Published online 14 January 2016 *Nucleic Acids Research, 2016, Vol. 44, No. 3* ***1247–1255***
doi: 10.1093/nar/gkw002

Alternative splicing creates two new architectures for human tyrosyl-tRNA synthetase

Zhiyi Wei[1,2,†], Zhiwen Xu[1,3,†], Xiaotian Liu[4], Wing-Sze Lo[1,3], Fei Ye[4], Ching-Fun Lau[1,3], Feng Wang[1,3], Jie J. Zhou[1,3], Leslie A. Nangle[5], Xiang-Lei Yang[1,6], Mingjie Zhang[1,4] and Paul Schimmel[1,7,8,*]

[1]IAS HKUST - Scripps R&D Laboratory, Institute for Advanced Study, Hong Kong University of Science and Technology, Clear Water Bay, Kowloon, Hong Kong, China, [2]Departmentof Biology, South University of Science and Technology of China, Shenzhen, Guangdong 518055, China, [3]Pangu Biopharma, Edinburgh Tower, The landmark, 15 Queen'sRoad Central, Hong Kong, China, [4]Division of Life Science, State Key Laboratory of Molecular Neuroscience Hong Kong University of Science and Technology, Clear Water Bay, Kowloon, Hong Kong, China, [5]aTyr Pharma, 3545 John Hopkins Court, Suite 250, San Diego, CA 92121, USA, [6]The Scripps Laboratories for tRNA Synthetase Research and the Departments of Chemical Physiology and of Cell and Molecular Biology, The Scripps Research Institute, La Jolla, CA 92037, USA, [7]The Scripps Laboratories for tRNA Synthetase Research and the Departments of Cell and Molecular Biology, and Chemistry, The Skaggs Institute for Chemical Biology, The Scripps Research Institute, La Jolla, CA 92037, USA and [8]The Scripps Laboratories for tRNA Synthetase Research and Departments of Metabolism & Aging, The Scripps Research Institute, Jupiter, FL 33458, USA

Received September 21, 2015; Revised December 12, 2015; Accepted January 03, 2016

ABSTRACT

Many human tRNA synthetases evolved alternative functions outside of protein synthesis. These functions are associated with over 200 splice variants (SVs), most of which are catalytic nulls that engender new biology. While known to regulate non-translational activities, little is known about structures resulting from natural internal ablations of any protein. Here, we report analysis of two closely related, internally deleted, SVs of homodimeric human tyrosyl-tRNA synthetase (TyrRS). In spite of both variants ablating a portion of the catalytic core and dimer-interface contacts of native TyrRS, each folded into a distinct stable structure. Biochemical and nuclear magnetic resonance (NMR) analysis showed that the internal deletion of TyrRSΔE2–4 SV gave an alternative, neomorphic dimer interface 'orthogonal' to that of native TyrRS. In contrast, the internal C-terminal splice site of TyrRSΔE2–3 prevented either dimerization interface from forming, and yielded a predominantly monomeric protein. Unlike ubiquitous TyrRS, the neomorphs showed clear tissue preferences, which were distinct from each other. The results demonstrate a sophisticated structural plasticity of a human tRNA synthetase for architectural reorganizations that are preferentially elicited in specific tissues.

INTRODUCTION

Recent investigations revealed essential non-translational functions of aminoacyl tRNA synthetases (AARS) in cytoplasmic as well as in nuclear and extracellular locations (1–11). These activities include major roles in regulating angiogenesis (11,12), inflammatory responses (1,13,14), mTOR signaling (15,16) and tumor growth (6,17). Importantly, alternative splicing creates over 200 new protein identities in the human AARS family (18). Most are catalytic nulls, i.e. they are catalytically inactive by virtue of deleted active site residues. These variants have been shown to have diverse extra-cellular, nuclear and cytoplasmic functions, which are idiosyncratic to the variant. Using known structures of native AARS, molecular modeling suggested that internally deleted splice variants (SVs) had major rearrangements. In spite of this, over 100 were expressed and purified as recombinant proteins, demonstrating that, in spite of any rearrangements, they folded into stable structures. The ability to form a stable structure was further confirmed by a high-resolution structural analysis of a SV of human histidyl-tRNA synthetase (HisRS) (19).

To better understand these structural rearrangements, we focused on two catalytic nulls of human TyrRS conserved between human and mouse. These are designated TyrRSΔE2–4 and TyrRSΔE2–3, which delete exons 2–4

*To whom correspondence should be addressed. Tel: +852 23585022; Fax: +852 27198158; Email: schimmel@scripps.edu
†These authors contributed equally to this work as the first authors.

and 2–3, respectively. Because these SVs ablate internal regions of the protein, they likely engender new conformations. Our characterization and structural analysis of the recombinant TyrRS SV proteins revealed that each had in common an ablation of the dimer interface of the native enzyme. However, the particulars of the sequences at the two distinct splice junctions engendered a dramatic reshaping that yielded distinct neomorphic structures, thus illustrating the remarkable plasticity of the tRNA synthetase architecture, which accommodates the disruptions of internal deletions.

MATERIALS AND METHODS

Deep sequencing of AARS-transcriptome enriched cDNA and identification of exon-skipping splicing events

The poly A+ RNA of Human tissues including adult brain, fetal brain and peripheral blood leukocytes were purchased from Clontech (Mountain View, CA, catalog No. 636102, 636106 and 636170). Total RNA of mouse leukemic macrophage-like RAW264.7 cells were extracted using PureLink™ RNA Mini kit (Invitrogen, Carlsbad, CA, USA), and analyzed by a NanoDrop 1000 spectrometer for quality and quantity. Genomic DNA was digested using TURBO DNase in the TURBO DNA-free Kit (Invitrogen). Messenger RNA (mRNA) was isolated from total RNA using the FastTrack MAG Maxi mRNA Isolation kit (Invitrogen).

The transcriptome of AARS genes were enriched and sequenced as previously described (18). Deep sequencing reads were mapped and counted using rSeq version 4 (20) for the number of sequencing reads mapped to alternatively spliced exon–exon junctions. Annotated exon splice sites of the AARS genes were obtained from RefGene of NCBI based on the human reference genome (NCBI version 36, hg18).

PCR (polymerase chain reaction) validation of the TyrRS splice variants

The first strand cDNA was synthesized from total RNA using oligo-dT primers. PCR (polymerase chain reaction) reaction was performed by primers targeting the 5′-UTR/Exon1 and 3′-UTR/Exon13 regions of the human TyrRS gene (FP and RP1), and the PCR products were validated by sequencing.

