ORIGINAL SCIENTIFIC PAPER

CROATICA CHEMICA ACTA

Croat. Chem. Acta **2026**, *99*(2), 165–174
Published online: June 27, 2026
DOI: 10.5562/cca4270

HKD 1926

Catalytic and Accessory Domains Jointly Ensure the Functional Fold of Isoleucyl-tRNA Synthetase

Igor Zivkovic, Melody Skulac, Ita Gruic-Sovulj*

University of Zagreb Faculty of Science, Department of Chemistry, Horvatovac 102a, 10000 Zagreb, Croatia

* Corresponding author's e-mail address: gruic@chem.pmf.unizg.hr

RECEIVED: February 24, 2026 * REVISED: May 11, 2026 * ACCEPTED: May 11, 2026

Abstract: Aminoacyl-tRNA synthetases (AARSs) covalently couple amino acids to their cognate tRNAs, providing aminoacyl-tRNAs for ribosomal protein synthesis. Given their essential role and the challenge of maintaining high fidelity despite the presence of similar amino acids and tRNAs in the cell, AARSs have acquired various domains, making them interesting models for studying protein modularity. Here, we used *Priestia megaterium* isoleucyl-tRNA synthetase (IleRS) to explore how the simultaneous removal of multiple domains distant from the aminoacylation active site affects activity. Yet we found that eliminating editing and C-terminal domains, with or without further removal of the CP2 domain, drastically increases IleRS propensity for aggregation. We succeeded in producing a small amount of the soluble truncated variant, which showed about a 4,000-fold decrease in activity. Our data suggest that during the evolution of IleRS, accessory domains couple with the rest of the protein scaffold to maintain the productive fold of the active site.

Keywords: aminoacyl-tRNA synthetases, multi-domain proteins, modular proteins, protein aggregation propensity, truncated proteins.

Abbreviations: AARS – aminoacyl-tRNA synthetase, AA-AMP – aminoacyl-adenylate, AMP – adenosine 5'-monophosphate, ATP – adenosine 5'-triphosphate, CP1 – connective peptide 1, CP2 – connective peptide 2, IleRS – isoleucyl-tRNA synthetase, LeuRS – leucyl-tRNA synthetase, PPi – pyrophosphate, SerRS – seryl-tRNA synthetase.

INTRODUCTION

AMINOACYL-tRNA synthetases (AARSs) play a crucial role in ribosomal protein biosynthesis. They serve as translators of the genetic code by pairing amino acids with their corresponding transfer RNAs (tRNA). This pairing, called aminoacylation, occurs through a two-step reaction at the synthetic active site (Figure 1a).[1–3] First, the amino acid is activated by the nucleophilic attack of the carboxyl oxygen on α-phosphorus of the ATP to form the aminoacyl-adenylate intermediate (AA-AMP) (Figure 1b). The inorganic pyrophosphatase enzymes rapidly hydrolyse the released pyrophosphate byproduct to ensure the directionality of the reaction.[4–6] In the second step, the 2' or 3' hydroxyl group of the tRNA terminal adenosine (A76) performs the nucleophilic attack at the AA-AMP carbonyl carbon to form the aminoacylated tRNA (Figure 1c). Aminoacyl-tRNA (AA-tRNA) is released from the AARS and carried by the elongation factor Tu to the ribosome, where the amino acid will be incorporated into the growing polypeptide chain.

Based on their structural similarities, AARSs are divided into two evolutionarily distinct classes, I and II.[7] The most notable difference between the classes is their active-site architecture. Class I AARS are characterised by the HUP domain design,[8,9] which contains two highly conserved catalytic motifs, HIGH and KMSKS. Both motifs participate in substrate binding at the ground or transition states. The Class II active site comprises three motifs, designated as 1, 2, and 3. Motif 1 is involved in protein dimerisation, while motifs 2 and 3 participate in the catalysis.[10,11] Motif 1 is moderately conserved, while Motifs 2 and 3 exhibit higher conservation.[12,13]

Contemporary AARS are modular enzymes. AARSs of both classes have a synthetic domain, where both activation and aminoacyl transfer occur, and a tRNA-binding domain, which, in most cases, recognises the anticodon. In addition to these domains, some AARSs have an editing domain that corrects errors. Furthermore, enzymes from both classes may contain auxiliary domains with diverse functions that may not be directly involved in catalysis.

This work is licensed under a Creative Commons Attribution 4.0 International License.

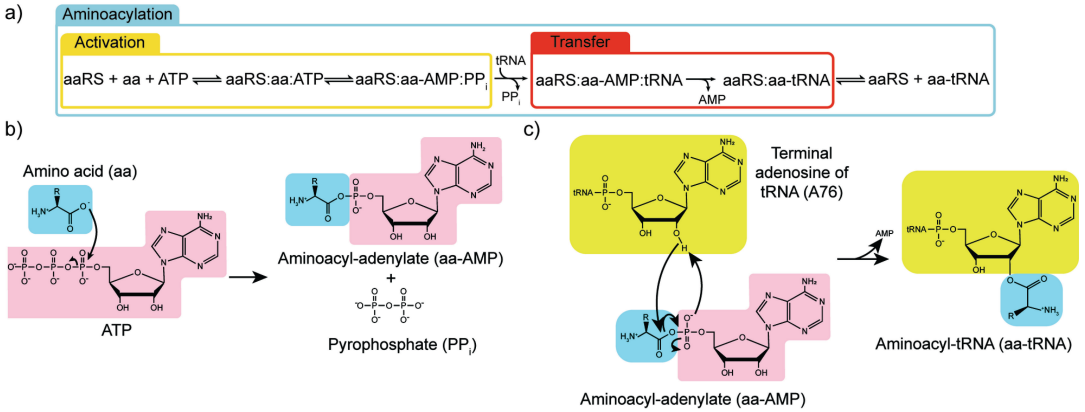


Figure 1. a) Reaction scheme of a two-step aminoacylation by aminoacyl-tRNA synthetases. b) Reaction mechanism of amino acid activation by ATP. c) Reaction mechanism of aminoacyl transfer to the tRNA.

How did AARSs evolve into such complex structures, and what did the "first AARS" look like? Schimmel's operational code hypothesis proposed that the primordial AARSs likely contained only synthetic domains and used a code that relied on recognition of the discriminatory base at the tRNA acceptor stems.[14–18] This code likely preceded the canonical genetic code, which uses tRNA anticodon for recognition. Possible remnants of this activity are reflected in the observation that contemporary AARSs of each class can aminoacylate tRNA minihelices, which comprise only the acceptor and TΨC stems and are considered evolutionary precursors of modern tRNAs.[14] The additional AARS domains, such as anticodon-binding or editing domains, likely arose later.

