition of diluted enzyme and then incubated at 50˚C for 5 min. Reactions were initiated by adding CH_3-H_4folate, then quenched by mixing with formic acid-HCl, followed by heating at 80˚C for 10 min. After cool-down, the reaction formed methenyltetrahydrofolate, and the concentration was determined using the extinction coefficient of 26.5 $x10^3$ M^{-1} cm^{-1} at 350 nm.

3.3.4 Crystallization of *Thermus thermophilus* MS

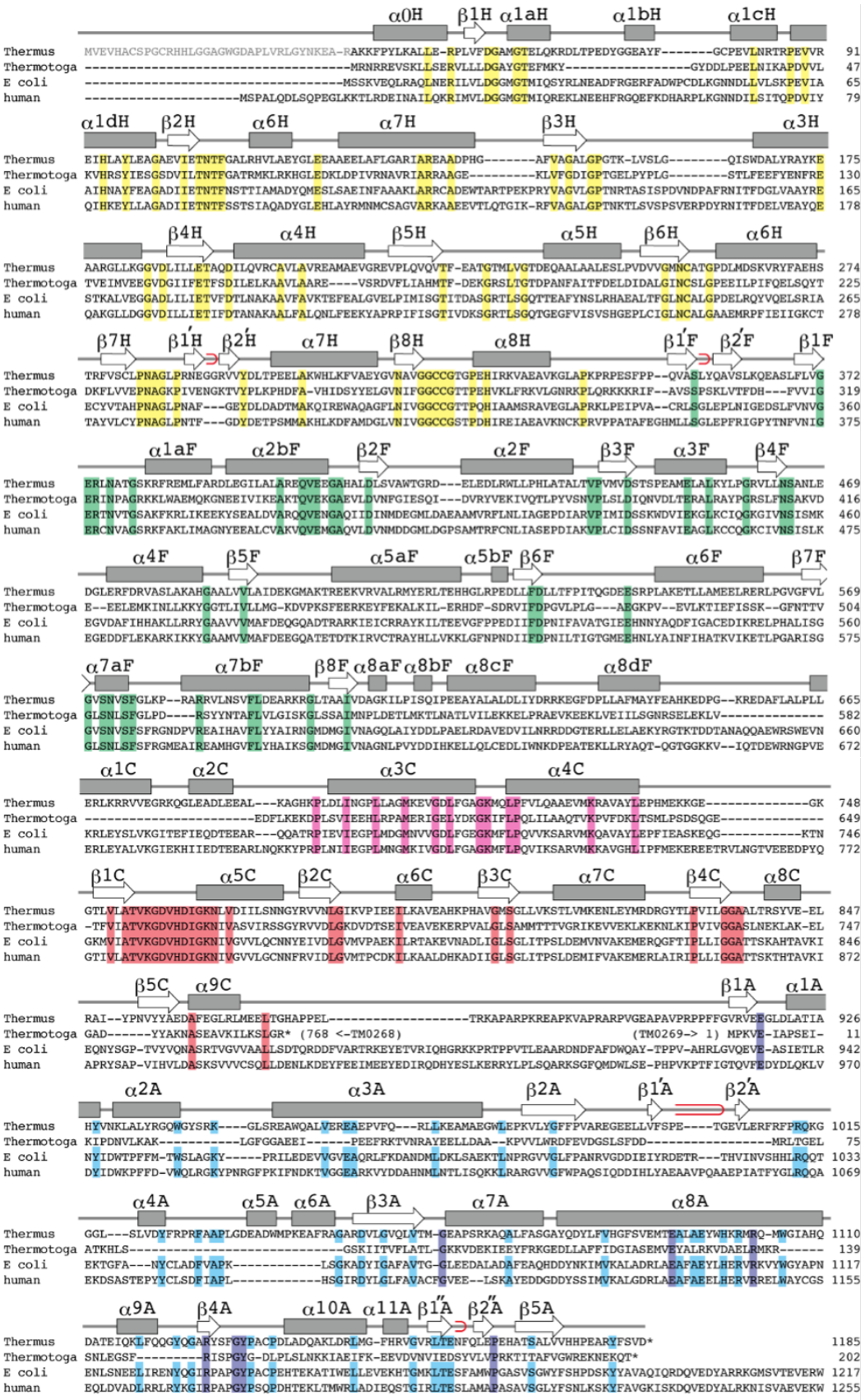
Crystallization. (#memo 2013Nov_trip, P13-2, crystal prep 130730 Index F2_A). To prepare crystals, the sitting drop vapor diffusion technique was employed. Purified apo-$tMS^{\Delta N35}$ (~15mg/ml) was mixed with 0.2M Trimethylamine N-oxide, 0.1 M Tris-HCl (pH 8.5), 20% (w/v) Polyethylene glycol monomethylether 2,000 in a 1:1 ratio and incubated at 4˚C. Crystals put on a fiber loop in the presence of glycerol as cryoprotectant were frozen in liquid nitrogen.

Data collection. X-ray data sets were collected at 100K on GM/CA CAT beamline 23-ID-B at the Advanced Photon Source, Argonne National Laboratory (Argonne, IL).

3.4 Supplementary Information

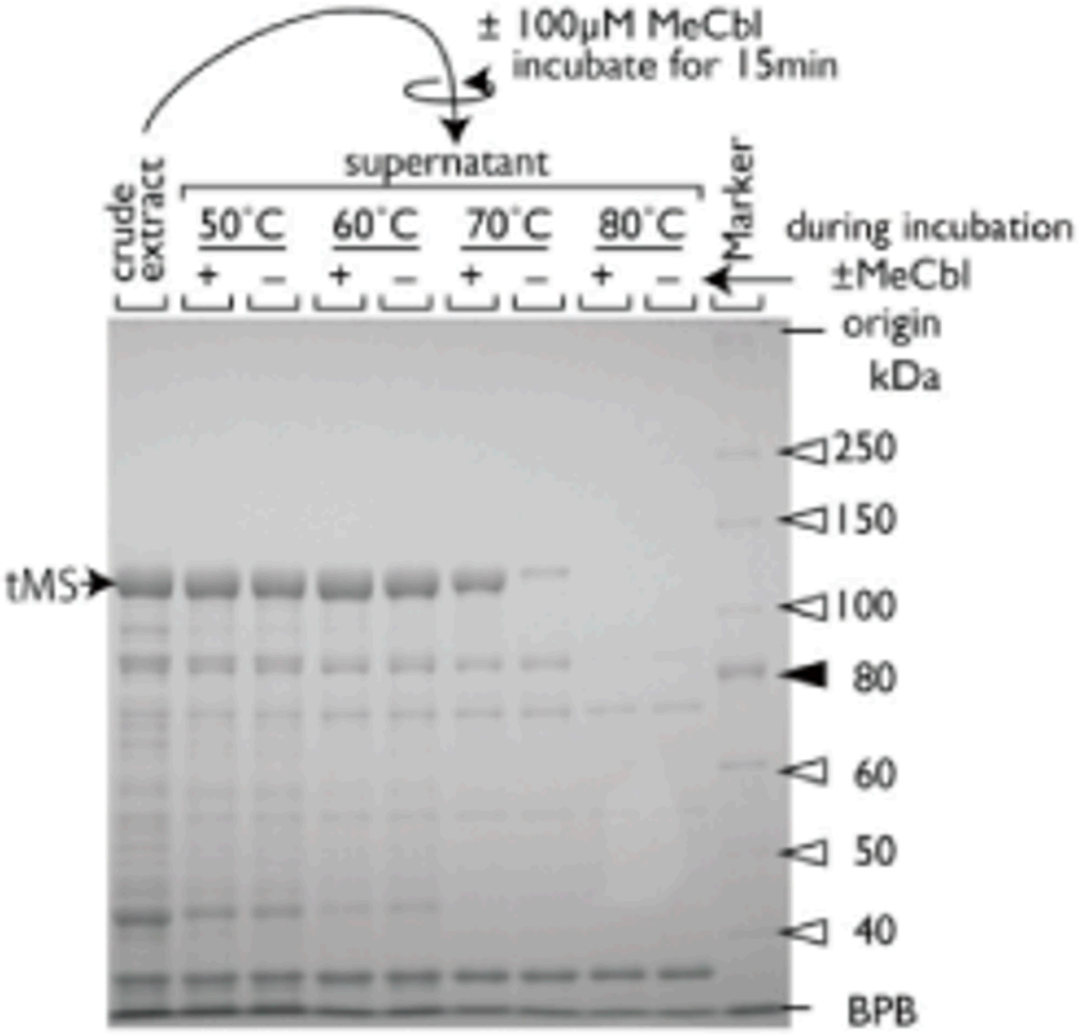


SI Figure 3.1: MS domain organization and PDB structures


α0H β1H α1aH α1bH α1cH
Thermus MVEVHACSPGCRHHLGGAGWGDAPLVRLGYNKEA-RAKKFPYLKALLE-RPLVFDGAMGTELQKRDLTPEDYGGEAYF-------GCPEVLNRTRPEVVR 91
Thermotoga ----------MRNRREVSKLLSERVLLLDGAYGTEFMKY----------GYDDLPEELNIKAPDVVL 47
E coli ----------MSSKVEQLRAQLNERILVLDGGMGTMIQSYRLNEADFRGERFADWPCDLKGNNDLLVLSKPEVIA 65
human ----------MSPALQDLSQPEGLKKTLRDEINAILQKRIMVLDGGMGTMIQRELKNEEHFRGQEFKDHARPLKGNNDILSITQPDVIY 79
α1dH β2H α6H α7H β3H α3H
Thermus EIHLAYLEAGAEVIETNTFGALRHVLAEYGLEEAAEELAFLGARIAREAADPHG-------AFVAGALGPGTK-LVSLG--------QISWDALYRAYKE 175
Thermotoga KVHRSYIESGSDVILTNTFGATRMKLRKHGLEDKLDPIVRNAVRIARRAAGE---------KLVFGDIGPTGELPYPLG--------STLFEEFYENFRE 130
E coli AIHNAYFEAGADIIETNTFNSTTIAMADYQMESLSAEINFAAAKLARRCADEWTARTPEKPRYVAGVLGPTNRTASISPDVNDPAFRNITFDGLVAAYRE 165
human QIHKEYLLAGADIIETNTFSSTSIAQADYGLEHLAYRMNMCSAGVARKAAEEVTLQTGIK-RFVAGALGPTNKTLSVSPSVERPDYRNITFDELVEAYQE 178
β4H α4H β5H α5H β6H α6H
Thermus AARGLLKGGVDLILLETAQDILQVRCAVLAVREAMAEVGREVPLQVQVTF-EATGTMLVGTDEQAALAALESLPVDVVGMNCATGPDLMDSKVRYFAEHS 274
Thermotoga TVEIMVEEGVDGIIFETFSDILELKAAVLAARE----VSRDVFLIAHMTF-DEKGRSLTGTDPANFAITFDELDIDALGINCSLGPEEILPIFQELSQYT 225
E coli STKALVEGGADLILIETVFDTLNAKAAVFAVKTEFEALGVELPIMISGTITDASGRTLSGQTTEAFYNSLRHAEALTFGLNCALGPDELRQYVQELSRIA 265
human QAKGLLDGGVDILLIETIFDTANAKAALFALQNLFEEKYAPRPIFISGTIVDKSGRTLSGQTGEGFVISVSHGEPLCIGLNCALGAAEMRPFIEIIGKCT 278
β7H β1′H β2′H α7H β8H α8H β1′F β2′F β1F
Thermus TRFVSCLPNAGLPRNEGGRVVYDLTPEELAKWHLKFVAEYGVNAVGGCCGTGPEHIRKVAEAVKGLAPKPRPESFPP--QVASLYQAVSLKQEASLFLVG 372
Thermotoga DKFLVVEPNAGKPIVENGKTVYPLKPHDFA-VHIDSYYELGVNIFGGCCGTTPEHVKLFRKVLGNRKPLQRKKKRIF--AVSSPSKLVTFDH---FVVIG 319
E coli ECYVTAHPNAGLPNAF---GEYDLDADTMAKQIREWAQAGFLNIVGGCCGTTPQHIAAMSRAVEGLAPRKLPEIPVA--CRLSGLEPLNIGEDSLFVNVG 360
human TAYVLCYPNAGLPNTF---GDYDETPSMMAKHLKDFAMDGLVNIVGGCCGSTPDHIREIAEAVKNCKPRVPPATAFEGHMLLSGLEPFRIGPYTNFVNIG 375
α1aF α2bF β2F α2F β3F α3F β4F
Thermus ERLNATGSKRFREMLFARDLEGILALAREQVEEGAHALDLSVAWTGRD---ELEDLRWLLPHLATALTVPVMVDSTSPEAMELALKYLPGRVLLNSANLE 469
Thermotoga ERINPAGRKKLWAEMQKGNEEIVIKEAKTQVEKGAEVLDVNFGIESQI---DVRYVEKIVQTLPYVSNVPLSLDIQNVDLTERALRAYPGRSLFNSAKVD 416
E coli ERTNVTGSAKFKRLIKEEKYSEALDVARQQVENGAQIIDINMDEGMLDAEAAMVRFLNLIAGEPDIARVPIMIDSSKWDVIEKGLKCIQGKGIVNSISMK 460
human ERCNVAGSRKFAKLIMAGNYEEALCVAKVQVEMGAQVLDVNMDDGMLDGPSAMTRFCNLIASEPDIAKVPLCIDSSNFAVIEAGLKCCQGKCIVNSISLK 475
α4F β5F α5aF α5bF β6F α6F β7F
Thermus DGLERFDRVASLAKAHGAALVVLAIDEKGMAKTREEKVRVALRMYERLTEHHGLRPEDLLFDLLTFPITQGDEESRPLAKETLLAMEELRERLPGVGFVL 569
Thermotoga E--EELEMKINLLKKYGGTLIVLLMG-KDVPKSFEERKEYFEKALKIL-ERHDF-SDRVIFDPGVLPLG---AEGKPV--EVLKTIEFISSK--GFNTTV 504
E coli EGVDAFIHHAKLLRRYGAAVVVMAFDEQGQADTRARKIEICRRAYKILTEEVGFPPEDIIFDPNIFAVATGIEEHNNYAQDFIGACEDIKRELPHALISG 560
human EGEDDFLEKARKIKKYGAAMVVMAFDEEGQATETDTKIRVCTRAYHLLVKKLGFNPNDIIFDPNILTIGTGMEEHNLYAINFIHATKVIKETLPGARISG 575
α7aF α7bF β8F α8aF α8bF α8cF α8dF
Thermus GVSNVSFGLKP--RARRVLNSVFLDEARKRGLTAAIVDAGKILPISQIPEEAYALALDLIYDRRKEGFDPLLAFMAYFEAHKEDPG--KREDAFLALPLL 665
Thermotoga GLSNLSFGLPD-----RSYYNTAFLVLGISKGLSSAIMNPLDETLMKTLNATLVILEKKELPRAEVKEEKLVEIILSGNRSELEKLV---------- 582
E coli GVSNVSFSFRGNDPVREAIHAVFLYYAIRNGMDMGIVNAGQLAIYDDLPAELRDAVEDVILNRRDDGTERLLELAEKYRGTKTDDTANAQQAEWRSWEVN 660
human GLSNLSFSFRGMEAIREAMHGVFLYHAIKSGMDMGIVNAGNLPVYDDIHKELLQLCEDLIWNKDPEATEKLLRYAQT-QGTGGKKV--IQTDEWRNGPVE 672
α1C α2C α3C α4C
Thermus ERLKRRVVEGRKQGLEADLEEAL---KAGHKPLDLINGPLLAGMKEVGDLFGAGKMQLPFVLQAAEVMKRAVAYLEPHMEKKGE----------GK 748
Thermotoga ----------EDFLKEKDPLSVIEEHLRPAMERIGELYDKGKIFLPQLILAAQTVKPVFDKLTSMLPSDSQGE---------- 649
E coli KRLEYSLVKGITEFIEQDTEEAR---QQATRPIEVIEGPLMDGMNVVGDLFGEGKMFLPQVVKSARVMKQAVAYLEPFIEASKEQG----------KTN 746
human ERLEYALVKGIEKHIIEDTEEARLNQKKYPRPLNIIEGPLMNGMKIVGDLFGAGKMFLPQVIKSARVMKKAVGHLIPFMEKEREETRVLNGTVEEEDPYQ 772
β1C α5C β2C α6C β3C α7C β4C α8C
Thermus GTLVLATVKGDVHDIGKNLVDIILSNNGYRVVNLGIKVPIEEILKAVEAHKPHAVGMSGLLVKSTLVMKENLEYMRDRGYTLPVILGGAALTRSYVE-EL 847
Thermotoga -TFVIATVKGDVHDIGKNIVASVIRSSGYRVVDLGKDVDTSEIVEAVEKERPVALGLSAMMTTTVGRIKEVVEKLKEKNLKIPVIVGGASLNEKLAK-EL 747
E coli GKMVIATVKGDVHDIGKNIVGVVLQCNNYEIVDLGVMVPAEKILRTAKEVNADLIGLSGLITPSLDEMVNVAKEMERQGFTIPLLIGGATTSKAHTAVKI 846
human GTIVLATVKGDVHDIGKNIVGVVLGCNNFRVIDLGVMTPCDKILKAALDHKADIIGLSGLITPSLDEMIFVAKEMERLAIRIPLLIGGATTSKTHTAVKI 872
β5C α9C β1A α1A
Thermus RAI--YPNVYYAEDAFEGLRLMEELTGHAPPEL-------------------TRKAPARPKREAPKVAPRARPVGEAPAVPRPPFFGVRVEEGLDLATIA 926
Thermotoga GAD------YYAKNASEAVKILKSLGR* (768 <-TM0268) (TM0269-> 1) MPKVE-IAPSEI- 11
E coli EQNYSGP-TVYVQNASRTVGVVAALLSDTQRDDFVARTRKEYETVRIQHGRKKPRTPPVTLEAARDNDFAFDWQAY-TPPV-AHRLGVQEVE-ASIETLR 942
human APRYSAP-VIHVLDASKSVVVCSQLLDENLKDEYFEEIMEEYEDIRQDHYESLKERRYLPLSQARKSGFQMDWLSE-PHPVKPTFIGTQVFEDYDLQKLV 970
α2A α3A β2A β1′A β2′A
Thermus HYVNKLALYRGQWGYSRK----GLSREAWQALVEREAEPVFQ---RLLKEAMAEGWLEPKVLYGFFPVAREGEELLVFSPE----TGEVLERFRFPRQKG 1015
Thermotoga KIPDNVLKAK-----------LGFGGAEEI------PEEFRKTVNRAYEELLDAA--KPVVLWRDFEVDGSLSFDD----------MRLTGEL 75
E coli NYIDWTPFFM-TWSLAGKY-----PRILEDEVVGVEAQRLFKDANDMLDKLSAEKTLNPRGVVGLFPANRVGDDIEIYRDETR---THVINVSHHLRQQT 1033
human DYIDWKPFFD-VWQLRGKYPNRGFPKIFNDKTVGGEARKVYDDAHNMLNTLISQKKLRARGVVGFWPAQSIQDDIHLYAEAAVPQAAEPIATFYGLRQQA 1069
α4A α5A α6A β3A α7A α8A
Thermus GGL---SLVDYFRPRFAAPLGDEADWMPKEAFRAGARDVLGVQLVTM-GEAPSRKAQALFASGAYQDYLFVHGFSVEMTEALAEYWHKRMRQ-MWGIAHQ 1110
Thermotoga ATKHLS----------------------------GSKIITVFLATL-GKKVDEKIEEYFRKGEDLLAFFIDGIASEMVEYALRKVDAELRMKR------ 139
E coli EKTGFA---NYCLADFVAPK------------LSGKADYIGAFAVTG-GLEEDALADAFEAQHDDYNKIMVKALADRLAEAFAEYLHERVRKVYWGYAPN 1117
human EKDSASTEPYYCLSDFIAPL------------HSGIRDYLGLFAVACFGVEE--LSKAYEDDGDDYSSIMVKALGDRLAEAFAEELHERVRRELWAYCGS 1155
α9A β4A α10A α11A β1″A β2″A β5A
Thermus DATEIQKLFQQGYQGARYSFGYPACPDLADQAKLDRLMG-FHRVGVRLTENFQLEPEHATSALVVHHPEARYFSVD* 1185
Thermotoga SNLEGSF---------RISPGYG-DLPLSLNKKIAEIFK-EEVDVNVIEDSYVLVPRKTITAFVGWREKNEKQT* 202
E coli ENLSNEELIRENYQGIRPAPGYPACPEHTEKATIWELLEVEKHTGMKLTESFAMWPGASVSGWYFSHPDSKYYAVAQIQRDQVEDYARRKGMSVTEVERW 1217
human EQLDVADLRRLRVKGIRPAPGYPSQPDHTEKLTMWRLADIEQSTGIRLTESLAMAPASAVSGLYFSNLKSKYFAVGKISKDQVEDYALRKNISVAEVEKW 1255

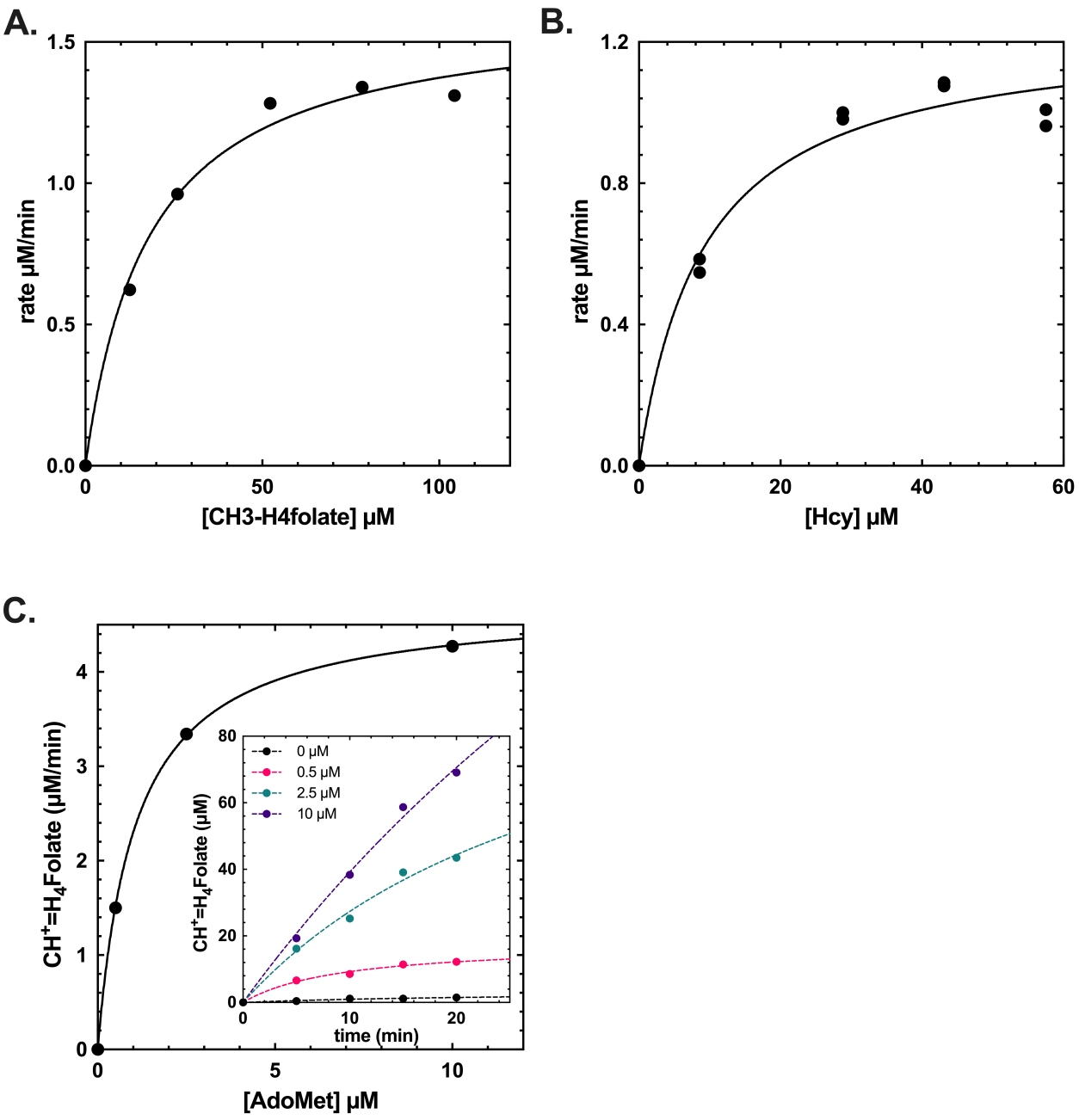
SI Figure 3.2: Methionine synthase sequence alignment

Schematic illustration of the secondary structure of tMS and alignments with the deduced amino acid sequence of MS from *Thermus thermophilus* [Genbank accession number NC_00646], *Thermotoga maritima* [NC_000853], *Escherichia coli* [J04975], and human [U73338]. α-helices and β-sheets are shown in boxes and arrows, respectively. Red loops indicate β-hairpins. The N-terminal residues absent in $tMS^{\Delta N35}$ are shown in grey. Conserved amino acid residues are highlighted by yellow, green, red, and blue for the homocysteine-, folate-, cobalamin-, and *S*-adenosylmethionine-binding domains, respectively. In the activation (adenosylmethionine-binding) domain, cyan is used to highlight conserved amino acid residues in *T. thermophilus*, *E. coli*, and human MS.



SI Figure 3.3: Thermostability of *t*MS

Thermostability. *E. coli* crude extract containing tMS was used. Extracts were incubated for 15 min in the presence or absence of 100 μM MeCbl. After centrifuging, the supernatant was subjected to denaturing polyacrylamide gel electrophoresis, then stained by Coomassie Brilliant Blue.



SI Figure 3.4: Biochemical characterization of *t*MS

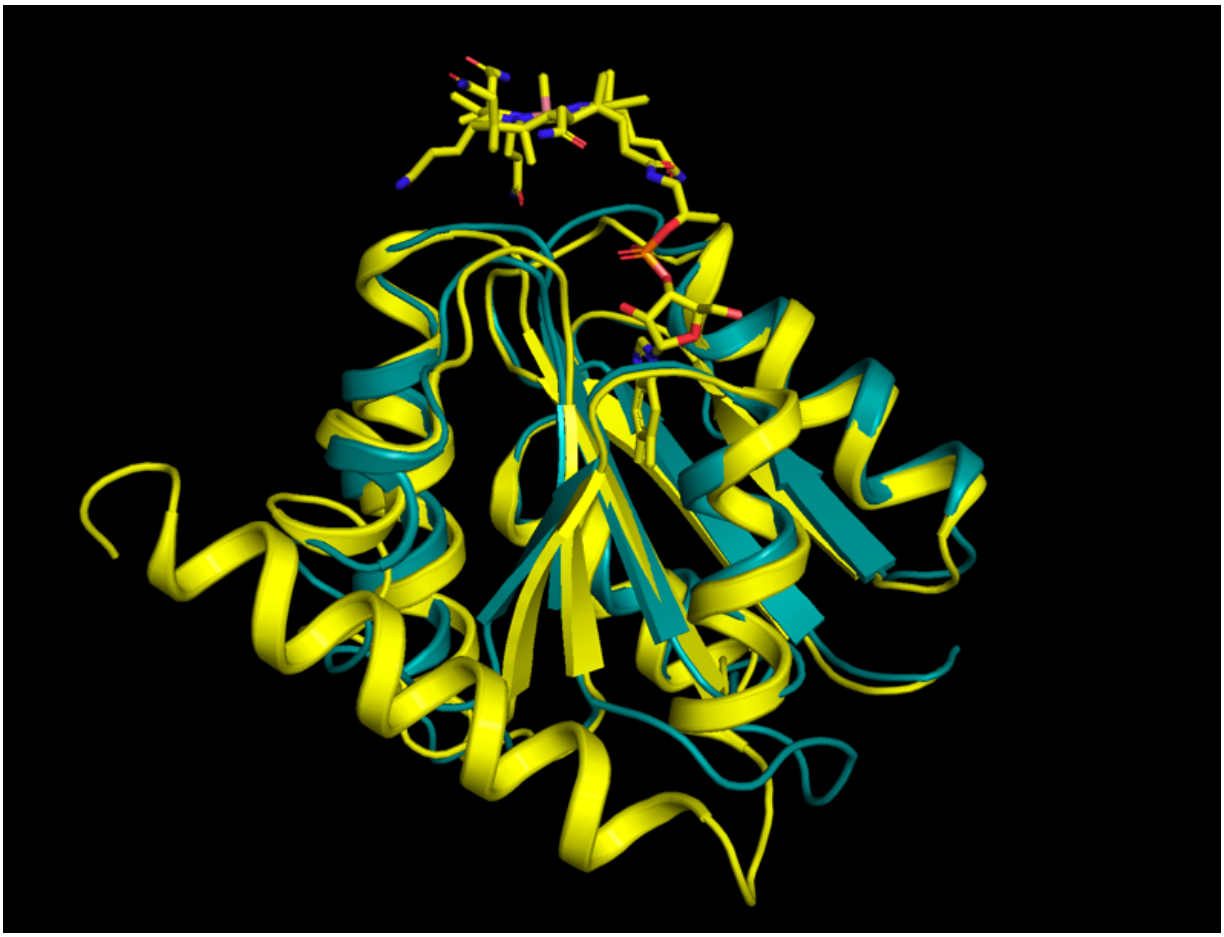
(A-B) Affinities for homocysteine and CH_3-H_4folate in the steady-state assay. Concentrations of 100 μM Homocysteine and 250μM CH_3-H_4folate were used in (A) and (B), respectively. The reaction was carried out at 50˚C. **(C)** AdoMet-dependency of enzyme activity. Product formation in a limited amount of AdoMet (0~10 μM) was measured for up to 20 min (inset). The initial rates were plotted against the initial concentration of AdoMet.

SI Table 3.1: Kinetic parameters of Methionine Synthase

Enzyme	Molecular Weight (kDa)	Specific Activity (μmol $min^{-1}mg^{-1}$ protein)	k_{cat} (min^{-1})	K_m (μM)	
				(6S)CH_3-H_4folate	Homocysteine
E. coli MS	133	11.6 [1]	1542 at 37°C 1128 at 25°C [2]	27.8	0.8
T. thermophilus MS	131	8.1	1062 at 50°C	18	9.3
S. scrofa MS (Pig)	150	1.7	255	12.6	2.16
R. rattus MS (Rat)	140	1.6	224	38	1.7

[1] Biochemistry 1988, 27, 8458-65

[2] Biochemistry 1990, 29, 11101-9

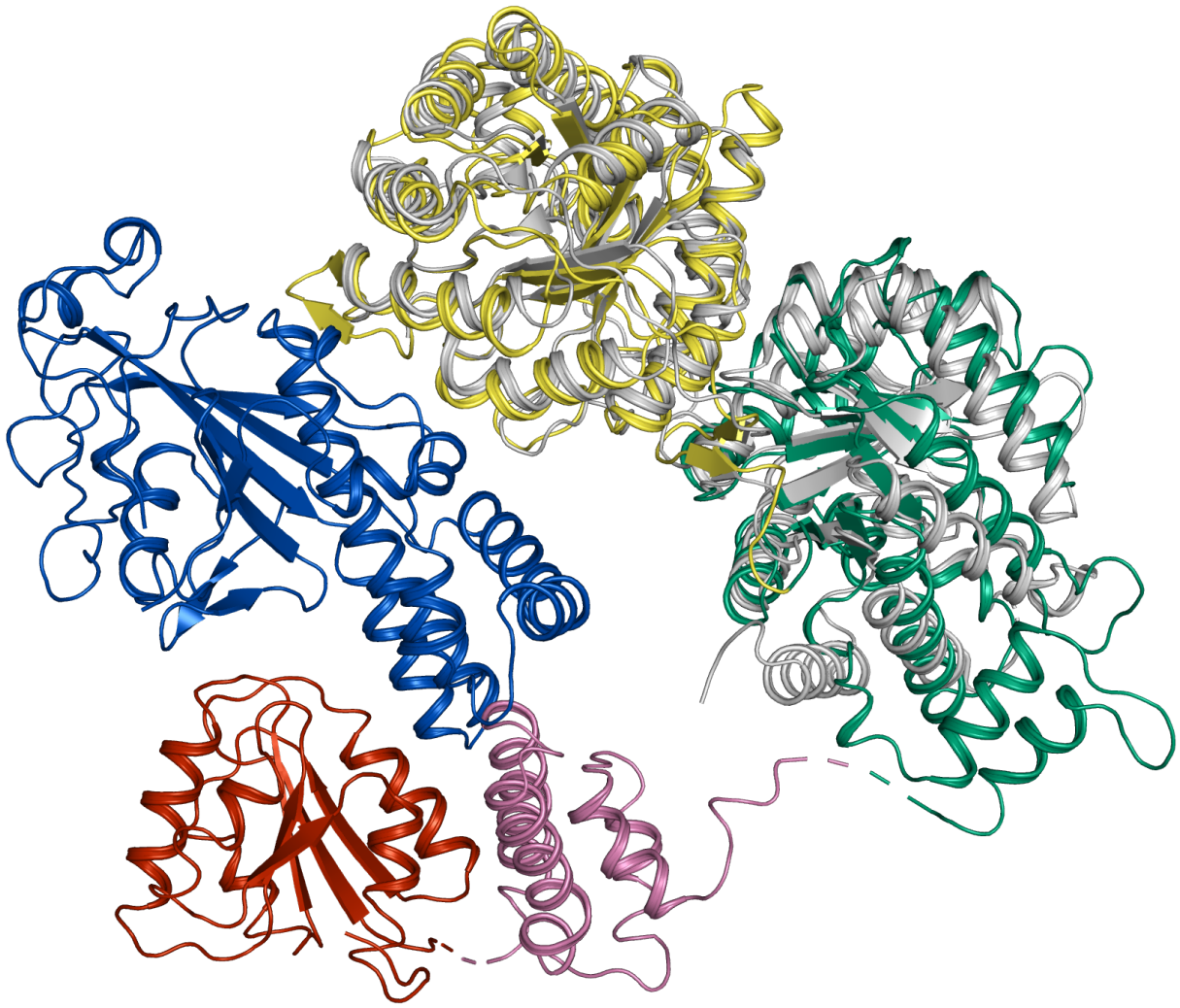


SI Figure 3.5: Cobalamin domain in full length MS and resting state structures

Alignment of the Cob-domain from full-length MS (teal) with Cob-domain from resting structure (1BMT, yellow).

