Structural Insights into RNA-dependent Cysteine Biosynthesis in Archaea

The universal genetic code of twenty amino acids is interpreted by the canonical set of aminoacyl-tRNA synthetases (aaRSs), which attach specific amino acids to their cognate tRNAs. However, methanogenic archaea lack the canonical cysteinyl-tRNA synthetase (CysRS) for the "direct" Cys-tRNA^{Cys} formation. In such organisms, Cys-tRNA^{Cys} is produced by the "indirect" pathway, in which non-canonical *O*-phosphoseryl-tRNA synthetase (SepRS) ligates a non-canonical amino acid, *O*-phosphoserine (Sep), to tRNA^{Cys}, and Sep-tRNA:Cys-tRNA synthase (SepCysS) converts the produced Sep-tRNA^{Cys} to Cys-tRNA^{Cys} (Fig. 1). The unique feature of *O*-phosphoserine is that it has divalent negative charges in the side-chain phosphate group, unlike the 20 canonical amino acids. The SepRS/SepCysS pathway is the sole route for cysteine biosynthesis in such organisms. The "indirect" pathway for Cys-tRNA^{Cys} formation by SepRS/SepCysS is ancient and may predate the "direct" pathway by CysRS. Therefore, elucidation of the structure of SepRS and its recognition mechanism for tRNA^{Cys} and *O*-phosphoserine may contribute to our understanding of the evolution of the genetic code table. To gain insight into the evolution of cysteine coding, we determined the crystal structures of the *Archaeoglobus fulgidus* SepRS·tRNA^{Cys}·*O*-phosphoserine ternary complex (Fig. 2 and Fig. 3) [1] and *A. fulgidus* SepCysS (Fig. 4) [2] at 2.6 and 2.4 Å resolutions, respectively, using **BL41XU** beamline.

SepRS forms an α_4 tetramer, which binds two tRNA^{Cys} molecules (Fig. 2(b)). This is the first structure of an aaRS that forms an α_4 tetramer. The SepRS monomer consists of four parts, the N-terminal extension, the catalytic domain, the inserted domain, and the C-terminal anticodon-binding domain, which are connected by linker loops (Fig. 2(a)). No structural neighbors of the inserted domain and the anticodon-binding domain were detected by a DALI search. Therefore, the two domains are novel structures. The aminoacylation catalytic domain recognizes *O*-phosphoserine uniquely (Fig. 3(a)). The phosphate moiety of *O*-phosphoserine is extensively recognized; each of the three non-bridging oxygen atoms is recognized via two hydrogen bonds. The N-terminal end of the conserved central helix, which is positively charged by the dipole moment of the α-helix, contributes to the recognition. Notably, the involvement of protein main-chain groups in the recognition of the amino-acid side chain of the substrate is unique to SepRS, and has never been observed in canonical aaRSs. The GCA anticodon bases of tRNA^{Cys} are located on the edge of the long two-stranded antiparallel β-sheet in the anticodon-binding domain and are recognized in a sequence-

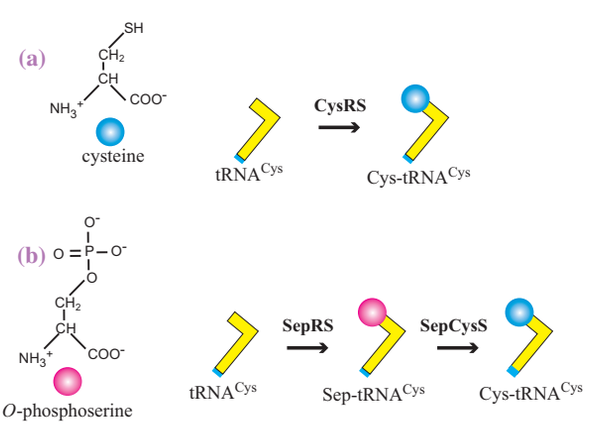

Fig. 1. **(a)** "Direct" Cys-tRNA^{Cys} formation by CysRS in normal organisms. **(b)** "Indirect" Cys-tRNA^{Cys} formation by SepRS and SepCysS in methanogenic archaea.

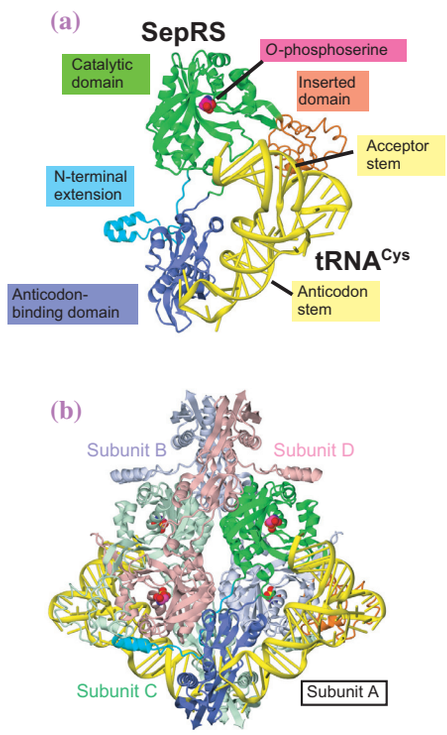

Fig. 2. **(a)** Overall structure of *A. fulgidus* SepRS·tRNA^{Cys}·*O*-phosphoserine complex. The tRNA^{Cys} and *O*-phosphoserine molecules are represented by yellow tubes and magenta ball-and-stick models, respectively. **(b)** *A. fulgidus* SepRS α_4 tetramer complexed with two tRNA^{Cys} molecules and four *O*-phosphoserine molecules. Color coding is as in **(a)** for SepRS subunit A. SepRS subunits B, C, and D are shown in light blue, light green, and light pink, respectively.