There has been some research on truncating AARS into smaller units. To explore minimal functional modules, Carter designed AARS urzymes, a module comprising fewer amino acids than a synthetic domain, whose catalytic activity could still be traced.[19–24] Another avenue of research produced domain-truncated AARSs to examine the roles of specific domains. This showed that deleting the editing domain of prolyl-tRNA synthetase led to a significant loss of the activation activity.[25,26] Opposite to that, the removal of the N-terminal tRNA-binding arm of seryl-tRNA synthetase (SerRS)[27] or the C-terminal tRNA-binding domain of leucyl-tRNA synthetase (LeuRS)[28] does not affect amino acid activation. However, as expected, removal of these domains reduces aminoacylation activity as it requires tRNA binding.[27,28]

Driven by the idea of going backwards towards a simpler protein structure and of exploring how domain composition influences AARS function, we used *Priestia megaterium* isoleucyl-tRNA synthetase type 2 (*Pm*IleRS2) as a model enzyme. IleRS2 is a class I enzyme of modular structure. Besides a synthetic domain (also called HUP), it contains a helix bundle (often called anticodon binding), C-terminal, CP1 (connective peptide 1), CP2 (connective peptide 2), and editing domains (Figure 2a). The CP1 domain participates in the aminoacylation reaction[29] while the CP2 domain has presumably a structural role. The editing domain, which is inserted into the CP1 domain, plays a well-understood role in post-transfer editing, namely the hydrolysis of misaminoacylated tRNAs.[30–34] The helix-bundle domain and the C-terminal domains participate in tRNA anticodon recognition.[31]

To this end, we constructed two truncated *Pm*IleRS2 variants; one lacking both the C-terminal and editing domains, and another lacking, in addition, the CP2 domain. Both variants were poorly expressed and showed a strong tendency to aggregate, despite painstaking optimisation of expression and purification protocols. Kinetic analysis showed that the truncated IleRS exhibits low activity in the activation step. Our data indicate that domains distal to the active site may play important roles in maintaining the protein functional fold required for catalysis.

EXPERIMENTAL SECTION

Construction of Truncated *Pm*IleRS2 Genes

Recombinant pET28b plasmids carrying truncated *Pm*IleRS2 (E.C. 6.1.1.5) genes with a His tag at the N-terminus were acquired from *TwistBioscience*. Two truncated variants of *Pm*IleRS2 were constructed: *i*) *Pm*IleRS2_ΔEDΔCD, lacking both the editing domain (ED, D199-S383) and C-terminal domain (CD, E786-C-end), and *ii*) *Pm*IleRS2_ΔEDΔCDΔCP2, having, in addition, removed the CP2 domain (S445-V513) (Figure 2b and 2c). In both constructs, Gly-Ser linkers were inserted in place of the editing and CP2 domains. Both genes were ordered as codon-optimised for expression in *E. coli*, while *Pm*IleRS2_ΔEDΔCDΔCP2 was, in addition, ordered in the original sequence.

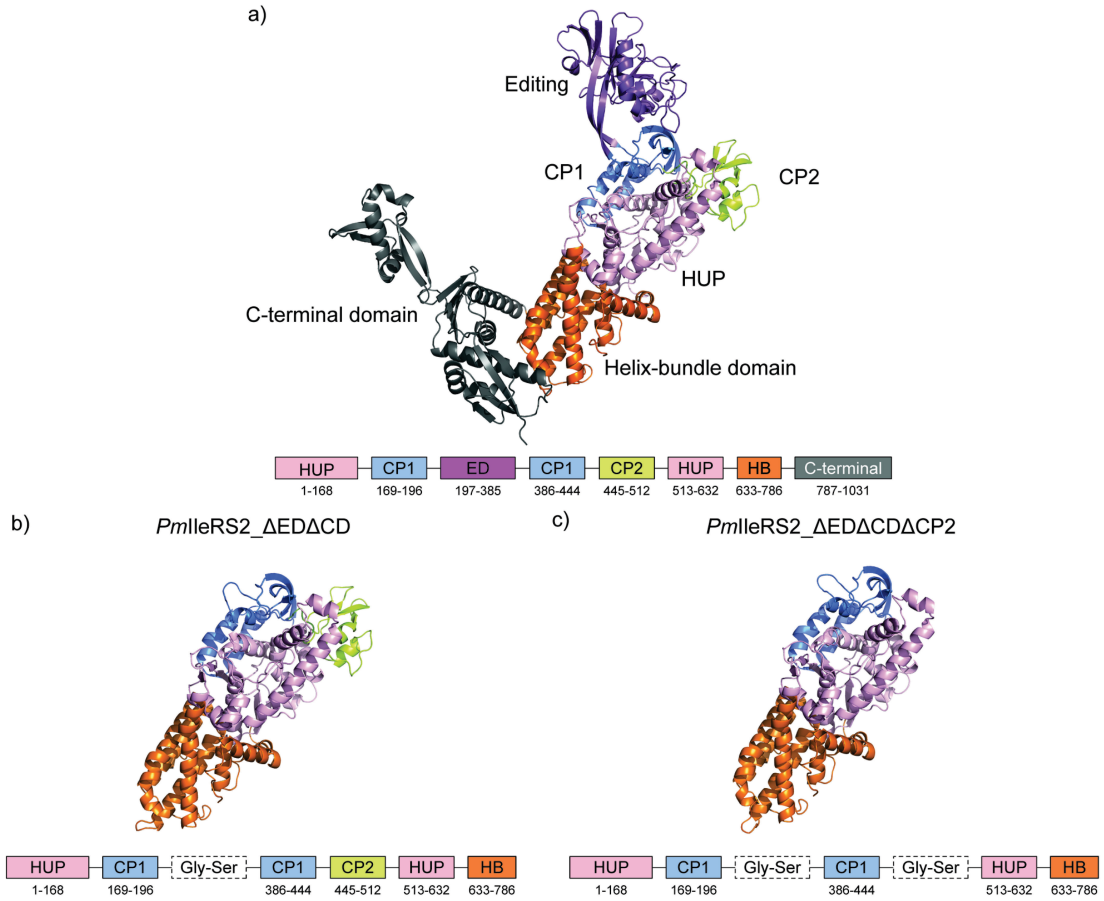


Figure 2. Structure of *P. megaterium* IleRS2 and its truncated variants. a) Structure of full-length *P. megaterium* IleRS2 (PDB ID: 8C8V). b) AlphaFold generated structure of *Pm*IleRS2_ΔEDΔCD. c) AlphaFold generated structure of *Pm*IleRS2_ΔEDΔCDΔCP2. The domains are highlighted in different colours: HUP domain in pink, CP1 domain in blue, editing domain in purple, CP2 domain in green, helix-bundle domain in orange and C-terminal domain in grey. The same colouring is used in the corresponding schematic representation of domain structure, with numbers representing the primary sequence range.

Table 1. Sequences of the primers used for amplification of truncated constructs from the pET28 plasmid and introducing desired restriction sites for subsequent cloning into pMAL-c2x and pGEX-6P-3 plasmids.

Targeted cloning plasmid	Truncated protein	Primer position	Restriction site introduction	Primer sequence
pMAL-c2x (MBP tag)	*Pm*IleRS2_ΔEDΔCD	Forward	EcoRI	5' GAGAGGAATTCATGAAGGAAGTCAATGTTCGCGA 3'
		Reverse	PstI	5' GAGAGCTGCAGTTAATTAACTTTAGTCTGGTCGC 3'
	*Pm*IleRS2_ΔEDΔCDΔCP2	Forward	EcoRI	5' GAGAGGAATTCATGAAGGAAGTCAATGTTCGCGA 3'
		Reverse	PstI	5' GAGAGCTGCAGTCAGTTCACTTTCGTTTGATCAC 3'
pGEX-6P-3 (GST tag)	*Pm*IleRS2_ΔEDΔCD	Forward	BamHI	5' GAGAGGGATCCATGAAAGAAGTGAATGTCCGAGA 3'
		Reverse	XhoI	5' TGGTGCTCGAGTTAATTAACTTTAGTCTGGTCGC 3'
	*Pm*IleRS2_ΔEDΔCDΔCP2	Forward	BamHI	5' GAGAGGGATCCATGAAAGAAGTGAATGTAAGGGA 3'
		Reverse	XhoI	5' TGGTGCTCGAGTCAGTTCACTTTCGTTTGATCAC 3'

Table 2. PCR program used for amplification of truncated constructs from the pET28 plasmid.