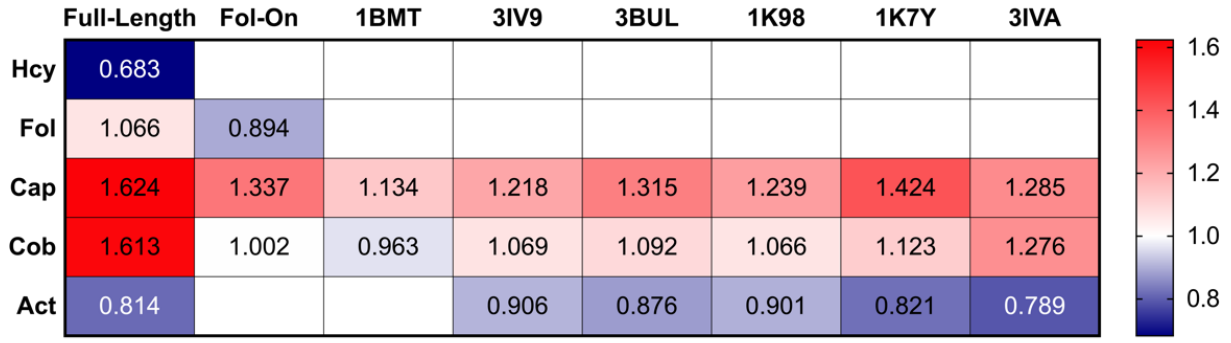
SI Figure 3.6: Displacement of Cap-domain in full length MS

Alignment of Cob and Cap from the full length structure (Cob: teal, Cap: light blue) with the Cob domain from 1BMT (yellow, resting, his-on) and the Cob and Cap from 3Iv9 (pink, reactivation, his-on) and 3BUL (gray, reactivation, his-off).



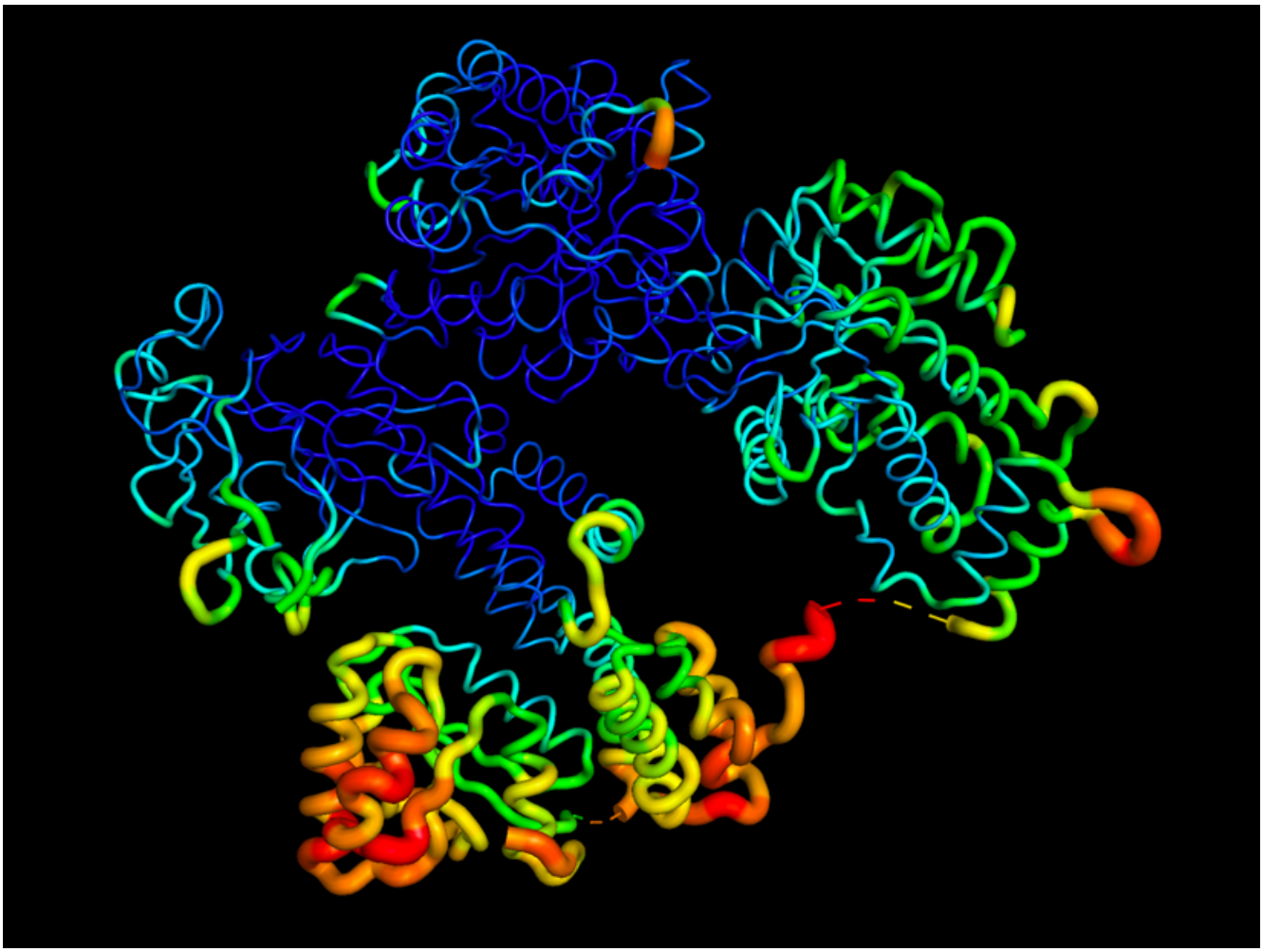
SI Figure 3.7: Full-length Methionine Synthase

Alignment of Fol and Hcy from 3BOL (gray) with Fol (green) and Hcy (yellow) in full length MS – RMSD =2.63 Å.



SI Figure 3.8: B factors of individual domains in multi-domain MS structures

1BMT is the resting state, and 3IV9 through 3IVA are various reactivation state structures. Bfactors of individual domains and the entire structures were calculated by BAverage in CCP4. The domain Bfactors were divided by the average B factor for the entire structure in order to normalize them for comparison between structures. Values shown above show how much larger or smaller the b factors for a given domain were compared for the average of the total structure. Red indicates greater b factors, with blue indicating lower B factors.



SI Figure 3.9: B-factor putty representation of full-length MS

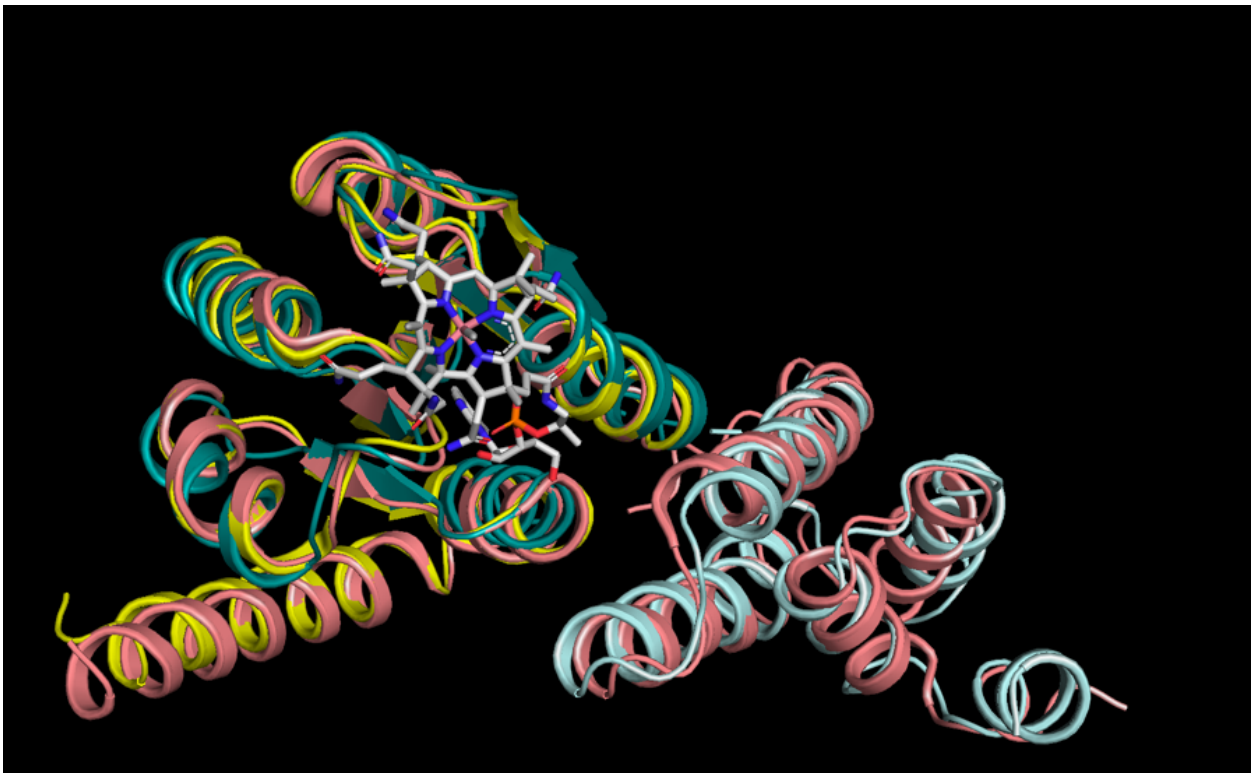
SI Table 3.2: PISA analysis of domain interactions

	Interactions	CSS
Fol-on		
Cobalamin:Cap	0	0
Cobalamin:Cob	4H	1
Cobalamin:Fol	4H	1
Fol:Cap	1H	0.284
Fol:Cob	3H/1S	0.353
Cob:Cap	1H	0.364
1BMT		
Cobalamin:Cap	1H	1
Cobalamin:Cob	11H	1
Cob:Cap	6H/2S	0.324
3iv9		
Cobalamin:Cap	0	0
Cobalamin:Cob	6H	1
Cobalamin:Act	2H	0.225
Cob:Cap	1H/1S	0.28
Cob:Act	2H/3S	0.423
Act:Cap	3H/5S	0.297
FL		
Cob:Cap	0	0
Cob:Act	2H	0
Cob:Fol	0	0
Cob:Hcy	0	0
Cap:Act	4H	0
Cap:Fol	0	0
Cap:Hcy	0	0
Act:Hcy	8H/6S	0
Act:Fol	3H	0
Hcy:Fol	6H/1S	0.1

Domain interfaces were analyzed via the PISA server. The number of interactions include hydrogen bonds (H) and salt bridges (S). Complex significance score (CSS) is an indicator of the strength of the interface.

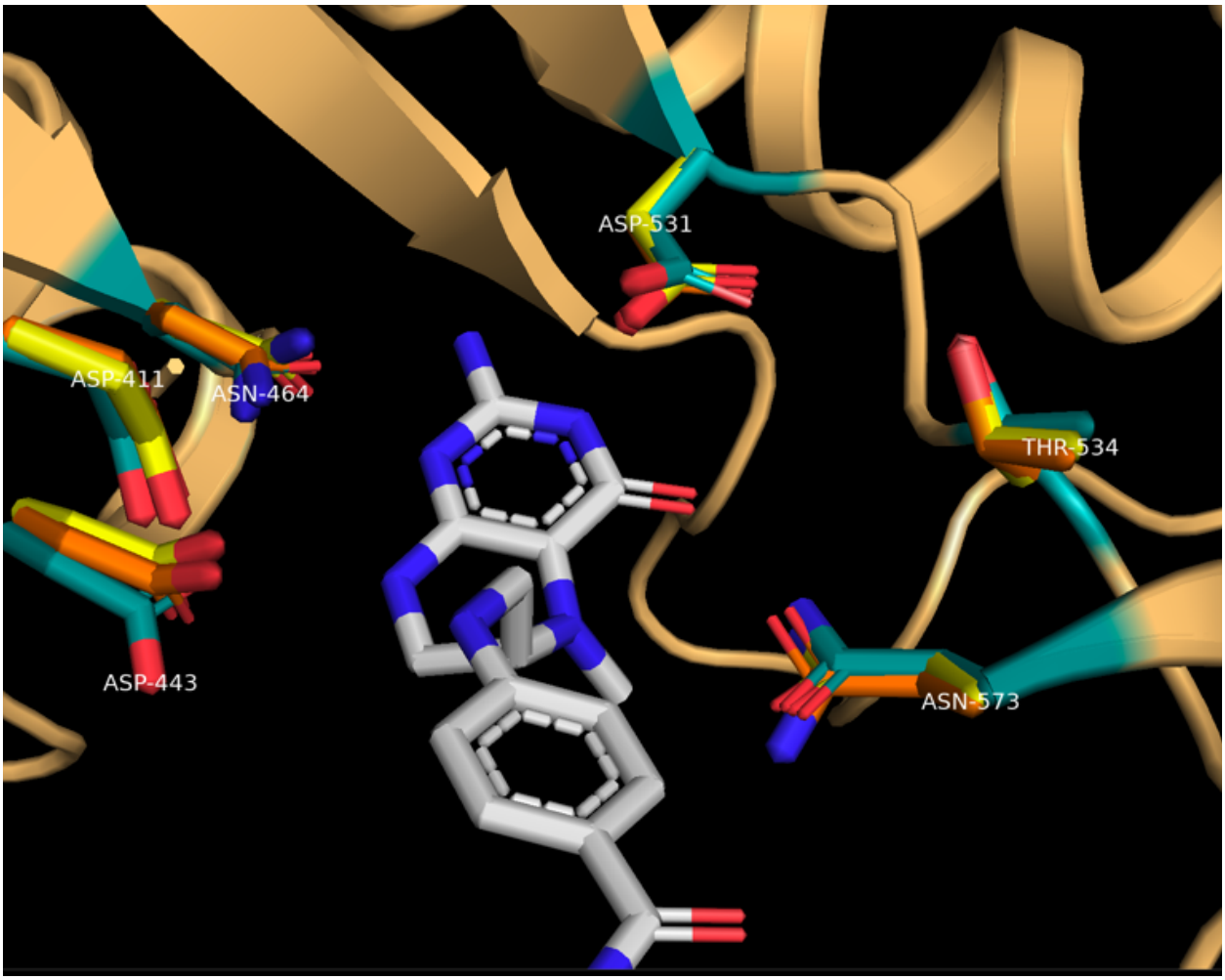
SI Figure 3.10: Alignment of Cob- and Cap- domain in Fol-on structure

Alignment of the Cob domain from 1BMT (yellow) and 3iv9 (pink) with the cob domain of the fol-on structure (teal). Cap alignment with 3iv9 (pink) with the fol-on structure (light blue). Cobalamin is from the fol-on structure (gray).



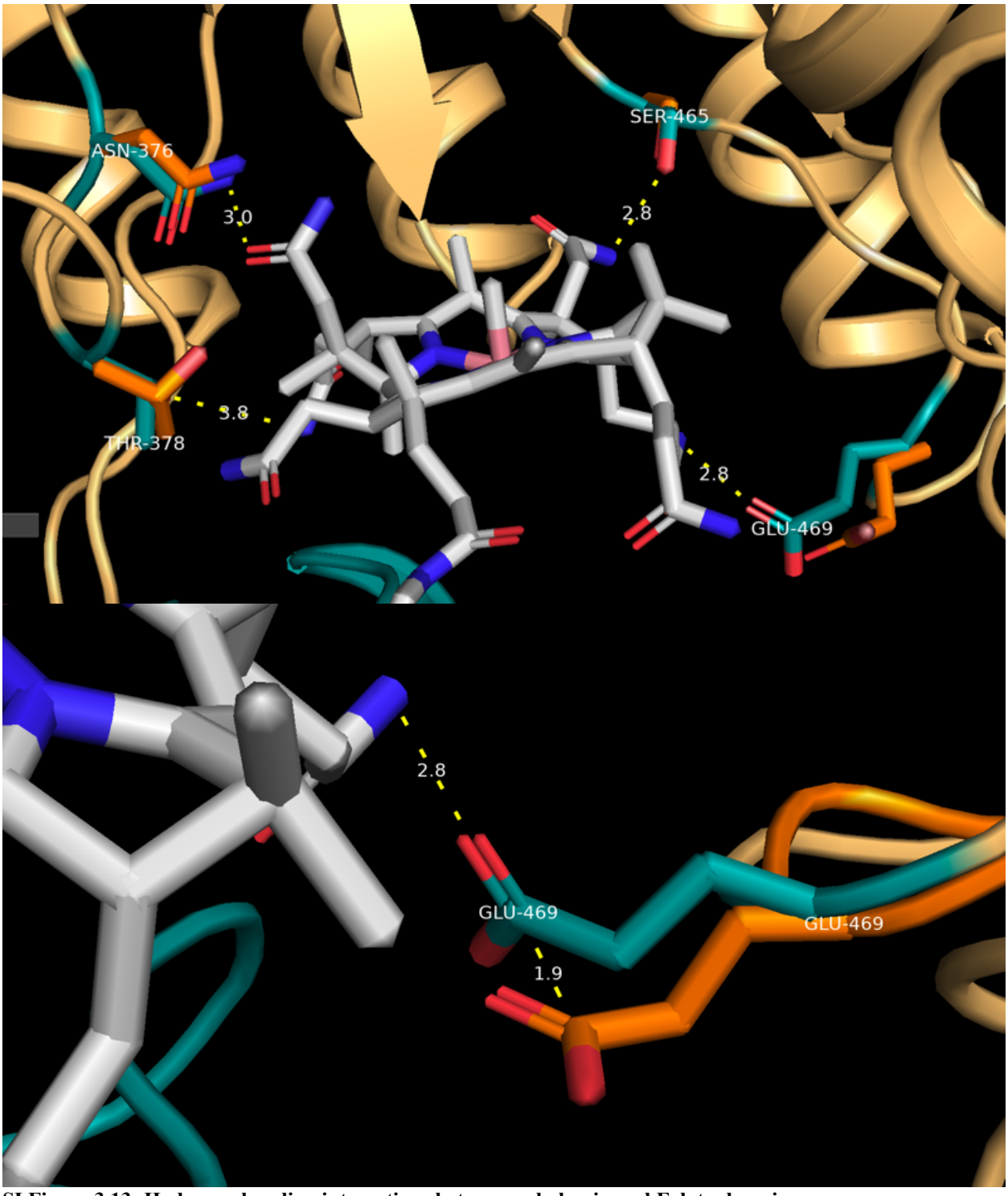
SI Figure 3.11: Top-down view of Cob/Cap alignment

Alignment of the Cob domain from 1BMT (yellow) and 3iv9 (pink) with the cob domain of the fol-on structure (teal). Cap alignment with 3iv9 (pink) with the fol-on structure (light blue). Cobalamin is from the fol-on structure (gray).



SI Figure 3.12: Folate-domain active site

Cartoon representation of Folate domain (wheat) from the Fol-on structure, with active site residues are shown in sticks (teal). Active site residues are also shown from the excised Folate domain structure (yellow, PDB: 5VON) and MTF-bound structure (orange, PDB: 5VOP). MTF shown in gray from 5VOP.



SI Figure 3.13: Hydrogen bonding interactions between cobalamin and Folate-domain

(top) Interactions between cobalamin (gray) and residues of the folate domain from the Fol-on structure (teal) and the MTF bound structure of the folate domain (orange, PDB: 5VOP). **(bottom)** Glu469 repositions from the MTF-bound structure (orange, sticks) to interact with cobalamin when the cofactor is present, seen in Fol-on structure (teal, sticks).

3.5 References

1. Matthews RG, Koutmos M, Datta S. Cobalamin-dependent and cobamide-dependent methyltransferases. Curr Opin Struct Biol. 2008 Dec;18(6):658–66.

2. Brown KL. Chemistry and Enzymology of Vitamin B12. Chem Rev. 2005 Jun 1;105(6):2075–150.

3. Banerjee R. Radical Carbon Skeleton Rearrangements: Catalysis by Coenzyme B12-Dependent Mutases. Chem Rev. 2003 Jun 1;103(6):2083–94.

4. Banerjee R, Ragsdale SW. The many faces of vitamin B12: catalysis by cobalamin-dependent enzymes. Annu Rev Biochem. 2003;72:209–47.

5. Gruber K, Puffer B, Kräutler B. Vitamin B12-derivatives-enzyme cofactors and ligands of proteins and nucleic acids. Chem Soc Rev. 2011 Aug;40(8):4346–63.

6. Roth JR, Lawrence JG, Bobik TA. Cobalamin (coenzyme B12): synthesis and biological significance. Annu Rev Microbiol. 1996;50:137–81.

7. Bridwell-Rabb J, Drennan CL. Vitamin B(12) in the spotlight again. Curr Opin Chem Biol. 2017 Apr;37:63–70.

8. Payne KAP, Quezada CP, Fisher K, Dunstan MS, Collins FA, Sjuts H, et al. Reductive dehalogenase structure suggests a mechanism for B12-dependent dehalogenation. Nature. 2015 Jan;517(7535):513–6.

9. Ludwig ML, Drennan CL, Matthews RG. The reactivity of B12 cofactors: the proteins make a difference. Structure. 1996 May 15;4(5):505–12.

10. Marsh EN. Coenzyme B12 (cobalamin)-dependent enzymes. Essays Biochem. 1999;34:139–54.

11. Giedyk M, Goliszewska K, Gryko D. Vitamin B12 catalysed reactions. Chem Soc Rev. 2015 Jun 7;44(11):3391–404.

12. Kräutler B. Organometallic chemistry of b(12) coenzymes. Met Ions Life Sci. 2009;6:1–51.

13. Banerjee RV, Matthews RG. Cobalamin-dependent methionine synthase. The FASEB Journal. 1990;4(5):1450–9.

14. Kräutler B. Biochemistry of B12-cofactors in human metabolism. Subcell Biochem. 2012;56:323–46.

15. O'Sullivan H, Jennings F, Ward K, McCann S, Scott JM, Weir DG. Human bone marrow biochemical function and megaloblastic hematopoiesis after nitrous oxide anesthesia. Anesthesiology. 1981 Dec;55(6):645–9.

16. Lassen HCA, Henriksen E, Neukirch F, Kristensen H. TREATMENT OF TETANUS: SEVERE BONE-MARROW DEPRESSION AFTER PROLONGED NITROUS-OXIDE ANESTHESIA. The Lancet. 1956 Apr 28;267(6922):527–30.

17. Layzer RobertB. MYELONEUROPATHY AFTER PROLONGED EXPOSURE TO NITROUS OXIDE. The Lancet. 1978 Dec 9;312(8102):1227–30.

18. Lenhert PG, Hodgkin DC. Structure of the 5,6-Dimethylbenzimidazolylcobamide Coenzyme. Nature. 1961 Dec;192(4806):937–8.

19. Schrauzer GN, Deutsch E, Windgassen RJ. The nucleophilicity of vitamin B12. J Am Chem Soc. 1968 Apr 24;90(9):2441–2.

20. Banerjee R. Chemistry and Biochemistry of B12. John Wiley & Sons; 1999. 960 p.

21. Schrauzer GN, Deutsch E. Reactions of cobalt(I) supernucleophiles. The alkylation of vitamin B12s, cobaloximes(I), and related compounds. J Am Chem Soc. 1969 Jun 1;91(12):3341–50.

22. Koutmos M, Pejchal R, Bomer TM, Matthews RG, Smith JL, Ludwig ML. Metal active site elasticity linked to activation of homocysteine in methionine synthases. Proceedings of the National Academy of Sciences. 2008 Mar 4;105(9):3286–91.

23. Peariso K, Goulding CW, Huang S, Matthews RG, Penner-Hahn JE. Characterization of the Zinc Binding Site in Methionine Synthase Enzymes of Escherichia coli: The Role of Zinc in the Methylation of Homocysteine. J Am Chem Soc. 1998 Aug 1;120(33):8410–6.

24. Goulding CW, Matthews RG. Cobalamin-Dependent Methionine Synthase from Escherichia coli: Involvement of Zinc in Homocysteine Activation. Biochemistry. 1997 Dec 1;36(50):15749–57.

25. Banerjee RV, Frasca V, Ballou DP, Matthews RG. Participation of Cob(I)alamin in the reaction catalyzed by methionine synthase from Escherichia coli: a steady-state and rapid reaction kinetic analysis. Biochemistry. 1990 Dec 18;29(50):11101–9.

26. Drummond JT, Huang S, Blumenthal RM, Matthews RG. Assignment of enzymic function to specific protein regions of cobalamin-dependent methionine synthase from Escherichia coli. Biochemistry. 1993 Sep 1;32(36):9290–5.

27. Fujii K, Huennekens FM. Activation of methionine synthetase by a reduced triphosphopyridine nucleotide-dependent flavoprotein system. J Biol Chem. 1974 Nov 10;249(21):6745–53.

28. Taylor RT, Weissbach H. N5-methyltetrahydrofolate-homocysteine transmethylase. Role of S-adenosylmethionine in vitamin B12-dependent methionine synthesis. J Biol Chem. 1967 Apr 10;242(7):1517–21.

29. Foster MA, Dilworth MJ, Woods DD. COBALAMIN AND THE SYNTHESIS OF METHIONINE BY ESCHERICHIA COLI. Nature. 1964 Jan 4;201:39–42.

30. Banerjee RV, Harder SR, Ragsdale SW, Matthews RG. Mechanism of reductive activation of cobalamin-dependent methionine synthase: an electron paramagnetic resonance spectroelectrochemical study. Biochemistry. 1990 Feb 6;29(5):1129–35.

31. Brown KL, Hakimi JM, Nuss DM, Montejano YD, Jacobsen DW. Acid-base properties of .alpha.-ribazole and the thermodynamics of dimethylbenzimidazole association in alkylcobalamins. Inorg Chem. 1984 May 1;23(10):1463–71.

32. Brown KL, Peck-Siler S. Heteronuclear NMR studies of cobalamins. 9. Temperature-dependent NMR of organocobalt corrins enriched in carbon-13 in the organic ligand and the thermodynamics of the base-on/base-off reaction. Inorg Chem. 1988 Oct 1;27(20):3548–55.

33. Fleischhacker AS, Matthews RG. Ligand Trans Influence Governs Conformation in Cobalamin-Dependent Methionine Synthase. Biochemistry. 2007 Oct 1;46(43):12382–92.

34. Bandarian V, Ludwig ML, Matthews RG. Factors modulating conformational equilibria in large modular proteins: A case study with cobalamin-dependent methionine synthase. Proc Natl Acad Sci U S A. 2003 Jul 8;100(14):8156–63.

35. Jarrett JT, Huang S, Matthews RG. Methionine Synthase Exists in Two Distinct Conformations That Differ in Reactivity toward Methyltetrahydrofolate, Adenosylmethionine, and Flavodoxin†. Biochemistry. 1998 Apr 21;37(16):5372–82.

36. Goulding CW, Postigo D, Matthews RG. Cobalamin-dependent methionine synthase is a modular protein with distinct regions for binding homocysteine, methyltetrahydrofolate, cobalamin, and adenosylmethionine. Biochemistry. 1997 Jul 1;36(26):8082–91.

37. Dixon MM, Huang S, Matthews RG, Ludwig M. The structure of the C-terminal domain of methionine synthase: presenting S-adenosylmethionine for reductive methylation of B12. Structure. 1996 Nov 15;4(11):1263–75.

38. Banerjee RV, Johnston NL, Sobeski JK, Datta P, Matthews RG. Cloning and sequence analysis of the Escherichia coli metH gene encoding cobalamin-dependent methionine synthase and isolation of a tryptic fragment containing the cobalamin-binding domain. J Biol Chem. 1989 Aug 15;264(23):13888–95.

39. Drennan CL, Huang S, Drummond JT, Matthews RG, Ludwig ML. How a Protein Binds B12: A 3.0 Å X-Ray Structure of B12-Binding Domains of Methionine Synthase. Science. 1994 Dec 9;266(5191):1669–74.

40. Drennan CL, Matthews RG, Ludwig ML. Cobalamin-dependent methionine synthase: the structure of a methylcobalamin-binding fragment and implications for other B12-dependent enzymes. Curr Opin Struct Biol. 1994 Dec;4(6):919–29.

41. Datta S, Koutmos M, Pattridge KA, Ludwig ML, Matthews RG. A disulfide-stabilized conformer of methionine synthase reveals an unexpected role for the histidine ligand of the cobalamin cofactor. Proc Natl Acad Sci U S A. 2008 Mar 18;105(11):4115–20.

42. Bandarian V, Pattridge KA, Lennon BW, Huddler DP, Matthews RG, Ludwig ML. Domain alternation switches B(12)-dependent methionine synthase to the activation conformation. Nat Struct Biol. 2002 Jan;9(1):53–6.

43. Koutmos M, Datta S, Pattridge KA, Smith JL, Matthews RG. Insights into the reactivation of cobalamin-dependent methionine synthase. Proc Natl Acad Sci USA. 2009 Nov 3;106(44):18527.

44. Fick RJ, Clay MC, Vander Lee L, Scheiner S, Al-Hashimi H, Trievel RC. Water-Mediated Carbon–Oxygen Hydrogen Bonding Facilitates S-Adenosylmethionine Recognition in the Reactivation Domain of Cobalamin-Dependent Methionine Synthase. Biochemistry. 2018 Jul 3;57(26):3733–40.

45. Evans JC, Huddler DP, Hilgers MT, Romanchuk G, Matthews RG, Ludwig ML. Structures of the N-terminal modules imply large domain motions during catalysis by methionine synthase. Proc Natl Acad Sci U S A. 2004 Mar 16;101(11):3729–36.

46. Yamada K, Koutmos M. The folate-binding module of *Thermus thermophilus* cobalamin-dependent methionine synthase displays a distinct variation of the classical TIM barrel: a TIM barrel with a `twist'. Acta Crystallogr D Struct Biol. 2018 Jan 1;74(1):41–51.

47. Kung Y, Ando N, Doukov TI, Blasiak LC, Bender G, Seravalli J, et al. Visualizing molecular juggling within a B12-dependent methyltransferase complex. Nature. 2012 Apr;484(7393):265–9.

48. Dekker K, Yamagata H, Sakaguchi K, Udaka S. Xylose (glucose) isomerase gene from the thermophile Thermus thermophilus: cloning, sequencing, and comparison with other thermostable xylose isomerases. J Bacteriol. 1991 May;173(10):3078–83.

49. Tanner JJ, Hecht RM, Krause KL. Determinants of enzyme thermostability observed in the molecular structure of Thermus aquaticus D-glyceraldehyde-3-phosphate dehydrogenase at 25 Angstroms Resolution. Biochemistry. 1996 Feb 27;35(8):2597–609.

