Structural basis for tRNA-dependent cysteine biosynthesis

Meirong Chen¹, Koji Kato¹,², Yume Kubo¹, Yoshikazu Tanaka¹,²,³, Yuchen Liu⁴, Feng Long⁵, William B. Whitman⁵, Pascal Lill⁶, Christos Gatsogiannis⁶, Stefan Raunser⁶, Nobutaka Shimizu⁷, Akira Shinoda², Akiyoshi Nakamura⁸, Isao Tanaka² & Min Yao¹,²

Cysteine can be synthesized by tRNA-dependent mechanism using a two-step indirect pathway, where O-phosphoseryl-tRNA synthetase (SepRS) catalyzes the ligation of a mismatching O-phosphoserine (Sep) to tRNA<sub>Cys</sub> followed by the conversion of tRNA-bounded Sep into cysteine by Sep-tRNA:Cys-tRNA synthase (SepCysS). In ancestral methanogens, a third protein SepCysE forms a bridge between the two enzymes to create a ternary complex named the transsulfursome. By combination of X-ray crystallography, SAXS and EM, together with biochemical evidences, here we show that the three domains of SepCysE each bind SepRS, SepCysS, and tRNA<sub>Cys</sub>, respectively, which mediates the dynamic architecture of the transsulfursome and thus enables a global long-range channeling of tRNA<sub>Cys</sub> between SepRS and SepCysS distant active sites. This channeling mechanism could facilitate the consecutive reactions of the two-step indirect pathway of Cys-tRNA<sub>Cys</sub> synthesis (tRNA-dependent cysteine biosynthesis) to prevent challenge of translational fidelity, and may reflect the mechanism that cysteine was originally added into genetic code.
The fidelity of protein synthesis is based on the accurate aminoacylation of tRNAs in which specific amino acids are attached to cognate tRNAs. A group of enzymes named aminoacyl-tRNA synthetases (aaRSs) are responsible for this reaction. It was believed that there is a matching aaRS for each of the 20 amino acids, which specifically ligates an amino acid to the 3’ terminal adenosine of its cognate tRNA. However, the advent of whole genome sequencing has revealed that several assigned aaRSs are absent in some organisms. In these cases, a non-cognate amino acid is initially attached to the tRNA by the first enzyme, which is subsequently converted to the cognate amino acid by the second enzyme in a tRNA-dependent manner. This indirect pathway is adopted for the synthesis of Gln-tRNA<sub>Gln</sub> and Asn-tRNA<sub>Asn</sub> in many bacteria and archaea. In most of the methanogenic archaea and Sec-tRNA<sub>Sec</sub> in many species across all three domains of life. Interestingly, the presence of the indirect pathways is accompanied with the absence of the enzymes responsible for canonical amino acids biosynthesis, where the indirect pathways are the source of the amino acids biosynthesis by tRNA-dependent mechanism.

The synthesis of Cys-tRNA<sub>Cys</sub> (and also cysteine) by an indirect pathway involves two enzymes, O-phosphoserine-tRNA synthetase (SepRS) and Sep-tRNA:Cys-tRNA synthase (SepCysS), which are respectively responsible for the attachment of a non-canonical amino acid O-phosphoserine (Sep) to tRNA<sub>Cys</sub> and the conversion of Sep-tRNA<sub>Cys</sub> to Cys-tRNA<sub>Cys</sub>. Phylogenetic studies have suggested that the indirect pathway for Cys-tRNA<sub>Cys</sub> synthesis is of ancient origin. This system also acts as the sole route for cysteine biosynthesis for these organisms. More recently, it was reported that a third protein SepCysE (SepRS/SepCysS pathway enhancer) is necessary for coupling the consecutive two reactions and for enhancing tRNA<sub>Cys</sub> binding. Thus, the SepRS, SepCysE, and SepCysS ternary complex (named the transsulfursome) is responsible for the whole process. SepCysE is preserved in class I methanogens, which evolved before other methanogens and may have appeared as the ancestral Euryarchaeota.

Therefore, elucidation of the functional mechanism of this system is of vital importance as it may explain the mechanism by which cysteine was added to the genetic code.

SepRS is a homotetramer and structurally characterized as a Class II aaRS, closely related to PheRS. SepCysS is a homodimer with pyridoxal-5′-phosphate (PLP) bound in the active sites that shares structural similarity to the PLP-dependent cysteine desulfurase. The structural analysis of SepCysS in complex with the N-terminal domain of SepCysE (SepCysE<sub>NTD</sub>) showed that the dimer of SepCysE is bound to the dimer of SepCysS with their dimer axes aligned. However, it is not yet known how SepRS is bound to this binary complex to form a ternary complex (transsulfursome) and how Sep-tRNA<sub>Cys</sub> is transferred between these two enzymes safely. Furthermore, while structural and biochemical analyses showed the recognition of tRNA<sub>Cys</sub> by SepRS and tRNA<sub>Cys</sub> is unclear, which is essential to guarantee the fidelity of the two-step indirect pathway in translation.

