Structure
Short Article

Internally Deleted Human tRNA Synthetase Suggests Evolutionary Pressure for Repurposing

Zhiwen Xu,[1,3,5] Zhiyi Wei,[1,2,5] Jie J. Zhou,[1,3] Fei Ye,[2] Wing-Sze Lo,[1,3] Feng Wang,[1,3] Ching-Fun Lau,[1,3] Jingjing Wu,[2] Leslie A. Nangle,[4] Kyle P. Chiang,[4] Xiang-Lei Yang,[1] Mingjie Zhang,[1,2] and Paul Schimmel[1,*]
[1]IAS HKUST - Scripps R&D Laboratory, Institute for Advanced Study
[2]Division of Life Science, State Key Laboratory of Molecular Neuroscience
Hong Kong University of Science and Technology, Clear Water Bay, Kowloon, Hong Kong, China
[3]Pangu Biopharma, Edinburgh Tower, The Landmark, 15 Queen's Road Central, Hong Kong, China
[4]aTyr Pharma, 3545 John Hopkins Court, Suite 250, San Diego, CA 92121, USA
[5]These authors contributed equally to this work
*Correspondence: schimmel@scripps.edu
http://dx.doi.org/10.1016/j.str.2012.08.001

SUMMARY

Aminoacyl-tRNA synthetases (AARSs) catalyze aminoacylation of tRNAs in the cytoplasm. Surprisingly, AARSs also have critical extracellular and nuclear functions. Evolutionary pressure for new functions might be manifested by splice variants that skip only an internal catalytic domain (CD) and link noncatalytic N- and C-terminal polypeptides. Using disease-associated histidyl-tRNA synthetase (HisRS) as an example, we found an expressed 171-amino acid protein (HisRSΔCD) that deleted the entire CD, and joined an N-terminal WHEP to the C-terminal anticodon-binding domain (ABD). X-ray crystallography and three-dimensional NMR revealed the structures of human HisRS and HisRSΔCD. In contrast to homodimeric HisRS, HisRSΔCD is monomeric, where rupture of the ABD's packing with CD resulted in a dumbbell-like structure of flexibly linked WHEP and ABD domains. In addition, the ABD of HisRSΔCD presents a distinct local conformation. This natural internally deleted HisRS suggests evolutionary pressure to reshape AARS tertiary and quaternary structures for repurposing.

INTRODUCTION

Although aminoacyl-tRNA synthetases (AARSs) catalyze the aminoacylation of tRNAs in the first step of protein synthesis in the cytoplasm, numerous reports document their activities in nuclear and extracellular locations, where translation does not occur (Fu et al., 2012; Guo et al., 2010; Kim et al., 2011; Martinis and Joy Pang, 2007; Park et al., 2008, 2012; Sajish et al., 2012; Xu et al., 2012). These activities include major roles in regulating angiogenesis (Xu et al., 2012; Yao et al., 2012), inflammatory responses (Arif et al., 2009; Fu et al., 2012; Lee et al., 2012), mTOR signaling (Bonfils et al., 2012; Han et al., 2012), and tumor growth (Dorrell et al., 2007; Park et al., 2012). In at least some instances, a fragment produced by natural proteolysis is the active factor. These fragments typically remove an external N- or a C-terminal peptide and leave intact all or most of the internal catalytic domain (CD). Indeed, a CD pocket is used by a natural tryptophanyl-tRNA synthetase (TrpRS) fragment to bind to the extracellular domain of VE-cadherin on endothelial cells to prevent the assembly of blood vessels (Zhou et al., 2010). These observations suggest evolutionary pressures to expropriate AARSs for functions outside of the cytoplasm, perhaps because of their close association with the origin and creation of the genetic code, and the latter's capacity to adopt functions and species in response to changes in the environment (Giegé, 2008).

With that consideration in mind, we thought that a clear manifestation of these selective pressures and their consequences would be the appearance of forms of AARSs that could not arise from proteolysis, but rather from alternative splicing that specifically removed the internal CD and little else. The one well-studied example of an AARS splice variant, mini TrpRS, removes the N-terminal 48 amino acids and leaves the entire CD intact (Wakasugi et al., 2002). It has full catalytic activity. An exquisitely tailored deletion that only excised the CD would suggest strong selective pressures to create forms that were catalytically inactive and therefore presumably designed for repurposing. The structural consequences of an internal deletion of this sort are unknown. To investigate this question, we chose HisRS, which is associated with idiopathic inflammatory myopathies (IIM) and interstitial lung disease (ILD) (Bernstein et al., 1984; Jura et al., 2007). Our rationale was that the roots of this disease association could be related to a variant of HisRS that was designed for another function, which itself was connected to inflammatory responses.

We identified the alternatively spliced forms of the gene for human HisRS by the high-throughput deep-sequencing method, and discovered a splice variant HisRSΔCD that skips exons encoding the entire CD. This splice variant encodes an endogenously expressed protein with the N-terminal WHEP domain joined to the C-terminal ABD. It showed enriched expression in human lung tissue and interacted with Jo-1 antibodies of human myositis patients, implicating the connection to the autoimmune diseases IIM and ILD. Employing crystallographic and NMR techniques, we revealed distinct structures of human HisRS

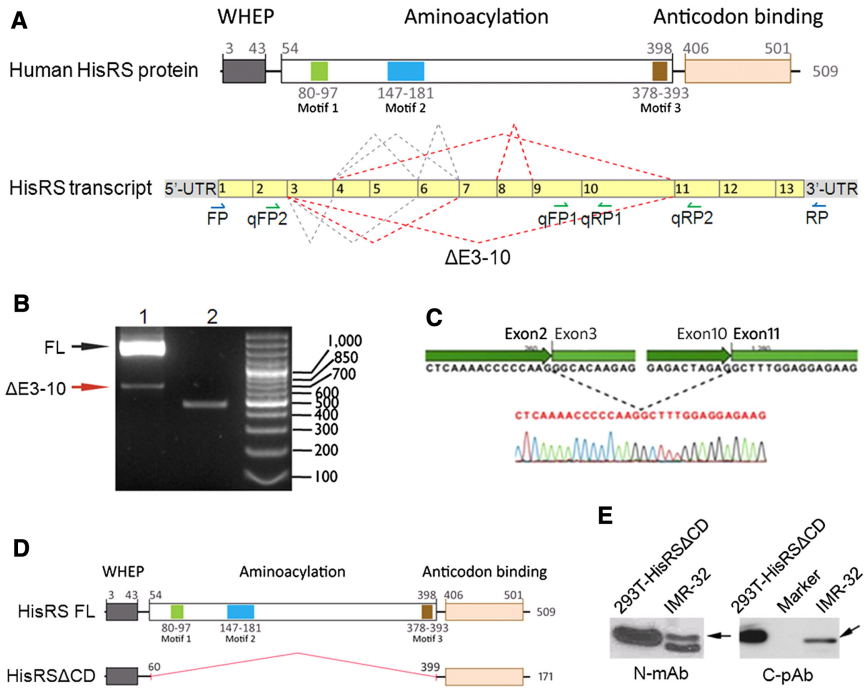

Figure 1. Identification and Validation of a HisRS Splice Variant that Skips the Entire Catalytic Domain