50. Zhou XX, Wang YB, Pan YJ, Li WF. Differences in amino acids composition and coupling patterns between mesophilic and thermophilic proteins. Amino Acids. 2008 Jan;34(1):25–33.

51. Fields PA, Dong Y, Meng X, Somero GN. Adaptations of protein structure and function to temperature: there is more than one way to "skin a cat." J Exp Biol. 2015 Jun;218(Pt 12):1801–11.

52. Somero GN. Adaptation of enzymes to temperature: searching for basic "strategies." Comp Biochem Physiol B Biochem Mol Biol. 2004 Nov;139(3):321–33.

53. Tollinger M, Eichmüller C, Konrat R, Huhta MS, Marsh ENG, Kräutler B. The B12-Binding Subunit of Glutamate Mutase from Clostridium tetanomorphum Traps the Nucleotide Moiety of Coenzyme B12. Journal of Molecular Biology. 2001 Jun 8;309(3):777–91.

54. Tollinger M, Konrat R, Hilbert BH, Marsh EN, Kräutler B. How a protein prepares for B12 binding: structure and dynamics of the B12-binding subunit of glutamate mutase from Clostridium tetanomorphum. Structure. 1998 Aug 15;6(8):1021–33.

55. Dorweiler JS, Finke RG, Matthews RG. Cobalamin-Dependent Methionine Synthase: Probing the Role of the Axial Base in Catalysis of Methyl Transfer between Methyltetrahydrofolate and Exogenous Cob(I)alamin or Cob(I)inamide. Biochemistry. 2003 Dec 1;42(49):14653–62.

56. Svetlitchnaia T, Svetlitchnyi V, Meyer O, Dobbek H. Structural insights into methyltransfer reactions of a corrinoid iron–sulfur protein involved in acetyl-CoA synthesis. Proceedings of the National Academy of Sciences. 2006 Sep 26;103(39):14331–6.

57. Yokoyama S, Hirota H, Kigawa T, Yabuki T, Shirouzu M, Terada T, et al. Structural genomics projects in Japan. Nat Struct Mol Biol. 2000 Nov;7(11):943–5.

58. Stols L, Gu M, Dieckman L, Raffen R, Collart FR, Donnelly MI. A new vector for high-throughput, ligation-independent cloning encoding a tobacco etch virus protease cleavage site. Protein Expr Purif. 2002 Jun;25(1):8–15.

59. Drummond JT, Jarrett J, Gonzalez JC, Huang S, Matthews RG. Characterization of Nonradioactive Assays for Cobalamin-Dependent and Cobalamin-Independent Methionine Synthase Enzymes. Analytical Biochemistry. 1995 Jul 1;228(2):323–9.

CHAPTER 4 Itaconyl-CoA Forms a Stable Biradical in Methylmalonyl-CoA Mutase and Derails Its Activity and Repair

Itaconate is an immunometabolite with both anti-inflammatory and bactericidal effects. Its coenzyme A (CoA) derivative, itaconyl-CoA, inhibits B_{12}-dependent methylmalonyl-CoA mutase (MCM) by an unknown mechanism. We demonstrate that itaconyl-CoA is a suicide inactivator of human and *Mycobacterium tuberculosis* MCM, which forms a markedly air-stable biradical adduct with the 5′-deoxyadenosyl moiety of the B_{12} coenzyme. Termination of the catalytic cycle in this way impairs communication between MCM and its auxiliary repair proteins. Crystallography and spectroscopy of the inhibited enzyme are consistent with a metal-centered cobalt radical ~6 angstroms away from the tertiary carbon-centered radical and suggest a means of controlling radical trajectories during MCM catalysis. Mycobacterial MCM thus joins enzymes in the glyoxylate shunt and the methylcitrate cycle as targets of itaconate in pathogen propionate metabolism.

This chapter was adapted with permission from Ruetz, M.; Campanello, G. C.; Purchal, M.; Shen, H.; McDevitt, L.; Gouda, H.; Wakabayashi, S.; Zhu, J.; Rubin, E. J.; Warncke, K.; Mootha, V. K.; Koutmos, M.; Banerjee, R. Itaconyl-CoA Forms a Stable Biradical in Methylmalonyl-CoA Mutase and Derails Its Activity and Repair. *Science* **2019**, *366* (6465), 589–593. © American Association for the Advancement of Science 2022.

Author Contributions: M.P. and M.K. were responsible for the crystallographic analyses (Figure 4.2 and SI Figure 4.10 and SI Table 4.1). M.R. and G.C.C. performed the majority of the biochemical studies (Figures 4.1, 4.3 and SI Figure 4.1, 4.5, 4.6, 4.8, 4.9, 4.11), H.G. did the kinetic analysis of hMCM (SI Figure 4.1), L.M. and M.R. cloned and expressed the Mtb genes used in this study, and K.W. was responsible for the EPR spectroscopic analysis (SI Figure 4.4, 4.5, 4.6). H.S. performed the macrophage experiments, and J.Z. and S.W. performed the Mtb growth experiments (Figure 4.4). M.R., G.C.C., and R.B. wrote the manuscript. All authors were involved with data analysis of experiments performed by them or in their laboratories and edited the manuscript.

4.1 Introduction

The immunomodulatory and antimicrobial effects of itaconate are evincing newfound interest in a compound historically used as a precursor in polymer synthesis. Upon activation, immune cells stimulate itaconate synthesis ~10-fold via aconitate decarboxylase (Irg1)–catalyzed decarboxylation of *cis*-aconitate, a tricarboxylic acid cycle intermediate (1). Itaconate activates Nrf2, inhibits succinate dehydrogenase, and blocks the transcription factor Iκβξ, leading to a switch from a pro- to an anti-inflammatory state (2). The antimicrobial activity of itaconate is purportedly due to its inhibition of two microbe-specific targets: isocitrate lyase in the glyoxylate shunt and methylcitrate lyase in the methylcitrate cycle, two enzymes that are needed for pathogen survival on acetate or propionate, respectively, as the sole carbon source (Figure 4.1A) (3). Propionyl-coenzyme A (CoA) is derived from cholesterol catabolism and is used by pathogens like *Mycobacterium tuberculosis* (Mtb) for biomass production in the glucose-limiting conditions found in phagosomes (4).

An unexpected intersection between itaconate and B_{12}-dependent propionate metabolism was revealed recently by the demonstration that itaconyl-CoA (I-CoA) is a potent inhibitor of human 5′-deoxyadenosylcobalamin (AdoCbl)–dependent methylmalonyl-CoA mutase (*h*MCM) (5). Itaconate can be cleared by a C_5-dicarboxylate pathway (6) via acylation to I-CoA, hydration to citramalyl-CoA, and cleavage to acetyl-CoA and pyruvate, a reaction catalyzed by citramalyl-CoA lyase, which is encoded by the recently deorphaned *citrate lyase beta-like* (CLYBL) gene (5) (Figure 4.1A). I-CoA (or methylenesuccinyl-CoA) is an analog of succinyl-CoA, which is interconverted to methylmalonyl-CoA (M-CoA) by the isomerase MCM. A metabolic connection between CLYBL and B_{12} was initially revealed in a genome-wide association study (7) that showed a correlation between the biallelic loss of CLYBL, which has an ~3 to 6% prevalence in

certain populations (8), and B_{12} deficiency. Although I-CoA inhibition of MCM provides a molecular link between mitochondrial B_{12} and C_5-dicarboxylate metabolism, it does not explain why inhibited MCM cannot be repaired by the auxiliary protein system that is dedicated for this function (9). Given the homology between bacterial and human MCM, host-derived itaconate could also potentially target MCM in pathogenic bacteria, in which it is required for lipid breakdown.

Itaconate induces electrophilic stress and modifies small molecules and protein targets by alkylating cysteine residues (10,11). Here, we report a radical suicide inactivation mechanism in which addition of the elusive 5′-deoxyadenosyl radical (dAdo•) to I-CoA in MCM leads to an air-stable biradical comprising a tertiary carbon radical coupled to the metal-centered cob(II)alamin radical. We visualized Mtb MCM bound to a dAdo adduct of I-CoA at 2.0-Å resolution by x-ray crystallography. In addition to identifying MCM as an antimicrobial target of itaconate and demonstrating its importance for Mtb growth on propionate, our study provides molecular insights into how MCM controls radical trajectories during catalysis with its normal substrates to promote the desired chemistry and suppress unwanted side reactions.

4.2 Results and Discussion

4.2.1 I-CoA inactivates *Mtb* MCM by forming an air-stable biradical

MCM catalyzes the reversible isomerization of M-CoA to succinyl-CoA via an AdoCbl-dependent radical mechanism. To test our hypothesis that bacterial MCM is also a target of the antimicrobial effect of itaconate, we cloned and expressed Mtb MCM and the two auxiliary proteins that load and repair AdoCbl (SI Figure 4.1A). The kinetic parameters of Mtb MCM are comparable to those of the human homolog (SI Figure 4.1, B and C). Addition of I-CoA to Mtb MCM-AdoCbl led to a rapid shift in the absorption maximum (λ_{max}), from 528 nm to 466 nm (Figure 4.1B), indicating stoichiometric binding (Figure 4.1B, inset), as was also seen with *h*MCM (Figure 4.1C) (5). Homolysis of the cobalt-carbon bond in AdoCbl, the first step in the MCM-catalyzed reaction, leads to formation of the radical pair dAdo• and cob(II)alamin, albeit with a different λ_{max} (474 nm). We therefore used electron paramagnetic resonance (EPR) spectroscopy to identify the 466-nm–absorbing products of I-CoA–inactivated MCM.

The EPR spectrum of MCM-bound cob(II) alamin displays the typical eight-line hyperfine splitting of the unpaired electron with the $I=^7/_2$ cobalt nucleus that is resolved in the highfield region of the spectrum (SI Figure 4.2, A and B). Additional superhyperfine splitting due to the coordinating $I= 1$ lower axial nitrogen from a histidine ligand is also observed. Notably, addition of I-CoA to human or Mtb MCM led to distinct spectra (Figure 4.1D, top, and SI Figure 4.2C) that had the hallmarks of a hybrid triplet system. The spectra were reminiscent of the EPR spectrum of a transient catalytic intermediate trapped during MCM turnover, exhibiting strong electron-electron spin-coupling between the product succinyl-CoA radical and the low-spin cob(II)alamin (12). Similar biradical intermediates have also been trapped in the AdoCbl-dependent enzymes

glutamate mutase (13) and 2-methyleneglutarate mutase (14). The hyperfine multiplicity and the substantial *g* anisotropy identified the cobalamin component in the triplet spin system.

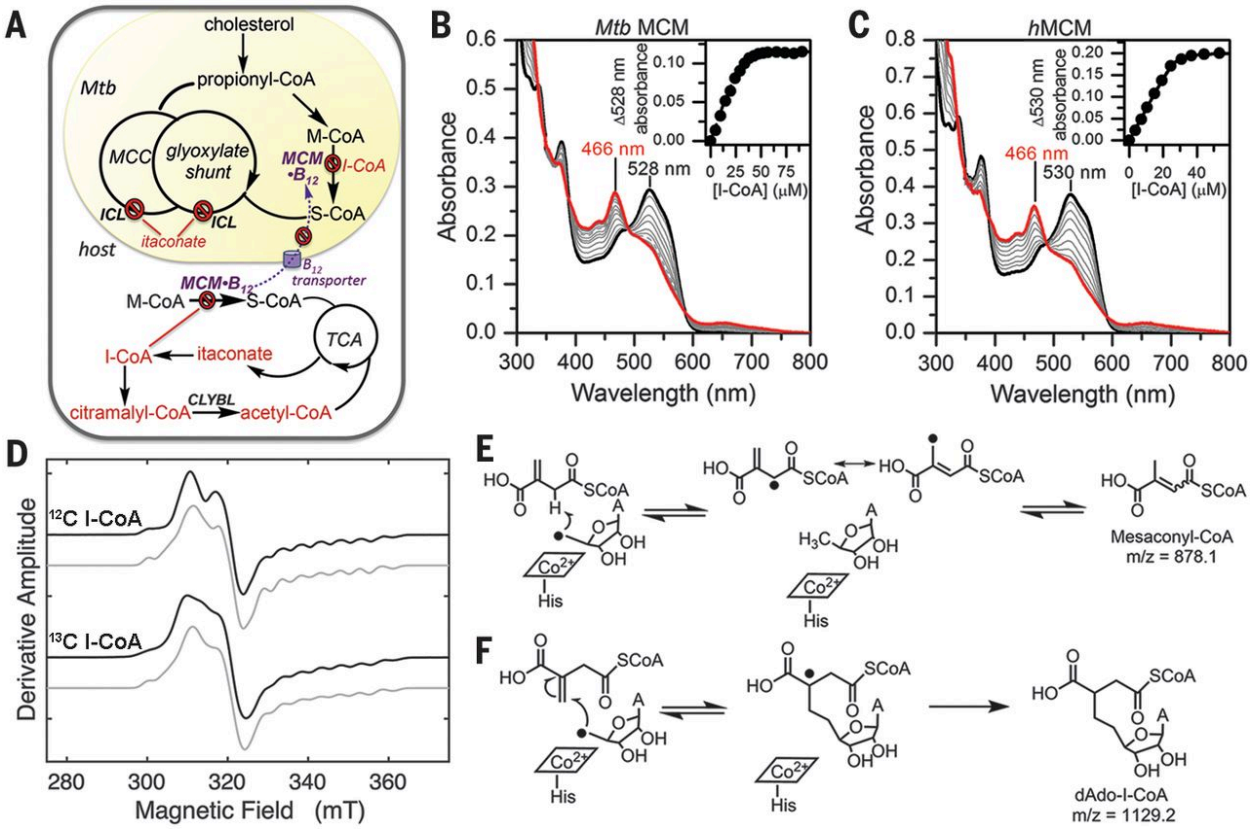


Figure 4.1: I-CoA inhibits human and *Mtb* MCM by forming an air-stable biradical

(A) Mtb pathways targeted by I-CoA. MCC, methylcitrate cycle; ICL, isocitrate lyase; SCoA, succinyl-CoA; TCA, tricarboxylic acid cycle. **(B and C)** Titration of holo-*Mtb* MCM (B) with 30 μM bound AdoCbl or holo-hMCM (C) with 40 μM bound AdoCbl (black traces) with increasing concentrations of I-CoA. The intermediate spectra (gray) were recorded after 5 min of equilibration. (Insets) Representative plots of Δ528 or Δ530 nm versus I-CoA indicated stoichiometric binding [n= 2 (*Mtb* MCM); n= 3 (*h*MCM)]. **(D)** EPR spectra of the 1 mM I-CoA-induced biradical on hMCM (375 μM) in the presence of natural abundance (top) or [13C]I-CoA (bottom). The experimental and simulated spectra are in black and gray, respectively. **(E and F)** Possible fates of dAdo• when I-CoA behaves as substrate (E), not observed, or inhibitor (F).

The identity of the organic radical component was further assessed with [^{13}C]I-CoA (uniformly labeled in the itaconate carbons) or [^{13}C]AdoCbl (uniformly labeled in the adenosyl moiety). Whereas ^{13}C-labeled AdoCbl had no effect on the EPR spectrum, [^{13}C]I-CoA led to inhomogeneous broadening throughout the line shape, which is consistent with appreciable mixing of cob(II)alamin and organic radical quantum states in the strong electron-electron coupling

regime (Figure 4.1D, bottom). These results indicate an absence of marked unpaired electron spin density on the dAdo moiety, which is in accord with the near-unit spin density inferred for the electron-^{13}C interaction in the radical pair formed from [^{13}C]I-CoA. Spectral simulations predict an interspin distance of 5 to 6 Å, while the Euler angles position the organic radical at an angle of 43° relative to the principal axis of the d^2_z orbital on cobalt. In other AdoCbl-dependent isomerases, the distance between the substrate and cobalamin radicals range between 5 to 6 Å and 8 to 12 Å, respectively, necessitating small or large movements of the initially formed dAdo• to reach the substrate hydrogen atom (H-atom) destined for abstraction (15).

The I-CoA–induced biradical was air stable for over 1 hour (SI Figures 4.3 and 4.4), unlike the transient biradical formed during catalytic turnover with M-CoA that was trapped by freeze-quenching (12). To understand the chemical basis of its unusual stability, we further investigated the identity of the organic radical.

4.2.2 The adenosyl radical is stabilized by addition to itaconyl-CoA

Following generation of the dAdo•-cob(II)alamin radical pair on MCM, the isomerization reaction is initiated by H-atom abstraction from M-CoA by dAdo•, forming a substrate-centered radical that undergoes rearrangement. The dAdo• is a primary and highly reactive alkyl radical that has eluded direct detection in all AdoCbl-dependent isomerases but recently was trapped in a radical S-adenosylmethionine enzyme (16). In principle, two potential reactions between dAdo• and I-CoA can be considered (Figure 4.1, E and F). To distinguish between these mechanistic possibilities, the reaction products from *Mtb* and *h*MCM inactivated by I-CoA under single-turnover and aerobic conditions were separated by high-performance liquid chromatography (SI Figure 4.5). Two major product peaks with retention times of 22.9 and 27.0 min (peaks 2 and 5) were identified; *Mtb* MCM showed an additional peak at 23.5 min (peak 2b).

Matrix-assisted laser desorption/ionization–mass spectrometry (MS) analysis confirmed the second reaction pathway leading to an addition product between dAdo and I-CoA (Figure 4.1G), indicating that dAdo• adds to the double bond in I-CoA, yielding a tertiary carbon radical additionally stabilized by delocalization onto the π-system of the adjacent carboxylate. MS analysis revealed that peaks 2a and 2b are isomers with the same mass/charge ratio (*m/z*) of 1145.4 (SI Figure 4.6), which is 16 mass units higher than the expected mass of the addition product (*m/z* 1129), indicating incorporation of an oxygen atom. We assign peak 2 as the hydroxyl derivative of the addition product and peak 5 (*m/z* 1099.3), which is 30 Da lighter, indicating the formal loss of formaldehyde to the oxidative decarboxylation product of peak 2 (SI Figure 4.7A). MS analysis of the *h*MCM samples yielded similar results (SI Figure 4.6, D and E).

Under anaerobic conditions, additional hydrophobic products were observed (SI Figure 4.8). Peak 6 with an *m/z* of 1129.4 represents the expected radical addition product, while peak 5 and the minor peaks 7 and 8, with *m/z* values that are two mass units lower (SI Figure 4.9), were assigned to intramolecular cyclization and/or elimination products (SI Figure 4.7, B and C). Ado• cyclization products have been reported during anaerobic photolysis of AdoCbl (17) and during the reaction of hydroxyl radicals with dAdo or deoxyguanosine (18), resulting in 5′,8-cyclopurine nucleosides.

4.2.3 Crystallographic capture of the biradical on MCM

Given the protracted air stability of the biradical, we attempted to visualize the inhibited form by soaking AdoCbl-reconstituted Mtb MCM crystals with I-CoA. We determined structures of the unsoaked and I-CoA–soaked enzyme at1.9-Å and 2.0-Å resolution, respectively (SI Table 4.1). *Mtb* MCM, like the *Propionibacterium shermanii* protein (19), is a heterodimer comprising an α subunit that binds B_{12} and a β subunit that is inactive (SI Figure 4.10A). In the native structure,

AdoCbl is bound with its endogenous dimethylbenzimidazole tail inserted in a side pocket while His-629 serves as the lower axial ligand (Figure 4.2A). On the opposite face, the 5′-carbon of the upper axial dAdo ligand is 2.5 Å away from the cobalt atom, while the adenine group is coplanar with the corrin ring and oriented above pyrrole rings A and B.

I-CoA binding induces a large conformational change in the AdoCbl-binding a subunit [Cα root mean square deviation (RMSD), 1.47 Å], whereas the small subunit is almost unchanged (Cα RMSD, 0.26 Å) (SI Figure 4.10A). Soaking in I-CoA did not affect crystal stability, as there are no crystal contacts in the region affected by its binding. The α subunit collapses around the I-CoA binding pocket, with the motion being largest at the periphery and smallest where the α and β subunits are proximal. The crystal structure of *h*MCM, which is a homodimer with two B_{12} binding subunits, similarly closes in on its substrate, M-CoA (20).

We assigned the electron density in the active site to the adduct between I-CoA and the 5′-carbon of dAdo (Figure 4.2B and SI Figure 4.10B). The 5′-carbon of the dAdo moiety is rotated almost 180° away from the cobalt, and the distance to the cobalt atom increases to 4.3 Å. The rotation places the 5′-carbon of dAdo 1.5 Å away from the methylene group of I-CoA, indicating the presence of a covalent bond between them. The geometry of the tertiary carbon of I-CoA is planar, as expected for an sp^2 carbon (Figure 4.2B and SI Figure 4.10, C and D). dAdo and corrin are shifted by 2.1 Å relative to the corrin ring in the structure without I-CoA. The acetamide group *a* on ring A is pushed up and in toward the adenine ring (Figure 4.2C), which causes the adenine to move from a parallel to an almost perpendicular position relative to the corrin plane (SI Figure 4.10E), as predicted computationally (21). A strong biradical EPR signal associated with *Mtb* MCM crystals soaked with I-CoA confirmed that the spin-coupled carbon- and metal-centered radical pair can form in crystals (Figure 4.2D). In the structure with the adduct, the tertiary carbon

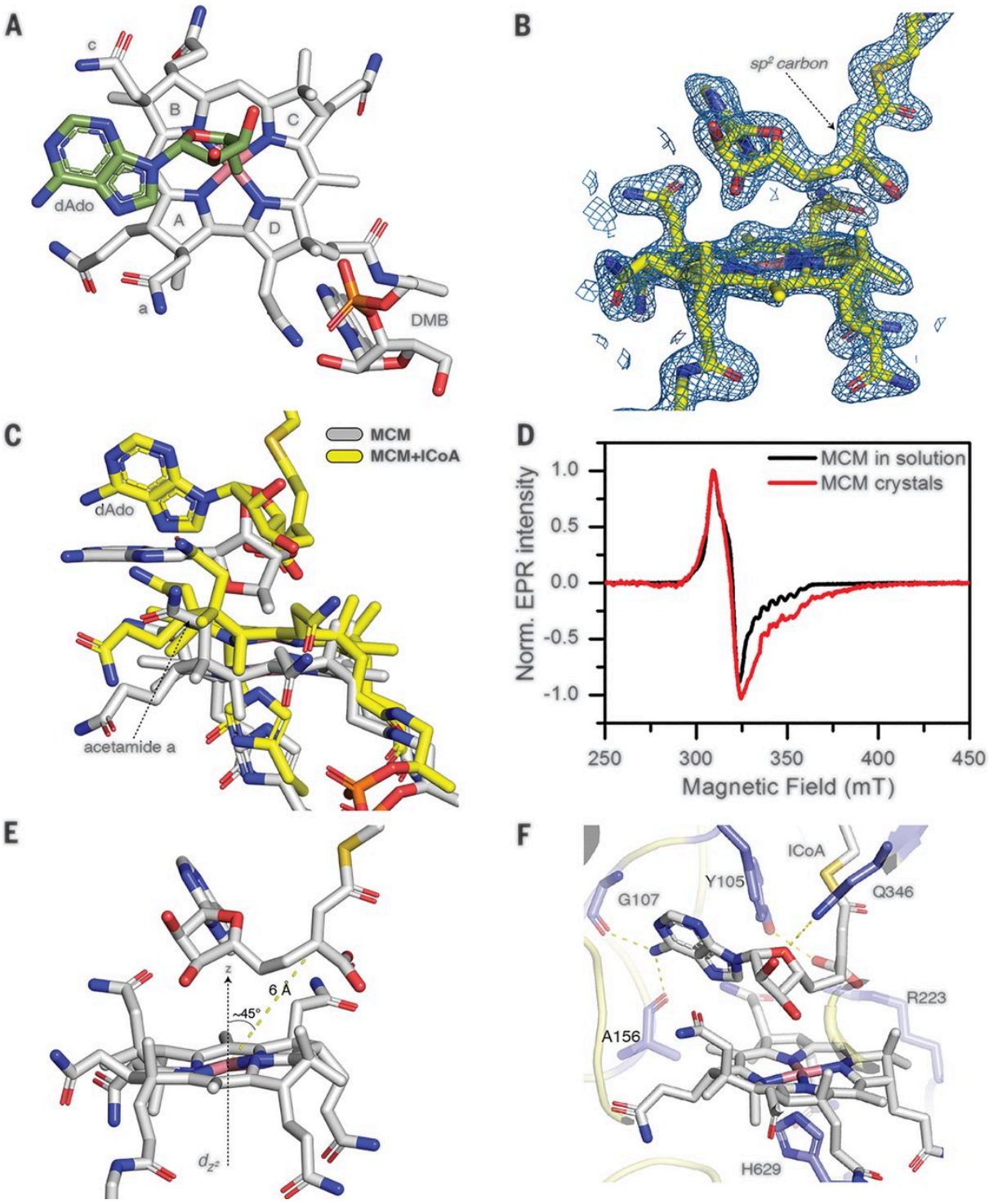


Figure 4.2: Crystallographic capture of a biradical in I-CoA-inactivated *Mtb* MCM

(A) Orientation of dAdo (green) in relation to the corrin ring (gray; pyrrole rings A to D and acetamides a and c are shown) in native Mtb MCM. **(B)** 2Fo-Fc omit maps (blue) around B_{12} and I-CoA contoured at 1.5s. **(C)** Shift in B12 and rotation of the adenine ring from the coplanar (gray) to perpendicular (yellow) position relative to the corrin ring. **(D)** EPR spectra of Mtb MCM + I-CoA. **(E)** Geometry of B12 and the I-CoA–dAdo adduct in crystal. **(F)** Hydrogen bonding interactions in the MCM–I-CoA structure.

is 6 Å from the cobalt and at an ~45° angle from the principal d_{z^2} orbital axis (Figure 4.2E), in excellent agreement with EPR simulations. To our knowledge, the only other enzyme-bound, carbon-centered radical that has been crystallized is the acetyl-thiazolium cation radical in pyruvate:ferredoxin oxidoreductase, which is formed via a one-electron transfer to an iron-sulfur cluster (22).

In the resting state, the C2′-OH in dAdo is engaged in a hydrogen-bonding interaction between Gln346 and a water-mediated hydrogen bond network to His260 and Arg223 (SI Figure 4.10F). In the I-CoA–inactivated structure, the C2′-OH maintains a hydrogen bond with only Gln346, and the adenine NH_2 group forms hydrogen bonds to the backbone carbonyls of Gly107 and Ala156 (Figure 4.2F). These interactions likely orient dAdo• for H-atom abstraction in the catalytic cycle when M-CoA is present; however, when I-CoA is present, it places dAdo• in close proximity to the double bond, setting up the radical addition reaction. Tyr105 forms a hydrogen bond with the terminal carboxylate of I-CoA and also flanks the dAdo moiety. In the resting enzyme, Tyr105 points in the opposite direction, i.e., away from the substrate. Arg223 also engages via electrostatic interactions with the terminal carboxylate of I-CoA and the acetamide side group *c* in the corrin ring. The structure has implications for how MCM controls the dAdo• radical trajectory to promote H-atom abstraction from M-CoA and concomitantly suppresses unwanted side reactions that could lead to radical extinction. As first predicted in computational studies (21), rotation and upward movement of the adenine ring (Figure 4.2C) position the C5′-carbon radical for H-atom abstraction from substrate. In contrast to MCM, glutamate mutase (23) and diol dehydratase (24) use ribose pseudorotation and N-glycosidic bond rotation, respectively, to bring the C5′-carbon of dAdo• to within Van der Waal's distance of the respective substrates.

4.2.4 I-CoA inhibits MCM repair

While cob(II)alamin is an intermediate in the MCM-catalyzed reaction, it also represents the inactive form of the enzyme when it becomes decoupled from the dAdo moiety, and thus it fails to re-form AdoCbl at the end of the catalytic cycle (9). Under these conditions, the auxiliary proteins CblA and ATR engage with MCM to off-load cob(II)alamin onto ATR for repair (Figure 4.3A). ATR catalyzes the adenosylation of cob(I)alamin to form AdoCbl and then transfers the cofactor to MCM to reconstitute the holoenzyme (25,26). Cofactor transfer in either direction between ATR and MCM requires the heterotrimeric guanine nucleotide binding protein chaperone CblA and is fueled by its guanosine triphosphatase (GTPase) activity (27). Because I-CoA leads to rapid inactivation of MCM and formation of cob(II)alamin, which cannot re-form AdoCbl, we assessed whether the enzyme can be repaired by the ATR-CblA system.

Addition of M-CoA, to initiate catalytic turnover by MCM in the presence of CblA-GDP (guanosine diphosphate), led to small (*h*MCM) or no (*Mtb* MCM) changes in the absorption spectra, indicating that these enzymes are resistant to oxidative inactivation (Figure 4.3, B and C). Consistent with this finding, the specific activities of both enzymes were reduced only ~15% after 1 hour of preincubation with M-CoA (Figure 4.3D). By contrast, the *P. shermanni* and *M. extorquens* MCM are much more prone to inactivation and accumulate aquocobalamin (OH_2Cbl) during turnover (28). Addition of the repair system (ATR, ATP, and GTP) led to complete recovery of MCM activity (Figure 4.3D), consistent with successful off-loading of inactive cob(II)alamin from MCM followed by reloading of AdoCbl from the assay mixture. In contrast to M-CoA, which supported catalytic turnover of MCM with minimal spectral changes, incubation of either human (Figure 4.3E) or *Mtb* (Figure 4.3F) MCM with I-CoA led to an immediate increase in absorption at 466 nm that did not change significantly over 1 hour and was correlated with complete loss of activity (Figure 4.3I). Addition of the respective human and *Mtb* repair proteins led to an increase

in absorbance at 350 to 356 nm and 530 to 534 nm, signaling oxidation of cob(II)alamin to OH_2Cbl (Figure 4.3, G and H). By contrast, addition of the repair systems to the same enzymes incubated with M-CoA did not induce cofactor oxidation (SI Figure 4.11). Following repair, only 23% (human) and 38% (*Mtb*) of the initial MCM activity was recovered (Figure 4.3I).

To gain insights into why the repair process is impeded by I-CoA but not M-CoA, we used *h*MCM, which forms a stable complex with CblA when it is in need of repair but is free in the active AdoCbl-bound state (SI Figure 4.12A) (27). In size exclusion chromatography profiles, *h*MCM is a stand-alone dimer (173 kDa) in the presence of M-CoA and the repair mixture (SI Figure 4.12B). However, in the presence of I-CoA, the MCM peak broadens and shifts to 211 kDa (SI Figure 4.12C), indicating the presence of a 1:1 *h*MCM:CblA complex (mass of 237 kDa) and free MCM. The 173- and 211-kDa fractions have spectra corresponding to AdoCbl and the biradical, respectively (SI Figure 4.12, D and F). The 82-kDa ATR fractions show that with M-CoA, very little cofactor is transferred to ATR (SI Figure 4.12E), because very little inactive MCM forms. By contrast, whereas I-CoA completely inactivates MCM, very little cofactor is offloaded to ATR (SI Figure 4.12G), indicating that the sustained presence of the dAdo-I-CoA adduct on MCM impedes cofactor repair.

Hobbling of the repair system helps explain why CLYBL deficiency is correlated with B_{12} deficiency. In the absence of available MCM active sites to off-load AdoCbl, ATR catalyzes an unusual sacrificial homolysis of the cobalt-carbon bond and sequesters cob(II)alamin, to which it binds more tightly than AdoCbl (25). We propose that CLYBL deficiency increases the propensity of I-CoA–dependent MCM inactivation and thereby leads to AdoCbl depletion (5). How a change in the mitochondrial B_{12} pool (AdoCbl) is signaled to the cytoplasm and affects B_{12} levels systemically is, however, not known.