Isolation of cytoplasmic and polyribosomal RNA of cultured cells

The THP-1 or Jurkat T cells were grown and maintained in RPMI 1640 medium supplemented with 10% FBS and 0.5% penicillin/streptomycin. The monocytic THP-1 cells were induced by phorbol 12-myristate 13-acetate (PMA, 10 ng/ml) for differentiation into macrophagic-like cells. For purification of cytoplasmic RNA, cells were firstly lysed in RLN-lysis buffer (Qiagen, Hilden, Germany) with 0.5% IGEPAL (Sigma-Aldrich, St. Louis, MO, USA), 40 mM dithiothreitol and 500 U/ml RNase inhibitor (ABI Biosystems, Foster City, CA, USA). Nuclei were removed from the cell lysates by centrifugation at 12 000 g for 10 s at 4°C, and supernatant was applied to an RNAeasy kit (Qiagen) for purification of the cytoplasmic RNA. Polysome-bound mRNA was isolated from cultured Jurkat T-cells as previously described (18).

Detection of the TyrRS transcripts in the cytoplasmic and polyribosomal RNA was carried out by PCR using primers targeting the 5′-UTR/Exon1 and Exon5/Exon6 regions of the TyrRS gene (FP and RP2).

Detection of TyrRS proteins by Western blotting

The Jurkat T and THP-1 cells or 293T cells transiently transfected with the TyrRS SVs were lysed by 50 mM Tris buffer (pH 8.0) containing 1% Triton X-100 and 5 mM EDTA. After incubation on ice for 30 min, lysed cells were centrifuged at 22 000 g for 15 min at 4°C, and the supernatant was collected and analyzed for protein concentration by Bio-Rad Protein Assay (Bio-Rad, Hercules, CA, USA). Whole cell lysates of Jurkat T or THP-1 cells (100 μg proteins per lane) and of 293T cells (5 μg proteins per lane) were loaded onto a NuPAGE 4–12% Bis-Tris gel (Invitrogen) for electrophoresis and transferred to a nitrocellulose membrane. The membranes were stained with polyclonal antibodies directed against N-terminal aa168–197 or C-terminal aa457–487 of TyrRS (Abgent, San Diego, CA, USA).

Protein expression and purification

The cDNA encoding the TyrRS SVs TyrRSΔE2–4 and TyrRSΔE2–3 were cloned into a modified pET32 vector and fused to the N-terminal thioredoxin-His_6-tags. The deletion mutants of the SVs were created using the standard PCR-based mutagenesis method. The fusion proteins were expressed in *Escherichia coli* BL21(DE3) and purified by Ni^{2+}-NTA affinity chromatography. Next, the thioredoxin-His_6-tag was removed by protease-3C digestion. The cleaved protein mixtures were further separated by a size-exclusion chromatography in a buffer containing 50 mM Tris (pH 7.8), 100 mM NaCl, 1 mM EDTA and 1 mM DTT.

Analytical gel filtration chromatography

Analytical gel filtration chromatography was carried out on an AKTA FPLC system (GE Healthcare, San Diego, CA, USA). Proteins were loaded onto a Superose 12 10/300 GL column (GE Healthcare) equilibrated with a buffer containing 50 mM Tris (pH 7.8), 100 mM NaCl, 1 mM EDTA and 1 mM DTT.

Coiled-coil prediction

The coiled-coil regions in protein were predicted by the COILS server (21). Window sizes were set at 14, 21 and 28 amino acids.

Multi-angle laser scattering spectroscopy

Protein sample at a concentration of 50 μM was injected by 200 μl into an AKTA FPLC system with a Superose 12

10/300 GL column (GE Healthcare) with the column buffer of 50 mM Tris-HCl pH 7.8, 100 mM NaCl, 1mM EDTA and 1 mM DTT. The chromatography system was coupled to a static light-scattering detector (miniDawn, Phoenix, Wyatt, CA, USA) and a differential refractive index detector (Optilab, Wyatt, CA, USA). Data were analyzed with ASTRA 6 (Wyatt, CA, USA).

NMR (Nuclear magnetic resonance) spectroscopy

The NMR (nuclear magnetic resonance) sample contained 0.3 mM protein in 50 mM Tris buffer (pH 7.0 or pH 7.6, with 50 mM NaCl, 1 mM DTT, 1 mM EDTA) in 90% H_2O/10% D_2O (vol/vol). The heteronuclear single quantum coherence (HSQC) spectra was acquired at 30°C on the Varian Inova 750-MHz NMR spectrometer, equipped with an actively z-gradient shielded triple resonance probe. Spectra were processed using the NMRPipe software (22) and Sparky (23).

RESULTS

Annotation of alternative SVs of human and mouse TyrRS

TyrRS is a class I tRNA synthetase, characterized by a Rossmann nucleotide binding fold that forms the active site for aminoacylation. The enzyme is composed of an N-terminal core catalytic domain, followed by an anticodon binding domain, and a C-terminal endothelial monocyte-activating polypeptide II (EMAPII)-like domain (Supplementary Figure S1A). The EMAPII-like domain was named for its high homology to the EMAPII protein, which was initially discovered as a tumor-secreted proinflammatory cytokine (24). Later, the precursor of EMAPII protein was found equivalent to the multi-tRNA synthetase complex protein p43 (25). This EMAPII-like domain was appended to TyrRS at the stage of insects and has been retained ever since (3).

The transcript of human full-length (FL) TyrRS contains 13 exons (the exon boundaries and amino acid sequences encoded by each exon are annotated in http://www.nextprot.org/db/entry/NX_P54577/exons), which has exactly the same exon composition and length as the mouse counterpart. The encoded human and mouse TyrRSs share 97% sequence identity (Supplementary Figure S2). The AceView database annotated a number of alternative mRNA variants of the gene for human and mouse TyrRSs (*YARS*) (26). These SVs 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. However, confirmation of the mRNAs and expression of either recombinant or natural versions of these potential SVs has never been investigated.

For a comprehensive identification of alternatively spliced forms of AARS, we developed an amplification-based transcriptome sequencing method as described (18). Briefly, the transcriptome for AARS was first enriched with RNA of human or mouse tissues and cells, followed by deep sequencing. This AARS-specific transcriptome approach allowed for higher read-depth of the exome of individual genes. The majority of the identified SVs have exon-skipping events. Specifically for the TyrRS gene, a total of 14 non-canonical exon-junctions were discovered in the RNA of human adult brain, fetal brain, total leukocytes or mouse RAW264.7 macrophages (Supplementary Table S1 and Supplementary Figure S1A). Eleven of these were new to that study and not annotated in the AceView database. The discovery of many more new exon junctions can be attributed to the high sensitivity of the approach we used to detect low-abundant SVs.

We were especially intrigued by three exon-skipping events, namely ΔE2, ΔE2–3 and ΔE2–4, found in both human and mouse mRNAs. Such species' conservativeness suggests the functional importance of these TyrRS variants (27). Among the other SVs, nine were only identified in human and two solely in mouse (Supplementary Table S1 and Supplementary Figure S1A). (We cannot exclude that they are present in tissues that have not yet been explored. Thus, it is not conclusive whether they are species-specific or could be shared by human and mouse.) Interestingly, 13 of the total 14 splicing events disrupted the aminoacylation and/or anti-codon binding domains. Hence, the canonical function of TyrRS is compromised in these isoforms. As such, the SV proteins could be repurposed for new functions. This propensity to remove core domains and retain the appended domains in SVs is common with other AARS (18).