In the present study, the structures of the binary complex of SepRS and SepCysE, and the ternary complex of SepCysE, SepCysS, and tRNA<sub>Cys</sub> from Methanocaldococcus jannaschii have been solved. Combined with the SAXS, EM, and biochemical results, the recognition of tRNA<sub>Cys</sub> by SepCysS was determined and the architecture of the transsulfursome was established. On the basis of these results, we propose a mechanism for tRNA channeling, whereby a hinge motion of the SepCysE-SepCysS complex relative to SepRS and a swinging motion of SepCysE (CTD) with bound tRNA<sub>Cys</sub> facilitate the transport of the mischarged Sep-tRNA<sub>Cys</sub> to the second, distant active site.

### Table 1 Summary of data collection and refinement statistics

|               | SepCysE-SepCysS-tRNA<sub>Cys</sub>    | Transsulfursome |
|---------------|--------------------------------------|-----------------|
| PDB ID        | 5X6B                                 | 5X6C            |
| Data collection | Beamline: SPRing-8 BL41XU             | Photon Factory BL-5A |
| Space group   | P6<sub>2</sub>2                      |                 |
| Unit cell parameters (Å) | a = b = 107.3, c = 551.1          | a = b = c = 279.8 |
| Resolution range (Å) | 4.26–2.60                         | 48.0–3.10       |
| R<sub>free</sub> (%) | 10.2 (97.5)                        | 14.0 (79.0)     |
| Completeness (%) | 99.7 (98.2)                        | 99.9 (99.5)     |
| Redundancy    | 11.4 (11.8)                         | 7.6 (7.5)       |
| No. of reflections | 59,370                             | 65,670          |
| B-factors (Å<sup>2</sup>) | 22.5/26.1                          | 18.5/20.9       |
| Bond lengths (Å) | 0.005                               | 0.014           |
| Bond angles (°) | 1.09                                | 1.35            |

Values in parentheses are for the highest resolution shell. R<sub>free</sub> was calculated using 5% of randomly selected test sets that were not used in the refinement.

### Results

**Structure of the SepCysE-SepCysS-tRNA<sub>Cys</sub> ternary complex.**

A crystal structure of the SepCysE-SepCysS-tRNA<sub>Cys</sub> ternary complex was determined at a resolution of 2.6 Å by a molecular replacement method (Table 1, Supplementary Fig. 1a). A Fo-Fc map calculated after several rounds of refinement cycle showed a bulk of density adjacent to the end of one SepCysE(NTD), which represents the linker chain and the C-terminal domain of SepCysE [SepCysE (CTD)], which was disordered in the binary complex previously determined. Thus, this crystal revealed most of the structural features of SepCysE (Supplementary Fig. 1b), as well as the binding manner with tRNA<sub>Cys</sub> and SepCysS (Fig. 1a). A dimer of the N-terminal domain of SepCysE [SepCysE(NTD), residues 35–101] was tightly bound to a dimer of SepCysS, with their twofold axes aligned. From one of the NTDs of the SepCysE dimer, a linker chain (residues 102–110) extended to the C-terminal domain of SepCysE [SepCysE(CTD), residues 111–213] to which tRNA<sub>Cys</sub> was bound with its CCA terminus in contact with SepCysS (Fig. 1a).

The SepCysE(CTD) of the second subunit was not visible in the crystal, probably because of the absence of bound tRNA<sub>Cys</sub>. The stoichiometry of SepCysE-SepCysStRNA<sub>Cys</sub> = 2:2:1 is in agreement with the result of electrophoresis mobility shift assay (EMSA).

The structure of the SepCysE(CTD) consisted of three-stranded parallel β sheets with four peripheral alpha helices (Supplementary Fig. 1b). The connectivity of the secondary structures was β1, α4, β2, α5, α6, β3, and α7. It is basically an α/β fold with short insertion helix α5. Although the primary sequence of SepCysE(CTD) is not related to any known structures in the PDB, a Dali server search based on the present structure identified that the fold of SepCysE(CTD) is related to the Topprim (topoisomerase-prime) domain, which is found in proteins...
related to DNA metabolism such as primases, topoisomerases, nucleases, and DNA repair proteins and believed to be an ancestral domain20.