(A) Schematic illustration of human HisRS protein and the identified exon-skipping splicing events. Human HisRS is composed of an N-terminal WHEP domain, a core catalytic aminoacylation domain and a C-terminal anticodon binding domain. The three conserved sequence motifs in its core active site are indicated by green, blue and brown bars, respectively. The mRNA transcript of the gene for human HisRS is shown below and aligns with its encoded protein. Canonical exons are drawn in scale to the nucleotide length and are labeled consecutively. The splicing events identified by deep sequencing in the current study are illustrated by dashed lines to indicate non-canonical exon junctions. Exon junctions that were not annotated in the AceView database, including ΔE3-10 (HisRSΔCD), are indicated by red lines, and the previously reported ones are shown in gray. Targeting sites of the PCR primers are indicated by blue arrows and those of qPCR primers by green arrows. The primer sequences are shown in Table S2.

(B) Validation by PCR of the splice variant that skips exons 3 to 10. PCR was performed using cDNA of IMR-32 neuronal cells and a pair of primers targeting 5′-UTR/Exon1 (FP) and 3′-UTR (RP) of the gene for HisRS. PCR products were separate by agarose gel electrophoresis. Lane 1: PCR by FP and RP, Lane 2: PCR by primers targeting GAPDH as control.

(C) Sequencing result specifically shows the Exon2-11 junction in the HisRSΔCD transcript.

(D) Schematic of protein products of human HisRS FL and HisRSΔCD. The protein product of splice variant HisRSΔCD has the entire aminoacylation domain (aa61-398) removed due to skipping of exons 3 to 10 and therefore directly connects the WHEP domain to the ABD.

(E) Detection of endogenous HisRSΔCD protein by western blot analysis. HisRSΔCD protein was detected in whole extracts of IMR-32 cells using antibodies against, separately, the N- and C terminus of HisRS (N-mAb, monoclonal antibody against HisRS aa1-97; C-pAb, polyclonal antibody against HisRS C terminus). Total lysates of 293T cells transiently transfected with a HisRSΔCD construct were run in parallel with IMR-32 cell extracts to serve as a control that shows the size of HisRSΔCD. The expected running position of the HisRSΔCD protein is indicated by an arrow.

See also Figure S1 and Tables S1 and S2.

and HisRSΔCD. Different from homodimeric HisRS, HisRSΔCD is monomeric. Release of the ABD's packing with CD resulted in a dumbbell-like structure of flexibly linked WHEP and ABD domains and the ABD presents a distinct local conformation, readily allowing specific interaction partners and non-conventional biological activities. Our study extends the understanding of structure and function of the AARS family of ancient enzymes and suggests evolutionary pressure to reshape AARS tertiary and quaternary structures for repurposing.

RESULTS AND DISCUSSION

Comprehensive Identification of Alternative Splice Variants of Human HisRS by Deep Sequencing of AARS-Transcriptome Enriched cDNA

Based on its sequence, the 509 amino acid human HisRS is a class II tRNA synthetase composed of a core CD, a C-terminal anticodon binding domain, and an N-terminal WHEP domain (Figure 1A). The catalytic aminoacylation domain is shared by all class II tRNA synthetases, which have a characteristic 7-stranded β-structure and flanking α helices, with three class-defining conserved sequence motifs (Carter, 1993; Li et al., 2011). The WHEP domain is a 50 amino acid, helix-turn-helix motif present in one or more copies in tryptophanyl-, histidyl-, glutamyl-prolyl-, glycyl-, and methionyl-tRNA synthetases, and takes its name from the first three (W, H, and EP). This domain was appended to specific AARSs as the tree of life ascended from lower to higher eukaryotes. It first appeared in HisRS in nematodes and was retained ever since (Guo et al., 2010). Although the functional role of the HisRS WHEP domain remains unclear, it harbors epitope(s) targeted by autoantibodies of patients with IIM/ILD (Jura et al., 2007; Martin et al., 1995; Raben et al., 1994). Until now, no crystal structures of human HisRS or expressed splice variant proteins have been reported.

For a comprehensive identification of alternatively spliced forms of HisRS, we employed the high-throughput transcriptome sequencing technique. Because whole transcriptome sequencing limits the read-depth of the exome of individual genes, we developed an amplification-based transcriptome sequencing method for a more thorough discovery of splice variants. Previous work annotated a number of alternative mRNA variants of the gene for human HisRS by sequencing of cDNA clones (Thierry-Mieg and Thierry-Mieg, 2006). However, as annotated in the AceView database, these splice variants include truncation of the 5′ or 3′ ends, overlapping exons with different boundaries, splicing versus retention of introns, and the presence or absence of cassette exons. Also, expression of either recombinant or natural versions of these potential splice variants has never

been investigated. The RNA was reverse transcribed to cDNA by primers specific to the target gene, and then amplified using primers targeting their exon regions at positions close to the exon-exon junctions. This method allowed sensitive detection of low-abundant splice variants and was mainly designed for discovery of splice variants having exon-skipping events.

A complete and more comprehensive description of the AARS-specific transcriptome approach will be published separately. In brief, the transcriptome for HisRS was first enriched with RNA of human adult brain, fetal brain, total leukocytes, Jurkat T, Raji B lymphocytes, or THP monocytes, and then followed with deep sequencing. When compared to other human exome sequencing efforts, this method was found to significantly enhance the sequencing depth, yielding a > 1000 fold increase in sequencing reads after enrichment. The exon-skipping events in the gene for HisRS were concentrated on the region of exons 3 to 10 that encode the aminoacylation domain (Figure 1A; Table S1 available online). If these splice variants give rise to protein products, the generated HisRS isoforms are expected to have partially or completely disrupted enzymatic activity. Thus, these splice variants may be endowed with novel biological functions through new domain compositions and structures. Possibly, they could be immunogenic or associated with pathologies when abnormally regulated or secreted.