Validation and expression analysis of the species-conserved splice variants of TyrRS

The PCR was used to validate the mRNA variants of TyrRS identified by deep sequencing. Because variants conserved between species were of major interest, and were detected with multiple reads in mouse macrophage cells, for validation we employed two immune cell lines – macrophage THP-1 cells and Jurkat T lymphocytes. To obtain a more complete sequence of the transcripts, PCR reactions were performed with a pair of primers targeting 5′ UTR/Exon1 and 3′-UTR/Exon13 of *YARS* (FP and RP1, Figure 1A). The results obtained from THP-1and Jurkat cells were similar (Figure 1B). The PCR product amplified from the FL transcript showed a length of 1.6 kb on gel electrophoresis. Two shorter bands, with lengths of about 1.2 and 1.3 kb, were also clearly identified. Sequencing of these PCR products confirmed the ΔE2–3 and ΔE2–4 splicing events, respectively, and retention of all remaining exons of *YARS* (Supplementary Figure S1B and C).

Considering that Exon2 contains only 90 nucleotides, the TyrRSΔE2 transcript was not expected to be well resolved (on gel electrophoresis) from the FL transcript. To better differentiate PCR products amplified from FL and SV mRNA transcripts, a reverse primer targeting Exon5/Exon6 (RP2) was used instead of RP1. Indeed, the SVs bands were then well resolved on gel electrophoresis (Figure 1C) and their sequences were confirmed. The three SVs were also detected when PCR reactions were performed using cDNA templates transcribed from cytoplasmic RNA of THP-1 and Jurkat cells. Analysis of polyribosomal RNA extracted from Jurkat cells gave evidence for the presence of the SVs and thus suggested that each can be translated (Figure 1C).

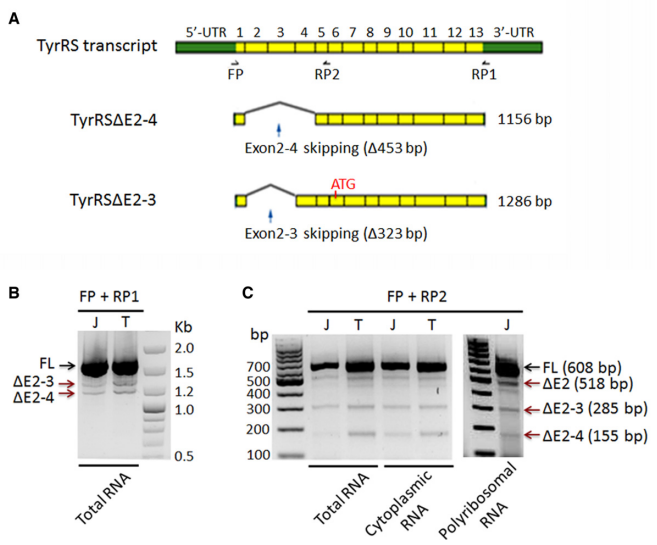

Figure 1. Validation of the TyrRS splice variants (SVs) that are conserved between human and mouse in total, cytoplasmic and polyribosomal RNA of human immune cells. (**A**) Schematic of transcripts of native TyrRS and the two SVs, TyrRSΔE2–4 and TyrRSΔE2–3. Positions of the PCR primers are also shown (FP: forward primer; RP: reverse primer). TyrRSΔE2–4 has an in-frame exon-skipping event, while deletion of Exons2–3 disrupts the reading frame and the TyrRSΔE2–3 transcript was predicted to translate from a start codon in Exon6 (indicated by 'ATG'). (**B**) Validation of the SVs by reverse transcription-PCR using a pair of primers FP and RP1 and the total RNA of Jurkat-T (J) and THP-1 (T) cells. (**C**) Detection of the SVs in total, cytoplasmic and polyribosomal RNA of Jurkat-T (J) or THP-1 (T) cells using primers FP and RP2.

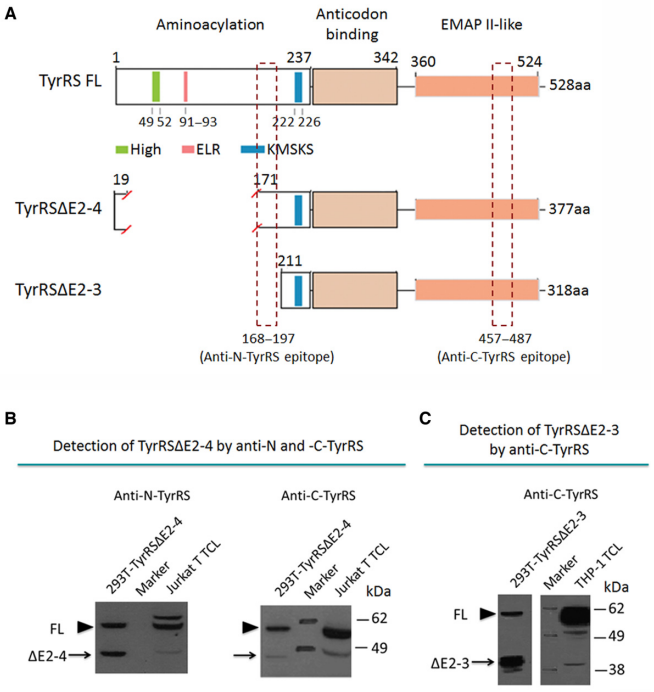

Figure 2. Western blot detection of endogenous TyrRS SV proteins in human immune cells. (**A**) Schematic of TyrRS TyrRSΔE2–4 and TyrRSΔE2–3 proteins. (**B**,**C**) Detection of endogenous TyrRS SV proteins by Western blot analysis using anti-TyrRS N-terminal antibody (epitope: aa168–197) and/or anti-C-terminal antibody (epitope: aa457–487). Arrow head: full-length TyrRS; Arrow: TyrRSΔE2–4 (B) or TyrRSΔE2–3 (C).