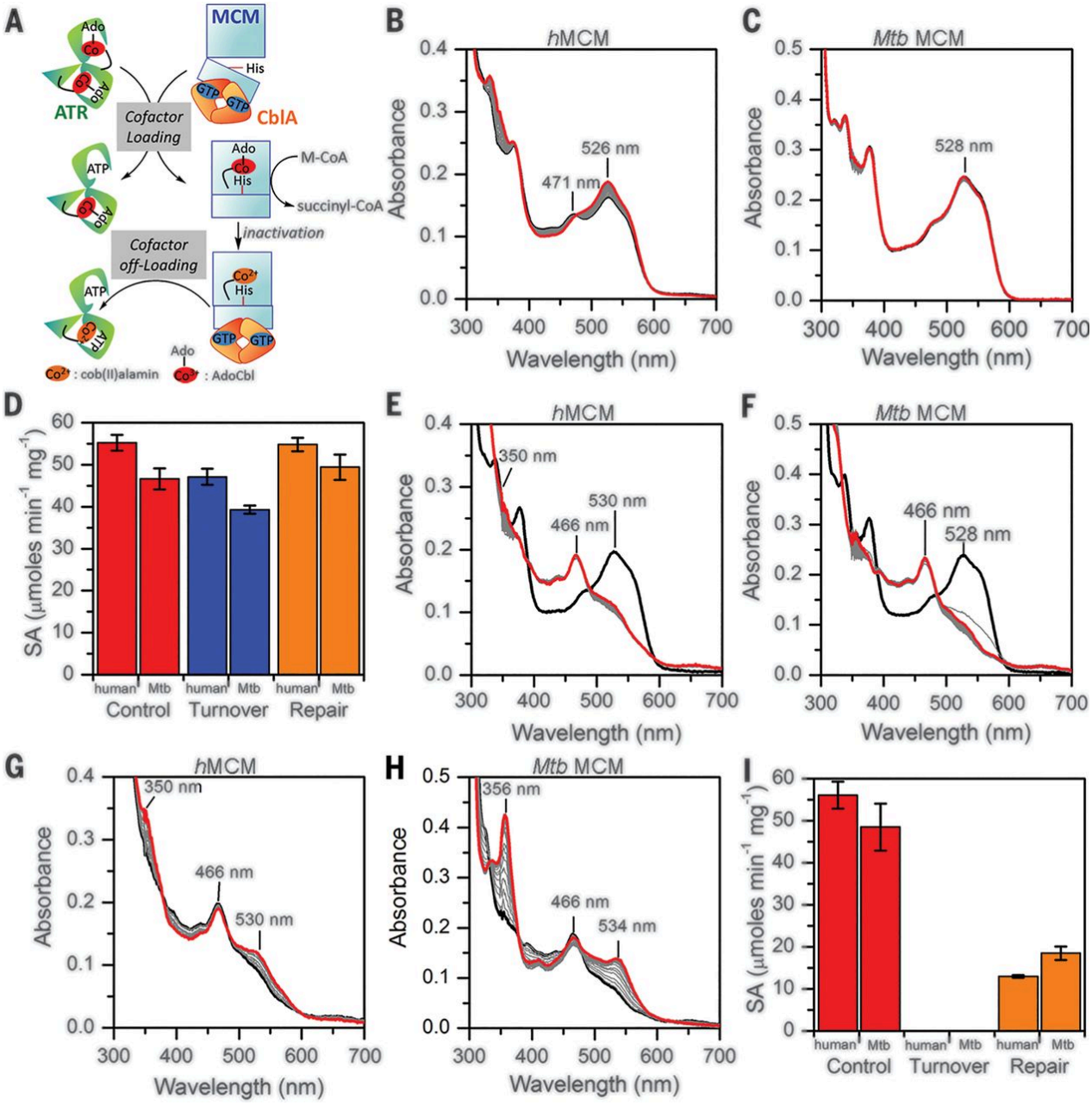


Figure 4.3: I-CoA inactivation impairs MCM repair

(A) Scheme showing the role of the auxiliary proteins in cofactor loading/off-loading to/ from MCM. **(B and C)** Enzyme-monitored turnover by human (B) and Mtb (C) MCM (black spectra) in the presence of M-CoA and human or Mtb CblA-GTP. Intermediate spectra (gray) were recorded every 2 min. Final spectra (red) were recorded at 1 hour. **(D)** Specific activity (SA) of human and Mtb MCM after 1-hour preincubation without or with M-CoA (red versus blue) and subsequent addition of the repair system (orange). **(E and F)** Addition of I-CoA to hMCM-AdoCbl [black, (E)] or Mtb MCM-AdoCbl [black, (F)] results in inactive enzyme (gray). Further incubation over 1 hour causes only modest spectral changes (red). **(G and H)** At the end of the experiments in (E) and (F), the repair system was added for 20 min to human (G) and Mtb (H) MCM. The increase in absorbance at 350 to 356 nm is indicative of OH_2Cbl formation (red). **(I)** Same as in (D) but with I-CoA; in both panels, data represent means ± SD (n= 3).

4.2.5 Itaconate inhibits vitamin B_{12}-stimulated *Mtb* growth on propionate

To corroborate the in vitro evidence that itaconate inhibits *Mtb* MCM, we directly tested whether exogenous itaconate can blunt B_{12}-stimulated *Mtb* growth on propionate as the sole carbon source. As reported previously, vitamin B_{12} supplementation at concentrations as low as 1 μg/ml stimulate growth of Mtb H37Rv on propionate (Figure 4.4, A and B), which has been attributed to the MCM-dependent pathway for propionate utilization (29). Growth stimulation was reduced in the presence of 1 mM itaconate and was completely inhibited at ≥5 mM itaconate (Figure 4.4C). Millimolar concentrations of itaconate are endogenously produced in activated macrophages (1), and Irg1-deficient mice, unlike controls, succumb early to *Mtb* infection (30), suggesting that such an inhibitory mechanism could be physiologically relevant.

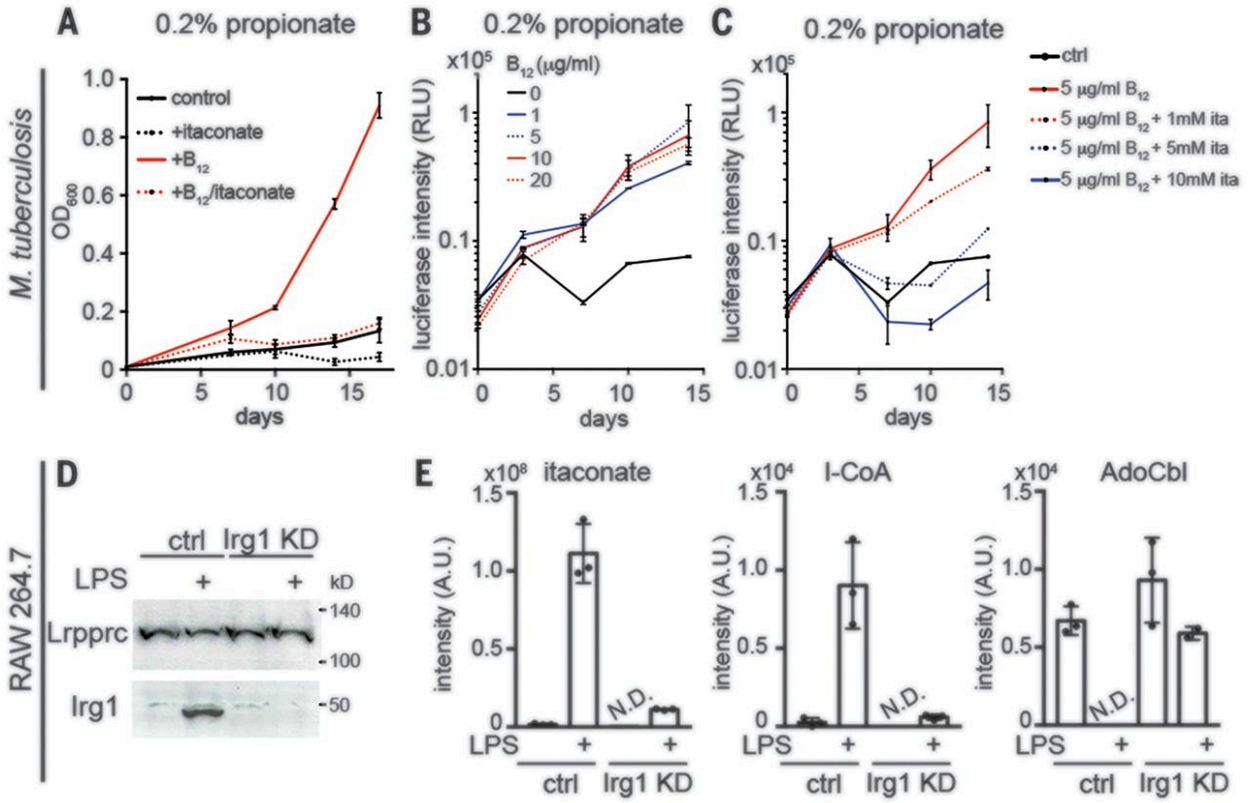


Figure 4.4: Itaconate inhibits B12-dependent Mtb and macrophage metabolism

(A) Vitamin B12 (10 μg/ml) stimulates growth of wild-type Mtb strain H37Rv on 0.2% propionate as the carbon source. OD, optical density. **(B and C)** B_{12} concentration dependence of Mtb growth and its inhibition by itaconate.

(D) Western blot of Irg1 in Irg1 CRISPR knockdown (KD) RAW264.7 cells with or without LPS (10 ng/ml) stimulation for 6 hours. Lrpprc, a mitochondrial protein, was used as the loading control**. (E)** Liquid chromatography–MS of itaconate, I-CoA, and AdoCbl in control and Irg1 KD RAW264.7 cells with or without LPS stimulation for 6 hours. Data represent means ± SD of three independent experiments. N.D., not detected.

4.2.6 MCM inhibition and AdoCbl depletion in macrophages require endogenous I-CoA synthesis

Mtb infection in mice elevates itaconate production in lungs (31), presumably through Irg1 induction. To recapitulate the metabolic consequence of endogenous itaconate production, we stimulated RAW264.7 cells with lipopolysaccharide (LPS), a potent activator of *Irg*1 transcription and itaconate production (1). We previously found that LPS stimulation depletes AdoCbl in macrophages (5). Using a CRISPR knockdown of Irg1 in RAW264.7 cells (Figure 4.4D), we observed that AdoCbl depletion is dependent on Irg1, which is transcriptionally upregulated in LPS-stimulated macrophages (1, 32). Treatment with LPS induced an increase in itaconate and I-CoA levels, which was significantly attenuated in Irg1 knockdown cells (Figure 4.4E). Although LPS stimulation reduced AdoCbl to undetectable levels in control cells, it was not significantly changed in the Irg1 knockdown cells (Figure 4.4E). Together, these data suggest that AdoCbl depletion and MCM inhibition in macrophages is caused by endogenous itaconate produced by Irg1 and induced during LPS stimulation.

4.3 Conclusions

AdoCbl is a radical initiator that generates the "working" dAdo• and "spectator" cob(II)alamin radical by homolytic cleavage of its cobalt-carbon bond. We found that I-CoA triggers homolytic cleavage of the cobalt-carbon bond in AdoCbl as in the normal MCM catalytic cycle, but proximity effects promote suicidal addition of dAdo• into its double bond. A chemically akin, albeit nonspecific, Michael addition mechanism has been invoked to explain itaconate-induced electrophilic stress (10, 11). The combined action of itaconate and I-CoA on *Mtb* propionate metabolism would be predicted to result in increased levels of toxic propionate/propionyl-CoA derived from cholesterol-dependent growth of this pathogen in host phagosomes (29). Although the conditions are not known under which the *Mtb* pathway for de novo B_{12} biosynthesis might be operative (33), *Mtb* can scavenge B_{12} from its host (34). Itaconate-induced B_{12} deficiency in host macrophages thus might be a strategy for restricting pathogen growth, in addition to targeting pathogen enzymes involved in propionate metabolism (Figure 4.1A); the relative importance of each inhibitory arm is unknown, however. We speculate that the *CLYBL* null background could boost the efficacy of this pathogen containment strategy, explaining the prevalence of the null genotype in human populations (5). In this context, it is noteworthy that the incidence of active tuberculosis is reported to be markedly lower in patients with B_{12} deficiency due to pernicious anemia (35).

4.4 Methods

4.4.1 Cloning, expression and purification of human mitochondrial B_{12} enzymes

The constructs used for expression of full-length human MCM and ATR (25), and CblA (27) and their purification have been described previously. The proteins were exchanged into 50 mM HEPES pH 7.5, 150 mM KCl, 2 mM $MgCl_2$, 2 mM TCEP, and 5% glycerol. The experiments were conducted in the same buffer unless indicated otherwise.

4.4.2 Construction of Mtb MCM, ATR and CblA expression plasmids

The genes encoding MCM_α (Rv1492) and MCM_β (Rv1493) were amplified by PCR using genomic DNA from *Mycobacterium tuberculosis* H37Rv (ATCC 25698), which was a generous gift from Dr. James Sacchettini (Texas A&M University), using the following primers:

MCM_α forward: 5′-ATGTGTGGATCC**TGTGTCCATTGATGTACCCGAGCGTGCC**-3′

MCM_α reverse: 5′-ACACCAAAGCTT**AGGCCCCCAACCGCGTCAGCAGAT**-3′

MCM_β forward: 5′-AGTGGTCATATG**ATGACAACCAAGACACCCG**-3′

MCM_β reverse: 5′ACAACCGGATCC**CTAATCCAGCGTGTACC**-3′

The restriction sites are underlined and the sequence in bold lettering represent the starting and terminal sequence of *Mtb* MCM. The PCR products were purified on a 1% agarose gel and extracted with a gel extraction kit (Quiagen). The small (MCMα) and large subunits (MCMβ) were each cloned into TOPO 2.1 vector (Thermo Fisher) to generate TOPO- MCM_α and TOPO- MCM_β, respectively. The TOPO- MCM_β plasmid was digested with NdeI and BamHI and after gel purification, cloned into multiple cloning site 2 in the petDuet vector (Novagen) to generate the petDuet-MCM_β vector. TOPO- MCM_α was digested with BamHI and HindIII and cloned into

multiple cloning site 1 in the petDuet-MCM$_\beta$ vector. A TEV cleavage site was introduced between the His tag and MCMα by PCR with the following primers (the TEV site is in bold lettering):

Forward: 5′-**GAGAACCTGTACTTCCAGAGCG**TGTCCATTGATGTACCCGAGCGTG-3′

Reverse; 5′-**GCTCTGGAAGTACAGGTTCT**CAGGATCCTGGCTGTGGTGATGATGG-3′

The *Mtb* ATR gene (Rv1314c) was amplified using the following primers:

5′-GGTGGTCATATG**GCAGTCCACCGGAACCGC**-3′

5′-ACAACAGCTAG**CTTAGCTCGCCGTCCTATCACCACCG**-3′

The PCR product was purified on a 1% agarose gel and ligated into the TOPO 2.1 vector. The TOPO-ATR vector was digested with NdeI and NheI HF and the ATR insert was purified on a 1% agarose gel and cloned into the pet28b vector. This construct generated His-tagged *Mtb* ATR with a thrombin cleavable his tag.

The *Mtb* CblA gene (Rv1496) was amplified from plasmid DNA pANT7 (DNASU plasmid repository) by PCR using the following primers:

5′-ACCACCGCTAGCTTAG**CTTACCTATCCGTTAGGTTAGCTATCTCC**-3′

5′-GGTGGTCATATG**ATGGCCGCATCCC**-3′

The PCR product was purified on a 1% agarose gel and ligated into the TOPO 2.1 vector. The TOPO-CblA vector was then digested with NdeI and NheI HF and cloned into the pet28b vector to generate a thrombin cleavable His-tagged Mtb CblA construct.The sequences of the *Mtb* MCM, ATR and CblA expression constructs were verified by nucleotide sequence analysis at the Genomics Core Facility (University of Michigan).

4.4.3 Purification of Mtb MCM

Mtb MCM (α and β subunits) was expressed in *Escherichia coli* BL21(DE3). The starting culture was grown overnight at 37 °C in Luria Bertani (LB) medium containing 100 μg/ml ampicillin. The initial culture (1 ml) was transferred to 1 L Terrific Broth (TB) medium containing 100 μg/ml ampicillin and grown at 37 °C until the OD_{600} reached 1.5 –1.8. Then, the temperature was reduced to 16 °C and 30 ml of DMSO was added (3 % v/v final concentration). Protein expression was induced with 500 μM isopropyl β-D-1-thiogalactopyranoside (IPTG) and the culture was grown for an additional 20–24 h at 16°C. The cells were harvested by centrifugation and stored at -80°C until further use.

The cell pellet was suspended (5 ml/g wet cell weight) in lysis buffer (50 mM Na_2HPO_4 pH 8.2, 300 mM NaCl, 10 mM imidazole, 10 mM β-mercaptoethanol and 5 % glycerol) containing 100 μM phenylmethylsulfonyl fluoride (PMSF) and 1 tablet protease inhibitor (Roche) and stirred for 30 min at 4°C. The cells were sonicated and the cell debris was removed by centrifugation at 38,500 x g for 45 min at 4 °C. The supernatant was loaded onto a Ni-NTA column (6 x 2.5 cm) and the column was washed with 500 ml lysis buffer. MCM was eluted from the column with 300 ml of a linear gradient containing 10 to 200 mM imidazole in the same buffer. Fractions containing MCM were pooled and concentrated to ~20 ml with an Amicon filter (50 kDa MWCO). TEV protease (0.02 mg TEV/mg protein) was added to the solution which was dialyzed against 50 mM Tris, pH 8.0, 300 mM NaCl, 1 mM DTT, 0.5 mM EDTA and 5% glycerol. The cleaved His tag and uncut MCM were removed by loading the solution onto a Ni-NTA column (6 x 2.5 cm). MCM present in the flow-through was concentrated with an Amicon filter (50 kDa MWCO) to ~20 ml and dialyzed against 50 mM HEPES pH 7.0, 50 mM NaCl, 5% glycerol. MCM was purified by ion exchange chromatography on a 2.5 x 10 cm Source Q column (Omnifit). Buffer A consisted of 50 mM HEPES, pH 7.0, 50 mM NaCl and 5% glycerol while Buffer B contained 50 mM

HEPES, pH 7.0, 500 mM NaCl and 5% glycerol. MCM was eluted with a linear gradient of 0–100 % Buffer B in 40 min at a flow rate of 4 ml/min. Fractions containing MCM were concentrated and dialyzed against buffer containing 50 mM HEPES, pH 7.5, 150 mM KCl, 2 mM $MgCl_2$ and 5% glycerol and flash frozen in liquid nitrogen.

4.4.4 Purification of Mtb CblA

Mtb CblA was expressed in *E. coli* BL21(DE3). The starting culture was grown overnight at 37°C in LB medium containing 50 μg/ml kanamycin. Then, 1 L of TB medium was inoculated with 1 ml of starting culture and grown at 37 °C until the OD_{600} reached 1.0–1.2. Protein expression was induced by addition of 500 μM IPTG and the culture was grown for an additional 18–20 h. The cells were harvested by centrifugation and stored at -80 °C.

The cell pellet was suspended (5ml/g wet cell pellet) in lysis buffer (50 mM Tris, pH 8.0, 500 mM NaCl, 10 mM imidazole and 5 % glycerol) containing 100 μM PMSF and 1 tablet protease inhibitor and stirred for 30 min at 4 °C. The cells were sonicated and the cell debris removed by centrifugation at 38,500 x g for 45 min at 4 °C. The supernatant was loaded onto a Ni-NTA column (5 x 2.5 cm). The column was washed with 500 ml lysis buffer. *Mtb* CblA was eluted from the column with a 300 ml of 10 to 300 mM linear imidazole gradient in the lysis buffer. Fractions containing CblA were pooled and concentrated to ~20 ml using an Amicon filter (30 kDa MWCO). Thrombin (5 U/mg protein) was added to the solution, which was dialyzed against 50 mM Tris, pH 8.0, 200 mM NaCl, 5% glycerol. The cleaved His tag and uncut CblA were removed by loading the solution onto a Ni-NTA column (5 x 2.5 cm). CblA present in the flow through was concentrated with an Amicon filter (30 kDa MWCO) to ~20 ml and loaded onto a benzamidine column (GE Healthscience) to remove thrombin. The column was washed with 150 ml lysis buffer.

Fractions containing Mtb CblA were pooled and dialyzed against 50 mM HEPES, pH 8.0, 300 mM KCl, 2 mM $MgCl_2$ and 5% glycerol. Aliquots were flash frozen in liquid nitrogen.

4.4.5 Purification of Mtb ATR

Mtb ATR was expressed in *E. coli* BL21(DE3). The starting culture was grown overnight at 37 °C in LB medium containing 50 μg/ml kanamycin. Then, 1 L of LB medium was inoculated with 1 ml starting culture and grown at 37 °C until the OD_{600} reached 0.5–0.7. Protein production was induced by addition of 500 μM IPTG and the culture was grown for an additional 4 h at 37 °C. The cells were harvested by centrifugation and stored at -80 °C.

The cell pellet was suspended (5 ml/g wet cell pellet) in lysis buffer (50 mM Tris, pH 8.0, 300 mM NaCl, 10 mM imidazole, 1 mM TCEP and 5 % glycerol) containing 100 μM PMSF and 1 tablet protease inhibitor and stirred for 30 min at 4°C. The cells were sonicated and the cell debris removed by centrifugation at 38,500 x g for 45 min at 4 °C. The supernatant was loaded onto a Ni-NTA column (5 x 2.5 cm). The column was washed with 500 ml lysis buffer. *Mtb* ATR was eluted from the column with 300 ml of a 10 to 300 mM linear imidazole gradient in the lysis buffer. Fractions containing ATR were pooled and concentrated to ~20 ml with an Amicon filter (10 kDa MWCO). Thrombin (5 U/mg protein) was added to the solution which was dialyzed into 50 mM Tris, pH 8.0, 300 mM NaCl, 1 mM TCEP, 5% glycerol. The cleaved His tag and uncut ATR were removed by loading the solution onto a Ni-NTA column (5 x 2.5 cm). ATR was eluted with 30 mM imidazole in buffer containing 50 mM Tris pH 8.0, 300 mM NaCl, and 5 % glycerol and was concentrated with using an Amicon filter (10 kDa MWCO) to ~30 ml. Thrombin was removed by a benzamidine column. The column was washed with 150 ml 50 mM Tris, pH 8.0, 300 mM NaCl, 5 % glycerol followed by 100 ml 50 mM Tris pH 8.0, 500 mM NaCl, 5 % glycerol. Fractions

containing Mtb ATR were pooled and dialyzed against 50 mM HEPES, pH 7.5, 150 mM KCl, 2 mM $MgCl_2$, 2 mM TCEP and 5% glycerol. Aliquots were flash frozen in liquid nitrogen.

4.4.6 Expression and purification of succinyl thiokinase

Thiokinase (*E. coli* succinyl-CoA synthase SucCD from Addgene) was expressed in BL21(DE3) *E. coli*. A single colony was used to inoculate a 50 ml LB culture (containing 50 µg/mL kanamycin) and grown overnight at 37 °C. The next day, 10 ml of the starting culture was added per liter of LB and grown at 37 °C until OD_{600} reached 0.4. At this point, the temperature was lowered to 27 °C and thiokinase expression was induced with 1 mM IPTG when OD_{600} was 0.8. The cells were grown for an additional 3 h and harvested by centrifugation at 2880 x g, the cell pellet was frozen at -80 °C.

The cell pellet from a 2 L culture was resuspended in 200 ml of 50 mM sodium phosphate, pH 8.0, 300 mM NaCl, and 10 mM imidazole, supplemented with 0.1 mM PMSF and one protease inhibitor tablet (Roche). Cells were lysed by sonication and loaded onto a 20 ml Ni(II)-NTA column. After washing with 150 ml of lysis buffer, thiokinase was eluted with a 150 ml linear gradient in lysis buffer containing 10-230 mM imidazole. Fractions containing thiokinase (30 and 42 kDa subunits) were concentrated, dialyzed overnight against 100 mM potassium phosphate, pH 7.5, then flash frozen, and stored at -80 °C until further use.

4.4.7 Repair of I-CoA inactivated hMCM

First, *h*MCM (11.5 µM dimer) was loaded with 23 µM AdoCbl and mixed with 40 µM *h*CblA (dimer) and 800 µM GDP in 50 mM HEPES, pH 7.5, 150 mM KCl, 2 mM $MgCl_2$, 2 mM TCEP, 5% glycerol (1 ml total volume) and incubated for 10 min at 30 °C. The solution was divided into three 350 µl aliquots: a “control” (no addition of I-CoA or the repair system), a “turnover” (addition

of I-CoA only) and a "repair" (addition of I-CoA followed by the repair system) sample. Next, the control and turnover samples were placed in a 30 °C temperature-controlled block and the repair sample in a cuvette was placed in a temperature-controlled cuvette holder (at 30 °C). Then, 10 µl of I-CoA (0.5 mM final concentration) was added to the "turnover" and "repair" samples, and an equal volume of buffer was added to the control sample. The samples were incubated for 60 min, and spectra were recorded every 2 min. Then, 40 µl of a premixed solution of *h*ATR (333 µM), ATP (50 mM), and GTP (70 mM) was added to the "repair" sample and an equal volume of buffer was added to the control and the "turnover" sample. The samples were incubated for additional 20 min. The final concentrations of reagents in the "repair" sample was: 10 µM *h*MCM dimer, 20 µM AdoCbl, 35 µM *h*CblA dimer, 33.3 µM *h*ATR trimer, 0.7 mM GDP, 7 mM GTP, 5 mM ATP and 0.5 mM I-CoA in 400 µl total volume. The samples were then assayed by the thiokinase assay described below to evaluate MCM activity. The same procedure was used to determine inactivation during turnover with 0.5 mM M-CoA replacing I-CoA.

Inactivation of *Mtb* MCM with I-CoA or M-CoA was tested in the same manner. *Mtb* MCM (25 µM) was mixed with 25 µM AdoCbl, 43.8 µM *Mtb* CblA and 0.875 mM GDP in 50 mM HEPES, pH 7.5, 150 mM KCl, 2 mM $MgCl_2$, 2 mM TCEP, 5% glycerol (700 µl total volume) and incubated at 30 °C for 10 min. The solution was split into three 197 µl samples: a "control", a "turnover" and a "repair" sample. Then, 3 µl I-CoA (0.5 mM final concentration) was added to the "turnover" and "repair "sample and incubated for 60 min. A premixed solution of *Mtb* ATR (167 µM), ATP (25 mM) and GTP (35 mM) in 50 µl total volume was added to the "repair" sample whereas an equal volume of buffer was added to the "control" and "turnover" sample. The final concentration in the control sample was 20 µM *Mtb* MCM, 20 µM *Mtb* AdoCbl, 35 µM *Mtb* CblA

(dimer), 33.3 μM *Mtb* ATR (trimer), 0.7 mM GDP, 7 mM GTP, 5 mM ATP, 0.5 mM I-CoA. The activity was then determined by the thiokinase assay described below.

4.4.8 Thiokinase assay for MCM activity

MCM activity was assessed using the thiokinase-coupled spectrophotometric assay as described previously (36). Briefly, the assay was conducted at 30 °C and concentrations of reagents in a final volume of 200 μl were: 2 nM *h*MCM (0.08 μg), 5 μM AdoCbl, 3 mM GDP, 100 mM potassium phosphate, pH 7.5, 0.5 mM M-CoA, 20 μg of thiokinase, and 70 μM 5,5′-dithiobis(2-nitrobenzene). First, thiokinase, GDP, and M-CoA were incubated for 5 min at 30 °C in a cuvette and the background thioesterase activity was recorded by the increase in absorbance at 412 nm. Next, the reaction was initiated by the addition of MCM and AdoCbl to the cuvette. The specific activity was calculated assuming that the amount of CoA detected is directly proportional to the concentration of succinyl-CoA formed by MCM. An extinction coefficient of 14,150 cm^{-1} M^{-1} was used for the TNB^- anion (37). When MCM activity in the "control", "turnover" and "repair" samples described in the preceding section were interrogated, 2 μl of the sample was added to the thiokinase reaction mixture instead of MCM. When the thiokinase assay was used to determine the kinetic parameters, K_m and V_{max}, 10 μM MCM and 20 μM AdoCbl were preincubated for 10 min at 30 °C prior to an aliquot being added to the thiokinase assay.

4.4.9 I-CoA synthesis and purification

The final concentrations of reagents used for I-CoA synthesis were: 7.5 mM itaconate, 7.5 mM ATP, 2.5 mM CoASH (~100 mgs), 10 mM $MgCl_2$ and 40-50 mg thiokinase in a final volume of 35 ml of 50 mM HEPES, pH 7.5, 300 mM NaCl and 10 mM $MgCl_2$. The reaction mixture was stirred gently at room temperature for 30 min and stopped by precipitating thiokinase with a 1:10

volume of 14% perchloric acid (1.4% final concentration). The precipitated protein was removed by using a 0.2 μm syringe filter. Approximately 7 x 5 ml sample was injected onto a preparative C18 column (SunFire, 10 μm) equilibrated with 0.1% trifluoroacetic acid (TFA). Buffer B contained 80% acetonitrile and 0.1% TFA. The following gradient was used to elute I-CoA flowing at a flow rate of 4 ml min^{-1} (time, %B): 0 min, 0%; 10 min, 0%; 30 min, 70%; 40 min, 70%; 41 min, 0%; 45 min, 0%. I-CoA eluted with a retention time of ~23 min while unreacted CoA eluted at 21 mins. Fractions containing I-CoA were frozen, lyophilized, and stored at -80 °C. The typical yield of I-CoA was 40-50%. Before use, the I-CoA powder was resuspended in water and an $\varepsilon_{258\ nm}$ of 19.84 cm^{-1} mM^{-1} for the CoA moiety was used to estimate concentration. The stock solution is acidic due to the carryover of TFA and stabilizes the thioester. Prior to the experiment, the stock solution was diluted 1:1 with 500 mM sodium phosphate, pH 7.5, to bring the pH to a neutral range and avoid protein precipitation.

4.4.10 Synthesis of [$^{13}C_5$]-itaconyl-C4-CoA

In a 1.5 ml tube, 0.5 mg [$^{13}C_5$]-itaconic acid (3.8 mM), CoA (4.6 mM), ATP (20 mM), $MgCl_2$ (3 mM) were mixed in with 100 mM HEPES, pH 7.5, 150 mM NaCl in a total volume of 1 ml and warmed to 30°C. Then, 40 μM thiokinase were added and the mixture was incubated at 30 °C for 30 min. The reaction was stopped with 100 μl 10 % TFA (1 % final concentration) and centrifuged for 10 min at 21,130 x g for 10 min. [$^{13}C_5$]-itaconyl-C4- CoA was purified as described above for the unlabeled compound. [$^{13}C_5$]-itaconic acid was provided by the NIH Common Fund Metabolite Standards Synthesis Core or purchased from Santa Cruz Biotechnology (sc-495554).

4.4.11 Expression and purification of methylmalonyl-CoA synthase (MCS)

The expression vector for MCS was provide by Dr. Chaitan Khosla (Stanford University). The protein was expressed in BL21(DE3) *E.coli*. Following a fresh transformation, starting 200 µl LB was added to the transformation mixture and allowed to shake for 1 h at 37 °C. Next, 100 µl of this culture was added to 2 x 40 ml LB medium containing 100 µg/mL amplicillin and grown overnight at 37 °C. The next day, 10 ml of the starting culture was added per liter of LB supplemented with ampicillin and grown until $OD_{600\ nm}$ was 0.5, at which point the temperature was dropped from 37 ° to 20 °C. Protein expression was induced with 0.5 mM IPTG and growth was continued overnight. Cells were harvested by centrifugation and stored at -80 °C. Cell lysis and purification was performed using the same protocol as described for thiokinase. Following elution from the Ni(II)-NTA column, MCS was concentrated and exchanged into 200 mM potassium phosphate, pH 7.5. The protein was immediately used for the M-CoA synthesis since it is unstable upon storage.

4.4.12 M-CoA synthesis

For the synthesis of M-CoA, the reaction was performed in 200 mM potassium phosphate, pH 7.5, with the following reagent concentrations: 10.6 mM ATP, 10.6 mM methylmalonate, 12.5 mM $MgCl_2$, 5.4 mM CoASH (100 mgs), and 50-60 mg of MCS with a reaction volume of 24 ml. Following addition of MCS, the reaction was usually >95% complete after 5 min based on an HPLC analysis of remaining CoA. The reaction was quenched with a 1:10 volume of 14% perchloric acid (1.4% final concentration, usually after 45 min). M-CoA was purified in the same manner as described above for I-CoA. Under these conditions, M-CoA showed a retention time of ~27.5 min. M-CoA containing fractions were pooled, lyophilized, and treated in the same manner as described above for I-CoA. The typical yields for the enzymatic synthesis of M-CoA was ~60-70%.