Validation and Expression Analysis of a Splice Variant HisRSΔCD that Skips the Entire Catalytic Domain

HisRSΔCD, the splice variant with the largest internal deletion, has the skipping of exons 3 to 10 (ΔE3-10) that encode the entire aminoacylation domain (Figure 1A). We found HisRSΔCD with 50 sequencing reads in human adult brain and seven reads in Jurkat T lymphocyte (Table S1). The putative protein product would carry no aminoacylation activity, but retained the N-terminal 60 amino acids and the C-terminal ABD. To further verify this splice variant and obtain a more complete sequence of its transcript, the PCR was performed. Because HisRSΔCD was identified with highest reads in the adult brain, we employed a human neuronal cell line IMR-32 for its validation. The PCR reaction with the cDNA template of mRNA from IMR-32 cells, together with a pair of primers targeting the 5′-UTR/Exon1 and 3′-UTR regions of the gene for HisRS (FP and RP), amplified a product with a size shorter than the expected band of the full-length (FL) transcript (Figures 1A and 1B; Table S2). This shorter PCR product was subjected to sequencing, and confirmed to bear the Exon2-11 junction of HisRSΔCD (Figure 1C). Based on the sequence of the PCR product, the HisRSΔCD transcript has exons 3 to 10 (1014 nt) removed, but still retains the 5′- and 3′-UTR region and the remaining exons of the FL transcript. It is therefore expected to translate into a protein with the in-frame deletion of the entire aminoacylation domain (aa61–398), and thereby join the N-terminal WHEP domain to the C-terminal ABD (Figure 1D).

We employed the SYBR green quantitative real-time PCR (qPCR) method to examine the mRNA expression level of native HisRS and of HisRSΔCD transcripts in various human tissues. We first optimized qPCR reactions to produce specific PCR products with high efficiency. All amplified products were designed to cover exon junctions and to exclude amplicons derived from intronic regions. After optimization, a pair of primers targeting Exon9 and Exon10 (qFP1 and qRP1) was used for amplification of the FL transcripts (Figure 1A; Table S2). For HisRSΔCD a pair of primers targeting Exon2 and Exon11 (qFP2 and qRP2) was employed in qPCR reactions having a short extension time (30 s), which thereby attenuated amplification of the longer FL transcript. Using the optimized qPCR reactions, the presence of the HisRS transcripts was analyzed across 13 human tissues, including those of the immune system (total leukocytes, bone marrow, spleen), circulatory system (lung, heart, kidney), digestive system (liver, pancreas, small intestine, colon), and others (thyroid, adipose cells, skeletal muscle). The FL transcript for HisRS was found to be evenly distributed among the various tissues that were tested, with no more than three times deviation from the median value (Figure S1A). In contrast, HisRSΔCD exhibited a differential mRNA expression profile, with the mRNA level highest in lung (about 11-fold above the median level, Figure S1B), and second highest in spleen (about 5-fold above the median). Although a causal relationship to IIM/ILD cannot be inferred from these results, it is nonetheless of interest to note that the mRNA expression of HisRSΔCD relative to HisRS FL is prominently higher in lung than in other tissues (Figure S1C). It has been suggested that the lung is the site of disease initiation in Jo-1 (HisRS) autoantibody associated ILD (Danoff and Casciola-Rosen, 2011).

Next, we used western blot methods to detect the HisRSΔCD splice variant. For this purpose, we used antibodies raised against the N- and, separately, the C-terminal region. In view of the relatively small amounts of the mRNAs that correspond to these splice variants, and due to the difficulty in obtaining adequate amounts of human tissues, human cell lines cultured in vitro were employed. Because the HisRSΔCD transcript was detected in the total RNA of IMR-32 cells, its protein product was probed using total cell extracts of IMR-32 cells with a monoclonal antibody raised against the N-terminus (aa1-97) of human HisRS, and a polyclonal antibody generated by a peptide from the C terminus. Both antibodies reacted with the same species having a molecular weight of about 20 kDa (Figure 1E; see also Figure S1D). The size of this protein corresponds to the recombinant HisRSΔCD protein overexpressed in HEK293T cells. Consistent with the relatively low amounts of mRNA that have been annotated, HisRSΔCD is clearly in much lower abundance (estimated as slightly above 1% of HisRS FL on western blots probed with the same antibodies).

Structure Determination of Human HisRS by X-Ray Crystallography

Crystal structures of *Escherichia coli* and *Thermus thermophilus* and of eukaryotic parasites *Trypanosoma brucei* and *Trypanosoma cruzi* in apo and histidine- or His-AMP- bound forms have been published (Aberg et al., 1997; Arnez et al., 1995; Merritt et al., 2010). All are α_2 dimers and, as expected for a class II synthetase, all have the characteristic and well conserved antiparallel β sheet fold flanked by α helices in the CD. The HisRSs all have an α/β fold in the anticodon binding domain. The adenine binding pocket and the topology of an extra domain inserted between the characteristic conserved motifs 2 and 3 of the class II AARS catalytic core is substantially different in bacterial and eukaryote parasitic forms of the enzyme. So far no structure has been reported for a higher eukaryote form of HisRS.