**C-terminal domain of SepCysE non-specifically binds tRNA.**

SepCysE(CTD) was bound to the elbow region of tRNA\(^{\text{Cys}}\) mainly via van der Waals and ionic interactions. The shape of the SepCysE(CTD) surface was highly complementary to tRNA\(^{\text{Cys}}\), and the central region of SepCysE(CTD) consisting of secondary structures of α, β, α5, and α6 interacted with the TΨC and D loops of tRNA\(^{\text{Cys}}\) (Fig. 1a). Especially prominent were the α4 and α6 helices in the groove of the tRNA, and the side-chains in these helices rich in lysine residues extended into this groove (Fig. 1b).

Although the poor density map at the region of α6 hindered building of precise model of the side chains, most of the residues...
involved in binding with tRNA\textsubscript{Cys} could be assigned. Alanine replacement mutations of these residues ([Asn117, Lys118, and Lys119], (Lys122, Lys123, Lys125, and Asn126) and (Lys159, Lys160 and Lys161]), reduced tRNA-binding ability (Supplementary Fig. 2a). Since most of the interactions of SepCysE(CTD) with tRNA\textsubscript{Cys} are with the phosphate backbone, SepCysE(CTD) is likely to be a non-specific tRNA binding domain. This idea was confirmed by a binding assay of SepCysE(CTD) with other tRNAs (tRNA\textsubscript{Sec}, tRNA\textsubscript{Phe} and tRNA\textsubscript{His}), which indicated that SepCysE(CTD) can bind these tRNAs with affinities almost at the same level as tRNA\textsubscript{Cys} (Supplementary Fig. 3a). Therefore, we characterized SepCysE(CTD) as an ancient, non-specific tRNA binding domain. However, we cannot rule out the possibility that SepCysE(CTD) exhibits binding preference to tRNA\textsubscript{Cys}. It was reported that, in E. coli, the interaction between ribose-phosphate backbone of tRNA\textsubscript{Cys} and CysRS is essential for recognition (indirect readout)\textsuperscript{31}. The same strategy may also be adopted by the present system.

It was previously noted that SepCysE possessed similar binding affinities as SepRS for tRNA\textsubscript{Cys}, but SepCysS did not show measurable binding to tRNA\textsubscript{Cys} and the formation of binary complex with SepCysE significantly increased the affinities of SepCysS for tRNA\textsubscript{Cys}\textsuperscript{11}. In the present ternary complex, tRNA\textsubscript{Cys} was in contact with both SepCysE(CTD) and SepCysS, while no direct interaction was observed between SepCysE(CTD) and SepCysS (Fig. 1a). Thus, it is possible that a mobile SepCysE(CTD) connected to the SepCysE(NTD)-SepCysS complex by a flexible linker, mediated the binding of tRNA\textsubscript{Cys} to SepCysS. Interactions of SepCysS to the acceptor stem of tRNA\textsubscript{Cys} may reinforce the binding of tRNA\textsubscript{Cys} to the SepCysE-SepCysS complex. Consistent with this model, the tRNA\textsubscript{Cys} binding affinity of the full complex (SepCysE-SepCysS: K\textsubscript{d} = 0.35 M\textsuperscript{-1}) is somewhat stronger than that of SepCysE(CTD) (K\textsubscript{d} = 0.67 M\textsuperscript{-1}), while the affinity of SepCysE(CTD)-SepCysS for tRNA\textsubscript{Cys} could not be detected (Supplementary Fig. 3b).

SepCysS recognizes the discriminator base U73 of tRNA\textsubscript{Cys}. SepCysS is a pyridoxal-5'-phosphate (PLP)-dependent enzyme and a homodimer with a large domain (about 300 residues, residues Gly43–Lys295) and a small domain (about 100 residues, residues Trp296–Lys396) per monomer. PLP was covalently linked to the side chain of the conserved Lys334 located deep in the cleft at the dimer interface. The inner surface of the cleft was highly positively charged, along which the negatively charged phosphate ions of tRNA\textsubscript{Cys}, the discriminator base U73 was cross-recognized by SepCysS. The fact that SepCysS also utilizes this recognition element. The discriminator for CysRS and SepCysS\textsuperscript{16}. Here we found that SepCysS also utilizes this recognition element. The fact that the discriminator base U73 is cross-recognized by SepRS and SepCysS may explain, at least partly, the relative insensitivity of SepRS to mutation of this nucleotide base compared with CysRS\textsuperscript{17}.