Organization of the splice variant transcripts

According to the nucleotide sequence of the TyrRSΔE2–4 transcript, its in-frame exon-skipping event results in an internal deletion of 151 amino acids (Δaa20–170) in TyrRS (Figure 2A). For the TyrRSΔE2–3 transcript that has a frame-disrupting splicing event, the putative protein product starts from an 'ATG' downstream in Exon6 and ends at the canonical stop codon (Figures 1A and 2A). Thus, this variant has a predicted protein sequence that is the same as the C-terminal aa211–528 of FL TyrRS. In the case of TyrRSΔE2, our sequencing only covered Exon1 to Exon6, and could not be extended further in a reliable way. Therefore, we focused on the TyrRSΔE2–4 and TyrRSΔE2–3 translation products in our subsequent studies. Both SVs removed most of the aminoacylation domain. This ablation included the HIGH motif that is embedded within the 11 amino acid signature sequence of class I tRNA synthetases (28). This motif is part of the structural unit that forms the ATP binding site and the site that docks the 3′-end of tRNA. Interestingly, the ELR motif is also deleted. This tripeptide is found in CXC cytokines, like IL-8, and is critical for the cell migration activity of a natural N-terminal fragment of TyrRS known as mini-TyrRS, which has retained the catalytic domain and is produced by natural proteolysis (29,30). Thus, sequence elements for neither aminoacylation nor cytokine-like activities of TyrRS are retained by the SVs. It was also noted that exon 5, which contains the specificity-determining helix and forms part of the active site that recognizes tyrosine, is uninterrupted in TyrRSΔE2–4 but is ablated in the TyrRSΔE2–3 variant. Both variants retain exon 10 that encompasses the KMSKS loop that is a signature for class I tRNA synthetases. Lastly, in the 3-D structure of TyrRS, a dimer is formed though contacts between the α6 and α8 helix. Most or all of the residues that make up these contacts are ablated in the SVs (Supplementary Figure S2).

Detection of endogenous SV proteins

To determine whether the predicted protein products of alternative splicing folded into stable proteins, we first constructed recombinant genes encoding TyrRSΔE2–4 and TyrRSΔE2–3 SVs. These genes were expressed in HEK 293T cells and yielded proteins of the predicted size that could be purified to homogeneity (Figure 2B and C). Thus, the splicing events yielded sequences that folded into stable proteins.

Next, we used whole cell lysates of Jurkat and THP-1 cells to investigate the presence of the endogenous proteins of the two SVs. These cells were chosen because they expressed the SV transcripts. Using Western blot methods, protein products were probed by polyclonal antibodies raised against aa168–197 (anti-N-TyrRS) or aa457–487 (anti-C-TyrRS) of human TyrRS (Figure 2A). In whole lysates of Jurkat cells, a species having a MW of about 45 kDa reacted with both the anti-N-TyrRS and anti-C-TyrRS antibodies (Figure 2B). The size of this protein matched that of the recombinant TyrRSΔE2–4 protein overexpressed in HEK 293T

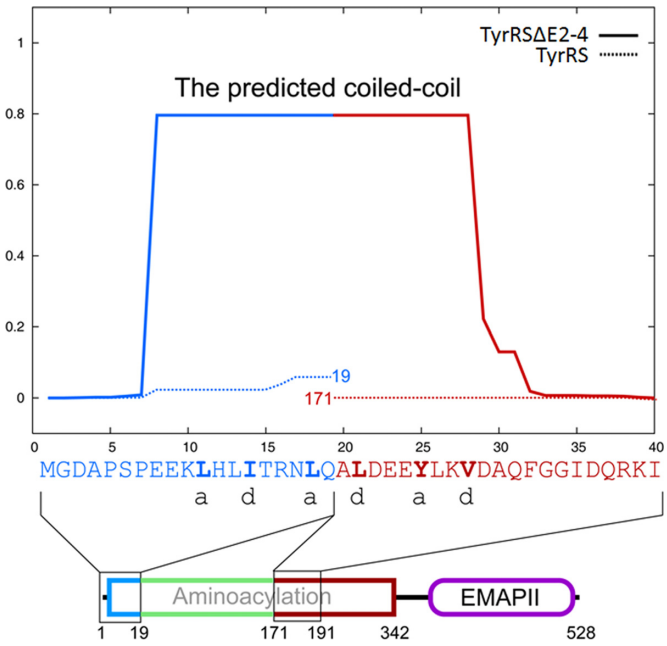

Figure 3. Structure prediction that internal deletion of Exon2–4 creates a new coiled-coil. The internal deletion of Exon2–4 was predicted to create a coiled-coil structure by joining aa1–19 to aa171–191 (solid line, window size = 21 amino acids). In comparison, aa1–19 and aa171–191 separately in native TyrRS were unlikely to form coiled-coils (dashed line, window size = 21). The amino acids and a cartoon illustrating the corresponding regions in FL TyrRS are shown below. The 'a' and 'd' indicate the first and fourth positions of the heptad repeats in the coiled-coil, respectively.

cells. In whole lysates of THP-1 cells, a species with a MW of about 40 kDa was detected, which showed a size similar to the recombinant TyrRSΔE2–3 protein overexpressed in HEK 293T cells (Figure 2C). As expected, the 40 kDa band was not detected by the anti-N-TyrRS (data not shown). Consistent with the relatively low amounts of mRNA that were annotated, the two proteins were in lower abundance compared to that of FL TyrRS (TyrRSΔE2–4 estimated as around 5% and TyrRSΔE2–3 as around 1% of TyrRS FL on Western blot analysis).

New coiled-coil regions created by exon skipping suggest at least one SV forms a new dimer architecture

Our interest turned to understanding the effect of deletion of the TyrRS dimerization domain on the quaternary structures of the SV proteins. TyrRSΔE2–4 and TyrRSΔE2–3 variants are closely related. The only difference lies in the 59-aa sequence present in the N-terminus of TyrRSΔE2–4 but absent in TyrRSΔE2–3. We carefully analyzed each sequence and noted several remaining helical regions (α1, α9 and part of α8, Supplementary Figure S2). Interestingly, these helices have a high potential to form one long helix, as part of a coiled-coil structural motif typically formed by two long alpha helices. As predicted by the COILS server, the probability of the coiled-coil formation was as high as 0.8 for a window size of 14 or 21 amino acids (Figure 3 and Supplementary Figure S3). This long helix arises from the skipping of Exon2–4 to fuse N-terminal residues 1–19 with internal residues171–191, and may form a coiled-coil structure with itself in a dimeric configuration. In contrast, these two regions of sequence in isolation, or when flanked by residues in native TyrRS, are unlikely to form coiled-coils (probability < 0.1).

Another predicted coiled-coil region (probability ∼0.9, window size = 14 or 21), composed of residues 231–249, was present in native TyrRS as well as in the two SVs (Supplementary Figure S3). In native TyrRS, residues 231–249 are not in spatial proximity across the dimer interface that is stabilized instead by the homotypic interaction between α6 and α8 of one monomer in the dimer and those of the other monomer. The 231–249 region contains the KMSKS motif that, in the context of an intact active site, is highly mobile to facilitate ATP binding and catalysis. However, in the context of the SV, the majority of the active site architecture restraint is removed, and the 231–249 region can now in principle be free to form a long helix and a coiled-coil structure with itself involved in dimerization. Thus, TyrRSΔE2–4 contains two probable coiled-coil regions, with one shared with TyrRSΔE2–3 and an additional one at its N-terminus. Because coiled-coil motifs favor formation of oligomers (dimers are the most common), we hypothesized that, while the dimer interface of native TyrRS is disrupted, at least one of the SV proteins, namely TyrRSΔE2–4, adopts an oligomeric quaternary structure through newly created coiled-coil(s).