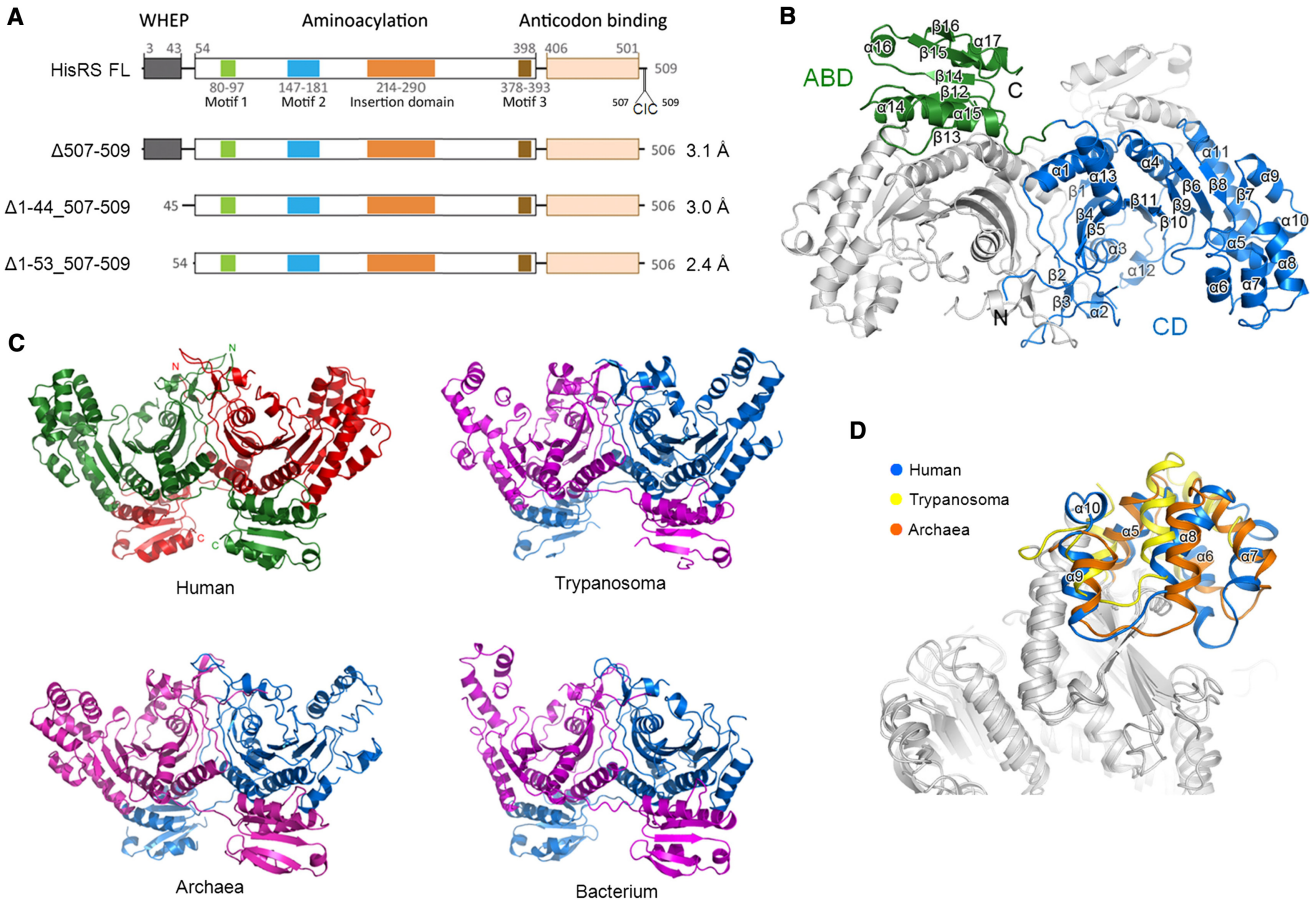

Figure 2. Structure Determination of Human HisRS by X-Ray Crystallography

(A) Optimization of the boundary of human HisRS for high quality crystals. The amino acid range included in each mutant and the corresponding crystal resolutions are shown. The three conserved sequence motifs as well as the insertion domain in the aminoacylation domain are indicated by colored bars and labeled below.

(B) Ribbon diagram of the 2.4 Å crystal structure of HisRS Δ1-53_507-509 (blue: CD, green: ABD). The secondary structure α helices and β strands are consecutively labeled, and the amino acids composing these structural elements are shown in detail in Figure S2D.

(C) Structure comparison of HisRSs of different species including human, trypanosoma (*T. brucei*, PDB: 3HRI), archaea (*T. acidophilum*, 1WU7), and bacterium (*S. aureus*, 1QE0).

(D) Superposition of the insertion domains of human, trypanosoma and archaea shows differences in the respective orientations of this domain. The α helices in the insertion domain of human HisRS are labeled, and the corresponding amino acid sequence and the alignment among different species are shown in Figure S2D.

See also Table 1 and Figure S2.

Purified recombinant human HisRS protein aggregated in the normal buffer conditions we used. By mapping cysteines in the human HisRS sequence to the known structures of *T. brucei* and *T. cruzi* HisRS, we found that two cysteines (C507 and C509) at the very C terminus may be solvent-exposed and non-specifically form intermolecular disulfide bonds. Therefore, we removed (from the HisRS-encoding recombinant gene) three residues at the C terminus (Figure 2A). Indeed, this truncation mutant Δ507-509 did not aggregate on PAGE gels, even without added reducing agent such as DTT. Using this C-terminal truncation mutant we obtained large crystals, which diffracted to 3.1 Å resolution (Figure S2A). The space group was determined to be $P4_12_12$ with the unit cell dimensions a = b = 100.4 Å, c = 257.1 Å (Table 1). The crystal structure was solved by molecular replacement using *T. brucei* HisRS (Protein Data Bank [PDB]: 3HRI) as the template.

In the crystal structure of human HisRS Δ507-509 (PDB: 4G84), the N-terminal WHEP domain is not visible, indicating that this domain is not tightly packed with the structural core (Figure S2B). The loose packing of the N-terminal domain is also supported by a previous study of the trypanosomal HisRS, which showed that its N-terminus was amenable to enzymatic cleavage during expression (Merritt et al., 2010). To further improve crystal quality, we optimized the boundary of HisRS by removing the flexible N-terminal region in different mutants and achieved a crystal structure at 2.4 Å resolution with the deletion mutant Δ1-53_507-509 (Figures 2A and S2A). The space group was determined to be $P4_12_12$ with the unit cell dimensions a = b = 93.5 Å, c = 254.5 Å (Table 1). This crystal form has the same space group and similar unit cell dimensions as the previous one, suggesting that the Δ1-53_507-509 mutant keeps the same conformation and crystal packing as the Δ507-509

Table 1. Statistics of Data Collection and Model Refinement of Human HisRS Crystal and NMR Structures