Step	θ / °C	t / s	Cycles
Initial denaturation	98	30	1
Denaturation	98	10	
Annealing	62	20	32
Elongation	72	55	
Final elongation	72	120	1

To replace His tag with MBP and GST, truncated *ileS2* constructs were amplified by PCR using pET28_*Pm*IleRS2_ΔEDΔCD or pET28_*Pm*IleRS2_ΔEDΔCDΔCP2 as templates (Table 1). PCR mixture comprised 0.002 U mL^{-1} Q5 High-Fidelity DNA polymerase (*New England Biolabs*), 2 ng µL^{-1} DNA template, 0.5 µmol dm^{-3} forward/reverse primer, 200 µmol dm^{-3} dNTPs, and 1 × Q5 buffer. The reaction program is shown in Table 2. PCR products were purified using the Qiagen PCR purification kit. Plasmids (pMAL-c2x for MBP tag on the N-terminus and pGEX-6P-3 for GST tag on the N-terminus) and purified PCR products were cleaved by the corresponding restriction endonucleases (Table 1). Cleaved DNAs were purified using the Qiagen PCR purification kit. Plasmids and PCR products were ligated using T4 DNA ligase (*Thermo*). Recombinant plasmids were identified by sequencing (*Macrogen*).

Truncated Protein Overexpression and Purification

E. coli BL21(DE3) carrying truncated *ileS2* constructs co-transformed with a plasmid carrying chaperone set 1 pG-KJE8 (dnaJ, dnaK, groES, groEL and tig)[35,36] were grown in LB media supplemented with 1 mmol dm^{-3} $ZnCl_2$, 1 mmol dm^{-3} $MgCl_2$, 30 µg mL^{-1} kanamycin and 34 µg mL^{-1} chloramphenicol at 200 RPM and 37 °C up to OD_{600} = 0.5 – 0.8. Cells were then cooled to 4 °C for 45 min, and overexpression was induced with 0.05 mmol dm^{-3} IPTG, 5 ng mL^{-1} tetracycline, and 0.5 mg mL^{-1} L-arabinose. The latter two are used to induce the overexpression of chaperones from the pG-KJE8 plasmid. Overexpression was left overnight, typically for 14 to 16 hours at 16 °C. Cells were pelleted and resuspended in the lysis buffer, 25 mmol dm^{-3} Tris, pH = 8.0, at 25 °C, 500 mmol dm^{-3} NaCl, 20 mmol dm^{-3} $MgCl_2$, 10 mmol dm^{-3} imidazole, 5 mmol dm^{-3} 2-mercaptoethanol, v/v = 1 % Tween20, 1 mmol dm^{-3} PMSF, and 3 µg mL^{-1} DNase I. Cells were lysed by sonication. Proteins with His tag were purified on Ni-NTA resin as described previously.[37,38] Constructs with the MBP tag were purified on the MBPTrap 1 mL HP column (*Cytiva*). Proteins were bound to the column in a buffer containing 20 mmol dm^{-3} Hepes, pH = 7.5, 200 mmol dm^{-3} NaCl, 5 mmol dm^{-3} 2-mercaptoethanol, and eluted with the addition of 10 mmol dm^{-3} maltose. GST tagged constructs were purified on the GSTrap 1 mL FF column (*Cytiva*). Proteins were bound to the column in a buffer containing 30 mmol dm^{-3} phosphate buffer, pH = 7.3, 140 mmol dm^{-3} NaCl, 500 mmol dm^{-3} KCl, and 10 mmol dm^{-3} 2-mercaptoethanol. Elution was performed with the following buffer: 50 mmol dm^{-3} Tris, pH = 8.0, 10 mmol dm^{-3} reduced glutathione, and 10 mmol dm^{-3} 2-mercaptoethanol. Fractions containing truncated variants were further separated from contaminations and soluble aggregates by size-exclusion chromatography at Superdex 200 10/300 (*Cytiva*) using 25 mmol dm^{-3} Tris, pH = 8.0 at 25 °C, 50 mmol dm^{-3} NaCl, 10 mmol dm^{-3} 2-mercaptoethanol, v/v = 10 % glycerol, and v/v = 1 % Tween20 as a mobile phase.

Labelling of tRNAs

P. megaterium $tRNA^{Ile}$ was a gift from V. Zanki. Preparation and purification are described.[39] Labelled [^{32}P]tRNA for aminoacylation experiments was prepared as described.[32] In short, 5 µmol dm^{-3} $tRNA^{Ile}$ was incubated for 1 min at 37 °C with 5 µmol dm^{-3} tRNA nucleotidyltransferase in a buffer containing 1 µmol dm^{-3} [α-^{32}P]ATP (specific activity of 3000 Ci mmol^{-1}), 200 mmol dm^{-3} Tris, pH = 8.0, 20 µmol dm^{-3} $MgCl_2$, 5 mmol dm^{-3} $Na_4P_2O_7$, and 0.5 mmol dm^{-3} DTT. Afterwards, 0.1 U µL^{-1} of thermostable inorganic pyrophosphatase (*Sigma*) was added to shift the equilibrium towards [α-^{32}P]ATP incorporation, and the mixture was incubated for 2 min at room temperature. $tRNA^{Ile}$ was purified by phenol/chloroform extraction followed by two consecutive chromatography steps on Bio-Spin P-30 columns (*Bio-Rad*) to remove the remaining [α-^{32}P]ATP. The [^{32}P]tRNA sample was dialysed against 10 mmol dm^{-3} Hepes, pH = 7.5.

Amino Acid Activation Using Radiolabelled [^{32}P]-γ-ATP

Amino acid activation was followed by the modified version of the ATP-PP_i exchange assay, [40–42] which uses radiolabelled [^{32}P]-γ-ATP instead of radiolabelled pyrophosphate.[43] The assay was performed at 30 °C in a buffer containing 50 mmol dm^{-3} HEPES, pH = 7.5, 40 mmol dm^{-3} $MgCl_2$, 25 mmol dm^{-3} ATP, spiked with [^{32}P]-γ-ATP, 0.01 mg ml^{-1} BSA, 5 mmol dm^{-3} DTT, 1 mmol dm^{-3} $Na_4P_2O_7$, 20 mmol dm^{-3} Ile, and 3.26 µmol dm^{-3} *Pm*IleRS2_ΔEDΔCDΔCP2. Reactions were quenched by mixing 1.5 µL of the reaction mixture with 3 µL of the quench solution (m/v = 0.15 % mass fraction SDS and 600 mmol dm^{-3} NaOAc, pH = 4.5, at 25 °C). Formed [^{32}P]-PP_i was separated from the remaining [^{32}P]-ATP by thin-layer chromatography on polyethyleneimine plates (Macherey-Nagel) in 4 mol dm^{-3} urea and 750 mmol dm^{-3} KH_2PO_4, pH = 3.5, at 25 °C. Signal visualisation was performed on a Typhoon Phosphoimager (*Cytiva*) and quantified with ImageQuant software as described.[44]