One splice variant—TyrRSΔE2–4—is largely dimeric while the other is mostly monomeric

To test this hypothesis, the SVs proteins were further characterized. The wild-type SVs were overexpressed in *E. coli* as recombinant proteins. However, probably due to lack of a natural expression environment (such as mammalian chaperons), both proteins heavily degraded during purification and could not be used for further analysis. This instability may be attributed to loose packing of the C-terminal domain with the N-terminal parts of the SVs (30,31). To stabilize the proteins, we removed the flexible linker (aa343–360) connecting the anticodon binding and EMAPII-like domains (Supplementary Figure S4A). With this modification, the deletion variants (TyrRSΔE2–4′ and TyrRSΔE2–3′) could be purified with much less degradation.

In size exclusion chromatography, both SVs showed peaks corresponding to the sizes of a monomer and a dimer (Supplementary Figure S4B–E). After collection of each peak fraction, they were separately analyzed by analytical gel-filtration. The results demonstrated that each fraction of TyrRSΔE2–4′ had a reversible monomer-dimer equilibrium (Figure 4A), which was further confirmed by multi-angle light scattering (MALS) analysis (Figure 4B). In contrast, TyrRSΔE2–3′, which lacks the N-terminal 59 amino acids of TyrRSΔE2–4′ and its newly created coiled-coil domain, was predominantly monomeric. This observation suggests the coiled-coil motif aa231–249, which is shared by both variants, by itself is insufficient to stabilize the dimeric form. Thus, the additional N-terminal extra peptide of TyrRSΔE2–4′ facilitated dimerization, most likely through its novel coiled-coil domain.

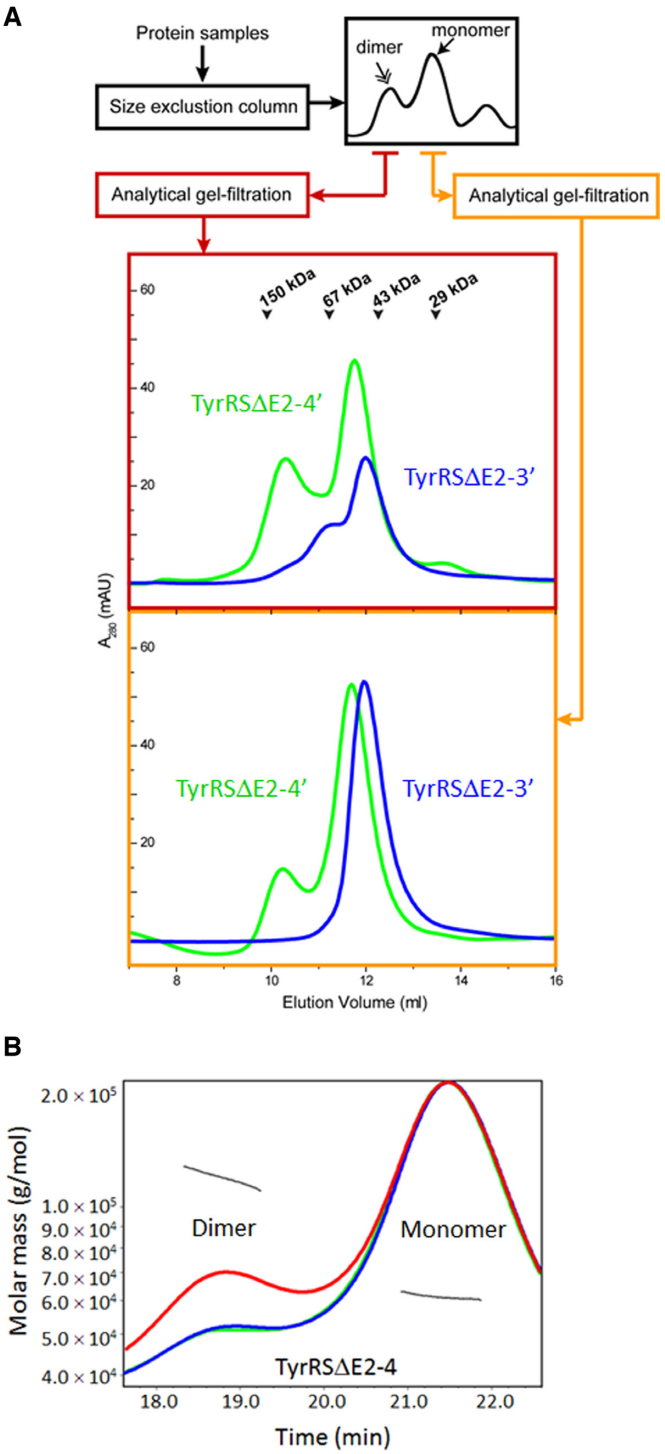

Figure 4. TyrRSΔE2–4′ showed a much higher tendency than TyrRSΔE2–3′ to form dimers. (**A**) Shown are the flow of protein purification and analysis, and results of the analytical gel-filtration. (**B**) MALS analysis of TyrRSΔE2–4. Red: light scattering; Green, UV; Blue: refractive index.