Crystal Structures of HisRS Variants			Family of 20 NMR Structures of HisRSΔCD[a]	
	HisRS Δ507-509	HisRS Δ1-53_Δ507-509	NMR Distance and Dihedral Constraints	
Data collection			Total NOE	2,397
Space group	$P4_12_12$	$P4_12_12$	Intra-residue	978
Unit cell parameters (Å)	a = b = 100.4, c = 257.1	a = b = 93.5, c = 254.5	Inter-residue	
Resolution range (Å)	50–3.1 (3.15–3.1)[b]	50–2.4 (2.44–2.4)	Sequential ($\|i-j\| = 1$)	491
No. of unique reflections	24,210 (1,191)	45,571 (2,212)	Medium-range ($\|i-j\| < 4$)	404
Redundancy	5.5 (5.6)	6.1 (5.8)	Long-range ($\|i-j\| > 5$)	929
I/σ	22.6 (2.4)	21.8 (2.3)	Hydrogen bonds	148
Completeness (%)	98.8 (99.9)	99.8 (99.9)	Total dihedral angle restraints	216
R_{merge} (%)[c]	6.9 (69.6)	7.4 (67.9)	Structure statistics	
Structure refinement			Violations (mean and s.d.)	
Resolution (Å)	50–3.1 (3.2–3.1)	50–2.4 (2.49–2.4)	Distance constraints (Å)	0.001 ± 0.001
R_{cryst} / R_{free} (%)[d]	27.1 (34.4) / 32.7 (40.0)	19.1 (26.8)/25.0 (35.1)	Dihedral angle constraints (°)	0.574 ± 0.063
Rmsd bonds (Å)/angles (°)	0.014/1.5	0.005/0.9	Deviations from idealized geometry	
Average B factor	55	54.4	Bond lengths (Å)	0.002 ± 0.000
No. of atoms			Bond angles (°)	0.429 ± 0.022
Protein atoms	6,430	6,907	Impropers (°)	0.347 ± 0.047
Water molecules		166	Ramachandran plot[e] (%)	
Other molecules		21	Most favorable regions	79.7
No. of reflections			Additional allowed regions	14.2
Working set	22,659	42,096	Generously allowed regions	4.3
Test set	1,213	2,124	Disallowed regions	1.8
Ramachandran plot			Coordinate precision	
Most favored regions (%)	88.2	92.7	Atomic rmsd[f] (Å)	
Additionally allowed (%)	11	7.3	Residues 1–45 for the WHEP domain	
Generously allowed (%)	0.8	0	Heavy	1.081
			Backbone	0.447
			Residues 69–165 for the ABD	
			Heavy	1.468
			Backbone	0.872

[a]None of the structures exhibits distance violations greater than 0.3 Å or dihedral angle violations greater than 4°.
[b]Numbers in parentheses represent the value for the highest resolution shell.
[c]$R_{merge} = \Sigma|Ii - Im|/\Sigma Ii$, where Ii is the intensity of the measured reflection and Im is the mean intensity of all symmetry related reflections.
[d]$R_{cryst} = \Sigma||Fobs| - |Fcalc||/\Sigma|Fobs|$, where Fobs and Fcalc are observed and calculated structure factors. $R_{free} = \Sigma T||Fobs| - |Fcalc||/\Sigma T|Fobs|$, where T is a test data set of about 5% of the total reflections randomly chosen and set aside prior to refinement.
[e]The program Procheck was used to assess the overall quality of the structures.
[f]The precision of the atomic coordinates is defined as the average Rmsd between 20 final structures and the mean coordinates of the protein.

mutant. Indeed, its structure is essentially identical to that of the HisRS Δ507-509 (Figures 2B and S2B), and shows a dimeric composition that agrees with the molecular weight determined by size exclusion chromatography (Figure S2C).

According to the structure of human HisRS Δ1-53_507-509 (PDB: 4G85), the overall fold of the CD and ABD of human HisRS is similar to its bacterial, archaeal and *T. brucei and T. cruzi* orthologs (Figure 2C). The most prominent difference among these structures is in the additional domain inserted between conserved motifs 2 and 3 of the class II catalytic core. This insertion domain increases in size from prokaryotes to eukaryotes, and is not conserved in either sequence or structure between prokaryotic and eukaryotic HisRSs (Figure S2D). With the exception of a missing α9 helix in the insertion domain of eukaryote parasite homologs, the secondary structure elements of human HisRS are similar to those of the parasite homologs. It was previously proposed that the insertion domain may contact the acceptor stem of the tRNA (Freist et al., 1999). Superposition of structural cores of human, *T. brucei, T. cruzi* and *T. thermophilus* HisRS also reveals the orientation difference of the insertion domains (Figure 2D).

Structure Determination of the Splice Variant HisRSΔCD by NMR Spectroscopy

Similar to native HisRS, wild-type HisRSΔCD forms oligomers even in the presence of 1 mM DTT (Figure S3A). To avoid the disulfide formation, the C-terminal Cys169 and Cys171 (corresponding to C507 and C509 in HisRS) were mutated to serines

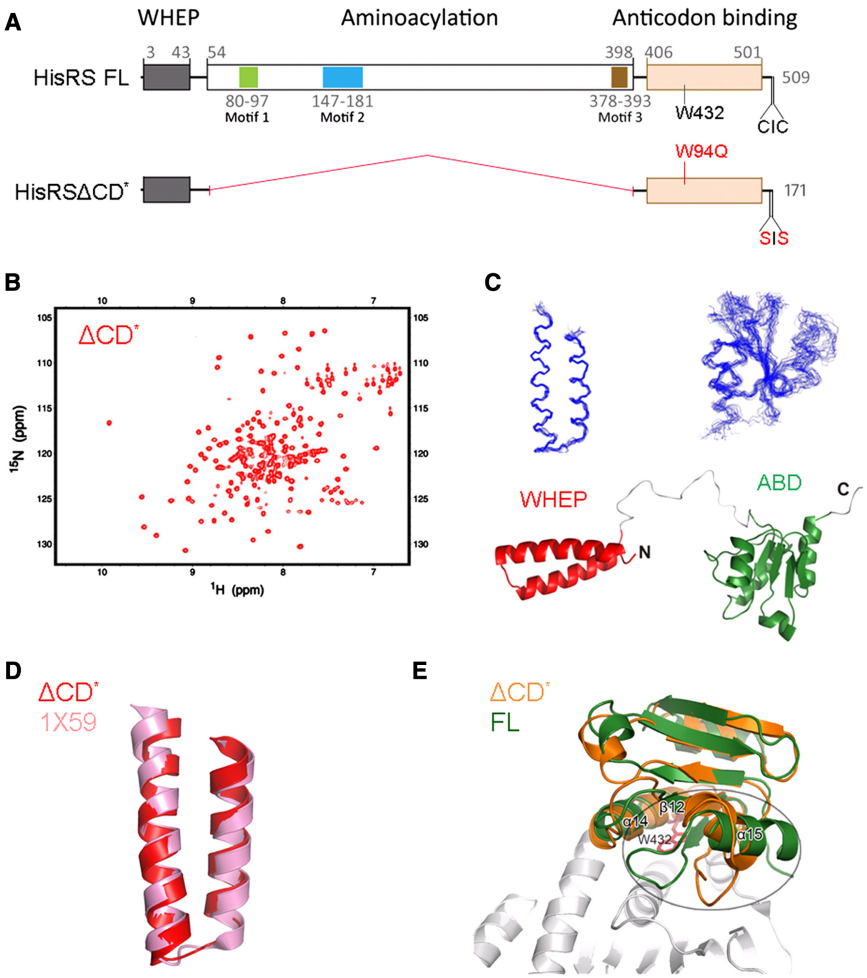

Figure 3. Structure Determination of the Splice Variant HisRSΔCD by NMR Spectroscopy

(A) Schematic of HisRSΔCD* (2C2S_W94Q) mutant employed for structural characterizations. The mutational sites are labeled in red and the corresponding C507, C509, and W432 residues in the native HisRS sequence are also indicated.