4.4.13 Size exclusion chromatography to assess *h*MCM-*h*CblA complex formation

The repair mixture contained the following final concentrations of reagents: 37.5 μM *h*MCM, 75 μM *h*AdoCbl, 131 μM *h*CblA, 0.7 mM GDP, 7 mM GTP, 5 mM ATP, 125 μM ATR, and 1 mM M-CoA or I-CoA in 50 mM HEPES, pH 7.5, 150 mM KCl, 2 mM $MgCl_2$, 2 mM TCEP, 5% glycerol. First, *h*MCM, *h*CblA, AdoCbl and GDP were mixed and incubated for 10 min at 30 °C using a temperature-controlled block. Then, M-CoA or I-CoA was added to the mixture incubation was continued for 60 min. Next, the rest of the repair system comprising ATR, ATP, and GTP were added and incubation was continued for 20 min. The reaction mixture (100 μl) was then immediately injected onto a Superdex 200 10/300 GL column (GE Life Sciences) equilibrated with 50 mM HEPES, pH 7.5, 150 mM KCl, 2 mM $MgCl_2$, 2 mM TCEP, 5% glycerol. The column was eluted with the same buffer at a flow rate of 0.4 ml min^{-1}. The absorbance of the eluate was monitored at 280, 350, and 465 nm. Fractions were also analyzed using SDS-PAGE and peaks corresponding to hMCM and hATR were collected separately and concentrated using a 10 kDa MWCO membrane (Nanosep) before recording absorption spectra.

4.4.14 EPR spectroscopy and simulations

EPR samples were prepared by mixing 375 μM *h*MCM and 500 μM AdoCbl in 50 mM HEPES, pH 7.5, 150 mM KCl, 2 mM $MgCl_2$, 2 mM TCEP, ~15% glycerol to which 1 mM (^{12}C or ^{13}C) I CoA was added. EPR spectra were collected on a Bruker E500 ElexSys EPR spectrometer equipped with a Bruker ER4123 SHQE cavity and a Bruker ER4131VT liquid nitrogen flow temperature control system. Spectra were acquired at 120 K with non-saturating microwave power. Acquisition parameters: microwave frequency, 9.46 GHz; microwave power, 2.0 mW; modulation amplitude, 1.0 mT; modulation frequency, 100 kHz; time constant, 40.96 ms; conversion time,

81.92 ms; spectra represent average of 8 scans collected over a full sweep range of 129.0 – 479.0 mT. The spectra were corrected by subtraction of a buffer baseline, acquired under the same conditions. Simulations of the cob(II)alamin-organic radical pair spectra were performed using the MATLAB (Mathworks, Natick, MA) programs that utilized the EasySpin (v. 5.2.23) tool box (38). The calculations were based on the following spin Hamiltonian:

$$\hat{H} = \beta_e \mathrm{B} g_1 \hat{S}_1 + \beta_e \mathrm{B} g_2 \hat{S}_2 + \beta_n \hat{S}_1 A_1 \hat{I}_1 + \beta_n \hat{S}_2 A_2 \hat{I}_2 + J\hat{S}_1 \hat{S}_2 + \hat{S} D \hat{S}$$

Equation 4.1

where β_e and β_n are the Bohr and nuclear magnetons, and the first two terms represent the electron Zeeman interaction of spin 1 (Co^{2+}, S=1/2) and spin 2 (radical, S=1/2), the third term represents the hyperfine interaction of spin 1 with the ^{59}Co nucleus (I=7/2), the fourth term represents the superhyperfine interaction of spin 2 with the ^{13}C nucleus (I=1/2; for the case of ^{13}C-labeled I-CoA), the fifth term represents the spin 1-spin 2 isotropic exchange interaction, and the last term represents the zero-field splitting (dipolar interaction of spins 1 and 2, in terms of the total electron spin operator). For the natural isotopic abundance radical pair, the fixed parameters included the following: g-tensor of Co^{2+} in cob(II)alamin, g_{1x}=2.30, g_{1y}=2.21, g_{1z}=2.00; g-tensor of radical, g_{2x}= g_{2y}= g_{2z}=2.00; $^{59}Co^{2+}$ hyperine coupling tensor, A_{1x} =28, A_{1y}=17, A_{1z}=294 MHz; line width, 1.0 mT. The $\boldsymbol{g}_1$ and $\boldsymbol{A}_1$ parameters for cob(II)alamin in MCM were as determined, previously (12). The principal axes of $\boldsymbol{g}_1$ and $\boldsymbol{A}_1$ were assumed to be aligned. The varied parameters included *J, **D**,* the Euler angles (*α, β, γ* ; *z-y-z* rotation) that relate the principal axis systems of $\boldsymbol{g}_1$ and ***D***, and a general isotropic broadening parameter, representing unresolved superhyperfine coupling. For the ^{13}C-labeled I-CoA substrate, the superhyperfine coupling tensor, $\boldsymbol{A}_2$, was varied, and assumed to be isotropic. In the simulations, an iterative global scanning of the varied parameters, followed by Nelder-Mead simplex optimization (performed on the as-collected, derivative lineshape), was used

to converge on a set of parameters, that satisfied the combination of visual and statistical (root mean square deviation, rmsd) criteria. The optimized values of the varied parameters for the natural abundance radical pair are as follows: J=6365 MHz; principal components of **D**, D_x =363, D_y =-8, D_z = -355 MHz; Euler angles, α=13°, β=43°, γ=147°. Simulation of the radical pair formed from ^{13}C -labeled I-CoA, and in particular, the broadening of the resolved high-field septet structure, was achieved by using the parameters for the natural isotopic abundance radical pair and the ^{13}C superhyperfine coupling, A_{2x}= A_{2y}= A_{2z}=90 MHz.

The stability of the biradical was followed by EPR spectroscopy. For this, *h*MCM (100 μM) was mixed with 50 μM AdoCbl in the buffer described above in an anaerobic chamber (O_2 < 0.3 ppm) and incubated for 10 min at room temperature. Then, 500 μM I-CoA were added and following 2 min incubation, was transferred to an EPR tube and frozen in liquid N_2. After the EPR spectrum was recorded, the sample was thawed and incubated at room temperature for the desired time (20, 30, 45 and 60 min) before being refrozen for recording the EPR spectrum. The same procedure was used for *Mtb* MCM, except that the sample contained 200 μM *Mtb* MCM, 100 μM AdoCbl and 500 μM I-CoA. These EPR spectra were recorded on a Bruker EMX 300 equipped with a Bruker 4201 cavity and a ColdEdge cryostat. The temperature was controlled by an Oxford Instruments MercuryiTC temperature controller. EPR spectra were recorded at 80 K using the following parameters: 9.44 GHz microwave frequency, power 2 mW, modulation amplitude 10 G, modulation frequency 100 kHz, 2000 G sweep width centered at 3500 G, conversion time 164 ms, time constant 82 ms. Five scans were collected per measurement.

4.4.15 EPR spectroscopy of Mtb MCM crystals treated with I-CoA

Mtb MCM was crystallized at 20 mg/ml in 0.4 M sodium phosphate monobasic dihydrate, 1.6 M potassium phosphate dibasic, 0.1 M imidazole/ HCl pH 8, and 0.2 M sodium chloride. Approximately 450 crystals were harvested and placed in a solution of 2.5 mM I-CoA, 0.4 M sodium phosphate monobasic dihydrate, 1.6 M potassium phosphate dibasic, 0.1 M imidazole/HCl, pH 8, and 0.2 M NaCl and allowed to soak for 1h. Glycerol was added to the solution for cryoprotection to a final concentration of 15% glycerol, 2.5 mM I-CoA, 0.32 M sodium phosphate monobasic, 1.3 M potassium phosphate dibasic, 0.082 imidazole/HCl, pH 8, 0.16 M sodium chloride. The entire solution was then transferred to an EPR tube and frozen in liquid N_2. The EPR spectrum was recorded at 80 K with the following parameters: 9.38 GHz microwave frequency, power 20 mW, modulation amplitude 10 G, modulation frequency 100 kHz, 2000 G sweep width centered at 3500 G, conversion time 328 ms, time constant 164 ms. 25 scans were accumulated.

4.4.16 Product analysis of I-CoA inactivated MCM

In a 500 µL reaction volume, 110 µM human or *Mtb* MCM was mixed with 220 µM AdoCbl and the sample was allowed to equilibrate for 10 min. Next, I-CoA was added to 220 µM and the reaction was allowed to progress for 30 min at room temperature before being quenched with a 1:10 volume of 7% perchloric acid (0.7% final concentration). After centrifuging to remove precipitated protein, the sample was injected onto a preparative C18 column (SunFire C-18, 10 µm 10 x 250 mm) and purified using the following Buffer A: 75 mM Na_3PO_4 pH 4.5, and Buffer B: 50% acetonitrile, and the following gradient at a flow rate of 4 ml min^{-1} (time, %B): 0 mins, 10%; 5 min, 10%; 40 min, 45%; 41 min, 100%; 46 min, 100%; 47 min, 10%; 52 min, 10%. Fractions of interest were lyophilized, and stored at -80 °C. For MALDI MS analysis, samples were resuspended in water and zip-tipped (C_4, Millipore) to remove salts. Samples were analyzed by

MALDI MS in the negative-ion mode (University of Michigan Chemistry MS Facility). Identical HPLC profiles (n=4) and MS (n=2) results were obtained with the human and *Mtb* MCM samples.

4.4.17 I-CoA titrations

*h*MCM (20 μM) was mixed with 40 μM AdoCbl in in 50 mM HEPES pH 7.5, 150 mM KCl, 2 mM $MgCl_2$, 2mM TCEP, 5% glycerol and the mixture was allowed to equilibrate for 10 min at 25 °C. An initial spectrum was recorded and after each addition of I-CoA (3.5 to 55 μM), the sample was allowed to equilibrate for 5 min before acquiring a spectrum. The change in absorbance at 530 nm was used to determine a K_d. The data were fit to a single I-CoA binding per MCM monomer using Dynafit (39).

Mtb MCM (30 μM) was mixed with AdoCbl (30 μM) in 50 mM HEPES pH 7.5, 150 mM KCl, 2 mM $MgCl_2$, 2 mM TCEP and 5 % glycerol and incubated at 25 °C for 10 min. I-CoA (5 – 200 μM) was added and the formation of inactivated species was monitored. The decrease in absorbance at 525 nm was plotted against I-CoA concentration and the data were fit to a one site binding model using Dynafit.

4.4.18 Crystallization of *Mtb* MCM

AdoCbl (600 μM) was mixed with *Mtb* MCM (500 μM) in 50 mM HEPES, pH 7.5, 150 mM KCl, 2 mM $MgCl_2$, 2 mM TCEP, 5% glycerol and incubated at 25 °C for 15 min. Unbound AdoCbl was removed by size exclusion chromatography on an S-200 column (1.6 x 100 cm) (GE Life Sciences), pre-equilibrated with the same buffer described above. Fractions containing *Mtb* MCM-AdoCbl were pooled and concentrated to 25 mg/ml with an Amicon filter (MWCO = 50 kDa).

Crystals of *Mtb* MCM-AdoCbl were obtained by the sitting-drop vapor-diffusion method by mixing 0.3 μl of protein solution (20 mg/ml) with 0.6 μl of the reservoir solution. *Mtb* MCM-

AdoCbl crystals were formed at 20 °C in 100 mM sodium cacodylate, HCl pH 6.5, 800 mM sodium citrate tribasic dihydrate, 0.05% w/v casein, 0.05% w/v hemoglobin, 0.0005% w/v pepsin, 0.0005% (w/v) protease, 0.0005% (w/v) proteinase K, 0.0005% (w/v) trypsin, and 0.002M HEPES sodium, pH 6.8. The crystals were cryoprotected in a solution consisting of 15% glycerol, 85 mM sodium cacodylate, and 680 mM sodium citrate tribasic dihydrate before being flash-cooled in liquid nitrogen. All procedures regarding crystal growth and manipulations were conducted in the dark.

Crystals of *Mtb* MCM-AdoCbl used for soaking with I-CoA were obtained by the sitting drop vapor-diffusion method by mixing 0.45 µl of protein solution (20 mg/mL) with 0.45 µl of the reservoir solution. *Mtb* MCM-AdoCbl crystals used for soaking were formed at 20 °C in 0.4 M sodium phosphate monobasic, 1.6 M potassium phosphate dibasic, 0.1 M imidazole/HCl pH 8, and 0.2 M sodium chloride. A 28 mM freshly prepared solution of I-CoA in water was mixed at a 1:1 ratio with 1M HEPES, pH 7.0, to obtain a 14 mM solution for each soaking experiment. I-CoA was then mixed with the reservoir solution to a final concentration of 2.5 mM. *Mtb* MCM-AdoCbl crystals were transferred to a solution of 2.5 mM I-CoA, 0.32 M sodium phosphate monobasic, 1.3 M potassium phosphate dibasic, 0.082 imidazole/HCl, pH 8, 0.16 M sodium chloride, 0.1 M HEPES, pH 7 and allowed to soak for 70 min. The crystals were cryoprotected in a solution of 15% glycerol, 2.5 mM I-CoA, 0.27 M sodium phosphate monobasic, 1.1 M potassium phosphate dibasic, 0.067 imidazole/HCl, pH 8, 0.13 M sodium chloride, 0.1 M HEPES, pH 7, before being flash-cooled in liquid N_2. All procedures regarding crystal growth and manipulations were conducted in the dark.

4.4.19 Crystal structure determination

Diffraction data were collected at 100 K on the GM/CA-CAT 23-ID-B (for *Mtb* MCM-AdoCbl) or 23-ID-D (for *Mtb* MCM-AdoCbl + I-CoA) beamline at Advanced Photon Source, Argonne National Laboratory using a DECTRIS EIGER16M detector for MCM-AdoCbl or a DECTRIS PILATUS3-6M detector for MCM-AdoCbl + I-CoA. All data sets were processed with *XDS* (40) to 1.9 Å and 2.0 Å, respectively for the structures without and with I-CoA. Data collection and processing statistics are summarized in Table S1. The crystals of MCM-AdoCbl were indexed to space group $P2_12_12_1$ (unit-cell parameters a = 77.8, b = 103.6, c = 211.2 Å) with one molecule in the asymmetric unit (Matthews coefficient V_M= 2.77 $Å^3$ Da^{-1}, 55.6 % solvent content) and the crystals of MCM-AdoCbl + I-CoA were indexed to space group $P2_12_12_1$ (unitcell parameters a = 76.6, b = 104.9, c = 194.1 Å) with one molecule in the asymmetric unit (Matthews coefficient V_M = 2.54 $Å^3$ Da^{-1}, 51.6 % solvent content).

4.4.20 Crystal structure refinement

Initial structure solutions were obtained by molecular replacement using as a search model substrate-free *P. shermanii* MCM (PDB: 2req) (41) for *Mtb* MCM-AdoCbl and *P. shermanii* MCM substrate complex (PDB: 4req) for *Mtb* MCM-AdoCbl + I-CoA and initial phases were calculated using *Phaser* (42).

Both structures were initially refined with *REFMAC5* (43) using isotropic individual *B* factors with maximum-likelihood targets where the Babinet model for bulk-solvent scaling was utilized. *PHENIX* (44) was then employed for subsequent refinements. Refinement included coordinate minimization and restrained individual *B*-factor adjustment with maximum-likelihood targets. Refinement was followed by model building and modification with *Coot* (45). Several iterative rounds of refinement followed by model building and modification were performed. In early rounds of refinement, ligands were not fitted into either structure.

Once the protein model was complete and the water molecules had been assigned, clear electron density was fitted and refined as B_{12} and dAdo ligands in both structures using *eLBOW* (46) to generate chemical restraints. Additionally, I-CoA was modeled into residual electron density in the *Mtb* MCM-AdoCbl + I-CoA structure. Restraints for I-CoA were generated using *ProDrug* (47) and I-CoA was inserted into the model using *ReadySet* (48) before subsequent refinement.

The final model was used to create an unbiased refined composite $2F_O$-F_C map with *PHENIX*. The refined iterative composite omit map was calculated using the corresponding program in Phenix (49). Crystallographic information as well as refinement statistics are provided in Table S1. The geometric quality of the model and its agreement with the structure factors were assessed with *MolProbity* (50). Figures displaying crystal structures were generated with *PyMOL* (51).

4.4.21 Inhibition of B_{12}-stimulated growth of Mtb on propionate

An auto-luminescent Mtb strain was constructed by transforming pJW276, a chromosomal-integration plasmid that allows expression of the bacterial *lux* operon in mycobacteria (a kind gift from Dr. Jeffrey Wagner, Harvard T.H. Chan School of Public Health), into wild-type *Mtb* (52). Wild-type *Mtb* or its luminescent derivative was cultured in sterile PETG inkwells containing 10ml 7H9 medium with 0.2% glycerol, 0.05% Tween-80 and 10% OADC enrichment. Exponentially growing bacteria (OD_{600} ~ 0.5-0.8) were pelleted at 4,000 x g, washed once with fresh Sauton's media (53), then suspended in modified Sauton's media (0.05% KH_2PO_4,0.05% $MgSO_4$, 0.2% citric acid, 0.005% ferric ammonium citrate, 0.05% $(NH_4)_2SO_4$, 1:100000 dilution of 1% $ZnSO_4$ stock solution, 1:2000 dilution of 10% Tyloxapol stock solution), supplemented with 0.2% (v/v) propionate to a final OD_{600} of 0.01. The diluted bacterial suspension was aliquoted into 96 well

plates supplemented with itaconate and/or vitamin B_{12} (cyanocobalamin) at the indicated concentrations, sealed with an air-permeable membrane (Breathe-Easy, Diversified Biotech), and incubated at 37 °C with constant shaking. A multi-channel plate reader (Synergy H1 Hybrid, BioTEK) was used to monitor OD_{600} or the luminescent signal at the indicated time points. Middlebrook 7H9 medium and OADC enrichment were purchased from BD Biosciences.

4.4.22 Effect of Irg knockdown on itaconate, I-Coa and AdoCbl concentration

RAW 264.7 cells (TIB-71) were obtained from ATCC. The cells were cultured in DMEM supplemented with 10 % FBS, penicillin and streptomycin (both at 100 U/ml) under 5% CO_2 at 37 °C. Cells were tested for mycoplasma contamination quarterly. The LC-MS based analysis for itaconate, I-CoA and AdoCbl were performed as previously reported (5).

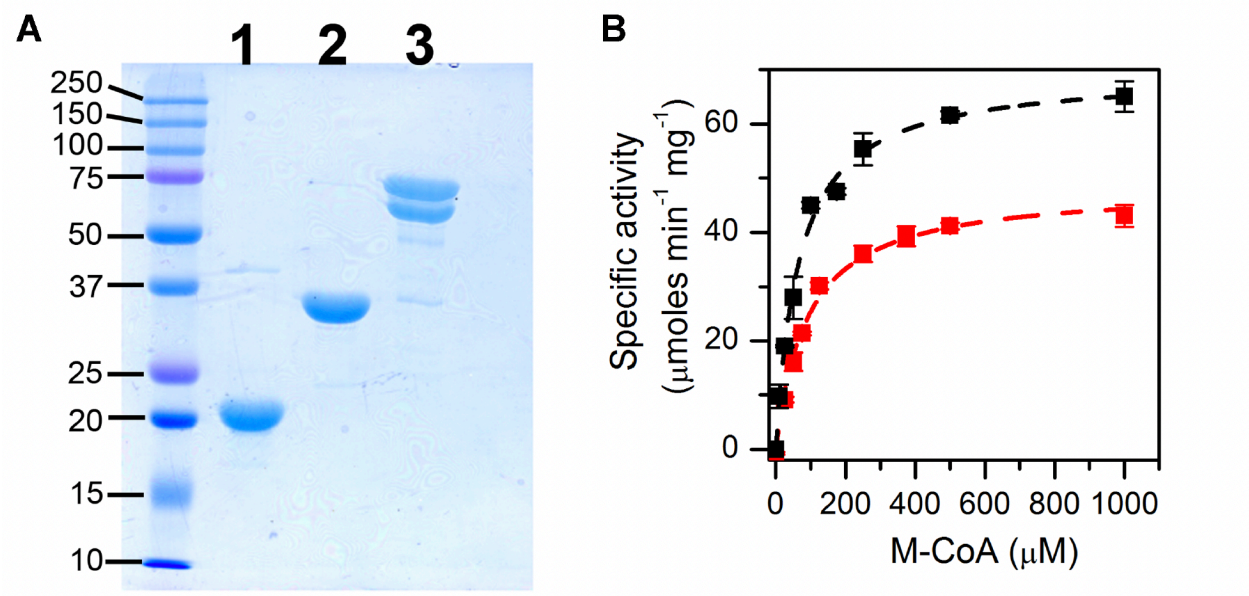
CRISPR KD RAW cells were generated by lentivirus infection as previously reported (5). Two sgRNAs per gene were cloned into pLentiCRISPRv2 plasmid (Addgene 52961). A lentiviral pool from the two guides was used for infection to achieve higher knockdown efficiency. The sequences of the sgRNAs are listed below:

Negative control: GCACTACCAGAGCTAACTCA

Irg1_sg1: TGAGCACGATTCTCTCGATG

Irg1_sg2: AAGCCAAAGACATACCAAAG

4.5 Supplementary Information

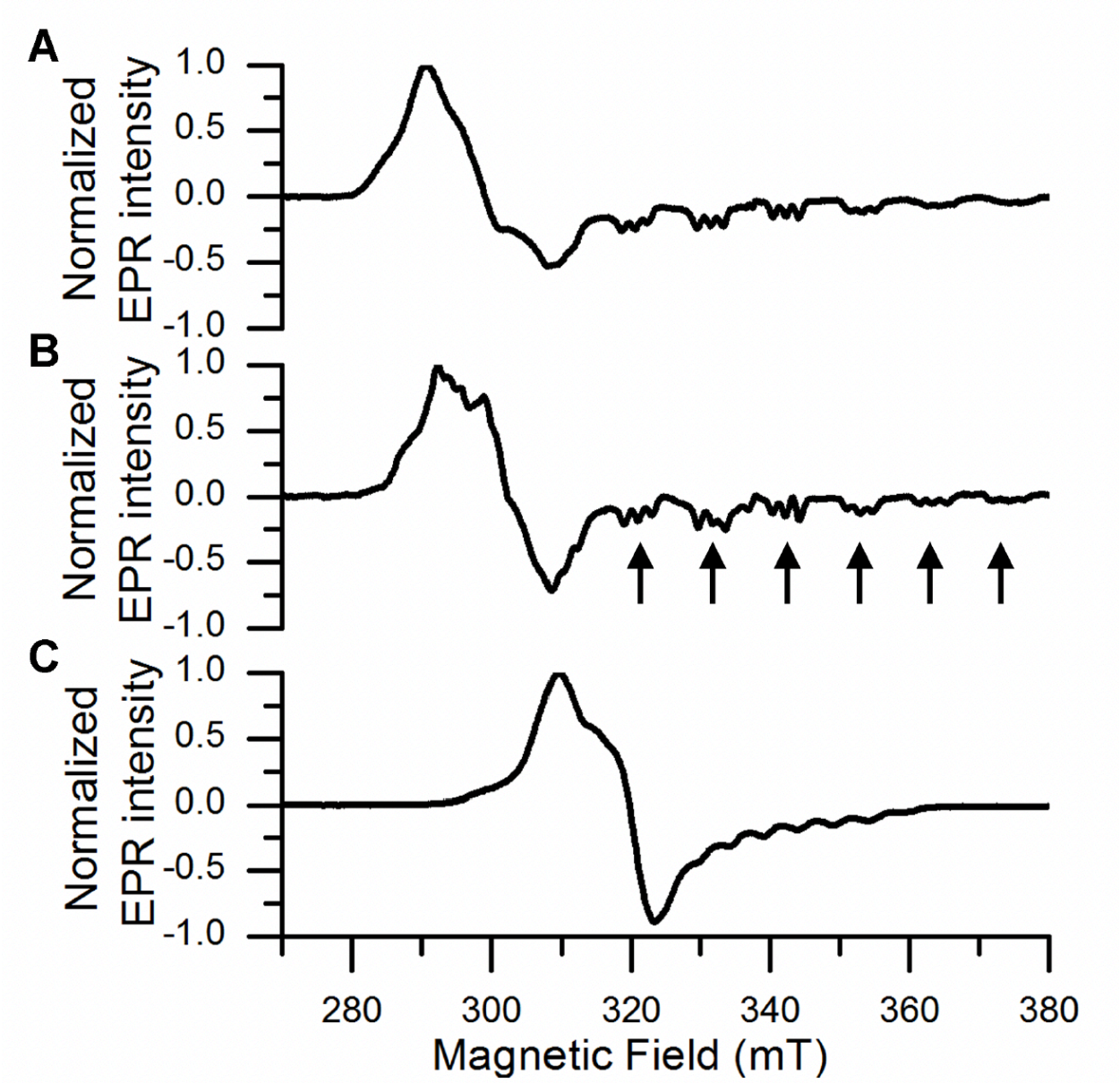


C

Enzyme	$K_{M(M\text{-}CoA)}$ (μM)	V_{max} (μmoles min^{-1} mg^{-1})
Human MCM	66.5 ± 6.1	69.4 ± 1.7
Mtb MCM	90.6 ± 7.5	48.4 ± 1.1

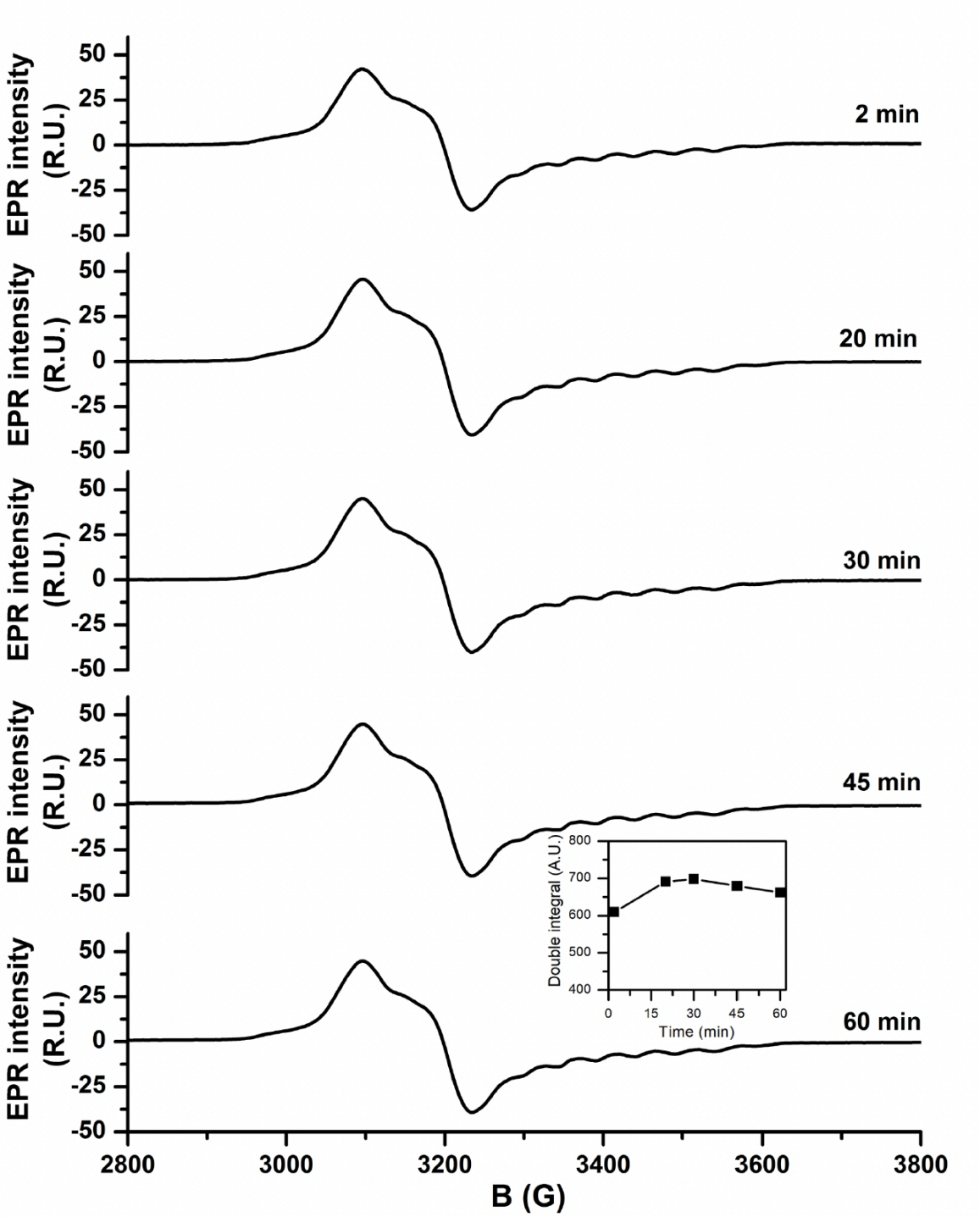
SI Figure 4.1: Characterization of Mtb MCM and its auxiliary proteins

(A) SDS-PAGE analysis demonstrating the purity of Mtb proteins: lane 1: ATR (27 kDa), lane 2: CblA (36 kDa), and land 3: MCM (α (66 kDa) and ß (80 kDa) subunits). **(B)** Michaelis-Menten kinetic analysis of hMCM (black trace) and Mtb MCM (red trace). MCM activity was determined in the thiokinase assay and performed in duplicate (hMCM) or triplicate (Mtb MCM); the data represent mean ± SD. **(C)** Summary of the kinetic parameters for human and Mtb MCM, which are homodimeric and heterodimeric, respectively.



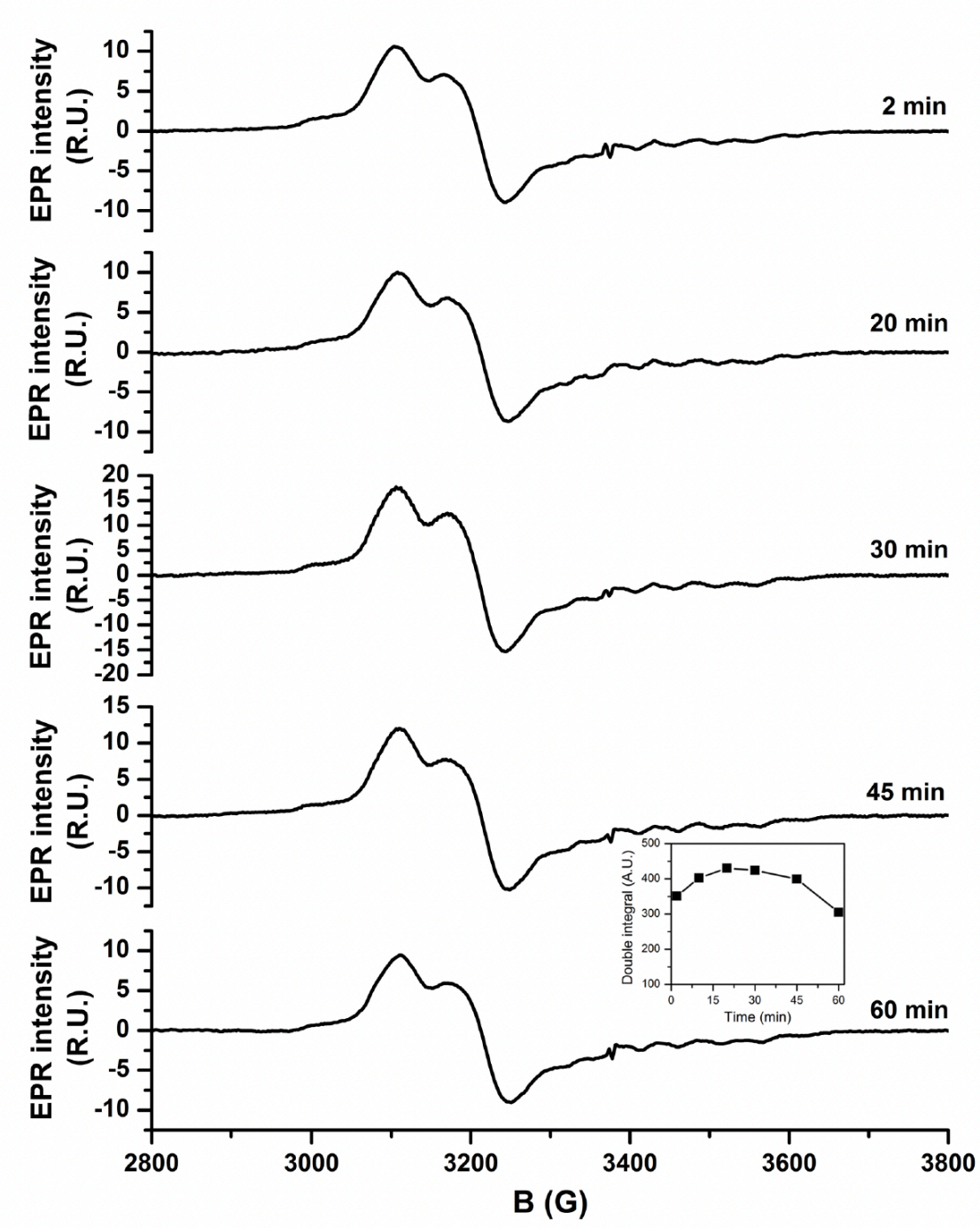
SI Figure 4.2: EPR spectra of Mtb MCM

(A) Spectrum of free 100 μM cob(II)alamin. **(B)** Spectrum of 100 μM cob(II)alamin bound to 200 μM Mtb MCM. **(C)** Spectrum of sample containing 100 μM AdoCbl bound to 200 μM Mtb MCM and incubated aerobically with 500 μM I-CoA for 30 min. All samples were prepared in 50 mM HEPES, pH 7.5, 150 mM KCl, 2 mM $MgCl_2$, 2 mM TCEP, and 10% glycerol. The arrows highlight the cobalt hyperfine line with additional triplet splitting due to 14N superhyperfine coupling.