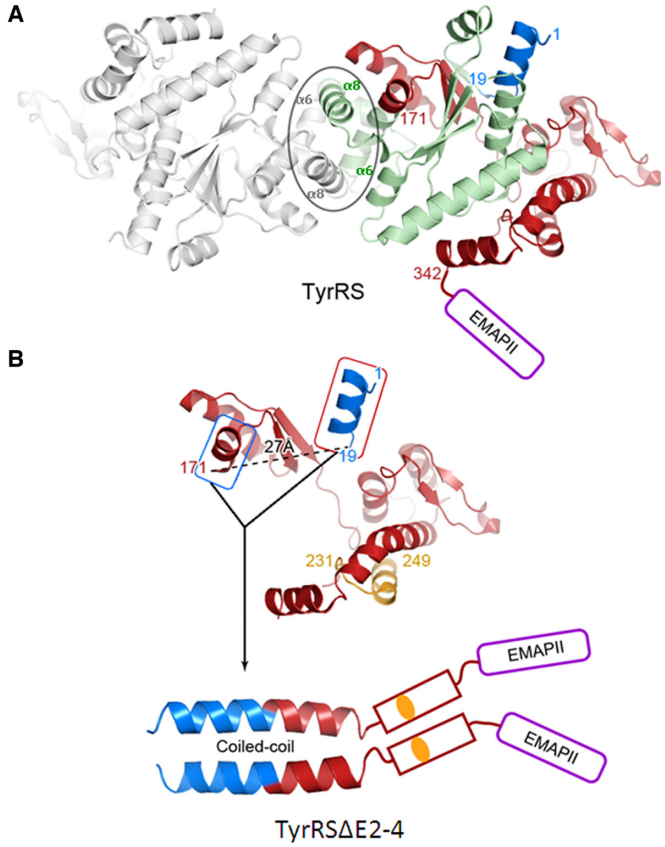

Figure 5. Structural representation of TyrRS and structural model for TyrRSΔE2–4. (**A**) Ribbon representation of the reported 3D structure of the TyrRS aa1–342 dimer (PDB: 1N3L). Circled area: dimer interface, which was deleted in TyrRSΔE2–4 and TyrRSΔE2–3. The structure uses the same color scheme as the cartoon in Figure 3. (**B**) Structural model of the new coiled-coil domain formed in TyrRSΔE2–4, through joining of helical regions of aa1–19 to aa171–191. Amino acids Q19 and A171 are 27Å apart in native TyrRS (upper, residues 20–170 encoded by Exon2–4 were removed), which were brought together in TyrRSΔE2–4 (lower). This newly created coiled-coil, which is absent in TyrRSΔE2–3, likely facilitates the dimer formation. Another coiled-coil was predicted to form by amino acids 231–249 (labeled yellow), which are present in both TyrRSΔE2–3 and TyrRSΔE2–4 (See also Supplementary Figure S3).

Structural model of TyrRSΔE2–4 reveals dimer architecture 'orthogonal' to TyrRS

For higher eukaryotes, the structure of FL TyrRS is unavailable. However, crystal structures of human mini-TyrRS catalytic N-terminal domain and, separately, of the C-terminal domain have been solved (30,31). In the 3-D structure, human mini-TyrRS is homodimeric. Also, both native human TyrRS and mini TyrRS are homodimers in solution. However, the C-terminal domain in isolation is monomeric. Thus, the dimeric structure of native TyrRS is undoubtedly determined by the α6-α8 helix interactions seen in the crystal structure of mini TyrRS. The dimerization interface formed by α6 and α8 helices from opposite chains is disrupted by removal of residues 20–170 in TyrRSΔE2–4 (Figure 5A).

We constructed a model for the TyrRSΔE2–4 protein that brings together residues 19 and 171 that are 27Å apart in native TyrRS to give a new coiled-coil interaction motif at its N-terminus (Figure 5B). In this model, residues 1–19 and 171–191 (in the numbering of FL TyrRS) are fused together to form a coiled-coil domain. Two polypeptide chains of TyrRSΔE2–4 are packed together through this

domain and may also be stabilized by the second coiled-coil composed of residues 231–249. The C-terminal domains in these two chains are more 'free-floating'. In contrast, the TyrRSΔE2–3 variant, which translates to a polypeptide as the C-terminal aa211–528 of FL-TyrRS (with the translation initiation site in Exon 6, Figures 1A and 2A), is largely monomeric. Consistently, it lacks both the dimer interface of FL-TyrRS and the first coiled-coil domain of TyrRSΔE2–4. Thus, the novel exploitation of the sequence elements at the splice junctions creates two distinct neomorphic forms.

Dimer model verification for TyrRSΔE2–4 by protein engineering and NMR analysis

To verify our structural model, we made further efforts to characterize the structure of TyrRSΔE2–4 by crystallization and NMR spectroscopy. Crystallization was first attempted with purified recombinant TyrRSΔE2–4′ protein, which showed a similar dimer fraction as wild-type TyrRSΔE2–4 in the analytical gel filtration analysis (Figure 6A). But extensive crystallization trials with this protein failed to yield well diffracting crystals. To expand the candidate proteins for structural determination, we made more truncation variants as shown in Figure 6A. Because verifying dimer formation of the first predicted coiled-coil (CC1) in TyrRSΔE2–4 was essential for our structure hypothesis, various truncation variants of TyrRSΔE2–4 were constructed with each preserving CC1. After several test runs and optimizations, we successfully expressed and purified three of these mutants, including CC1, CC1ext and CC1_2 (Figure 6A). The monomer/dimer formation of the purified mutants was also compared to the wild-type TyrRSΔE2–4 by analytical gel filtration, and only the mutant CC1_2, containing both coiled-coils showed a comparable portion of the dimer fraction (Supplementary Figure S4F). Thus, while the first coiled-coil in TyrRSΔE2–4 is required for dimer formation, the second coiled-coil also appears to contribute to stabilize the dimer.

We employed the truncation variant TyrRSΔE2–4-CC1_2 for further NMR-based structural characterizations. We obtained the ^{15}N-^{1}H HSQC spectrum of the protein (Figure 6B). This protein contains 102 amino acids. We observed 98 individual backbone amide peaks at pH 7.0, indicating that the majority of backbone amide groups can be detected in this spectrum and thus allowed further characterizations. Notably, chemical shifts of most of the amide peaks are located in a narrow region from 7.7 to 8.8 ppm. This observation suggests that the protein may contain α-helical elements (32) or be totally unstructured (because the backbone amide chemical shifts of residues adopting either a helical or random coil structure are similar). To differentiate the above two possibilities, we recorded a ^{15}N-^{1}H HSQC spectrum of CC1_2 at an elevated pH value (pH 7.6). At pH 7.6, the amide peaks for residues that adopt a random coil structure disappear due to their rapid exchange with solvent. In contrast, the amide peaks for residues adopting a helical structure can still be detected due to the formation of amide hydrogen bonds in the helices. Indeed, 62 amide peaks could still be detected for CC1_2 at pH 7.6, and thereby supports the helical conformation of the protein. Having ∼60% of the residues of CC1_2 showing protected amide exchange profiles is consistent with the partial helical structure of the protein shown by the CD spectrum (Figure 6C), in which θ_{220} is significantly less than θ_{208}. At pH 8.0, even those residues that were detectable at pH 7.6 also became undetectable, or nearly undetectable (data not shown). This behavior is likely due to transient helical structures and dynamic conformational equilibria of CC1_2, which is consistent with gel filtration analysis (Figure 4) and the CD spectrum of the protein (Figure 6C).

TyrRS SVs are specifically enriched in lymphocytes and lung tissue

As reported in our previous study (18), a majority of AARS SVs that ablate the catalytic domain, including TyrRSΔE2–4, were expressed in a tissue-specific way. In that study, the SYBR green quantitative real-time PCR (qPCR) method was employed to analyze the mRNA distribution of various AARS transcripts in 13 human tissues and 5 cell types. The tested human tissues include those of the immune system (total leukocytes, bone marrow, spleen), circulatory system (lung, heart, kidney), digestive system (liver, pancreas, small intestine, colon), among others (thyroid, adipose cells, skeletal muscle). The human cell types included Raji B and Jurkat T lymphocytes, monocytic and macrophagic THP-1 cells, and a neuronal cell line IMR-32. The mRNA transcripts encoding FL TyrRS were found distributed evenly across various tissues and cells, with the normalized expression levels all below 3 times of median (3M, Supplementary Figure S5A). In contrast, TyrRSΔE2–4 showed an enriched expression in Jurkat T and Raji B lymphocytes (∼6 and 14 times of median respectively, Supplementary Figure S5B).