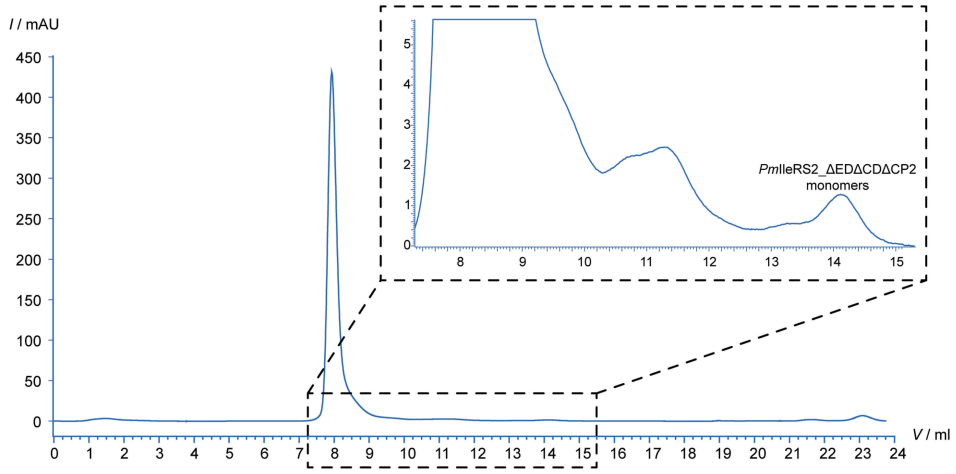


Figure 3. Purification of N-His-tag_*Pm*IleRS2_ΔEDΔCDΔCP2 construct after HisTrap by size exclusion chromatography on Superdex 200 Increase GL 10/300 column (*Cytiva*). The dominant peak at ca. 8 mL represents soluble aggregate formation, which elutes at the void volume. Truncated protein monomers elute around 14 mL and are shown in the inset.

Steady-state Aminoacylation

Two-step aminoacylation of $tRNA^{Ile}$ was followed at 30 °C in a buffer containing 20 mmol dm^{-3} Hepes, pH = 7.5, 10 mmol dm^{-3} $MgCl_2$, 150 mmol dm^{-3} NH_4Cl, *v*/*v* = 1 % Tween20, 2 mmol dm^{-3} ATP, 0.008 U/µl TIPP, 0.01 mg ml^{-1} BSA, 50 µmol dm^{-3} $tRNA^{Ile}$ spiked with $[^{32}P]$-$tRNA^{Ile}$, 20 mmol dm^{-3} Ile, and 2 µmol dm^{-3} *Pm*IleRS2_ΔEDΔCDΔCP2. The reactions were stopped by mixing 1.5 µL aliquots of the reaction with 3 µL of the quench solution (*m*/*v* = 0.15 % mass fraction SDS and 600 mmol dm^{-3} NaOAc, pH = 4.5 at 25 °C). 1.5 µL of the quench reaction mixture was mixed with 3 µL of P1 nuclease (*Sigma*) (≥ 0.01 U/µl in 300 mmol dm^{-3} NaOAc, pH = 5.0 at 25 °C, and 0.15 mmol dm^{-3} $ZnCl_2$) and incubated for 1 hour at room temperature to cleave phosphodiester bonds. Formed $[^{32}P]$-AMP and aa-$[^{32}P]$-AMP, reflecting non-aminoacylated and aminoacylated tRNA, respectively, were separated by thin-layer chromatography using polyethyleneimine plates (*Macherey-Nagel*) in a mobile phase of 100 mmol dm^{-3} NaOAc and *v*/*v* = 5 % acetic acid. Signal visualisation was performed on a Typhoon Phosphoimager (*Cytiva*) and quantified with ImageQuant (*Cytiva*) software as described.[44]

Complementation Assay

E. coli Bl21(DE3) strains transformed with one of the following: pET28_*Pm*IleRS2_ΔEDΔCD, pET28_*Pm*IleRS2_ΔEDΔCDΔCP2, pET28b_*E.coli*_IleRS (mupirocin susceptible; negative control), pET28_*Pm*IleRS2 (mupirocin resistant; positive control) and empty pET28 (negative control) were grown overnight in LB medium containing 30 µg mL^{-1} kanamycin. Bacterial cultures were diluted to an OD_{600} = 1.7. From this, serial dilutions ranging from 10-fold to 10^4-fold were made in MiliQ. Starting culture and corresponding dilutions were spotted (10 µL) on an LB agar plate containing 30 µg mL^{-1} kanamycin, 0 – 0.2 mmol dm^{-3} mupirocin, 1 mmol dm^{-3} $ZnCl_2$, and 0.1 mmol dm^{-3} IPTG. Due to the increased propensity of truncated proteins for aggregation, cells were incubated at 20 °C and 30 °C for 2 days and analysed. Alternatively, the assay was performed using overnight cultures in which truncated variants were overexpressed with 0.05 mmol dm^{-3} IPTG at 16 °C prior to plating the cells on mupirocin-containing plates. Both experimental designs gave the same results.

RESULTS

Optimisation of Truncated *Pm*IleRS2 Variants Overexpression and Purification

We designed two truncated variants of *Pm*IleRS2: *Pm*IleRS2_ΔEDΔCD, lacking editing and the C-terminal domains, and *Pm*IleRS2_ΔEDΔCDΔCP2, lacking the CP2 domain in addition. Both genes were codon-optimised for expression in *E. coli*. To explore whether codon usage may play a role in truncated protein expression, we also ordered *Pm*IleRS2_ΔEDΔCDΔCP2 with the original sequence, which proved to be better expressed than the codon-optimised variant. PmIleRS2 was chosen for two reasons: *i*) it has an experimentally solved structure and *ii*) it is resistant to the antibiotic mupirocin, which enabled us to test the functionality of truncated variants *in vivo* by complementation assay (see *Complementation Assay* in previous chapter).