(B) The ^{1}H-^{15}N HSQC spectrum of HisRSΔCD* used for structure determination.

(C) Backbone superimposition of 20 calculated lowest-energy structures of the WHEP domain and the ABD of HisRSΔCD* are shown, and the HisRSΔCD* structure is presented below by ribbon diagram. WHEP domain (red) and ABD (green) are well-folded and linked by a flexible loop.

(D) Superposition of NMR structures of HisRSΔCD* and of the WHEP domain alone (PDB: 1X59). Structures are shown in ribbon diagram. Red: WHEP domain of HisRSΔCD*, pink: 1X59.

(E) Superposition of human HisRS Δ1-53_507-509 crystal structure and HisRSΔCD* NMR structure. The W432 in HisRS FL was labeled in red. The circled area including helix α15 and the preceding loop had the most prominent differences (see also Figure S2D). Structures are shown in ribbon diagram. Orange: ABD of HisRSΔCD*, Green: ABD of HisRS FL, gray: CD of HisRS FL.

See also Table 1 and Figures S2 and S3.

(2C2S, Figure 3A). The HisRSΔCD_2C2S proteins were mostly monomeric in solution (Figures S3A and S3B), and the ^{1}H-^{15}N heteronuclear single quantum coherence (HSQC) spectrum showed an increase in peak count and more uniform peak shape compared to that of the wild-type HisRSΔCD (Figures S3C and S3D). But the peak number was still less than expected and, together with the presence of broadened peaks, we concluded that this protein still had nonspecific interactions and was not sufficiently homogeneous for structure determination. Based on the solved crystal structure of HisRS, it seemed likely that the absence of the CD in HisRSΔCD exposed the hydrophobic inter-domain interface of the ABD, leading to non-specific hydrophobic interactions and thereby introducing heterogeneity. To decrease the hydrophobicity, Trp94 (corresponding to Trp432 in the numbering of the sequence of HisRS and located at the center of the hydrophobic interface) was substituted by the more hydrophilic Gln (Figure 3A). This substitution greatly improved the protein homogeneity as demonstrated by the ^{1}H-^{15}N HSQC spectrum, which displays well dispersed peaks with a peak yield > 95% (Figure 3B). In addition, comparing HisRSΔCD 2C2S_W94Q mutant (HisRSΔCD*) to the wild-type and 2C_2S mutant, the shared HSQC peaks of the proteins exhibited no obvious chemical shifts, thus indicating that the substitutions did not alter the protein conformation (Figures S3C and S3E). Therefore, the HisRSΔCD* mutant was used for further structural characterizations.

We first tried to solve the structure of HisRSΔCD* by X-ray crystallography, but extensive trials failed to obtain well diffracting crystals. Because the ^{15}N-^{1}H HSQC spectrum of HisRSΔCD* is well dispersed and contains only one set of peaks, showing that this protein forms a well-folded structure in solution, we could determine its structure to a high resolution by NMR spectroscopy (Figure 3B; Table 1; PDB: 2LW7). In contrast to homodimeric HisRS, HisRSΔCD* behaved as a monomer in solution, which is probably due to the lack of the CD that facilitates dimerization of HisRS. The ensemble of 20 NMR structures of HisRSΔCD* are well defined, with a root-mean-square deviation (RMSD) of 0.447 Å for backbone atoms and 1.081 Å for heavy atoms of the WHEP domain, and a RMSD of 0.872 Å for backbone atoms and 1.468 Å for heavy atoms of the ABD (Figure 3C). Its WHEP domain adopts an antiparallel bi-helical structure. (An NMR structure of a 73 amino acid peptide, comprised of the first 60 amino acids of human HisRS and 13 amino acids derived from a recombinant plasmid construct that was used, is documented in the PDB [1X59]. According to our solution structure, the WHEP domain of this peptide is identical to that of the WHEP domain of HisRSΔCD* that is shown in Figure 3D.) The ABD of HisRSΔCD* forms a compact mixed α/β fold. The WHEP and ABD domains are connected by a 27 amino acid, highly flexible linker (Figure 3C). No long-distance NOE couplings between the residues of the two domains were found in the NMR spectra. The lack of couplings indicates that the two

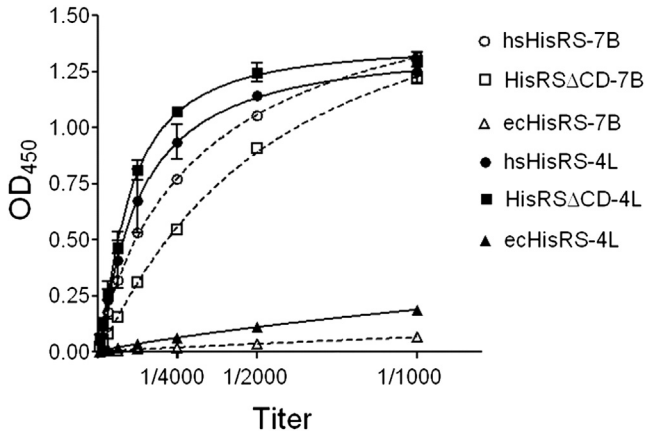

Figure 4. Possible Association of HisRSΔCD with IIM/ILD

Jo-1 antibodies from two different IIM patients react with recombinant human HisRS FL (hsHisRS) and HisRSΔCD, but not with *E. coli* HisRS (ecHisRS). The optical density at 450 nm was used to monitor the formation of antibody complexes in the ELISA. The "7B" stands for lot 7B04507 of Jo-1 antibodies and "4L" stands for lot 4L34811. The titration assay by ELISA was performed in duplicate and data are represented as mean ± SEM.

domains do not make contacts. As a further support, we purified the ABD (aa398-506) alone and, when comparing its HSQC spectrum with that of the ABD in HisRSΔCD*, these amino acids showed largely the same chemical shifts (Figure S3F). Thus, HisRSΔCD* appears as a dumbbell-like structure with "free-floating" N- and C-terminal domains. This structure appears poised to interact with two different partners and, with the flexible linker, to be able to adapt to a variable separation of these partners.