We speculated that, given their distinct quaternary structures, TyrRSΔ2–4 and TyrRSΔ2–3 might have a particular cell type distribution. With these considerations in mind, we also probed TyrRSΔE2–3 transcripts by SYBR green qPCR, using optimized SV-specific primers. A distinct tissue distribution pattern was observed for TyrRSΔE2–3, which exhibited the highest mRNA level in lung (∼7 times of median, Supplementary Figure S5C). These results raise the possibility that, in addition to their distinctive quaternary structures, the two TyrRS SVs are differentially enriched in lymphocytes or lung for separate biological functions.

DISCUSSION

Our analysis was prompted by the discovery of internally deleted catalytic nulls of human tRNA synthetases, which have diverse functions in cell-based and animal studies (8,18). These internal deletions most often ablate the catalytic domain and thereby redirect the resulting protein away from protein synthesis. While molecular modeling showed that these deletions within the catalytic core of the TyrRS must create a major disruption in the resulting protein, no actual structural analysis had been attempted. In contrast to these internal deletions of TyrRS within the core, the earlier-investigated large internal deletion of HisRS removed the entire core, and simply linked together the N- and C-terminal domains (19).

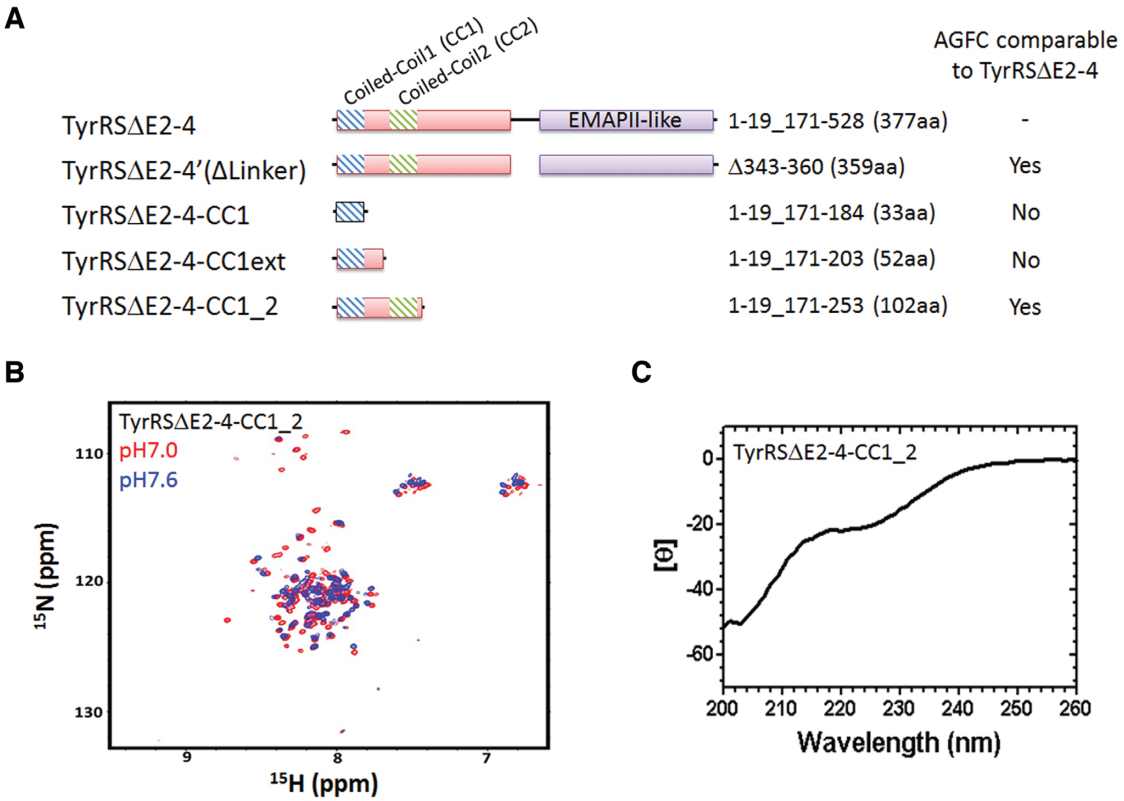

Figure 6. Structural characterizations of TyrRSΔE2–4. (**A**) Schematic of truncation variants of TyrRSΔE2–4. The amino acid composition is indicated and numbered based on the FL-TyrRS sequence. In the brackets is the length of each truncation variant. Their dimer/monomer formation were compared to wild-type by analytical gel filtration chromatography (AGFC). (**B**) Overlay of 2D ^{1}H-^{15}N heteronuclear single quantum coherence (HSQC) spectra of TyrRSΔE2–4-CC1_2 in solution of pH 7.0 (red) and pH 7.6 (blue). (**C**) Circular dichroism spectrum of the TyrRSΔE2–4-CC1_2 protein.

Prior work on internal deletions of proteins focused on engineering to create artificial ablations of loops that protrude from a core structure (33,34). Such ablations are not suspected to alter the core protein but, even in these instances, structural information of any sort is sparse. In this work, we focused on natural internal deletions that, by their very nature, are highly disruptive to the core, and have to be exquisitely designed (by evolution) in order to yield a stable protein.

During evolution AARS added novel domains, in a way that is progressive and accretive as the Tree of Life is ascended. These domains are dispensable for aminoacylation, but essential for the novel intracellular, extracellular and nuclear activities associated with AARSs. Importantly, while most SVs disrupt the catalytic core, virtually all of the more than 200 SVs of human AARS retain the novel domains (18). In the present study, the main new structural motif, added at the time of insects, is the EMAPII-like domain at the C-terminal end of the protein (Figure 2). A high-resolution structure of this domain revealed an OB fold, which harbors a critical hexapeptide motif for cell signaling activity (31,35). This domain is also important for the nuclear interactions of TyrRS with PARP-1 (8). While the two SVs appear to not disrupt the EMAPII-like domain, the alternative quaternary structures they present would change the spatial orientation of this signaling element, and is a subject for future investigation. It is worth noting that, in another context, quaternary structure changes have been associated with functional switches (36,37).

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

FUNDING

Innovation and Technology Fund of Hong Kong [UIM181, UIM192 and UIM199]; Research Grants Council of Hong Kong [16100015]; aTyr Pharma; Fellowship from the National Foundation for Cancer Research; National Cancer Institute [CA92577]; National Institutes of Health [R01 NS085092]; South University of Science and Technology of China; Science, Technology and Innovation Commission of Shenzhen Municipality [JCYJ20140417105742705]; National Natural Science Foundation of China [31570741]. Funding for open access charge: Hong Kong Research Grants Council [16100015].
Conflict of interest statement. None declared.