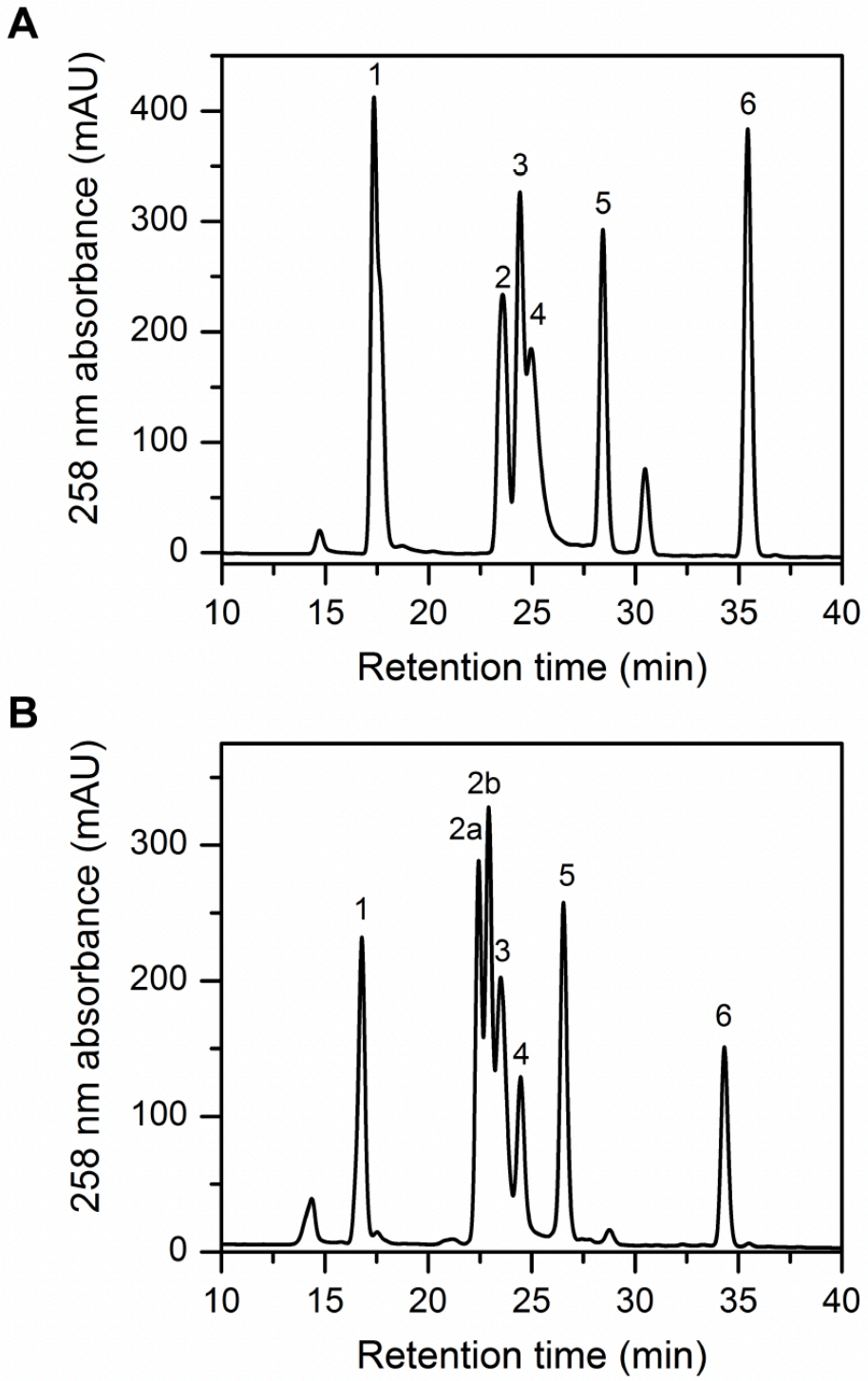
SI Figure 4.3: Aerobic stability of the I-CoA derived biradical on Mtb MCM

Mtb MCM (200 μM) was mixed with AdoCbl (100 μM) in 50 mM HEPES, pH 7.5, 150 mM KCl, 2 mM MgCl2, 2 mM TCEP, 10% glycerol. The biradical was formed upon addition of I-CoA (500 μM). Following 2 min incubation at room temperature, the sample was frozen in liquid N2 and an EPR spectrum was recorded. The sample was then quickly thawed and incubated for the indicated time at room temperature before being frozen in liquid N2 to record the EPR spectrum. The inset shows the time dependence of the double integral of the EPR signal over time.



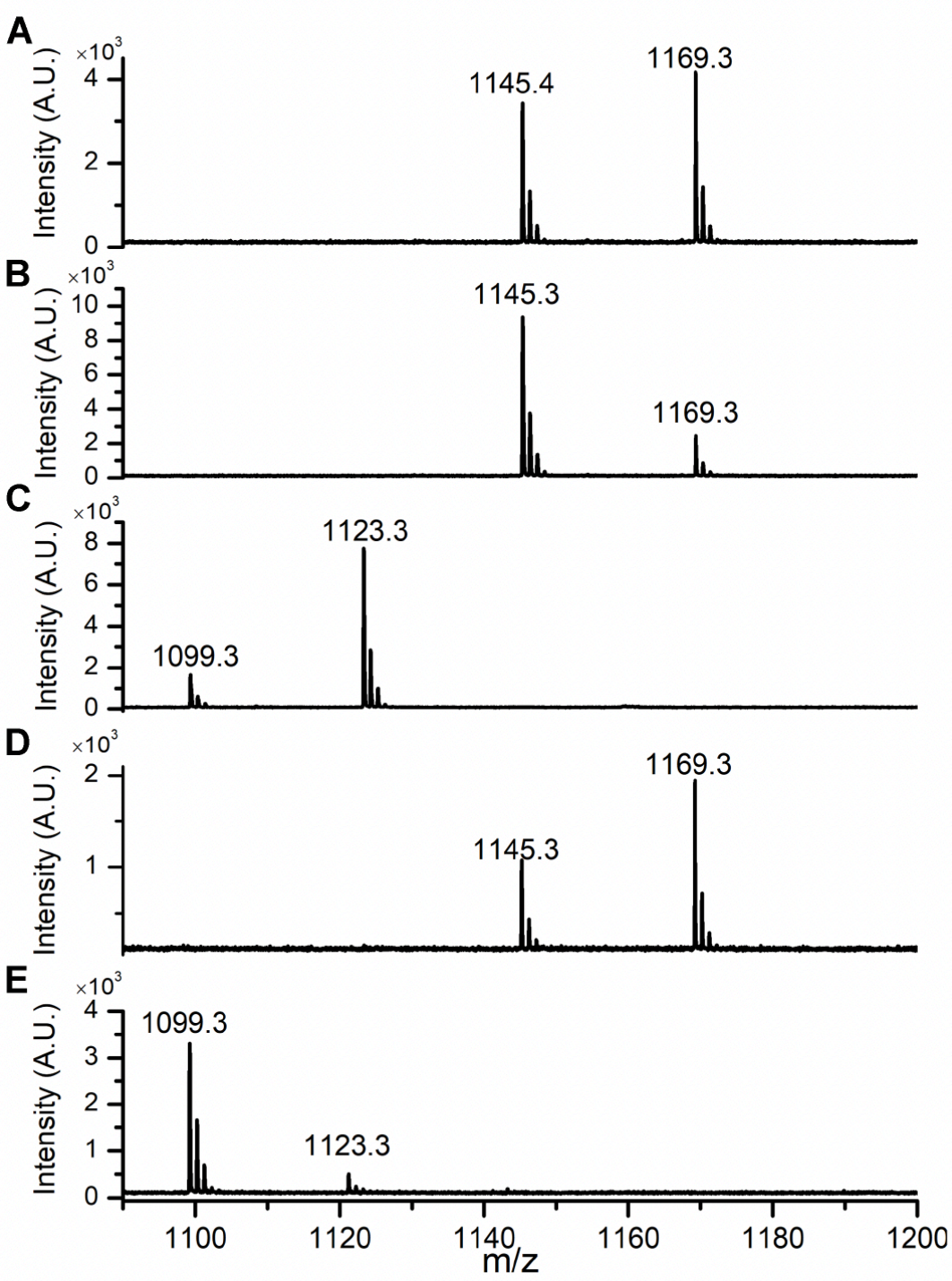
SI Figure 4.4: Aerobic stability of the I-CoA-derived biradical on hMCM

hMCM (100 μM) was mixed with AdoCbl (50 μM) in 50 mM HEPES, pH 7.5, 150 mM KCl, 2 mM MgCl2, 2 mM TCEP and 10% glycerol. The biradical was formed upon addition of I-CoA (500 μM). Following incubation for 2 min at room temperature, the sample was frozen in liquid N2 and the EPR spectrum was recorded. The sample was then quickly thawed and incubated at room temperature for the indicated time before being frozen in liquid N2 to record the EPR spectrum. The inset shows the time dependence of the double integral of the EPR signal over time.



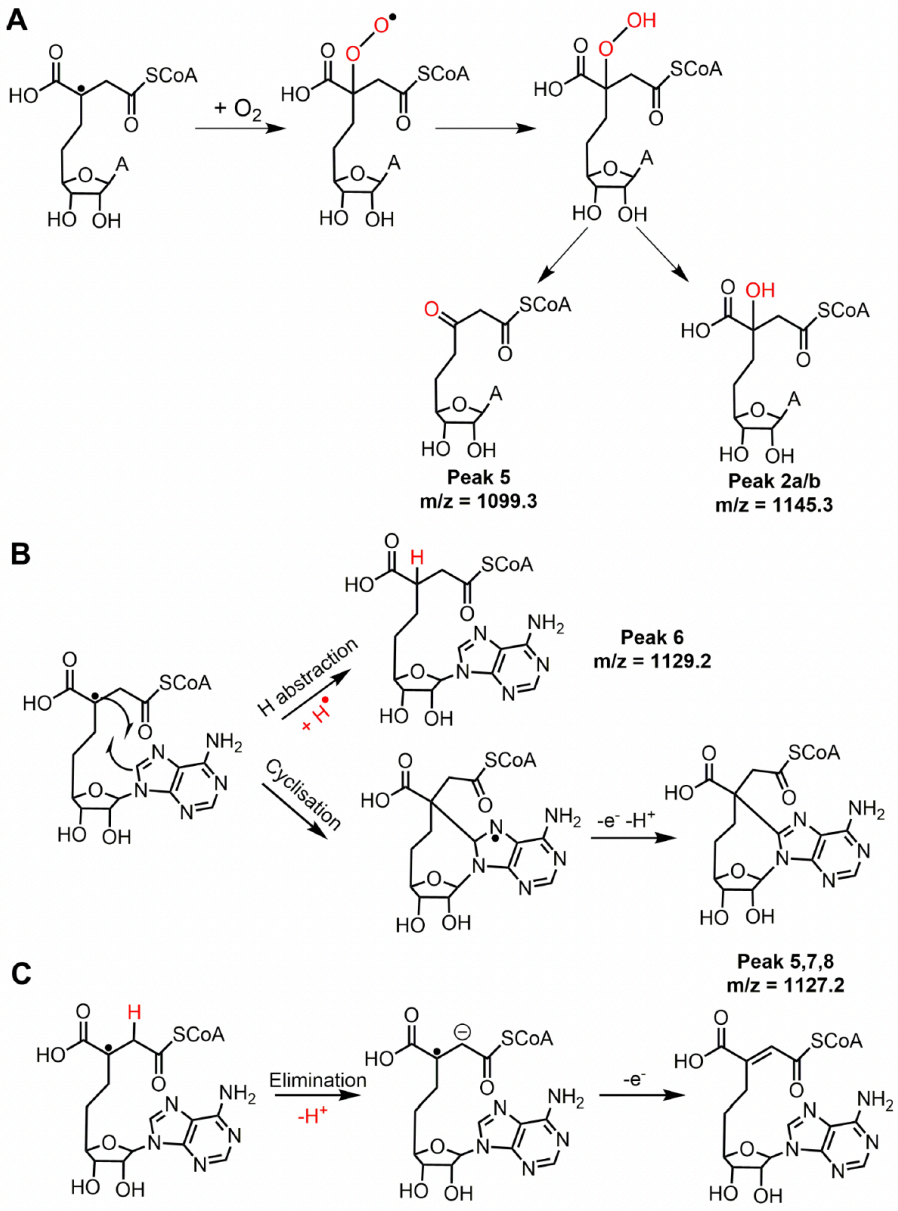
SI Figure 4.5: Chromatograms of aerobic reaction mixtures containing I-CoA-inactivated MCM

HPLC chromatogram of products formed during the reaction of I-CoA with **(A)** hMCM-AdoCbl or **(B)** Mtb MCM-AdoCbl under aerobic conditions. Products were identified by either negativeion MALDI/MS or co-migration with standards (OH_2Cbl and AdoCbl) with known retention times. The B_{12} peaks were also confirmed by their absorbance at 350 and 525 nm. The peaks were identified as follows: 1-CoA, 2a,b-hydroxylated adenosine adduct, 3-OH_2Cbl, 4-I-CoA, 5- decarboxylated adenosine adduct, 6-AdoCbl.



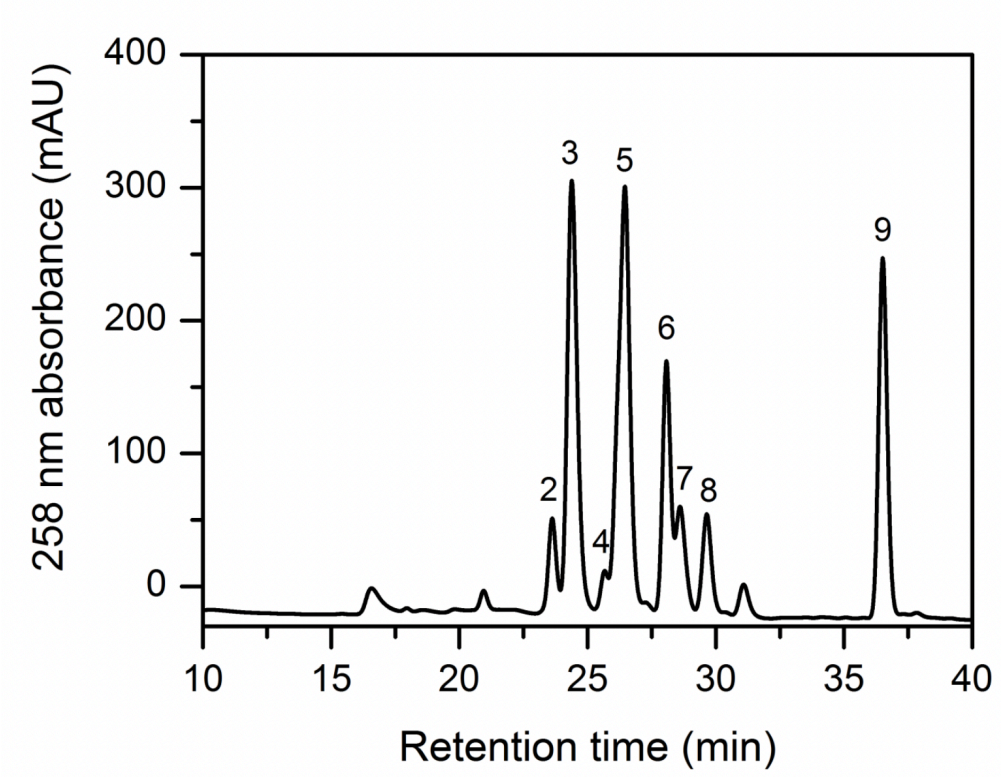
SI Figure 4.6: MALDI MS (negative ion) analysis of products from the aerobic reaction of MCM with I-CoA

The HPLC-purified products (SI Figure 4.5B) from fractions 2a **(A)**, 2b **(B)** and 5 **(C)**, were subjected to MS analysis after desalting. The identity of the +24 peaks (m/z = 1169.4, 1123.3) present in all the samples is not known.



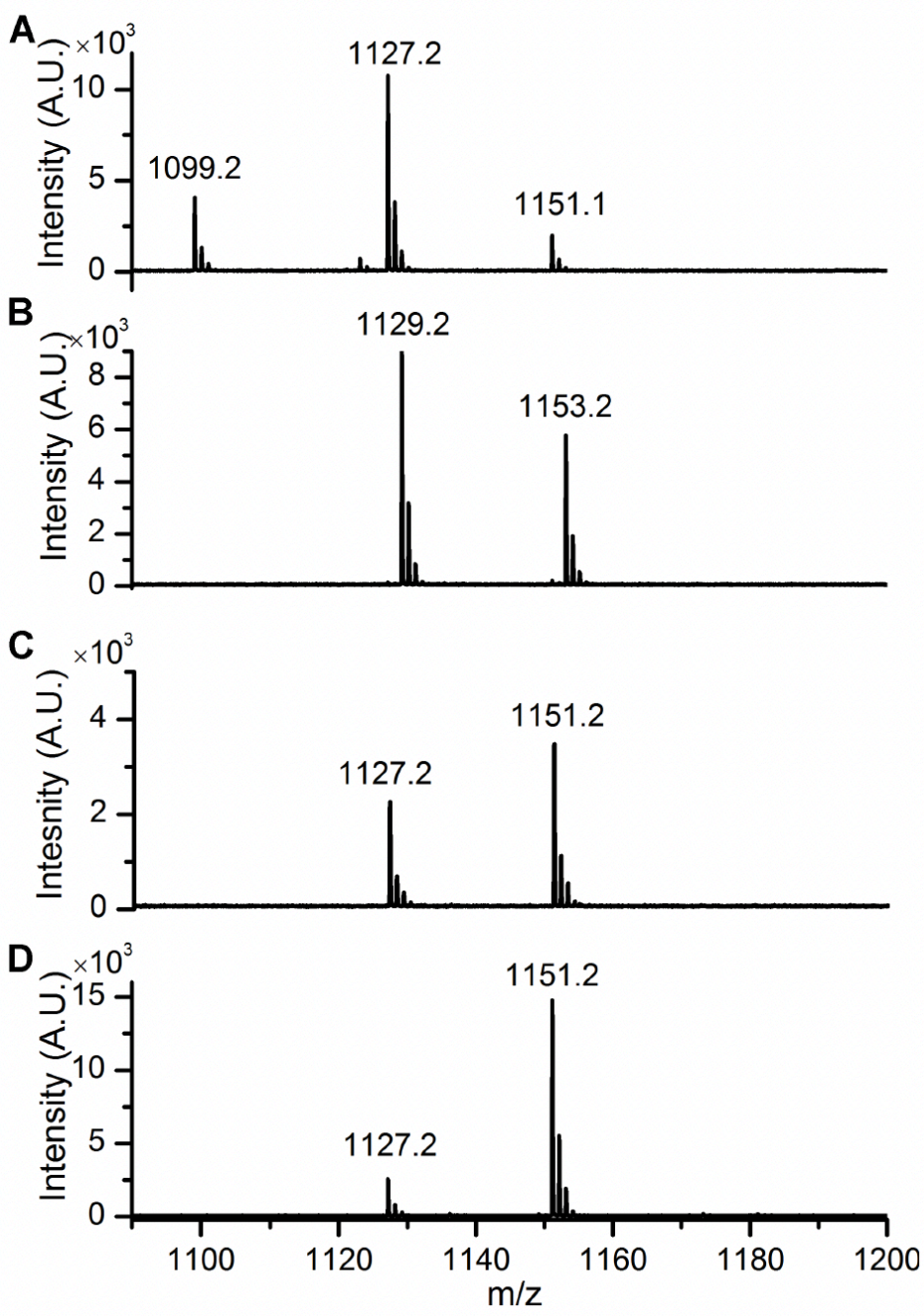
SI Figure 4.7: Proposed mechanism for the differential reactivity of Ado• under aerobic and anaerobic conditions

(A) Postulated mechanism for oxygen addition to the initial tertiary radical upon release into solution, forming a hydroperoxy radical intermediate, which degrades to the observed hydroxyl derivative. The second major product is proposed to form by oxidative decarboxylation of the α-hydroxy acid as described for model compounds (1,2). dAdo radical addition products have been reported previously (3,4). **(B)** Postulated mechanism for the intramolecular cyclization in the initially formed tertiary radical adduct and the 5′-dAdo moiety under anaerobic conditions. While the loss of two mass units is assigned to an intramolecular cyclization between the tertiary radical and the adenosine moiety derived from AdoCbl **(B)**, it could in principle, involve the 3′-phosphoadenosine group in I-CoA instead. **(C)** Postulated mechanism to account for Peaks 7 and 8 observed by HPLC (SI Figure 4.8) with m/z = 1127.2 (SI Figure 4.9). Proton loss could be facilitated by the higher acidity of the hydrogen atom adjacent to a radical center (than in a closed-shell counterpart (5,6)). Deprotonation at the methylene group adjacent to the thioester moiety followed by loss of an electron (e.g. to cob(II)alamin) would also lead to a product with two mass units less than the expected addition product.



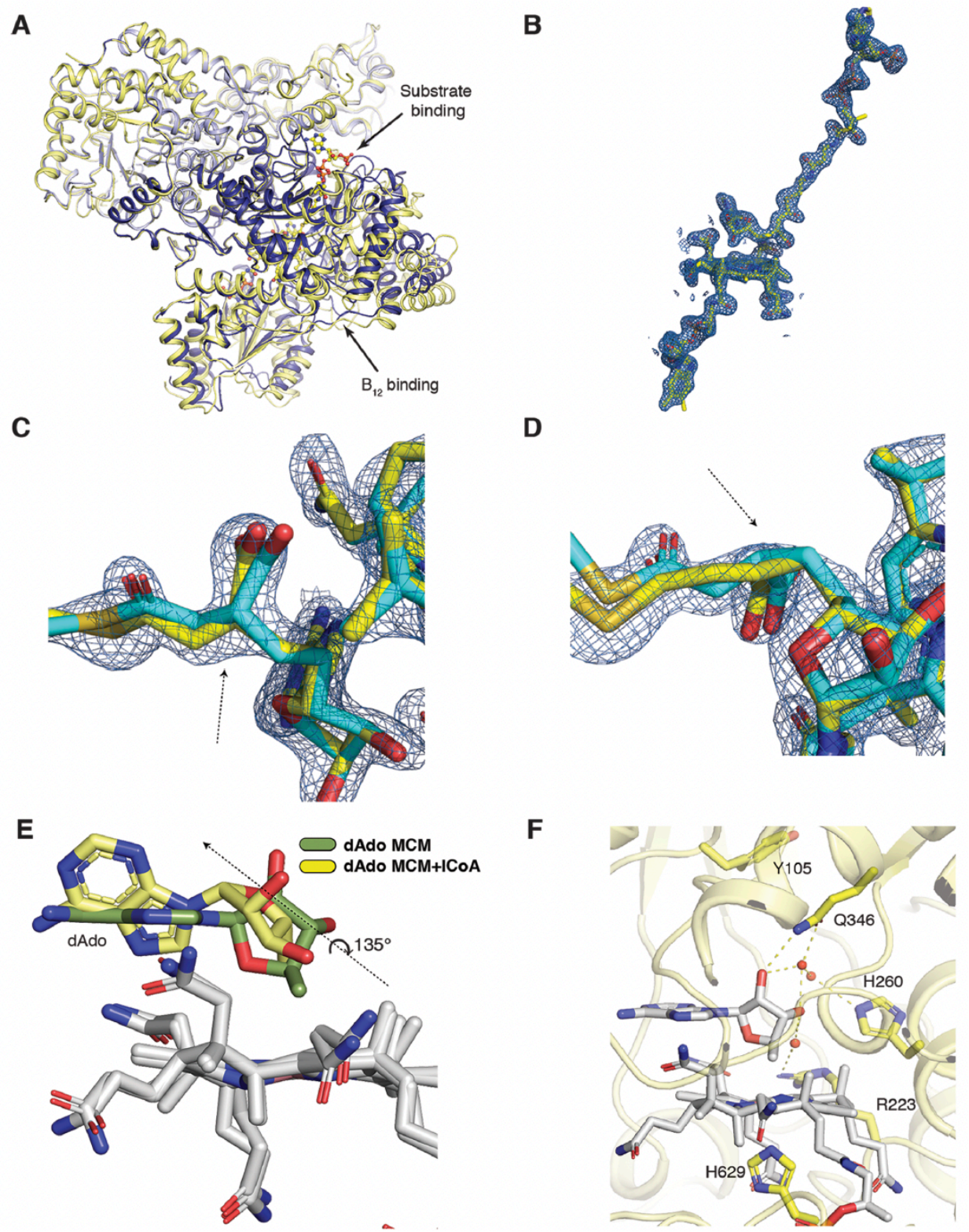
SI Figure 4.8: Chromatograms of anaerobic reaction mixtures containing I-CoA-inactivated Mtb MCM

HPLC chromatogram of products formed during the reaction of I-CoA with Mtb MCM under anaerobic conditions. Products were identified by either negative-ion MALDI/MS or by co-migration with standards (OH2Cbl and AdoCbl) with known retention times. The identified peaks are as follows: 2-hydroxylated-adenosine adduct, 3-OH2Cbl, 4-I-CoA, 5 cyclic adenosine adduct, 6-I-CoA-adenosine adduct, 7,8-cyclic adducts, 9-AdoCbl. The peaks are numbered starting from 2 to show their correspondence to the peaks obtained under aerobic conditions (SI Figure 4.5).



SI Figure 4.9: MALDI MS (negative ion) analysis of products from the anaerobic reaction of Mtb MCM with I-CoA

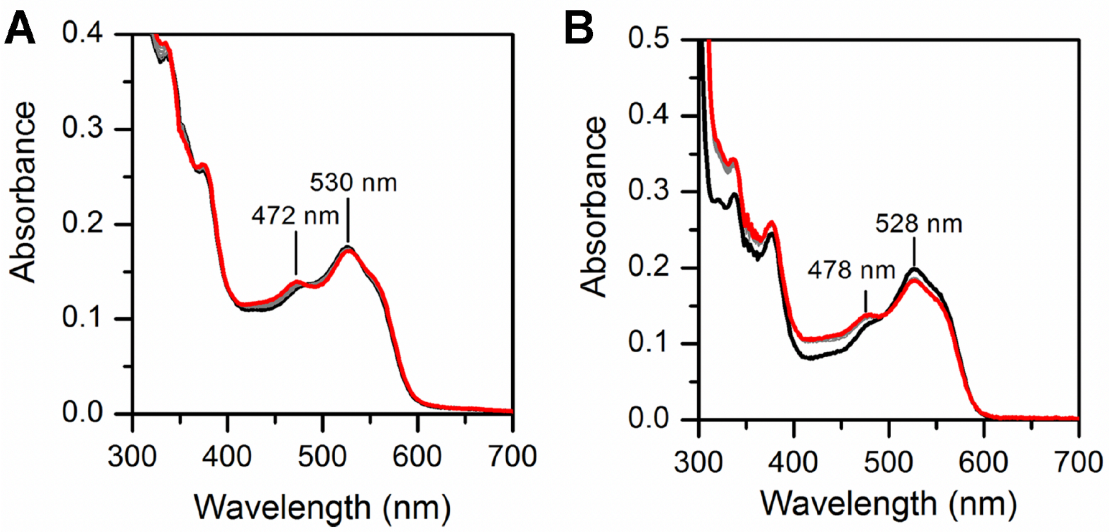
The HPLC-purified products (SI Figure 4.8) from fractions 5 **(A)**, 6 **(B)**, 7 **(C)** and 8 **(D)** were subjected to MS analysis.



SI Figure 4.10: Crystal structures of Mtb MCM with and without I-CoA

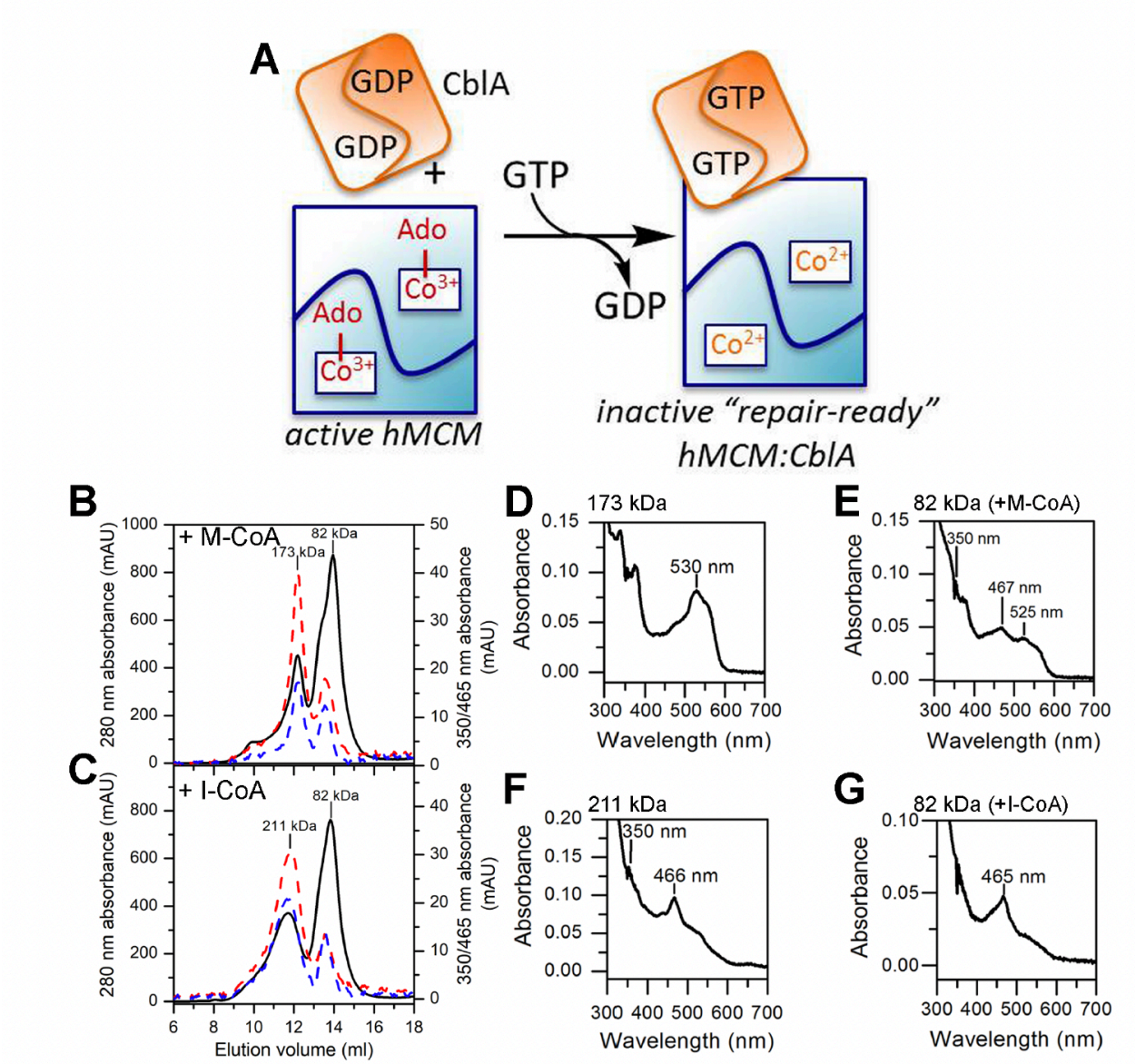
(A) Superposition of the I-CoA-bound (blue) and native Mtb MCM (yellow) structures. The inactive ß-subunit (shown in gray or light blue) shows virtually no conformational change upon binding I-CoA. The larger, B12-binding α-subunit

(blue for I-CoA bound, yellow for native) shows a substantial conformational change upon I-CoA binding. B_{12} and I-CoA are shown in stick representation with carbons in yellow, oxygen in red, nitrogen in blue, and phosphorous in orange. **(B)** 2Fo-Fc omit map showing electron density (blue) around B_{12} (yellow) and I-CoA (yellow) contoured at 1.5σ. **(C and D)** Two views of the fit of I-CoA ligands with an sp^2- (yellow) or an sp^3- (blue) hybridized carbon geometry refined with MCM. The resulting I-CoA models were overlaid on the composite omit map (contoured at 1.5 σ). The two views clearly demonstrate the superior fit of the I-CoA model with sp^2 geometry to the omit density map. **(E)** Superposition of B_{12} (gray for both structures) and dAdo in the native (green) and I-CoA bound (yellow) structures. The corrin rings were aligned for the superposition. **(F)** Interactions between dAdo and active site residues in the native structure. The 2′-OH of dAdo engages in hydrogen-bonding interactions with Q346 and a water-mediated hydrogen-bonding network with H260 and R223.