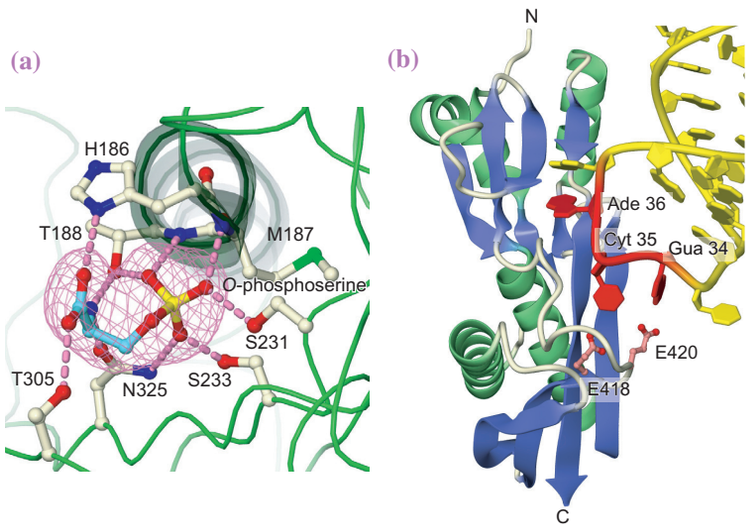

Fig. 3. **(a)** Recognition mechanism for *O*-phosphoserine in SepRS catalytic domain. **(b)** Recognition mechanism for GCA anticodon of $tRNA^{Cys}$ in SepRS C-terminal anticodon-binding domain.

specific manner (Fig. 3(b)).

SepCysS forms an α_2 dimer (Fig. 4(a)). The SepCysS monomer is composed of large and small domains. The large domain has a seven-stranded β-sheet, which is typical of pyridoxal 5'-phosphate (PLP)-dependent enzymes. In the active site, which is located near the dimer interface, PLP is covalently bound to the side chain of the conserved Lys209 by an aldimine bond. Lys209 is located on a short α-helix inserted into the seven-stranded β-sheet. PLP is deeply bound within the active site cleft near the dimer boundary. The active site is sufficiently wide to accommodate the 3'-end of Sep-$tRNA^{Cys}$.

On the basis of the two determined structures, we were able to construct the model structure of the SepRS·$tRNA^{Cys}$·SepCysS ternary complex (Fig. 4(b)). In the ternary complex, the phosphoserylated 3'-terminus of $tRNA^{Cys}$ can be transferred directly from SepRS to SepCysS, for conversion to the cysteinylated form. Such a mechanism may prevent the intermediate Sep-$tRNA^{Cys}$ from being delivered to the ribosome and prevent mistranslation.

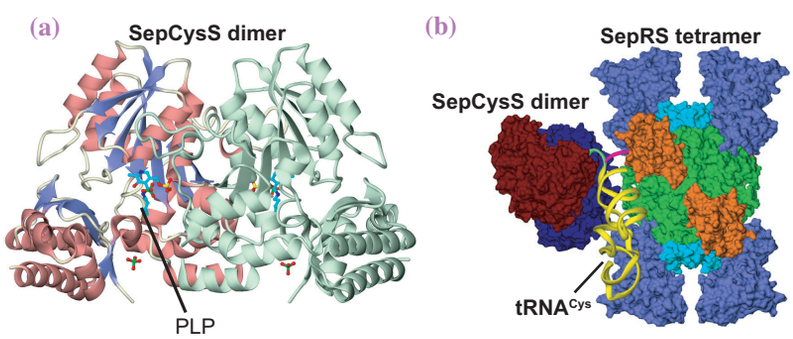

Fig. 4. **(a)** Overall structure of *A. fulgidus* SepCysS dimer. The α-helices and the β-strands of subunit A are shown in pink and blue, respectively. Subunit B is shown in light green. The internal aldimine Lys209-PLP is represented by a cyan ball-and-stick model. **(b)** Modeled ternary complex of SepRS·$tRNA^{Cys}$·SepCysS. The four 3'-terminal nucleotides of $tRNA^{Cys}$ are modeled for the SepCysS active site (green) and SepRS active site (magenta).

Ryuya Fukunaga[a,b,*,†] and Shigeyuki Yokoyama[a,b]

[a] Dept. of Biophysics and Biochemistry, The University of Tokyo
[b] RIKEN Genomic Sciences Center, Yokohama

*E-mail: fukunaga@berkeley.edu

[†] Present address: Dept. of Molecular and Cell Biology, University of California, Berkeley, USA

References

[1] *R. Fukunaga and S. Yokoyama: Nat. Struct. Mol. Biol.* ***14*** *(2007) 272.*
[2] *R. Fukunaga and S. Yokoyama: J. Mol. Biol.* ***370*** *(2007) 128.*