In HisRSΔCD*, the packing interactions of the ABD with the CD have been released. Comparing the ABDs in the NMR structure of HisRSΔCD* with that of the crystal structure of HisRS Δ1-53_507-509, the overall folds remain the same (Figure 3E). However, a prominent difference was found at helix α15 and the loop preceding it (Figure S2D), which are rigidly packed with the CD in HisRS. In HisRSΔCD* this region becomes flexible and moves inward, due to the lack the interaction with the CD.

Possible Association of HisRSΔCD with IIM and ILD

As stated earlier, HisRS is associated with IIM and ILD. HisRS or its constituent peptides have been implicated in the etiology of these diseases. For patients with IIM or ILD, Jo-1 autoantibodies target, in part, the N-terminal region of HisRS (Jura et al., 2007; Martin et al., 1995; Raben et al., 1994). As might be expected from its structure, we observed that Jo-1 antibodies from patients react with HisRSΔCD (Figure 4). In addition, Granzyme B cleavage of autoantigens is suggested to contribute to the initiation and propagation of autoimmunity, and it is noteworthy that HisRS was amenable to Granzyme B cleavage in vitro at the N-terminal sequence LGPD[48], a tetrapeptide that is also encompassed by HisRSΔCD (Darrah and Rosen, 2010; Levine et al., 2007). While the purpose of this study was not to investigate the etiology of IIM and ILD that is associated with HisRS, these results suggest that HisRSΔCD should at least be further considered in this regard.

Conclusions

We discovered an internal deletion splice variant of human HisRS that completely and precisely ablates the CD and little else. From investigations of the structures of HisRSΔCD and its parent HisRS, three prominent differences were noted. First, in contrast to the dimeric structure shared by HisRSs throughout evolution, HisRSΔCD is a monomer. Second, the tight domain-domain packing interactions seen in HisRS are absent in HisRSΔCD, which presents two "free-floating" domains loosely tethered together by a 27 amino acid flexible linker. Third, the release of packing interactions for the ABD of HisRSΔCD engenders a specific conformational change in the ABD. The existence of an AARS variant that lacks a CD, and that has undergone demonstrable changes in the tertiary and quaternary structures, is highly suggestive of strong selective pressures to repurpose HisRS.

EXPERIMENTAL PROCEDURES

The exon-skipping splicing events in the gene for human HisRS were identified by deep sequencing of AARS-transcriptome enriched cDNA of human tissues and cells. The splice variant HisRSΔCD was validated by PCR, and the endogenous protein was detected by western blot analysis. The mRNA expression level of HisRSΔCD in various human tissues was analyzed by the qPCR method. Recombinant human HisRS, HisRSΔCD, and their variants were overexpressed in *E. coli*, purified, and subjected to structural characterizations using X-ray crystallography and NMR techniques. The binding of the HisRS proteins to patient Jo-1 antibodies were analyzed by an enzyme-linked immunosorbent assay (ELISA). The detailed description of the experimental procedures is provided in Supplemental Experimental Procedures.

ACCESSION NUMBERS

The atomic coordinates and structure factors of HisRS Δ507-509 and Δ1-53_507-509 have been deposited in the Protein Data Bank under the accession codes 4G84 and 4G85, respectively. The atomic coordinates, chemical shifts and restraints of HisRSΔCD* have been deposited in the Protein Data Bank under the accession code 2LW7.

SUPPLEMENTAL INFORMATION

Supplemental Information includes three figures, two tables, and Supplemental Experimental Procedures and can be found with this article online at http://dx.doi.org/10.1016/j.str.2012.08.001.

ACKNOWLEDGMENTS

We thank the BL17U1 beamline of the Shanghai Synchrotron Radiation Facility for the X-ray beamline time. The work was supported in part by the Innovation and Technology Fund of Hong Kong (UIM181, UIM192, and UIM199); aTyr Pharma; Grants GM 15539, GM 23562, GM 88278, and CA 92577 from the NIH; and by a fellowship from the National Foundation for Cancer Research. We also thank Drs. Elisabeth Gardiner and Melissa Ashlock for helpful comments and advice. P.S., M.Z., and X.Y. have a financial benefit from aTyr Pharma in the form of compensation or stock ownership. This work was partially supported by a grant from aTyr Pharma. The rest of the authors have affiliations that are annotated, and these affiliations include aTyr Pharma and Pangu Biopharma.

Received: June 18, 2012
Revised: July 28, 2012
Accepted: August 3, 2012
Published: September 4, 2012

REFERENCES

Aberg, A., Yaremchuk, A., Tukalo, M., Rasmussen, B., and Cusack, S. (1997). Crystal structure analysis of the activation of histidine by Thermus thermophilus histidyl-tRNA synthetase. Biochemistry *36*, 3084–3094.

Arif, A., Jia, J., Mukhopadhyay, R., Willard, B., Kinter, M., and Fox, P.L. (2009). Two-site phosphorylation of EPRS coordinates multimodal regulation of noncanonical translational control activity. Mol. Cell *35*, 164–180.

Arnez, J.G., Harris, D.C., Mitschler, A., Rees, B., Francklyn, C.S., and Moras, D. (1995). Crystal structure of histidyl-tRNA synthetase from Escherichia coli complexed with histidyl-adenylate. EMBO J. *14*, 4143–4155.

Bernstein, R.M., Morgan, S.H., Chapman, J., Bunn, C.C., Mathews, M.B., Turner-Warwick, M., and Hughes, G.R.V. (1984). Anti-Jo-1 antibody: a marker for myositis with interstitial lung disease. Br. Med. J. (Clin. Res. Ed.) *289*, 151–152.

Bonfils, G., Jaquenoud, M., Bontron, S., Ostrowicz, C., Ungermann, C., and De Virgilio, C. (2012). Leucyl-tRNA synthetase controls TORC1 via the EGO complex. Mol. Cell *46*, 105–110.

Carter, C.W., Jr. (1993). Cognition, mechanism, and evolutionary relationships in aminoacyl-tRNA synthetases. Annu. Rev. Biochem. *62*, 715–748.

Danoff, S.K., and Casciola-Rosen, L. (2011). The lung as a possible target for the immune reaction in myositis. Arthritis Res. Ther. *13*, 230.