SI Figure 4.11: Spectra of MCM following catalytic turnover and upon addition of the repair system

Spectra of hMCM **(A)** or Mtb MCM **(B)** incubated with M-CoA for 60 min (black) to allow turnover, followed by the addition of the corresponding human and Mtb repair enzymes comprising CblA and ATR (gray spectra). The final spectra (red) were recorded after 20 min.



SI Figure 4.12: I-CoA impedes MCM repair by inhibiting CblA-MCM complex formation

(A) hMCM interacts with hCblA when cob(II)alamin but not AdoCbl is bound. **(B)** Following catalysis in the presence of M-CoA (+CblA and GDP), ATR (+GTP and ATP) addition promotes off-loading of cob(II)alamin from hMCM to ATR. The black, blue and red lines indicate absorbances at 280, 350 and 465 nm, respectively. The MWs of hMCM, CblA and ATR are 160 kDa, 77 kDa and 72 kDa, respectively. **(C)** I-CoA-inactivated hMCM exhibits a broad peak at 211 kDa. **(D and E)** Spectra of the 173 and 82 kDa peaks in (B) show AdoCbl bound to MCM (530 nm (D)), and a mixture of 4-coordinate cob(II)alamin (465 nm) and OH2Cbl (350, 525 nm) bound to ATR (E)). **(F and G)** Spectra of the 211 and 82 kDa peaks in (C) show cob(II)alamin bound to hMCM (466 nm) with a trace of OH2Cbl (350 nm, F) and 4-coordinate cob(II)alamin (465 nm) bound to ATR (G). Analysis of the 82 kDa peak (C) thus reveals that the repair system transferred some cob(II)alamin from I-CoA-inactivated hMCM to ATR (465 nm).

SI Table 4.1: X-Ray Crystallography Data Collection and Refinement Statistics

	***Mtb* MCM**	***Mtb* MCM + I-CoA**
Data collection	(208)	(2111)
Beamline	APS, GMCA 23-IDB	APS, GMCA 23-IDD
Wavelength (Å)	1.0332	1.0332
Temperature (K)	100	100
Resolution (Å)	29.1-1.9 (1.93-1.90)	29.8-2.0 (2.03-2.00)
Space group	$P2_12_12_1$	$P2_12_12_1$
Cell dimensions (Å)	a = 77.8, b = 103.6, c = 211.2	a = 76.6, b = 104.9, c = 194.1
Cell dimensions (°)	α = β = γ = 90	α = β = γ = 90
Observed reflections	537,061 (24,946)	476,084 (23,408)
Unique reflections	129,672 (6,274)	104,126 (5,172)
R_{meas} (%)	16.0 (110.4)	12.7 (124.3)
R_{merge} (%)	12.6 (88.7)	10.3 (99.7)
R_{pim} (%)	9.7 (64.9)	7.5 (73.2)
<I/σ>	7.9 (1.7)	8.9 (1.3)
CC(1/2)	0.993 (0.636)	0.996 (0.622)
Multiplicity	4.1 (4.0)	4.6 (4.5)
Completeness (%)	96.5 (95.4)	98.4 (99.8)
Overall *B* (Å^2) (Wilson plot)	24.6	28.3
Refinement		
Resolution range	29.0-1.9	29.8-2.0
Number of reflections (work/test set)	129,574/6,579	98,704/5,352
R_{work}/R_{free} (%)	16.2/20.0	16.5/20.6
No. of non-H atoms		
Protein	19,823	19,874
Water	1,390	1,127
Ligand: B_{12}	178	178
Ado	31	31
I-CoA	-	56
B-factors (Å^2)		
Protein	29.4	35.0
Water	35.5	37.8
Ligand: B_{12}	23.4	28.3
Ado	37.4	25.6
I-CoA	-	29.9
Rmsd deviations		
Bond lengths (Å)	0.011	0.008
Bond angles (°)	1.254	1.623
Ramachandran plot		
Favored/allowed/outliers	98.3/1.7/0.0	98.3/1.4/0.1
MolProbity Score	0.99 (100th percentile)	1.2 (100th percentile)
PDB	6OXC	6OXD

4.6 References

1. Michelucci A, Cordes T, Ghelfi J, Pailot A, Reiling N, Goldmann O, et al. Immune-responsive gene 1 protein links metabolism to immunity by catalyzing itaconic acid production. Proc Natl Acad Sci U S A. 2013 May 7;110(19):7820–5.

2. O'Neill LAJ, Artyomov MN. Itaconate: the poster child of metabolic reprogramming in macrophage function. Nat Rev Immunol. 2019 May;19(5):273–81.

3. Kumar R. Glyoxylate Shunt: Combating Mycobacterium at forefront. Int J Integr Biol. 2009;7(2):69–72.

4. McKinney JD, Höner zu Bentrup K, Muñoz-Elías EJ, Miczak A, Chen B, Chan WT, et al. Persistence of Mycobacterium tuberculosis in macrophages and mice requires the glyoxylate shunt enzyme isocitrate lyase. Nature. 2000 Aug 17;406(6797):735–8.

5. Shen H, Campanello GC, Flicker D, Grabarek Z, Hu J, Luo C, et al. The Human Knockout Gene CLYBL Connects Itaconate to Vitamin B12. Cell. 2017 Nov 2;171(4):771-782.e11.

6. Sasikaran J, Ziemski M, Zadora PK, Fleig A, Berg IA. Bacterial itaconate degradation promotes pathogenicity. Nat Chem Biol. 2014 May;10(5):371–7.

7. Grarup N, Sulem P, Sandholt CH, Thorleifsson G, Ahluwalia TS, Steinthorsdottir V, et al. Genetic architecture of vitamin B12 and folate levels uncovered applying deeply sequenced large datasets. PLoS Genet. 2013 Jun;9(6):e1003530.

8. Lek M, Karczewski KJ, Minikel EV, Samocha KE, Banks E, Fennell T, et al. Analysis of protein-coding genetic variation in 60,706 humans. Nature. 2016 Aug 18;536(7616):285–91.

9. Padovani D, Banerjee R. A G-protein editor gates coenzyme B12 loading and is corrupted in methylmalonic aciduria. Proc Natl Acad Sci U S A. 2009 Dec 22;106(51):21567–72.

10. Mills EL, Ryan DG, Prag HA, Dikovskaya D, Menon D, Zaslona Z, et al. Itaconate is an anti-inflammatory metabolite that activates Nrf2 via alkylation of KEAP1. Nature. 2018 Apr 5;556(7699):113–7.

11. Bambouskova M, Gorvel L, Lampropoulou V, Sergushichev A, Loginicheva E, Johnson K, et al. Electrophilic properties of itaconate and derivatives regulate the IκBζ-ATF3 inflammatory axis. Nature. 2018 Apr;556(7702):501–4.

12. Mansoorabadi SO, Padmakumar R, Fazliddinova N, Vlasie M, Banerjee R, Reed GH. Characterization of a succinyl-CoA radical-cob(II)alamin spin triplet intermediate in the reaction catalyzed by adenosylcobalamin-dependent methylmalonyl-CoA mutase. Biochemistry. 2005 Mar 8;44(9):3153–8.

13. Bothe H, Darley DJ, Albracht SP, Gerfen GJ, Golding BT, Buckel W. Identification of the 4-glutamyl radical as an intermediate in the carbon skeleton rearrangement catalyzed by

coenzyme B12-dependent glutamate mutase from Clostridium cochlearium. Biochemistry. 1998 Mar 24;37(12):4105–13.

14. Buckel W, Pierik AJ, Plett S, Alhapel A, Suarez D, Tu S min, et al. Mechanism-Based Inactivation of Coenzyme B12-Dependent 2-Methyleneglutarate Mutase by (Z)-Glutaconate and Buta-1,3-diene-2,3-dicarboxylate. European Journal of Inorganic Chemistry. 2006;2006(18):3622–6.

15. Banerjee R. Radical peregrinations catalyzed by coenzyme B12-dependent enzymes. Biochemistry. 2001 May 29;40(21):6191–8.

16. Yang H, McDaniel EC, Impano S, Byer AS, Jodts RJ, Yokoyama K, et al. The Elusive 5'-Deoxyadenosyl Radical: Captured and Characterized by Electron Paramagnetic Resonance and Electron Nuclear Double Resonance Spectroscopies. J Am Chem Soc. 2019 Jul 31;141(30):12139–46.

17. Hogenkamp HPC. A Cyclic Nucleoside Derived from Coenzyme B12. Journal of Biological Chemistry. 1963 Jan 1;238(1):477–80.

18. Jaruga P, Dizdaroglu M. 8,5'-Cyclopurine-2'-deoxynucleosides in DNA: mechanisms of formation, measurement, repair and biological effects. DNA Repair (Amst). 2008 Sep 1;7(9):1413–25.

19. Mancia F, Keep NH, Nakagawa A, Leadlay PF, McSweeney S, Rasmussen B, et al. How coenzyme B12 radicals are generated: the crystal structure of methylmalonyl-coenzyme A mutase at 2 A resolution. Structure. 1996 Mar 15;4(3):339–50.

20. Froese DS, Kochan G, Muniz JRC, Wu X, Gileadi C, Ugochukwu E, et al. Structures of the human GTPase MMAA and vitamin B12-dependent methylmalonyl-CoA mutase and insight into their complex formation. J Biol Chem. 2010 Dec 3;285(49):38204–13.

21. Kwiecien RA, Khavrutskii IV, Musaev DG, Morokuma K, Banerjee R, Paneth P. Computational insights into the mechanism of radical generation in B12-dependent methylmalonyl-CoA mutase. J Am Chem Soc. 2006 Feb 1;128(4):1287–92.

22. Chabrière E, Vernède X, Guigliarelli B, Charon MH, Hatchikian EC, Fontecilla-Camps JC. Crystal structure of the free radical intermediate of pyruvate:ferredoxin oxidoreductase. Science. 2001 Dec 21;294(5551):2559–63.

23. Gruber K, Reitzer R, Kratky C. Radical Shuttling in a Protein: Ribose Pseudorotation Controls Alkyl-Radical Transfer in the Coenzyme B12 Dependent Enzyme Glutamate Mutase. Angewandte Chemie International Edition. 2001;40(18):3377–80.

24. Masuda J, Shibata N, Morimoto Y, Toraya T, Yasuoka N. How a protein generates a catalytic radical from coenzyme B(12): X-ray structure of a diol-dehydratase-adeninylpentylcobalamin complex. Structure. 2000 Jul 15;8(7):775–88.

25. Campanello GC, Ruetz M, Dodge GJ, Gouda H, Gupta A, Twahir UT, et al. Sacrificial Cobalt-Carbon Bond Homolysis in Coenzyme B12 as a Cofactor Conservation Strategy. J Am Chem Soc. 2018 Oct 17;140(41):13205–8.

26. Padovani D, Banerjee R. Assembly and protection of the radical enzyme, methylmalonyl-CoA mutase, by its chaperone. Biochemistry. 2006 Aug 1;45(30):9300–6.

27. Ruetz M, Campanello GC, McDevitt L, Yokom AL, Yadav PK, Watkins D, et al. Allosteric Regulation of Oligomerization by a B12 Trafficking G-Protein Is Corrupted in Methylmalonic Aciduria. Cell Chem Biol. 2019 Jul 18;26(7):960-969.e4.

28. Padovani D, Labunska T, Palfey BA, Ballou DP, Banerjee R. Adenosyltransferase tailors and delivers coenzyme B12. Nat Chem Biol. 2008 Mar;4(3):194–6.

29. Savvi S, Warner DF, Kana BD, McKinney JD, Mizrahi V, Dawes SS. Functional characterization of a vitamin B12-dependent methylmalonyl pathway in Mycobacterium tuberculosis: implications for propionate metabolism during growth on fatty acids. J Bacteriol. 2008 Jun;190(11):3886–95.

30. Nair S, Huynh JP, Lampropoulou V, Loginicheva E, Esaulova E, Gounder AP, et al. Irg1 expression in myeloid cells prevents immunopathology during M. tuberculosis infection. J Exp Med. 2018 Apr 2;215(4):1035–45.

31. Shin JH, Yang JY, Jeon BY, Yoon YJ, Cho SN, Kang YH, et al. (1)H NMR-based metabolomic profiling in mice infected with Mycobacterium tuberculosis. J Proteome Res. 2011 May 6;10(5):2238–47.

32. Jha AK, Huang SCC, Sergushichev A, Lampropoulou V, Ivanova Y, Loginicheva E, et al. Network integration of parallel metabolic and transcriptional data reveals metabolic modules that regulate macrophage polarization. Immunity. 2015 Mar 17;42(3):419–30.

33. Gopinath K, Moosa A, Mizrahi V, Warner DF. Vitamin B(12) metabolism in Mycobacterium tuberculosis. Future Microbiol. 2013 Nov;8(11):1405–18.

34. Gopinath K, Venclovas C, Ioerger TR, Sacchettini JC, McKinney JD, Mizrahi V, et al. A vitamin B_{12} transporter in Mycobacterium tuberculosis. Open Biol. 2013 Feb 13;3(2):120175.

35. BARRON M. PERNICIOUS ANEMIA AND TUBERCULOSIS: IS THERE AN ANTAGONISM?: A REVIEW OF THE OCCURRENCE OF TUBERCULOSIS IN NINETY-THREE CASES OF PERNICIOUS ANEMIA AS FOUND IN 16,600 POSTMORTEM EXAMINATIONS, WITH A REPORT OF TWO CLINICAL CASES. Journal of the American Medical Association. 1933 May 20;100(20):1590–2.

36. Taoka S, Padmakumar R, Lai MT, Liu HW, Banerjee R. Inhibition of the human methylmalonyl-CoA mutase by various CoA-esters. Journal of Biological Chemistry. 1994 Dec 16;269(50):31630–4.

37. Ellman GL. Tissue sulfhydryl groups. Arch Biochem Biophys. 1959 May;82(1):70–7.

38. Stoll S, Schweiger A. EasySpin, a comprehensive software package for spectral simulation and analysis in EPR. J Magn Reson. 2006 Jan;178(1):42–55.

39. Kuzmic P. Program DYNAFIT for the analysis of enzyme kinetic data: application to HIV proteinase. Anal Biochem. 1996 Jun 1;237(2):260–73.

40. Kabsch W. XDS. Acta Crystallogr D Biol Crystallogr. 2010 Feb 1;66(Pt 2):125–32.

41. Mancia F, Evans PR. Conformational changes on substrate binding to methylmalonyl CoA mutase and new insights into the free radical mechanism. Structure. 1998 Jun 15;6(6):711–20.

42. McCoy AJ, Grosse-Kunstleve RW, Adams PD, Winn MD, Storoni LC, Read RJ. Phaser crystallographic software. J Appl Cryst. 2007 Aug 1;40(4):658–74.

43. Murshudov GN, Skubák P, Lebedev AA, Pannu NS, Steiner RA, Nicholls RA, et al. REFMAC5 for the refinement of macromolecular crystal structures. Acta Crystallogr D Biol Crystallogr. 2011 Apr;67(Pt 4):355–67.

44. Adams PD, Grosse-Kunstleve RW, Hung LW, Ioerger TR, McCoy AJ, Moriarty NW, et al. PHENIX: building new software for automated crystallographic structure determination. Acta Crystallogr D Biol Crystallogr. 2002 Nov;58(Pt 11):1948–54.

45. Emsley P, Lohkamp B, Scott WG, Cowtan K. Features and development of Coot. Acta Crystallogr D Biol Crystallogr. 2010 Apr;66(Pt 4):486–501.

46. Moriarty NW, Grosse-Kunstleve RW, Adams PD. electronic Ligand Builder and Optimization Workbench (eLBOW): a tool for ligand coordinate and restraint generation. Acta Crystallogr D Biol Crystallogr. 2009 Oct;65(Pt 10):1074–80.

47. Schüttelkopf AW, van Aalten DMF. PRODRG: a tool for high-throughput crystallography of protein-ligand complexes. Acta Crystallogr D Biol Crystallogr. 2004 Aug;60(Pt 8):1355–63.

48. Moriarty NW, Draizen EJ, Adams PD. An editor for the generation and customization of geometry restraints. Acta Crystallogr D Struct Biol. 2017 Feb 1;73(Pt 2):123–30.

49. Terwilliger TC, Adams PD, Read RJ, McCoy AJ, Moriarty NW, Grosse-Kunstleve RW, et al. Decision-making in structure solution using Bayesian estimates of map quality: the PHENIX AutoSol wizard. Acta Crystallogr D Biol Crystallogr. 2009 Jun;65(Pt 6):582–601.

50. Chen VB, Arendall WB, Headd JJ, Keedy DA, Immormino RM, Kapral GJ, et al. MolProbity: all-atom structure validation for macromolecular crystallography. Acta Crystallogr D Biol Crystallogr. 2010 Jan;66(Pt 1):12–21.

51. Schrödinger, LLC. The PyMOL Molecular Graphics System, Version 1.8. 2015.

52. Andreu N, Zelmer A, Fletcher T, Elkington PT, Ward TH, Ripoll J, et al. Optimisation of Bioluminescent Reporters for Use with Mycobacteria. PLOS ONE. 2010 May 24;5(5):e10777.

53. Allen BW. Mycobacteria. General culture methodology and safety considerations. Methods Mol Biol. 1998;101:15–30.

CHAPTER 5 Conclusions and Future Directions

5.1 Conclusions

Everything can be described by the structure-function relationship, where the basic units of a system define its function. However, this deceivingly simple idea is foundational to understanding natural processes from the atomic level to the scale of whole populations. Indeed, the interrogation of structure-function relationships in biological systems is a powerful approach to determine the molecular-level mechanisms that give rise to complex life. Such studies represent the backbone of scientific development, as advancements in biotechnology and biomedicine rely on the foundational principles established by basic research. Thus, this work integrates structural analyses with biochemical characterization to establish the structural context and molecular basis of remarkable chemistries performed by three unique enzymes: Pseudouridine Synthase 7 (Pus7), Methionine Synthase (MS), and Methylmalonyl-CoA Mutase (MCM). The findings reported here serve to advance the understanding of RNA modification machinery, and the molecular dynamics and control of B_{12}- dependent catalysis.

Pseudouridine and pseudouridine synthases are emerging as key players in cellular processes and adapting to rapidly changing environments. Understanding the determinants of Ψ formation, occurrence, and distribution is critical to be able to deconvolute the role of this seemingly simple modification in the cell. Thus, Chapter 2 centers on the determinants of substrate

selection in Pus7, a pseudouridine synthase known for its multi-site specificity and highly active towards mRNA. The structure of *Saccharomyces cerevisiae* Pus7 is one of the first structures depicting a full length eukaryotic pseudouridine synthase. This structure provides invaluable insight into the architecture and function of the unique eukaryotic insertions and informs on how these insertions could contribute to substrate selectivity and fine-tuning Pus7 activity. Further, we investigated the contributions of substrate structure and the consensus sequence to Pus7 modification. Pus7 is remarkably promiscuous and modified every tRNA, mRNA, and non-natural RNA substrate that contained the consensus sequence, and overall, Pus7 is relatively unaffected by secondary structure around the site of modification.

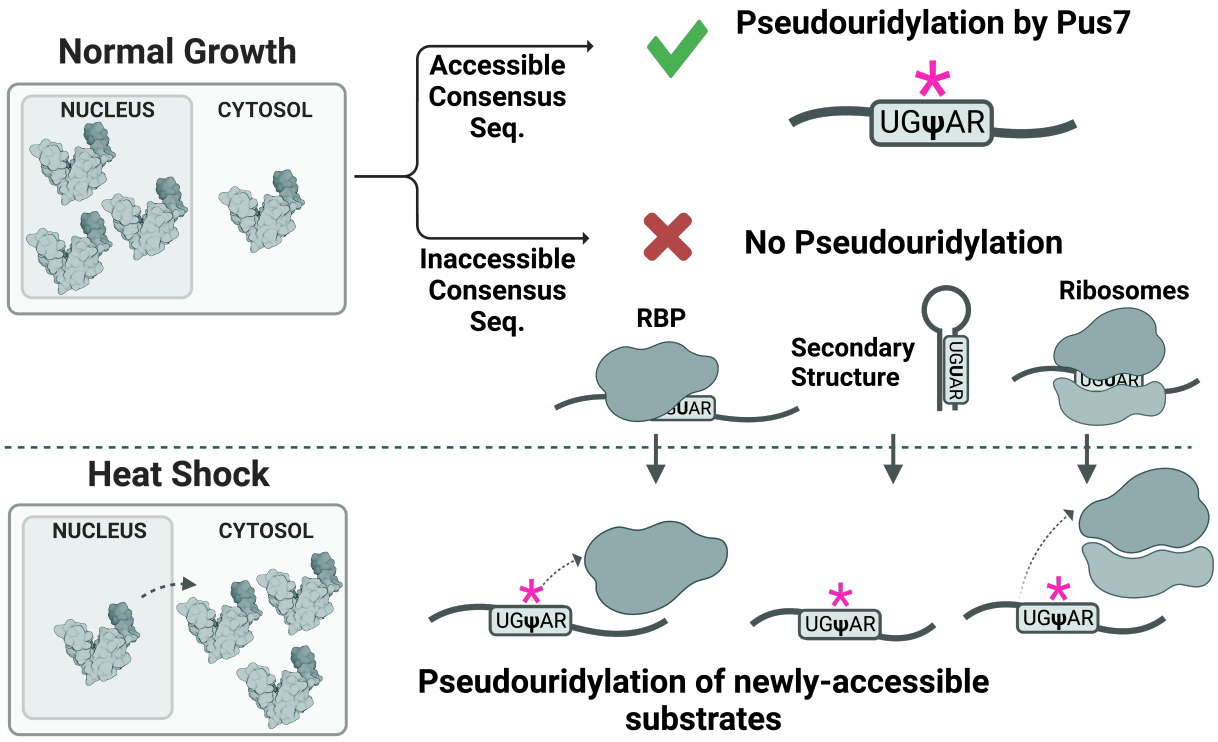


Figure 5.1: Opportunistic model of Pus7 selectivity

(top) Under normal growth conditions, Pus7 is located in the nucleus and will bind and modify RNA substrates with the consensus sequence available. Many sites that could be modified are inaccessible to Pus7 due to secondary structure, ribosomes, or RNA binding proteins. **(bottom)** Under heat shock conditions, Pus7 relocalizes from the nucleus to the cytoplasm and modifies many additional sites that are now accessible.

These findings, in combination with increased Pus7 activity and redistribution under heat-shock conditions, lead us to propose an opportunistic model of Pus7 selectivity wherein selectivity is not solely governed by enzymatic factors inherent to Pus7 itself (Figure 5.1). Rather, Pus7 selectivity is shaped by the accessibility of substrates to modify and affected by factors including cellular localization, RNA structure, and competition with other RNA binding proteins. In total, these findings support a role for Pus7 in providing cells with a mechanism to rapidly alter protein synthesis in response to cellular conditions.

In contrast to Pus7, which does not require a cofactor for catalysis, Chapter 3 and Chapter 4 focus on MS and MCM, respectively, two B_{12}-dependent enzymes that harness the powerful reactivity of cobalt to support essential metabolic processes in mammals. Progress in the biochemical and structural characterization of MS has been slow over the years; stymied by the difficulties associated with the expression and manipulation of the enzyme. The work presented here offers a solution to these longstanding barriers by identifying *Thermus thermophilus* MS (*t*MS), which is highly amenable to protein engineering and biochemical manipulation. Using this model system, we have solved the first full-length structure of MS, visualizing the relative orientation of each domain in 3D space, and providing insight into the mechanism of B_{12} incorporation. Furthermore, this structure demonstrates the validity of the 'divide-and-conquer' approach to structural characterization of MS, as excised structures were highly similar to the same domains in the full-length context. Additionally, we captured a catalytic orientation of MS (Fol-on, Fol:Cob) by taking advantage of the unique ability of *t*MS to incorporate non-natural cofactor analogs. The active site of the folate domain is positioned within reaction distance to the cobalt center, confirmed by modeling MTF into the active site. Thus, this structure may represent the first catalytically competent orientation captured in MS. Finally, to explore the basis of control over

each MS orientation we analyzed the domain interfaces in available multi-domain MS structures, in comparison to our Full-length and Fol-on catalytic structures. We propose that the cofactor serves as the primary mediator of domain interactions, rather than other protein:protein interfaces, which allows for each domain to be stabilized above the cobalamin-binding domain at the right time in the catalytic cycle.

Finally, the work done with MCM in Chapter 4 is a classic example of leveraging the structure-function relationship to discern mechanistic details about enzyme catalysis. MCM binds 5'deoxyadenosyl-cobalamin which undergoes homolytic radical cleavage to catalyze the interconversion between succinyl-CoA and methylmalonyl-CoA. It was unknown, however, how MCM controlled the radical cleavage products to direct catalysis and to prevent the radical from unwanted side-reactions. Recently, Itaconyl-CoA (I-CoA), a close analog of succinyl-CoA, was identified to bind to MCM and act as a potent inhibitor and, as we reasoned, a chemical probe to investigate MCM radical control. Indeed, the co-crystal structure of I-CoA bound MCM demonstrated the suicide addition of the 5'-deoxyadenosyl radical to the double bond of I-CoA, resulting in and air-stable bi-radical pair and an inactivated enzyme. Visualizing the rotation of the 5'deoxyadenosyl moiety demonstrated the mechanism of I-CoA inactivation and suggested the basis of control for radical trajectories to ensure the fidelity of catalysis.

In summary, this exploration of structure-function relationships in RNA-modifying Pus7 and B_{12}-dependent MS and MCM has yielded important insights into the molecular basis of their enzymatic activities and contributes to our understanding of how they might affect biological processes.

5.2 Future Directions

5.2.1 Pseudouridine synthases

This work integrates structural and biochemical analyses to establish the basis of Pus7 selectivity and provides a foundation for further interrogation of Pus7-RNA interactions. Pseudouridine synthases have grown in intrigue over the past decades as Pus-enzymes have been associated with a variety of different pathologies and processes (1,2). Furthermore, some Pus-enzymes have even been suggested as therapeutic targets (3–5), which may signal a biomedical potential of these enzymes. To be able to address these questions, several key areas should be further investigated to understand the basis of selectivity in these enzymes.

First, structural characterization of eukaryotic Pus enzymes is severely lacking. The only complete structures of eukaryotic Pus-enzymes include Pus7 (yeast (6) and human (7)) and Pus10 (human)(8), though the structure catalytic domain of Pus1 has also been reported (9–11). The unique insertions present only in eukaryotic pseudouridine synthases are proposed to contribute to expanded substrate selectivity, consistent with our findings. However, without structural information detailing the architecture of these domains, the position, orientation, and function of eukaryotic insertions in other Pus enzymes remains to be defined. Further, co-crystal structures of Pus complexed with an RNA substrate will be especially important to discern the contribution of these domains and determine if/how they interact with RNAs. To date, there are no reported structures of a eukaryotic pseudouridine synthase with RNA in the active site (11), though many have been reported for the prokaryotic pseudouridine synthases (12–20). This represents a significant gap in our understanding of the molecular level determinants of Pus-RNA interactions and selectivity.

Second, the finding that *E. coli* TruB exhibited 'chaperone like' behavior raised the possibility that catalyzing Ψ-formation may not be the only function/role of pseudouridine synthases in the cell (21–23). Indeed, pseudouridine synthases have be implicated in a variety of important cellular activities (1,2), some of which are suggested to be independent of Pus catalytic-function (21,24,25), further highlighting the potential involvement of Pus-enzymes in non-modification based biological processes.

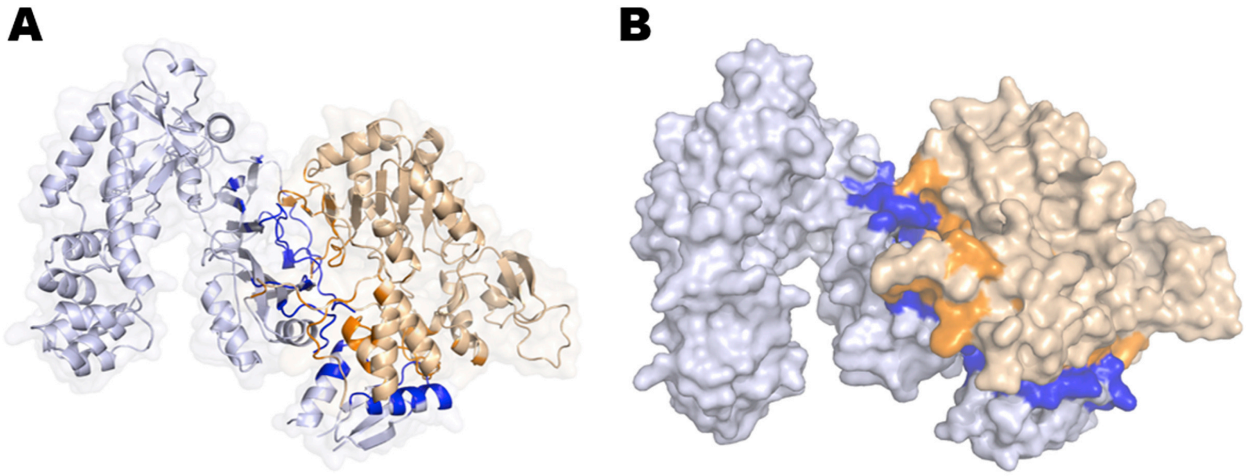


Figure 5.2: Docking simulation of Pus7-SIRT1 interaction

(A) Cartoon representation of the PatchDock model 8 of the SIRT1 (wheat) - PUS7 (grey) complex. **(B)** Surface representation of the PatchDock model 8 of the SIRT1 (wheat) - PUS7 (grey) complex. Adapted from (26).

Characterization of Pus interactome with other cellular components will be important to deconvolute the potential role(s) of Pus enzymes in the cell and if/how catalytic function affects these outcomes. Indeed, a combination of proteomic studies and pulldown assays (26,27), have identified Pus7 as an interaction partner of sirtuin1 (SIRT1), a deacetylase involved in a dynamic regulatory network whose dysfunction has been implicated in obesity-associated metabolic diseases, cancers, and development (28). Biochemical analysis informed by docking simulations demonstrated the Pus7:SIRT1 interaction is primarily mediated through the N-terminal region (1-255), which encompasses the largest eukaryotic insertion (Figure 5.2) (26). An extensive interface

is formed outside of the catalytic core of Pus7, found to be enriched with many residues highly conserved in eukaryotes. Further investigation into this interaction is required to establish if/how Pus7 influences SIRT1 activity (or vice-versa) (29), and establish the contributions of the conserved insertions to supporting protein interaction partners. These interactions could be especially interesting if modifications or modifying enzymes are found to impact the activity of other RNA-binding proteins, as has been suggested (29). Regarding the characterization of Pus-functionality throughout the cell, we know only a small portion of what is likely a diverse and complicated network of Pus protein partners/interactions and the coming years will begin to reveal a more complete picture of Pus cellular function.