First, we generated AlphaFold[45] structures of the truncated PmIleRS2 variants to explore how domain removal may influence the overall fold. We did not observe significant structural distortions related to the absence of

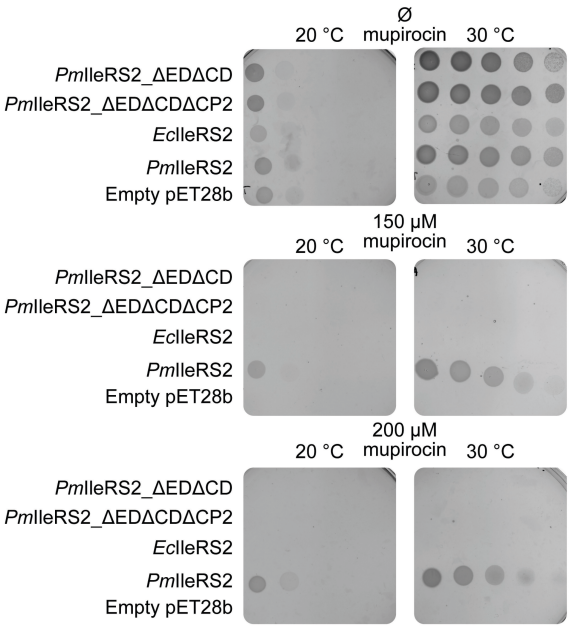


Figure 4. *In vivo* complementation assay plates for survival of *E. coli* BL21(DE3) cells on LB agar supplemented with 150 – 200 μmol dm^{-3} mupirocin and 0.1 mmol dm^{-3} IPTG, incubated at 20 and 30 °C. Different *E. coli* strains were used: pET28_*Pm*IleRS2_ΔEDΔCD, pET28_*Pm*IleRS2_ΔEDΔCDΔCP2, pET28b_*E. coli*_IleRS (mupirocin susceptible; negative control), pET28_*Pm*IleRS2 (mupirocin resistant; positive control) and empty pET28 (negative control). Plates without mupirocin serve as a control, indicating the functionality of all strains. Plates with mupirocin show survival only for the strain overexpressing PmIleRS2.

various domain combinations. The chosen variants (*Pm*IleRS2_ΔEDΔCD and *Pm*IleRS2_ΔEDΔCDΔCP2) have predicted structures that largely retain the original fold in the remaining protein scaffold (Figure 2).

However, producing tag-less, soluble, truncated *Pm*IleRS2 variants in reasonable yield has proved nearly unattainable. During the optimisation process, multiple conditions were explored (Table 3). We tried to overexpress truncated constructs tagged with His, GST, and MBP. The latter two were used for their benefits in increasing protein solubility. The GST tag has not affected protein solubility, while the MBP tag has increased it. Additionally, we used PROSS, a protein-stability design method,[46] to enhance the solubility of truncated IleRS variants. PROSS uses a primary sequence and a 3D structure (in our case, predicted structures) as input and outputs a series of primary sequences with a certain number of mutations intended to increase protein stability. Residues comprising the synthetic active site and their primary contacts were excluded from *in silico* mutagenesis. We opted for three PROSS constructs of *Pm*IleRS2_ΔEDΔCDΔCP2 variant with 19, 42, and 57 mutations, respectively. However, upon expression in *E. coli* and purification, we did not obtain a higher yield compared to the original variant. Finally, building on the increased stability of proteins from extremophiles, we also constructed, expressed, and purified *T. thermophilus* IleRS2_ΔEDΔCDΔCP2. As with PROSS, we have not observed any increase in the preference for the soluble state.

Regarding the expression system, several *E. coli* strains were used. Standard BL21(DE3), Rosetta(DE3) to

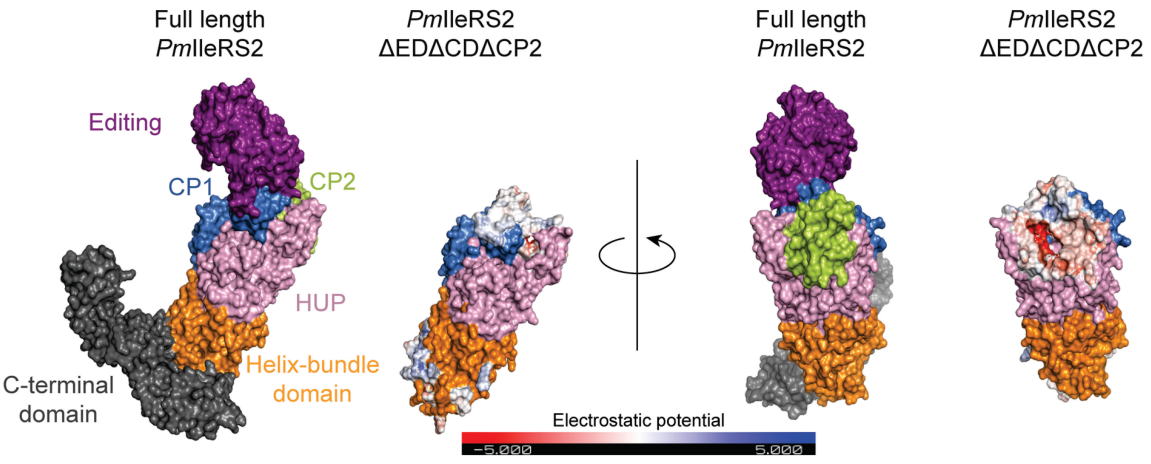


Figure 5. Structure of *Pm*IleRS2 and AlphaFold model of *Pm*IleRS2_ΔEDΔCDΔCP2 shown in surface mode. Domains are depicted in different colours: HUP domain in pink, CP1 domain in blue, editing domain in purple, CP2 domain in green, helix-bundle domain in orange and C-terminal domain in grey. Electrostatic potentials were calculated using PyMol APBS plugin with default settings, for the surfaces exposed upon domain removal in *Pm*IleRS2_ΔEDΔCDΔCP2. Residues used for calculation had at least one atom within 5 Å distance from any deleted domain.

Table 3. Explored conditions for overexpression of truncated *Pm*IleRS2 constructs.

Constructs	*E. coli* strains	Expression temperature (and time)	Media additives/specific conditions	Protein coexpression	Lysis method	Purification additives
N-His-tag_*Pm*IleRS2_ ΔEDΔCD/ΔEDΔCDΔCP2	**BL21(DE3)**	4 °C (70 h)	0.05/0.1/**0.25** mmol dm^{-3}	**dnaK**	**Sonication**	50 mmol dm^{-3}
N-GST-tag-His-tag_*Pm*IleRS2_ ΔEDΔCD/ΔEDΔCDΔCP2	Rosetta(DE3)	**16 °C (14-16 h)**	**IPTG**	**dnaJ**	BugBuster	arginine + 50 mmol dm^{-3} glutamate
N-MBP-tag-His-tag_*Pm*IleRS2_ ΔEDΔCD/ΔEDΔCDΔCP2	Tuner(DE3)	25 °C (3-5 h)	*v*/*v* = 5% EtOH (added before IPTG)	**grpE**	High-pressure	0.5 mmol dm^{-3} urea
	SoluBL21	30 °C (3-5 h)	Cold shock[a]	**groES**	homogenisation	***v*/*v* = 1% Tween 20**
		37 °C (2-4 h)	Heat shock[b]	**groEL**		*v*/*v* = 1% Tween 80
			1 mmol dm^{-3} $ZnCl_2$	tig		
			1 mmol dm^{-3} $MgCl_2$			

The listed parameters were tested in multiple combinations during expression optimization and do not represent one-to-one experimental conditions by row. Conditions that proved to yield the best results are bolded.

[a] Cold shock was performed by cooling the culture on ice in the cold room at 4 °C for 30 – 60 min.