Darrah, E., and Rosen, A. (2010). Granzyme B cleavage of autoantigens in autoimmunity. Cell Death Differ. *17*, 624–632.

Dorrell, M.I., Aguilar, E., Scheppke, L., Barnett, F.H., and Friedlander, M. (2007). Combination angiostatic therapy completely inhibits ocular and tumor angiogenesis. Proc. Natl. Acad. Sci. USA *104*, 967–972.

Freist, W., Verhey, J.F., Rühlmann, A., Gauss, D.H., and Arnez, J.G. (1999). Histidyl-tRNA synthetase. Biol. Chem. *380*, 623–646.

Fu, G., Xu, T., Shi, Y., Wei, N., and Yang, X.L. (2012). tRNA-controlled nuclear import of a human tRNA synthetase. J. Biol. Chem. *287*, 9330–9334.

Giegé, R. (2008). Toward a more complete view of tRNA biology. Nat. Struct. Mol. Biol. *15*, 1007–1014.

Guo, M., Yang, X.L., and Schimmel, P. (2010). New functions of aminoacyl-tRNA synthetases beyond translation. Nat. Rev. Mol. Cell Biol. *11*, 668–674.

Han, J.M., Jeong, S.J., Park, M.C., Kim, G., Kwon, N.H., Kim, H.K., Ha, S.H., Ryu, S.H., and Kim, S. (2012). Leucyl-tRNA synthetase is an intracellular leucine sensor for the mTORC1-signaling pathway. Cell *149*, 410–424.

Jura, M., Rychlewski, L., and Barciszewski, J. (2007). Comprehensive insight into human aminoacyl-tRNA synthetases as autoantigens in idiopathic inflammatory myopathies. Crit. Rev. Immunol. *27*, 559–572.

Kim, S., You, S., and Hwang, D. (2011). Aminoacyl-tRNA synthetases and tumorigenesis: more than housekeeping. Nat. Rev. Cancer *11*, 708–718.

Lee, P.S., Zhang, H.M., Marshall, A.G., Yang, X.L., and Schimmel, P. (2012). Uncovering of a short internal peptide activates a tRNA synthetase procytokine. J. Biol. Chem. *287*, 20504–20508.

Levine, S.M., Raben, N., Xie, D., Askin, F.B., Tuder, R., Mullins, M., Rosen, A., and Casciola-Rosen, L.A. (2007). Novel conformation of histidyl-transfer RNA synthetase in the lung: the target tissue in Jo-1 autoantibody-associated myositis. Arthritis Rheum. *56*, 2729–2739.

Li, L., Weinreb, V., Francklyn, C., and Carter, C.W., Jr. (2011). Histidyl-tRNA synthetase urzymes: Class I and II aminoacyl tRNA synthetase urzymes have comparable catalytic activities for cognate amino acid activation. J. Biol. Chem. *286*, 10387–10395.

Martin, A., Shulman, M.J., and Tsui, F.W. (1995). Epitope studies indicate that histidyl-tRNA synthetase is a stimulating antigen in idiopathic myositis. FASEB J. *9*, 1226–1233.

Martinis, S.A., and Joy Pang, Y.L. (2007). Jekyll & Hyde: evolution of a superfamily. Chem. Biol. *14*, 1307–1308.

Merritt, E.A., Arakaki, T.L., Gillespie, J.R., Larson, E.T., Kelley, A., Mueller, N., Napuli, A.J., Kim, J., Zhang, L., Verlinde, C.L.M.J., et al. (2010). Crystal structures of trypanosomal histidyl-tRNA synthetase illuminate differences between eukaryotic and prokaryotic homologs. J. Mol. Biol. *397*, 481–494.

Park, M.C., Kang, T., Jin, D., Han, J.M., Kim, S.B., Park, Y.J., Cho, K., Park, Y.W., Guo, M., He, W., et al. (2012). Secreted human glycyl-tRNA synthetase implicated in defense against ERK-activated tumorigenesis. Proc. Natl. Acad. Sci. USA *109*, E640–E647.

Park, S.G., Schimmel, P., and Kim, S. (2008). Aminoacyl tRNA synthetases and their connections to disease. Proc. Natl. Acad. Sci. USA *105*, 11043–11049.

Raben, N., Nichols, R., Dohlman, J., McPhie, P., Sridhar, V., Hyde, C., Leff, R., and Plotz, P. (1994). A motif in human histidyl-tRNA synthetase which is shared among several aminoacyl-tRNA synthetases is a coiled-coil that is essential for enzymatic activity and contains the major autoantigenic epitope. J. Biol. Chem. *269*, 24277–24283.

Sajish, M., Zhou, Q.S., Kishi, S., Valdez, D.M., Jr., Kapoor, M., Guo, M., Lee, S., Kim, S., Yang, X.L., and Schimmel, P. (2012). Trp-tRNA synthetase bridges DNA-PKcs to PARP-1 to link IFN-γ and p53 signaling. Nat. Chem. Biol. *8*, 547–554.

Thierry-Mieg, D., and Thierry-Mieg, J. (2006). AceView: a comprehensive cDNA-supported gene and transcripts annotation. Genome Biol. *7* (*Suppl 1*), S12, 1–14.

Wakasugi, K., Slike, B.M., Hood, J., Otani, A., Ewalt, K.L., Friedlander, M., Cheresh, D.A., and Schimmel, P. (2002). A human aminoacyl-tRNA synthetase as a regulator of angiogenesis. Proc. Natl. Acad. Sci. USA *99*, 173–177.

Xu, X.L., Shi, Y., Zhang, H.M., Swindell, E.C., Marshall, A.G., Guo, M., Kishi, S., and Yang, X.L. (2012). Unique domain appended to vertebrate tRNA synthetase is essential for vascular development. Nat. Commun. Published online February 21, 2012. http://dx.doi.org/10.1038/ncomms1686.

Yao, P., Potdar, A.A., Arif, A., Ray, P.S., Mukhopadhyay, R., Willard, B., Xu, Y.C., Yan, J., Saidel, G.M., and Fox, P.L. (2012). Coding region polyadenylation generates a truncated tRNA synthetase that counters translation repression. Cell *149*, 88–100.

Zhou, Q.S., Kapoor, M., Guo, M., Belani, R., Xu, X.L., Kiosses, W.B., Hanan, M., Park, C., Armour, E., Do, M.H., et al. (2010). Orthogonal use of a human tRNA synthetase active site to achieve multifunctionality. Nat. Struct. Mol. Biol. *17*, 57–61.