REFERENCES

1. 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.
2. Guo,M. and Schimmel,P. (2013) Essential nontranslational functions of tRNA synthetases. *Nat. Chem. Biol.*, **9**, 145–153.

3. 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.
4. Kim,S., You,S. and Hwang,D. (2011) Aminoacyl-tRNA synthetases and tumorigenesis: more than housekeeping. *Nat. Rev. Cancer*, **11**, 708–718.
5. Martinis,S.A. and Joy Pang,Y.L. (2007) Jekyll & Hyde: evolution of a superfamily. *Chem. Biol.*, **14**, 1307–1308.
6. 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. U.S.A.*, **109**, E640–E647.
7. Park,S.G., Schimmel,P. and Kim,S. (2008) Aminoacyl tRNA synthetases and their connections to disease. *Proc. Natl. Acad. Sci. U.S.A.*, **105**, 11043–11049.
8. Sajish,M. and Schimmel,P. (2015) A human tRNA synthetase is a potent PARP1-activating effector target for resveratrol. *Nature*, **519**, 370–373.
9. Sajish,M., Zhou,Q., 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-gamma and p53 signaling. *Nat. Chem. Biol.*, **8**, 547–554.
10. Wei,N., Shi,Y., Truong,L.N., Fisch,K.M., Xu,T., Gardiner,E., Fu,G., Hsu,Y.S., Kishi,S., Su,A.I. *et al.* (2014) Oxidative stress diverts tRNA synthetase to nucleus for protection against DNA damage. *Mol. Cell*, **56**, 323–332.
11. Xu,X., 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.*, **3**, Article number: 681, 1–9.
12. Yao,P., Potdar,A.A., Arif,A., Ray,P.S., Mukhopadhyay,R., Willard,B., Xu,Y., 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.
13. 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.
14. 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.
15. 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.
16. 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.
17. 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. U.S.A.*, **104**, 967–972.
18. Lo,W.S., Gardiner,E., Xu,Z., Lau,C.F., Wang,F., Zhou,J.J., Mendlein,J.D., Nangle,L.A., Chiang,K.P., Yang,X.L. *et al.* (2014) Human tRNA synthetase catalytic nulls with diverse functions. *Science*, **345**, 328–332.
19. Xu,Z., Wei,Z., Zhou,J.J., Ye,F., Lo,W.S., Wang,F., Lau,C.F., Wu,J., Nangle,L.A., Chiang,K.P. *et al.* (2012) Internally deleted human tRNA synthetase suggests evolutionary pressure for repurposing. *Structure*, **20**, 1470–1477.
20. Jiang,H. and Wong,W.H. (2009) Statistical inferences for isoform expression in RNA-Seq. *Bioinformatics*, **25**, 1026–1032.
21. Lupas,A., Van Dyke,M. and Stock,J. (1991) Predicting coiled coils from protein sequences. *Science*, **252**, 1162–1164.
22. Delaglio,F., Grzesiek,S., Vuister,G.W., Zhu,G., Pfeifer,J. and Bax,A. (1995) NMRPipe: a multidimensional spectral processing system based on UNIX pipes. *J. Biomol. NMR*, **6**, 277–293.
23. Goddard,T.D. and Kneller,D.G. (2008) *SPARKY 3*. University of California Press, San Francisco.
24. Kao,J., Ryan,J., Brett,G., Chen,J., Shen,H., Fan,Y.G., Godman,G., Familletti,P.C., Wang,F., Pan,Y.C. *et al.* (1992) Endothelial monocyte-activating polypeptide II. A novel tumor-derived polypeptide that activates host-response mechanisms. *J. Biol. Chem.*, **267**, 20239–20247.
25. Quevillon,S., Agou,F., Robinson,J.C. and Mirande,M. (1997) The p43 component of the mammalian multi-synthetase complex is likely to be the precursor of the endothelial monocyte-activating polypeptide II cytokine. *J. Biol. Chem.*, **272**, 32573–32579.
26. Thierry-Mieg,D. and Thierry-Mieg,J. (2006) AceView: a comprehensive cDNA-supported gene and transcripts annotation. *Genome Biol.*, **7** (Suppl 1), S12, 11–14.
27. Sorek,R., Shamir,R. and Ast,G. (2004) How prevalent is functional alternative splicing in the human genome? *Trends Genet.*, **20**, 68–71.
28. Ludmerer,S.W. and Schimmel,P. (1987) Gene for yeast glutamine tRNA synthetase encodes a large amino-terminal extension and provides a strong confirmation of the signature sequence for a group of the aminoacyl-tRNA synthetases. *J. Biol. Chem.*, **262**, 10801–10806.
29. Wakasugi,K. and Schimmel,P. (1999) Two distinct cytokines released from a human aminoacyl-tRNA synthetase. *Science*, **284**, 147–151.
30. Yang,X.L., Skene,R.J., McRee,D.E. and Schimmel,P. (2002) Crystal structure of a human aminoacyl-tRNA synthetase cytokine. *Proc. Natl. Acad. Sci. U.S.A.*, **99**, 15369–15374.
31. Yang,X.L., Liu,F.M., Skene,R.J., McRee,D.E. and Schimmel,P. (2003) Crystal structure of an EMAP-II-like cytokine released from a human tRNA synthetase. *Helv. Chim. Acta*, **86**, 1246–1257.
32. Lindhout,D.A., Litowski,J.R., Mercier,P., Hodges,R.S. and Sykes,B.D. (2004) NMR solution structure of a highly stable de novo heterodimeric coiled-coil. *Biopolymers*, **75**, 367–375.
33. Li,X., Zhang,G., Ngo,N., Zhao,X., Kain,S.R. and Huang,C.C. (1997) Deletions of the Aequorea victoria green fluorescent protein define the minimal domain required for fluorescence. *J. Biol. Chem.*, **272**, 28545–28549.
34. Starzyk,R.M., Webster,T.A. and Schimmel,P. (1987) Evidence for dispensable sequences inserted into a nucleotide fold. *Science*, **237**, 1614–1618.
35. Wakasugi,K., Slike,B.M., Hood,J., Ewalt,K.L., Cheresh,D.A. and Schimmel,P. (2002) Induction of angiogenesis by a fragment of human tyrosyl-tRNA synthetase. *J. Biol. Chem.*, **277**, 20124–20126.
36. Morais,M.A., Giuseppe,P.O., Souza,T.A., Alegria,T.G., Oliveira,M.A., Netto,L.E. and Murakami,M.T. (2015) How pH modulates the dimer-decamer interconversion of 2-Cys peroxiredoxins from the Prx1 subfamily. *J. Biol. Chem.*, **290**, 8582–8590.
37. Vo,M.N., Yang,X.L. and Schimmel,P. (2011) Dissociating quaternary structure regulates cell-signaling functions of a secreted human tRNA synthetase. *J. Biol. Chem.*, **286**, 11563–11568.