[b] Heat shock was performed by heating the cultures for 5 min at 47 °C, adding IPTG, and then incubating the culture for an additional 30 min at 47 °C before transferring it to the desired overexpression temperature.

overcome rare codons in codon non-optimised gene (construct *Pm*IleRS2_ΔEDΔCDΔCP2), Tuner(DE3) for better control over IPTG concentration and expression level, and SoluBL21, which was developed specifically for the production of highly insoluble proteins. Standard BL21(DE3) gave the best compromise between yield and quality. As expression rate and protein folding are correlated to temperature, overexpression was carried out at 10 °C, 16 °C, 30 °C, and 37 °C. Higher temperatures led to more inclusion bodies, while expression levels at 10 °C were too low to give reasonable yields; thus, we opted for 16 °C. Overexpression was induced with 0.05, 0.1, or 0.25 mmol dm^{-3} IPTG, and we did not observe significant differences in expression levels or protein quality. To facilitate protein folding in some experiments, we also performed cold shock, heat shock, the addition of *v*/*v* = 3% ethanol, and, finally, the addition of 1 mmol dm^{-3} $MgCl_2$.[47] Neither of these resulted in noticeably better solubility. We also attempted coexpression of truncated proteins with chaperones, including dnaJ, dnaK, groES, groEL, grpE and tig, in various combinations (sets 1–5).[35,36] This proved crucial as set 1 (dnaJ, dnaK, groES, groEL, grpE) indeed helped generate stable monomers. Proteins were expressed for 4–70 hours, and various cell lysis methods were explored, including BugBuster (*Merck*), sonication, and cell disruption. The duration of overexpression had little effect on folding quality, but it did affect the yield. Sonication was selected as the method of choice because it was most compatible with the chosen lysis buffer. The composition of the lysis buffer and mobile phases for affinity chromatography was also varied. Most notably, we used several additives previously reported to help prevent protein aggregation: arginine, glutamate, urea, Tween 20, and Tween 80.[47] To summarise, in basically all tested conditions, overexpressed proteins were exclusively found in the cell precipitate. Moreover, the soluble proteins were mainly present within soluble aggregates (Figure 3, peak at ca. 8 mL). The conditions that yielded the best results were constructs with or without the MBP tag, the BL21(DE3) strain, IPTG at 0.25 mmol dm^{-3}, overexpression at 14–16 h at 4 °C, coexpression with chaperone set 1, sonication lysis, and the addition of *v*/*v* = 1 % Tween 20. Even though initial purification of MBP constructs resulted in better gain, removal of the MBP tag by the TEV protease and subsequent separation of the truncated protein resulted in the same yield as that obtained through the purification of the construct without the MBP tag. Thus, we opted for the latter. So, despite extensive optimisation of overexpression and purification of truncated *Pm*IleRS2 variants, none of the procedures yields monomeric proteins at even remotely high yields. The approximate yield was below 0.1 mg per 1 L of cell culture.

*Pm*IleRS2_ΔEDΔCDΔCP2 is Catalytically Active

For kinetic characterisation, we chose the variant *Pm*IleRS2_ΔEDΔCDΔCP2 because it was expressed at the highest yield. We first followed the amino acid activation step and observed a very low rate (0.0161 s^{-1}, Table 4), which precluded the determination of k_{cat} and K_m. Interestingly, the observed rate is still 10^9-fold faster than the analogous non-enzymatic rate, which is around 10^{-11} s^{-1}.[21] Thus, the truncated variant exhibits considerable enzymatic activity, albeit 4000-fold lower than that of the full-length *Pm*IleRS2.[39] As none of the excluded domains contains side chains that directly participate in amino acid activation or form part of the first-contact shell of the active site, the drastic decrease in enzymatic activity of the truncated variant relates to the indirect role of the distal domains in maintaining proper catalytic fold.

Regarding the two-step aminoacylation, the truncated variant exhibits a 4800-fold lower activity than the full-length *Pm*IleRS2.[39] The additional decrease in the aminoacylation relative to the amino activation step can be attributed to the absence of the C-terminal domain, which is involved in anticodon recognition.

Table 4. Kinetic data obtained for activation and aminoacylation of isoleucine by truncated *P. megaterium* isoleucyl-tRNA synthetase constructs.

	$k_{observed}$ / s^{-1}	
	Activation	Aminoacylation
PmIeRS2_ΔED_ΔCD_ΔCP2	0.0161 ± 0.0033	0.0002 ± 0.0001

The values represent the average ± SEM from at least two independent experiments.

Truncated IleRS Variants do not Support the *in vivo* Aminoacylation

To explore the ability of truncated enzymes to perform aminoacylation *in vivo*, we used a complementation assay in which plasmid-born *Pm*IleRS2 can functionally replace endogenous *E. coli* IleRS when inhibited by the antibiotic mupirocin. The IPTG-induced overexpression of mupirocin-resistant *Pm*IleRS2[39,48,49] enables *E. coli* to grow on mupirocin plates. We reasoned that if truncated proteins exhibit some activity *in vivo,* a weak complementation should be observed. To test that, we grew *E. coli* on plates (Figure 4), with varying concentrations of mupirocin. Two different temperatures (20 and 30 °C) were used to slow expression and facilitate folding of the truncated *Pm*IleRS2 variants. We could not find evidence that the truncated *Pm*IleRS2 variants support the growth of *E. coli* on mupirocin plates, suggesting they cannot uphold protein translation *in vivo*, in agreement with minimal amino-acylation activity observed *in vitro*.

DISCUSSION

To investigate the impact of domain removal on protein activity, we prepared two truncated variants of *Pm*IleRS2. The first, *Pm*IleRS2_ΔEDΔCD, lacks the editing and C-terminal domains, and the second, *Pm*IleRS2_ΔEDΔCDΔCP, lacks the editing, C-terminal, and CP2 domains. We did not anticipate major problems in producing *Pm*IleRS2_ΔEDΔCD, as its AlphaFold model[45] acquired a canonical fold, and the structural data (Figure 2) indicate that the editing and C-terminal domains are distal to the protein core and establish relatively few contacts with the part of the structure retained after truncation. We were more concerned about *Pm*IleRS2_ΔEDΔCDΔCP2, as CP2, a zinc-binding domain, has a structural role and is closer to the active site.

However, both truncated variants were difficult to overexpress and purify. Several avenues were pursued to increase yield. We tried *i*) different protein tags, *ii*) PROSS-guided mutations for higher solubility, *iii*) IleRS from thermophiles, *iv*) different expression strains, *v*) various expression conditions (temperature, length and IPTG concentration), *vi*) various additives, and *vii*) different lysis and purification methods. Even after painstaking optimisation, we still struggled to produce soluble truncated variants in reasonable amounts. Thus, it appears that removing the editing, C-terminal, and, optionally, CP2 domains dramatically increases protein propensity to aggregate, likely by disrupting proper folding. The final protocol for purifying monomers included coexpression with chaperones and purification of truncated proteins with the addition of Tween 20 (Table 3). Here, we were particularly surprised that the introduction of PROSS-guided mutations did not improve the yield of protein monomer form. Most of the PROSS-introduced mutations were distal to the active site and located on the enzyme surface. Previous work by the Åqvist group has shown that introducing distal surface mutations can modulate enzyme flexibility, thereby enhancing activity in some cases.[50] However, in our case, the negative effects of domain removal appear to have outweighed any potential benefits of distal surface mutations introduced by PROSS.