Third, how the regulated, site-specific incorporation of Ψ in mRNAs is achieved is a persistent question and an important area of investigation to understand gene expression. Despite advances in Ψ-mapping techniques, effectively capturing the regulation, occurrence, and distribution of Ψ in mRNA is proving to be far more complex than initially thought. Only a small proportion of the entire landscape of Ψ in mRNAs has likely been identified, especially considering unique cell- and condition-dependent Ψ-distributions. Previous detection techniques for mapping Ψ rely on its reaction with *N*-cyclohexyl-*N*′-(2-morpholinoethyl)carbodiimide methyl-*p*-toluenesulfonate (CMC) to form Ψ-CMC adducts that generate detectable stops after reverse transcription (CMC-RT) (Figure 5.3) (30). Variations of this technique, referred to as Psi-seq or CeU-seq, which still rely on CMC-adduct formation, have been the predominant approach to mapping Ψ in mRNAs (31–34). However, the results are often inconsistent – comparison between these studies reveals very little overlap in identified sites of Ψ – and do not offer quantitative information regarding the stoichiometry of Ψ at a particular site. Furthermore, while mapping studies take a conservative approach to weed out false positives, focusing only on high

confidence sites may artificially bias our interpretation of what constitutes the most important/fundamental determinants for substrate selectivity in Pus enzymes.

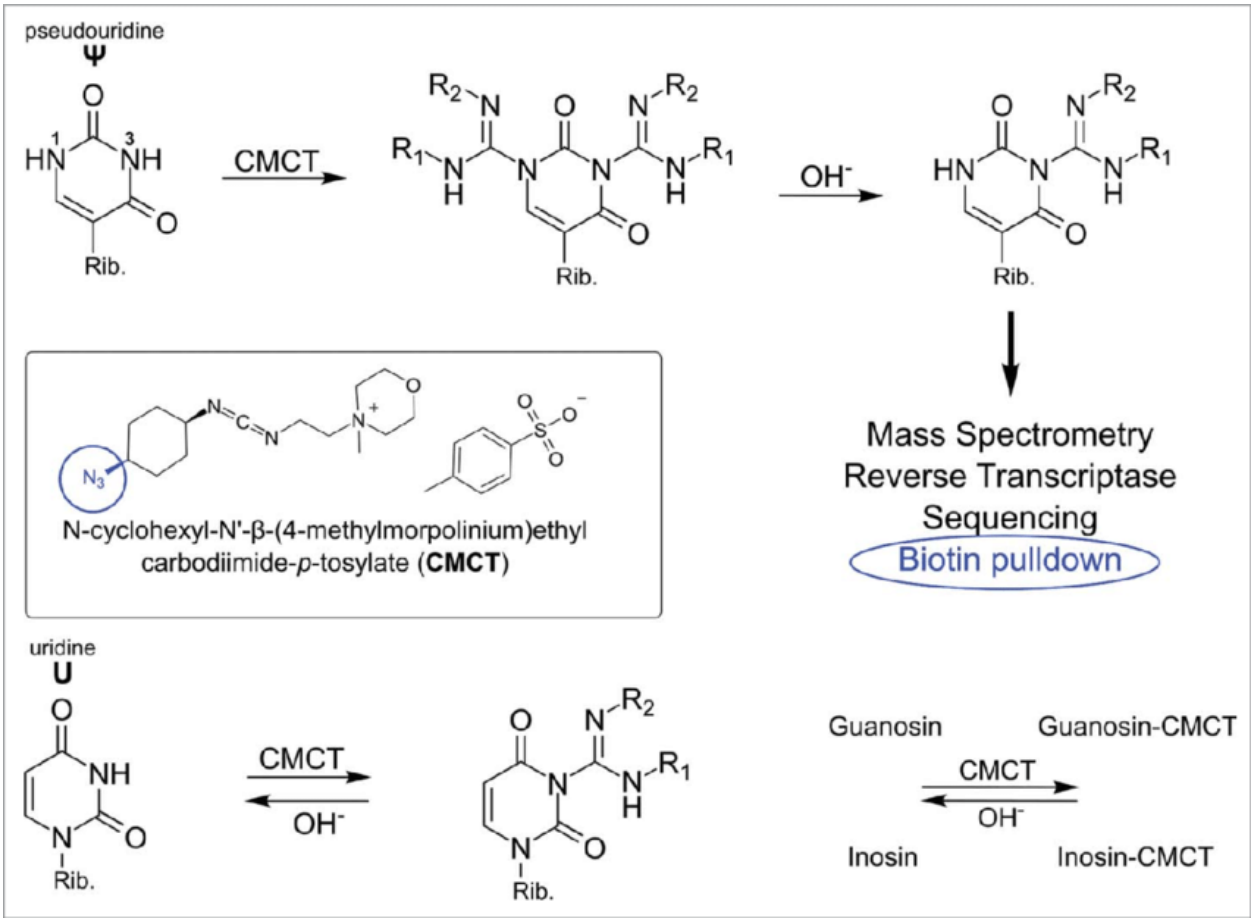


Figure 5.3: Reaction of carbodiimides with pseudouridine and uridine

(top) acylation of pseudouridine (Ψ) with the carbodiimide CMCT (full name in text). The carbodiimide group reacts with both N1 and N3, but after alkaline treatment the N1-CMCT is cleaved. The remaining CMC at N3 enables detection by mass spectrometry, reverse transcription, sequencing or even biotin pulldown in case of the CMC-azide (encircled, blue). (**bottom**) Like pseudouridine, uridine gets labeled at the N3 position, which is removed after alkaline treatment. The same was found to be true for guanosine and inosine. Insert: Structure of CMCT. Adapted from (35).

Moreover, any CMC-RT based assay will fall victim to a similar bias. CMC does selectively derivatize Ψ at high pH (~10.4), but at lower pH (~8.5) CMC will interact with unmodified uridines, guanosine, inosine, as well as other modified bases including 4-thouridine, 5-methyluridine, and 2-methlythio-6-isopentenyladenosine (ms^2i^6A) (Figure 5.3) (36,37). Thus, the selective labeling of Ψ relies on slight chemo-electronic differences between the bases, which

results in the incomplete derivatization of modified bases, as well as labeling of non-Ψ residues – overall contributing to misinterpretation or incorrect identification of Ψ sites (38,39). And, because this method relies on RT-stops, detecting additional Ψs downstream of a site of modification in a single transcript is not possible. Furthermore, because CMC labeling relies on such minute electronic differences, local structure or sequence motifs may affect the affinity of CMC for Ψ. This could result in over- or underrepresentation of specific sequence or structural motifs in identified Ψ sites, which skews subsequent analysis of the elements required for substrate recognition and selectivity.

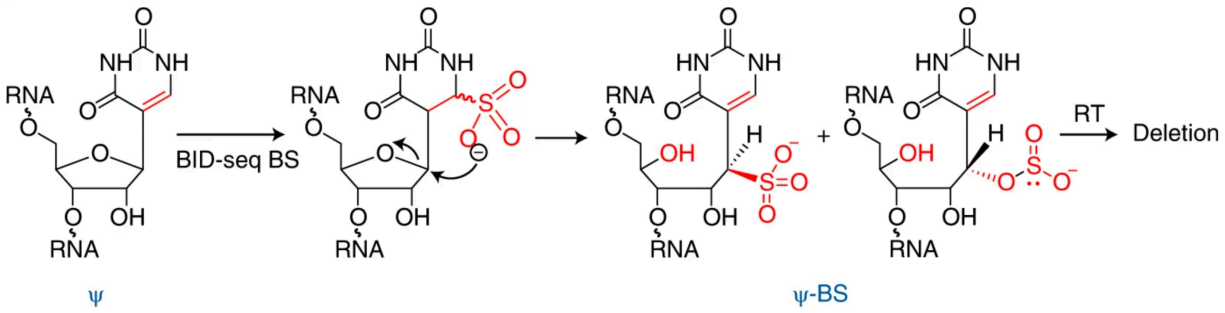


Figure 5.4: BID-seq

Chemical structure of the Ψ-BS adduct after bisulfite treatment. BS selectively reacts with Ψ and completely converts it into the Ψ-BS adduct under optimized conditions, without affecting normal C or U bases in RNA. Adapted from (54).

Fortunately, the past five years have seen significant developments in Ψ detection methods that do not rely on formation of a CMC-Ψ adduct (35,40). Many computational methods have been reported (41–52), but the most promising methods for the quantitative detection of Ψ at base-resolution include nanopore sequencing, primarily NanoPsu (53), and bisulfite-induced deletion sequencing (BID-seq) (54). NanoPsu enables the direct interrogation of RNA modifications in native transcripts and both signal strength and dwell time are used for the identification of Ψ (55). The strength of this technique relies on the quality of training data from known sites of modification, and NanoPsu profiling studies recently identified interferon induced Ψ in human

mRNAs (53) and alternate nanopore sequencing methods identified Ψ sites in SARS-CoV-2 (55). Alternatively, BID-seq has been developed for the detection and quantification of abundant Ψ sites at base-resolution (54). BID-seq relies on the reactivity of bisulfate (BS) towards Ψ (40,56) that results in two stable aromatic end products, both of which are monobisulfite adducts at C1′ and have a ring-opened ribose, which generate a detectable RT deletion at the site of modification (54,55). This method was recently applied to comprehensively map Ψ across the transcriptome at single-base resolution in a variety of human cell lines and mouse tissues, where they found stark differences in Ψ profiles in different cell types, suggesting tissue-specific roles for Ψ modification (54). In total, these methodologies represent significant advancements in detecting Ψ at single-base resolution. Accurate mapping of Ψ is important not only to be able to deconvolute the role of Ψ in diverse biological processes, but also as a diagnostic tool as pseudouridylation may also represent new biomarkers differentiating healthy and diseased states (57,58).

5.2.2 B_{12} dependent enzymes

The importance of cobalamin-dependent methionine synthase stems not only from the chemically improbable methylation that it catalyzes, but also from the unique capabilities of its cobalamin cofactor. During catalysis, cobalamin forms one of the only metal-carbon bonds found in nature (Co-C) to serve as an intermediate methyl carrier coupling two independent methyl-transfer reactions. Given its modular structure, MS presents an intriguing model to study the interplay between structural rearrangements, cofactor states, and reactivity.

Despite decades of investigation into the dynamic behavior of MS, the precise mechanistic details of each methylation reaction remain elusive. Furthermore, the molecular rearrangements required to support these chemistries and the dynamics that govern conformational

interconversions are largely unknown. Structural characterization of each catalytic orientation, ideally in a full-length context, will provide insight into the molecular and structural basis of catalysis. Furthermore, capturing each orientation in the context of the full-length enzyme will establish the dynamic range of motions/positions accessible to each module. Indeed, even comparison of the Full-length and Fol-on structures suggest significant structural rearrangements (Figure 5.5). The active site of the Fol-domain in the Full-length structure is orientated away from the Cob-domain (red), on the distal side of the TIM-barrel. To come in proximity of the cofactor for methyl-transfer, the Folate-domain has to travel over 85 Å may undergo a 'corkscrew' type rotation to bring the active site into proximity of the cobalamin cofactor. CryoEM will also be valuable for the structural characterization of MS conformational dynamics and accessible orientations. Furthermore, Cryo-EM can be used to determine the effect of substrate concentration or cofactor state on MS conformational distribution, and potentially inform on the mechanisms dictating the transitions between conformations.

Due to the strict requirement for MeCbl present at the time of expression for cofactor incorporation in traditional MS homologs (*H. sapiens* and *E. coli*), the ability of MS to bind and perform catalysis with anything other than the canonical cofactor (MeCbl, in this case) is relatively unexplored. Thus, given that *apo-t*MS is amenable to reconstitution with a host of different cobalamin analogs, *t*MS presents an ideal model to interrogate how different cobalamin analogs with diverse ligands/chemical groups in the upper/lower axial positions affect cofactor incorporation and catalysis by *t*MS. Additionally, as demonstrated in Chapter 3, cobalamin analogs can also be strategically applied to foster the structural characterization of specific MS conformations, either by stabilizing or destabilizing canonical domain:cofactor interactions in favor of a particular domain being positioned over the cofactor.

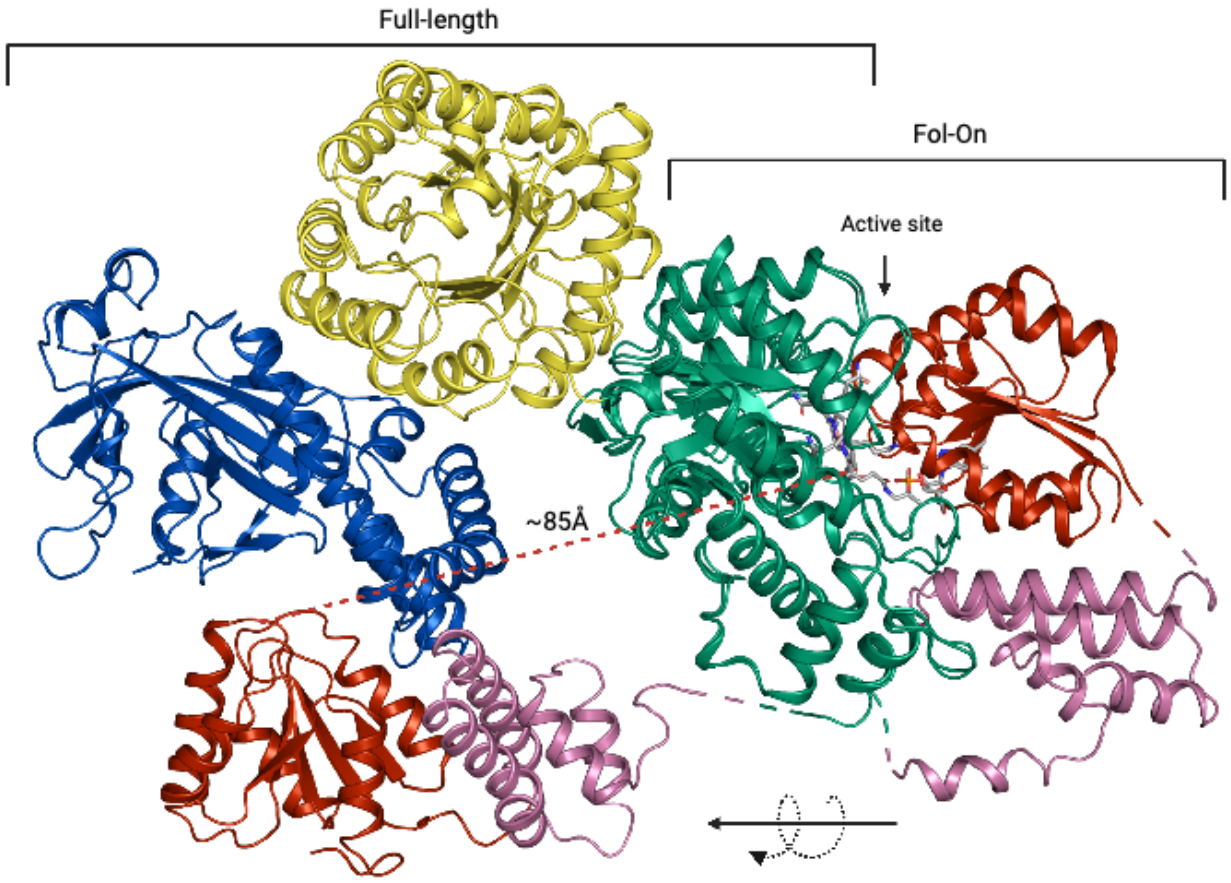


Figure 5.5: Alignment of Folate domain suggests large motions

Alignment of the Folate domain (green) captured in the full-length MS (left) structure with the folate domain (green, right) in the Fol-on structure hints at the remarkable dynamics that support catalysis.

Further, B_{12}-mimics with different metals can be used to probe the requirement of a cobalt ion at the center of the corrin, potentially serving as molecular probes to more broadly characterize B_{12} catalysis and biology. For example, nickel(II)-balamin (Nibl) is a structural B_{12}-mimic for cob(I)alamin species, due to its four coordinate planar geometry resembling that of Co(I) complexes (59–61). However, in contrast to the extraordinary reactivity of Co(I) in B12-catalysis, the ring-planarity of Nibl is thought to make this analog inert to the organometallic transformations typically required of enzyme-bound cob(I)alamin cofactors (62). Therefore, Nibl and other B_{12}-mimics could be used to probe the atomic-level and electronic requirements of MS catalysis, as well as the dependence of MS' dynamic behavior on the identity of cofactor metal-center (63–67).

Given the recent influx of papers reporting novel Ni-analogs and B12-mimics (68,69), this may represent a new direction in which to approach characterization of methyl-transfer in MS, and B_{12}-dependent processes as a whole (62).

5.3 References

1. Rintala-Dempsey AC, Kothe U. Eukaryotic stand-alone pseudouridine synthases – RNA modifying enzymes and emerging regulators of gene expression? null. 2017 Sep 2;14(9):1185–96.

2. Borchardt EK, Martinez NM, Gilbert WV. Regulation and Function of RNA Pseudouridylation in Human Cells. Annu Rev Genet. 2020 Nov 23;54(1):309–36.

3. Cui Q, Yin K, Zhang X, Ye P, Chen X, Chao J, et al. Targeting PUS7 suppresses tRNA pseudouridylation and glioblastoma tumorigenesis. Nat Cancer. 2021 Sep;2(9):932–49.

4. Afolabi R, Chinedu S, Ajamma Y, Adam Y, Koenig R, Adebiyi E. Computational identification of Plasmodium falciparum RNA pseudouridylate synthase as a viable drug target, its physicochemical properties, 3D structure prediction and prediction of potential inhibitors. Infection, Genetics and Evolution. 2022 Jan 1;97:105194.

5. Armando RG, Mengual Gómez DL, Juritz EI, Lorenzano Menna P, Gomez DE. Homology Model and Docking-Based Virtual Screening for Ligands of Human Dyskerin as New Inhibitors of Telomerase for Cancer Treatment. Int J Mol Sci. 2018 Oct 18;19(10).

6. Purchal MK, Eyler DE, Tardu M, Franco MK, Korn MM, Khan T, et al. Pseudouridine synthase 7 is an opportunistic enzyme that binds and modifies substrates with diverse sequences and structures. Proc Natl Acad Sci U S A. 2022 Jan 25;119(4):e2109708119.

7. Bank RPD. RCSB PDB - 5KKP: Crystal Structure of Human Pseudouridylate Synthase 7 [Internet]. [cited 2021 Oct 21]. Available from: https://www.rcsb.org/structure/5KKP

8. McCleverty CJ, Hornsby M, Spraggon G, Kreusch A. Crystal structure of human Pus10, a novel pseudouridine synthase. J Mol Biol. 2007 Nov 9;373(5):1243–54.

9. Czudnochowski N, Wang AL, Finer-Moore J, Stroud RM. In human pseudouridine synthase 1 (hPus1), a C-terminal helical insert blocks tRNA from binding in the same orientation as in the Pus1 bacterial homologue TruA, consistent with their different target selectivities. J Mol Biol. 2013 Oct 23;425(20):3875–87.

10. Huet T, Miannay FA, Patton JR, Thore S. Steroid receptor RNA activator (SRA) modification by the human pseudouridine synthase 1 (hPus1p): RNA binding, activity, and atomic model. PLoS One. 2014;9(4):e94610.

11. Grünberg S, Doyle LA, Dai N, Corrêa IR, Yigit E, Stoddard BL. The structural basis of mRNA recognition and binding by eukaryotic pseudouridine synthase PUS1 [Internet]. bioRxiv; 2021 [cited 2022 Sep 8]. p. 2021.12.08.471817. Available from: https://www.biorxiv.org/content/10.1101/2021.12.08.471817v1

12. Hoang C, Ferré-D'Amaré AR. Cocrystal structure of a tRNA Psi55 pseudouridine synthase: nucleotide flipping by an RNA-modifying enzyme. Cell. 2001 Dec 28;107(7):929–39.

13. Pan H, Agarwalla S, Moustakas DT, Finer-Moore J, Stroud RM. Structure of tRNA pseudouridine synthase TruB and its RNA complex: RNA recognition through a combination of rigid docking and induced fit. Proc Natl Acad Sci U S A. 2003 Oct 28;100(22):12648–53.

14. Phannachet K, Huang RH. Conformational change of pseudouridine 55 synthase upon its association with RNA substrate. Nucleic Acids Res. 2004;32(4):1422–9.

15. Hoang C, Chen J, Vizthum CA, Kandel JM, Hamilton CS, Mueller EG, et al. Crystal structure of pseudouridine synthase RluA: indirect sequence readout through protein-induced RNA structure. Mol Cell. 2006 Nov 17;24(4):535–45.

16. Phannachet K, Elias Y, Huang RH. Dissecting the roles of a strictly conserved tyrosine in substrate recognition and catalysis by pseudouridine 55 synthase. Biochemistry. 2005 Nov 29;44(47):15488–94.

17. Hur S, Stroud RM. How U38, 39, and 40 of many tRNAs become the targets for pseudouridylation by TruA. Mol Cell. 2007 Apr 27;26(2):189–203.

18. Hoang C, Hamilton CS, Mueller EG, Ferré-D'Amaré AR. Precursor complex structure of pseudouridine synthase TruB suggests coupling of active site perturbations to an RNA-sequestering peripheral protein domain. Protein Sci. 2005 Aug;14(8):2201–6.

19. Alian A, DeGiovanni A, Griner SL, Finer-Moore JS, Stroud RM. Crystal structure of an RluF-RNA complex: a base-pair rearrangement is the key to selectivity of RluF for U2604 of the ribosome. J Mol Biol. 2009 May 15;388(4):785–800.

20. Liang B, Zhou J, Kahen E, Terns RM, Terns MP, Li H. Structure of a functional ribonucleoprotein pseudouridine synthase bound to a substrate RNA. Nat Struct Mol Biol. 2009 Jul;16(7):740–6.

21. Keffer-Wilkes LC, Veerareddygari GR, Kothe U. RNA modification enzyme TruB is a tRNA chaperone. Proc Natl Acad Sci U S A. 2016 Dec 13;113(50):14306–11.

22. Gutgsell N, Englund N, Niu L, Kaya Y, Lane BG, Ofengand J. Deletion of the Escherichia coli pseudouridine synthase gene truB blocks formation of pseudouridine 55 in tRNA in vivo, does not affect exponential growth, but confers a strong selective disadvantage in competition with wild-type cells. RNA. 2000 Dec;6(12):1870–81.

23. Wolin SL. Two for the price of one: RNA modification enzymes as chaperones. Proc Natl Acad Sci USA. 2016 Dec 13;113(50):14176–8.

24. Song J, Zhuang Y, Zhu C, Meng H, Lu B, Xie B, et al. Differential roles of human PUS10 in miRNA processing and tRNA pseudouridylation. Nat Chem Biol. 2020 Feb;16(2):160–9.

25. Song D, Guo M, Xu S, Song X, Bai B, Li Z, et al. HSP90-dependent PUS7 overexpression facilitates the metastasis of colorectal cancer cells by regulating LASP1 abundance. J Exp Clin Cancer Res. 2021 May 14;40(1):170.

26. Dalal S, Deshmukh P, Unni S, Padavattan S, Padmanabhan B. Biochemical insight into pseudouridine synthase 7 (PUS7) as a novel interactor of sirtuin, SIRT1. Biochem Biophys Res Commun. 2019 Oct 20;518(3):598–604.

27. Armour SM, Bennett EJ, Braun CR, Zhang XY, McMahon SB, Gygi SP, et al. A high-confidence interaction map identifies SIRT1 as a mediator of acetylation of USP22 and the SAGA coactivator complex. Molecular and cellular biology. 2013;33(8):1487–502.

28. Rahman S, Islam R. Mammalian Sirt1: insights on its biological functions. Cell Commun Signal. 2011 May 8;9:11.

29. Vaidyanathan PP, AlSadhan I, Merriman DK, Al-Hashimi HM, Herschlag D. Pseudouridine and N6-methyladenosine modifications weaken PUF protein/RNA interactions. RNA. 2017 May;23(5):611–8.

30. Bakin A, Ofengand J. Four newly located pseudouridylate residues in Escherichia coli 23S ribosomal RNA are all at the peptidyltransferase center: Analysis by the application of a new sequencing technique. Biochemistry. 1993 Sep 21;32(37):9754–62.

31. Carlile TM, Rojas-Duran MF, Zinshteyn B, Shin H, Bartoli KM, Gilbert WV. Pseudouridine profiling reveals regulated mRNA pseudouridylation in yeast and human cells. Nature. 2014 Nov 6;515(7525):143–6.

32. Schwartz S, Bernstein DA, Mumbach MR, Jovanovic M, Herbst RH, León-Ricardo BX, et al. Transcriptome-wide mapping reveals widespread dynamic-regulated pseudouridylation of ncRNA and mRNA. Cell. 2014 Sep 25;159(1):148–62.

33. Lovejoy AF, Riordan DP, Brown PO. Transcriptome-wide mapping of pseudouridines: pseudouridine synthases modify specific mRNAs in S. cerevisiae. PLoS One. 2014;9(10):e110799.

34. Li X, Zhu P, Ma S, Song J, Bai J, Sun F, et al. Chemical pulldown reveals dynamic pseudouridylation of the mammalian transcriptome. Nat Chem Biol. 2015 Aug;11(8):592–7.

35. Heiss M, Kellner S. Detection of nucleic acid modifications by chemical reagents. RNA Biology. 2017 Sep 2;14(9):1166–74.

36. Ofengand J, Del Campo M, Kaya Y. Mapping Pseudouridines in RNA Molecules. Methods. 2001 Nov 1;25(3):365–73.

37. Gilham P, Ho NW. Reaction of pseudouridine and inosine with N-cyclohexyl-N'-β-(4-methylmorpholinium) ethylcarbodiimide. Biochemistry. 1971;10(20):3651–7.

38. Durairaj A, Limbach PA. Improving CMC-derivatization of pseudouridine in RNA for mass spectrometric detection. Analytica Chimica Acta. 2008 Apr 7;612(2):173–81.

39. Campo MD, Recinos C, Yanez G, Pomerantz SC, Guymon R, Crain PF, et al. Number, position, and significance of the pseudouridines in the large subunit ribosomal RNA of Haloarcula marismortui and Deinococcus radiodurans. RNA. 2005 Feb 1;11(2):210–9.

40. Khoddami V, Yerra A, Mosbruger TL, Fleming AM, Burrows CJ, Cairns BR. Transcriptome-wide profiling of multiple RNA modifications simultaneously at single-base resolution. Proceedings of the National Academy of Sciences. 2019 Apr 2;116(14):6784–9.

41. Li F, Guo X, Jin P, Chen J, Xiang D, Song J, et al. Porpoise: a new approach for accurate prediction of RNA pseudouridine sites. Briefings in Bioinformatics. 2021 Nov 1;22(6):bbab245.

42. Hassan D, Acevedo D, Daulatabad SV, Mir Q, Janga SC. Penguin: A tool for predicting pseudouridine sites in direct RNA nanopore sequencing data. Methods. 2022 Jul;203:478–87.

43. Li YH, Zhang G, Cui Q. PPUS: a web server to predict PUS-specific pseudouridine sites. Bioinformatics. 2015 Oct 15;31(20):3362–4.

44. Chen W, Tang H, Ye J, Lin H, Chou KC. iRNA-PseU: Identifying RNA pseudouridine sites. Mol Ther Nucleic Acids. 2016;5(7):e332.

45. He J, Fang T, Zhang Z, Huang B, Zhu X, Xiong Y. PseUI: Pseudouridine sites identification based on RNA sequence information. BMC Bioinformatics. 2018 Aug 29;19(1):306.

46. Tahir M, Tayara H, Chong KT. iPseU-CNN: Identifying RNA Pseudouridine Sites Using Convolutional Neural Networks. Molecular Therapy - Nucleic Acids. 2019 Jun 7;16:463–70.

47. Liu K, Chen W, Lin H. XG-PseU: an eXtreme Gradient Boosting based method for identifying pseudouridine sites. Mol Genet Genomics. 2020 Jan 1;295(1):13–21.

48. Bi Y, Jin D, Jia C. EnsemPseU: Identifying Pseudouridine Sites With an Ensemble Approach. IEEE Access. 2020;8:79376–82.

49. Lv Z, Zhang J, Ding H, Zou Q. RF-PseU: A Random Forest Predictor for RNA Pseudouridine Sites. Frontiers in Bioengineering and Biotechnology [Internet]. 2020 [cited 2022 Nov 3];8. Available from: https://www.frontiersin.org/articles/10.3389/fbioe.2020.00134

50. Khan SM, He F, Wang D, Chen Y, Xu D. MU-PseUDeep: A deep learning method for prediction of pseudouridine sites. Comput Struct Biotechnol J. 2020;18:1877–83.

51. Song B, Tang Y, Wei Z, Liu G, Su J, Meng J, et al. PIANO: A Web Server for Pseudouridine-Site (Ψ) Identification and Functional Annotation. Front Genet. 2020;11:88.

52. Song B, Chen K, Tang Y, Ma J, Meng J, Wei Z. PSI-MOUSE: Predicting Mouse Pseudouridine Sites From Sequence and Genome-Derived Features. Evol Bioinform Online. 2020 Jan 1;16:1176934320925752.

53. Huang S, Zhang W, Katanski CD, Dersh D, Dai Q, Lolans K, et al. Interferon inducible pseudouridine modification in human mRNA by quantitative nanopore profiling. Genome Biology. 2021 Dec 6;22(1):330.

54. Dai Q, Zhang LS, Sun HL, Pajdzik K, Yang L, Ye C, et al. Quantitative sequencing using BID-seq uncovers abundant pseudouridines in mammalian mRNA at base resolution. Nat Biotechnol. 2022 Oct 27;1–11.

55. Fleming AM, Mathewson NJ, Howpay Manage SA, Burrows CJ. Nanopore Dwell Time Analysis Permits Sequencing and Conformational Assignment of Pseudouridine in SARS-CoV-2. ACS Cent Sci. 2021 Oct 27;7(10):1707–17.

56. Fleming AM, Alenko A, Kitt JP, Orendt AM, Flynn PF, Harris JM, et al. Structural Elucidation of Bisulfite Adducts to Pseudouridine That Result in Deletion Signatures during Reverse Transcription of RNA. J Am Chem Soc. 2019 Oct 16;141(41):16450–60.

57. Cui XL, Nie J, Ku J, Dougherty U, West-Szymanski DC, Collin F, et al. A human tissue map of 5-hydroxymethylcytosines exhibits tissue specificity through gene and enhancer modulation. Nature communications. 2020;11(1):1–11.

58. Li K, Cai J, Zhang M, Zhang X, Xiong X, Meng H, et al. Landscape and regulation of m6A and m6Am methylome across human and mouse tissues. Molecular cell. 2020;77(2):426–40.

59. Metzler-Nolte N, Kraatz H. Concepts and models in bioinorganic chemistry. Wiley-VCH.; 2006.

60. Berg JM. Principles of bioinorganic chemistry. University Science Books; 1994.

61. Cotton FA, Wilkinson G, Murillo CA, Bochmann M. Advanced inorganic chemistry. John Wiley and Sons, Inc.; 1999.

62. Kieninger C, Wurst K, Podewitz M, Stanley M, Deery E, Lawrence AD, et al. Replacement of the Cobalt Center of Vitamin B12 by Nickel: Nibalamin and Nibyric Acid Prepared from Metal-Free B12 Ligands Hydrogenobalamin and Hydrogenobyric Acid. Angewandte Chemie. 2020;132(45):20304–11.

63. Zelder F. Recent trends in the development of vitamin B 12 derivatives for medicinal applications. Chemical Communications. 2015;51(74):14004–17.

64. Ferreira GC, Kadish KM, Smith KM, Guilard R. The Handbook of Porphyrin Science. Vol. 27. World Scientific; 2013.

65. Ruetz M, Gherasim C, Gruber K, Fedosov S, Banerjee R, Kräutler B. Access to Organometallic Arylcobaltcorrins through Radical Synthesis: 4-Ethylphenylcobalamin, a Potential “Antivitamin B12.” Angewandte Chemie International Edition. 2013;52(9):2606–10.

66. Ruetz M, Shanmuganathan A, Gherasim C, Karasik A, Salchner R, Kieninger C, et al. Antivitamin B12 Inhibition of the Human B12-Processing Enzyme CblC: Crystal Structure of an Inactive Ternary Complex with Glutathione as the Cosubstrate. Angewandte Chemie International Edition. 2017;56(26):7387–92.

67. Mutti E, Ruetz M, Birn H, Kräutler B, Nexo E. 4-ethylphenyl-cobalamin impairs tissue uptake of vitamin B12 and causes vitamin B12 deficiency in mice. PloS one. 2013;8(9):e75312.

68. Holze G, Inhoffen HH. Erstmalige chemische Partialsynthese des Nickelkomplexes eines Cobyrinsäure-Derivates. Angewandte Chemie. 1985;97(10):887–8.

69. Brenig C, Prieto L, Oetterli R, Zelder F. A Nickel (II)-Containing Vitamin B12 Derivative with a Cofactor-F430-type π-System. Angewandte Chemie International Edition. 2018;57(50):16308–12.