To further explore the reasons for variants' increased aggregation propensity, we visualised the exposed surfaces of the AlphaFold models of *Pm*IleRS2_ΔEDΔCD and *Pm*IleRS2_ΔEDΔCDΔCP2 (Figure 5). This analysis showed that removing the editing and C-terminal domains exposed 18 residues. In contrast, deletion of the CP2 domain appears to have a much larger effect (59 residues). Calculating the surface electrostatic potential using PyMol's APBS plugin with default settings revealed that the removal of the editing domain results in exposed residues that are predominantly neutral. This is not observed for the C-terminal or CP2 domains, where a larger fraction of charged residues is found in the regions exposed by domain deletion. Therefore, assuming that the truncated variants adopt structures similar to their AlphaFold models, the instability and aggregation propensity of the truncated proteins are likely due to the exposure of hydrophobic amino acids, particularly in place of the editing domain. This observation is consistent with previous findings that hydrophobic interactions primarily drive inter-domain stabilisation in multi-domain proteins.[51] However, limitations of structure prediction algorithms in folding truncated variants cannot be ruled out, and the folded structures of truncated IleRSs may differ significantly from the predicted models. To further explore this hypothesis, we analysed SerRS and LeuRS, in which the removal of the N-terminal and C-terminal domains, respectively, does not affect amino acid activation. In both enzymes, these deletions do not expose new residues. In the SerRS Δ35–97 construct, this is because part of the N-terminal region remains preserved, whereas in LeuRS the C-terminal region (D798–end) is spatially separated from the rest of the protein. These observations suggest that the deleted domains are

structurally independent and/or distant from the enzyme core, allowing the catalytic site to retain its proper fold and activity in amino acid activation.

Kinetic analysis shows that the *Pm*IleRS2_ΔEDΔCDΔCP2 variant exhibits a 4,000-fold decrease in activation rate relative to the full-length enzyme. Interestingly, this is on par with the catalytic rates observed for AARS urzymes, the minimal catalytic module of the synthetic domain,[52] suggesting that removing domains distal to the synthetic core has a detrimental effect on protein activity. It appears that once these domains have been incorporated into ancient IleRS variants, they co-evolve with the rest of the protein and subsequently become essential for maintaining the functional protein fold.

Acknowledgment. This paper was supported by the: Croatian Science Foundation under the project number HRZZ-IP-2022-10-1400 (IGS), the European Regional Development Fund (infrastructural project CIuK) [KK.01.1.1.02.0016], and the European Union – NextGenerationEU through the National Recovery and Resilience Plan 2021-2026. via Institutional grants of the University of Zagreb Faculty of Science (NextGenChem).

REFERENCES

[1] J. J. Perona, I. Gruic-Sovulj, *Top. Curr. Chem.* **2014**, *344*, 1–41. https://doi.org/10.1007/128_2013_456

[2] R. W. Alexander, T. L. Hendrickson in *Biology of Aminoacyl-TRNA Synthetases* (Eds.: L. Ribas de Pouplana, L. S. Kaguni), Academic Press, **2020**, pp. 39–68.

[3] I. Zivkovic, M. Dulic, I. Gruic-Sovulj, *FEBS Lett.* **2025**, *599*, 3404–3405. https://doi.org/10.1002/1873-3468.70098

[4] J. D. Dignam, M. P. Deutscher, *Biochemistry* **1979**, *18*, 3165–3170. https://doi.org/10.1021/bi00581a039

[5] J. R. Sampson, M. E. Saks, *Nucleic Acids Res.* **1993**, *21*, 4467–4475. https://doi.org/10.1093/nar/21.19.4467

[6] A. Khvorova, Y. Motorin, A. D. Wolfson, A. N. Bakh, *Nucleic Acids Res.* **1999**, *27*, 4451–4456. https://doi.org/10.1093/nar/27.22.4451

[7] G. Eriani, M. Delarue, O. Poch, J. Gangloff, D. Moras, *Nature* **1990**, *347*, 203–206. https://doi.org/10.1038/347203a0

[8] L. Aravind, V. Anantharaman, E. V Koonin, *Proteins: Structure, Function and Genetics* **2002**, *48*, 1–14. https://doi.org/10.1002/prot.10064

[9] I. Gruic-Sovulj, L. M. Longo, J. Jabłońska, D. S. Tawfik, *Crit. Rev. Biochem. Mol. Biol.* **2022**, *57*, 1–15. https://doi.org/10.1080/10409238.2021.1957764

[10] G. Eriani, J. Cavarellit, F. Martin, G. Dirheimer, D. Morast, J. Gangloff, *Proc. Natl. Acad. Sci. USA* **1993**, *90*, 10816–10820. https://doi.org/10.1073/pnas.90.22.10816

[11] A. Aberg, A. Yaremchuk, M. Tukalo, B. Rasmussen, S. Cusack, *Biochemistry* **1997**, 3084–3094. https://doi.org/10.1021/bi9618373

[12] T. F. Smith, H. Hartman, *FEBS Lett.* **2015**, *589*, 3499–3507. https://doi.org/10.1016/j.febslet.2015.10.006

[13] A. Chaliotis, P. Vlastaridis, D. Mossialos, M. Ibba, H. D. Becker, C. Stathopoulos, G. D. Amoutzias, *Nucleic Acids Res.* **2017**, *45*, 1059–1068. https://doi.org/10.1093/nar/gkw1182

[14] P. Schimmel, R. Giege, D. Morast, S. Yokoyamat, *Proc. Natl. Acad. Sci. USA* **1993**, *90*, 8763–8768. https://doi.org/10.1073/pnas.90.19.8763

[15] D. Hipps, K. Shibat, B. Henderson, P. Schimmel, *Proc. Natl. Acad. Sci. U. S. A.* **1995**, *92*, 5550–5552. https://doi.org/10.1073/pnas.92.12.5550

[16] P. Schimmel, *J. Mol. Evol.* **1995**, *40*, 531–536. https://doi.org/10.1007/BF00166621

[17] P. Schimmel, L. Ribas de Pouplana, *Cell* **1995**, *81*, 983–986. https://doi.org/10.1016/S0092-8674(05)80002-9

[18] L. R. De Pouplana, P. Schimmel, *J. Biol. Chem.* **2001**, *276*, 6881–6884. https://doi.org/10.1074/jbc.R000032200

[19] L. Li, V. Weinreb, C. Francklyn, C. W. Carter, *J. Biol. Chem.* **2011**, *286*, 10387–10395. https://doi.org/10.1074/jbc.M110.198929

[20] L. Li, C. W. Carter, *J. Biol. Chem.* **2013**, *288*, 34736–34745. https://doi.org/10.1074/jbc.M113.510958

[21] C. W. Carter, L. Li, V. Weinreb, M. Collier, K. Gonzalez-Rivera, M. Jimenez-Rodriguez, O. Erdogan, B. Kuhlman, X. Ambroggio, T. Williams, S. N. Chandrasekharan, *Biol. Direct* **2014**, *9*. 11. https://doi.org/10.1186/1745-6150-9-11

[22] J. J. Hobson, Z. Li, H. Hu, C. W. Carter, *Int. J. Mol. Sci.* **2022**, *23*, https://doi.org/10.3390/ijms23084229

[23] G. Q. Tang, J. J. H. Elder, J. Douglas, C. W. Carter, *Nucleic Acids Res.* **2023**, *51*, 8070–8084. https://doi.org/10.1093/nar/gkad590

[24] S. K. Patra, J. Douglas, P. R. Wills, L. Betts, T. G. Qing, C. W. Carter, *Nucleic Acids Res.* **2024**, *52*, 13305–13324. https://doi.org/10.1093/nar/gkae992

[25] F. C. Wong, P. J. Beuning, C. Silvers, K. Musier-Forsyth, *J. Biol. Chem.* **2003**, *278*, 52857–52864. https://doi.org/10.1074/jbc.M309627200

[26] S. Hati, B. Ziervogel, J. Sternjohn, F. C. Wong, M. C. Nagan, A. E. Rosen, P. G. Siliciano, J. W. Chihade, K. Musier-Forsyth, *J. Biol. Chem.* **2006**, *281*, 27862–27872. https://doi.org/10.1074/jbc.M605856200

[27] F. Borel, C. Vincent, R. Leberman, M. Hartlein, *Nucleic Acids Res.* **1994**, *22*, 2963–2969. https://doi.org/10.1093/nar/22.15.2963
[28] J. L. Hsu, S. B. Rho, K. M. Vannella, S. A. Martinis, *J. Biol. Chem.* **2006**, *281*, 23075–23082. https://doi.org/10.1074/jbc.M601606200
[29] L. Pang, V. Zanki, S. V. Strelkov, A. Van Aerschot, I. Gruic-Sovulj, S. D. Weeks, *Commun. Biol.* **2022**, *5*, https://doi.org/10.1038/s42003-022-03825-8
[30] L. Lin, S. P. Hale, P. Schimmel, *Nature* **1996**, *384*, 33–34. https://doi.org/10.1038/384033b0
[31] L. F. Silvian, J. Wang, T. A. Steitz, *Science* **1999**, *285*, 1074–1077. https://doi.org/10.1093/nar/gkac207
[32] I. Zivkovic, K. Ivkovic, N. Cvetesic, A. Marsavelski, I. Gruic-Sovulj, *Nucleic Acids Res.* **2022**, *50*, 4029–4041. https://doi.org/10.1093/nar/gkac207
[33] T. L. Hendrickson, T. K. Nomanbhoy, de Crecy-Lagard, S. Fukai, O. Nureki, S. Yokoyama, P. Schimmel, *Mol. Cell* **2002**, *9*, 353–362. https://doi.org/10.1016/S1097-2765(02)00449-5
[34] O. Nureki, D. G. Vassylyev, M. Tateno, A. Shimada, T. Nakama, S. Fukai, M. Konno, T. L. Hendrickson, P. Schimmel, S. Yokoyama, *Science* **1998**, *280*, 578–582. https://doi.org/10.1126/science.280.5363.578
[35] K. Nishihara, M. Kanemori, M. Kitagawa, H. Yanagi, T. Yura, *Appl. Environ. Microbiol.* **1998**, *64*, 1694–1699. https://doi.org/10.1128/AEM.64.5.1694-1699.1998
[36] K. Nishihara, M. Kanemori, H. Yanagi, T. Yura, *Appl. Environ. Microbiol.* **2000**, *66*, 884–889. https://doi.org/10.1128/AEM.66.3.884-889.2000
[37] M. Dulic, N. Cvetesic, J. J. Perona, I. Gruic-Sovulj, *J. Biol. Chem.* **2010**, *285*, 23799–23809. https://doi.org/10.1074/jbc.M110.133553
[38] N. Cvetesic, J. J. Perona, I. Gruic-Sovulj, *J. Biol. Chem.* **2012**, *287*, 25381–25394. https://doi.org/10.1074/jbc.M112.372151
[39] V. Zanki, B. Bozic, M. Mocibob, N. Ban, I. Gruic-Sovulj, *Protein Sci.* **2022**, *31*, https://doi.org/10.1002/pro.4418
[40] R. Calendar, P. Berg, *Biochemistry* **1966**, *5*, 1681–1690. https://doi.org/10.1021/bi00869a033
[41] F. X. Cole, P. R. Schimmel, *Biochemistry* **1970**, *9*, 480–489. https://doi.org/10.1021/bi00805a005
[42] C. S. Francklyn, E. A. First, J. J. Perona, Y. M. Hou, *Methods* **2008**, *44*, 100–118. https://doi.org/10.1016/j.ymeth.2007.09.007
[43] I. Živković, M. Dulic, P. Kozulic, M. Mocibob, I. Gruic-Sovulj, *FEBS Open Bio* **2025**, *15*, 580–586. https://doi.org/10.1002/2211-5463.13903
[44] N. Cvetesic, I. Gruic-Sovulj, *Methods* **2017**, *113*, 13–26. https://doi.org/10.1016/j.ymeth.2016.09.015
[45] J. Abramson, J. Adler, J. Dunger, R. Evans, T. Green, A. Pritzel, O. Ronneberger, L. Willmore, A. J. Ballard, J. Bambrick, S. W. Bodenstein, D. A. Evans, C. C. Hung, M. O'Neill, D. Reiman, K. Tunyasuvunakool, Z. Wu, A. Žemgulytė, E. Arvaniti, C. Beattie, O. Bertolli, A. Bridgland, A. Cherepanov, M. Congreve, A. I. Cowen-Rivers, A. Cowie, M. Figurnov, F. B. Fuchs, H. Gladman, R. Jain, Y. A. Khan, C. M. R. Low, K. Perlin, A. Potapenko, P. Savy, S. Singh, A. Stecula, A. Thillaisundaram, C. Tong, S. Yakneen, E. D. Zhong, M. Zielinski, A. Žídek, V. Bapst, P. Kohli, M. Jaderberg, D. Hassabis, J. M. Jumper, *Nature* **2024**, *630*, 493–500. https://doi.org/10.1038/s41586-024-07487-w
[46] A. Goldenzweig, M. Goldsmith, S. E. Hill, O. Gertman, P. Laurino, Y. Ashani, O. Dym, T. Unger, S. Albeck, J. Prilusky, R. L. Lieberman, A. Aharoni, I. Silman, J. L. Sussman, D. S. Tawfik, S. J. Fleishman, *Mol. Cell* **2016**, *63*, 337–346. https://doi.org/10.1016/j.molcel.2016.06.012
[47] M. Lebendiker, T. Danieli, *FEBS Lett.* **2014**, *588*, 236–246. https://doi.org/10.1016/j.febslet.2013.10.044
[48] A. Brkic, M. Leibundgut, J. Jablonska, V. Zanki, Z. Car, V. Petrovic Perokovic, A. Marsavelski, N. Ban, I. Gruic-Sovulj, *Nat. Commun.* **2023**, *14*, 5498. https://doi.org/10.1038/s41467-023-41244-3
[49] I. Zivkovic, I. Gruic-Sovulj, *Biochem. Soc. Trans.* **2024**, *52*, 1109–1120. https://doi.org/10.1042/BST20230581
[50] J. Sočan, G. V. Isaksen, B. O. Brandsdal, J. Åqvist, *Sci. Rep.* **2019**, *9*, 19147. https://doi.org/10.1038/s41598-019-55697-4
[51] R. M. Bhaskara, N. Srinivasan, *Sci. Rep.* **2011**, *1*, 40. https://doi.org/10.1038/srep00040
[52] L. Li, C. Francklyn, C. W. Carter, *J. Biol. Chem.* **2013**, *288*, 26856–26863. https://doi.org/10.1074/jbc.M113.496125