A binding assay of SepCysE-SepCysS to tRNA\textsubscript{Cys} variants confirmed the importance of U73. The tRNA\textsubscript{Cys}(U73G) had a weakened binding affinity to SepCysE-SepCysS, while tRNA\textsubscript{Phe} (A73U) acquired binding ability towards SepCysE-SepCysS (Fig. 1d), suggesting that U73, but not the anticodon of tRNA\textsubscript{Cys}, was the major identity element for SepCysS. U73 in tRNA\textsubscript{Cys} is conserved in the three domains of life, and has been reported as a discriminator for CysRS and SepRS\textsuperscript{16}. Here we found that SepCysS also utilizes this recognition element. The fact that the discriminator base U73 is cross-recognized by SepRS and SepCysS may explain, at least partly, the relative insensitivity of SepRS to mutation of this nucleotide base compared with CysRS\textsuperscript{17}. 

**Fig. 1** Structure of the SepCysE-SepCysS-tRNA\textsubscript{Cys} ternary complex. a Orthogonal views of the ternary complex drawn in ribbon model. A dimer of the SepCysE N-terminal domain (SepCysE(NTD), yellow orange), and the SepCysS (SepCysE(CTD), red), which in turn binds to the elbrow region of tRNA\textsubscript{Cys} (orange). The CCA terminus of tRNA\textsubscript{Cys} is in contact with the active site of SepCysS. A stick representation of PLP is also drawn at the active site. Black and red dashed lines represent flexible loops connected to the disordered N-helix and CTD of SepCysE, respectively. A schematic representation of SepCysE with domain boundaries is given at the bottom. b Close-up views of interactions between SepCysE(CTD) and tRNA\textsubscript{Cys}. Interactions are mainly nonspecific between main-chain phosphates of tRNA\textsubscript{Cys} and lysyl side-chains at α4 and α6 of SepCysE(CTD). Hydrogen bonds are indicated by dashed lines. c A detailed view of the recognition of tRNA\textsubscript{Cys} by SepCysS. The tRNA\textsubscript{Cys} is bound to SepCysS at two regions marked by yellow/green circles in a. d Binding assays of the SepCysE-SepCysS complex to tRNA variants by gel filtration. The result shows the importance of the discriminator base U73 in blue. Red and blue lines represent the absorption at a wavelength of 280 nm and of 260 nm, respectively.
SepCysE dimer binds to SepRS by a pair of N-terminal helices.

During the study of the quaternary complex SepRS-SepCysE-SepCysS-tRNACys, we serendipitously obtained the crystal of the ternary complex SepRS-SepCysE-SepCysS. The structure was solved by the molecular replacement method (Table 1). In the electron density map, whole SepRS and the N-terminal helix (Met1–Lys24) of SepCysE(N-helix) were visible (Supplementary Fig. 6a). However, the remaining components of SepCysE and SepCysS were disordered in the crystal, although there was enough space in the unit cell (Supplementary Fig. 6b) to accommodate the entire complex. SepRS was organized as a tetramer with 222 local symmetry (one of the two-fold axes is crystallographic), as previously observed. A pair of N-terminal helices (N-helices) of the dimeric SepCysE wound around each other to form a coiled-coil structure and were tightly bound to the SepRS with their two-fold axes aligned (Fig. 2a). This N-helix was not visible in the structures of the ternary complex SepCysE-SepCysS-tRNACys, suggesting that an intervening stretch of about 10 residues (Ile25–Ile34) between N-helix and NTD was a flexible linker. This view was consistent with a secondary structure prediction that a long loop connects N-helix and NTD.

The highly conserved hydrophobic residues Leu9, Ile10, Ile17, and Leu20 (Supplementary Fig. 4b) in the N-helix contributed to the hydrophobic interactions within the left-handed coiled-coil structure. Most of these residues (Leu9, Ile10, and Ile17) plus Ile14 were clustered with the hydrophobic residues in three pairs of loops ((loop 1: 376–380 ILDEF), (loop 2: 411–414 FNGE), and (loop 3: 494–498 ALVSN)) locating at the interface of two anticodon domains of the SepRS dimer (Fig. 2b) and contributed to the formation of the complex together with some hydrogen bonds (Fig. 2c). To verify the physiological importance of these interactions, several deletion mutants were made and the binding ability was checked by gel filtration experiments. Whereas deletion mutants of each of the three loops of SepRS showed no significant effect on binding, the double-loop deletion mutants did not bind SepCysE, suggesting these loops together contribute to the interaction with SepCysE (Supplementary Fig. 7). Likewise, N-helix deletion mutant SepCysE(ΔN-helix) completely lost the ability to bind SepRS (Supplementary Fig. 7). This structural and mutational evidence suggested that the SepCysE dimer tightly binds to the SepRS dimer by inserting its coiled-coil N-terminal helices into the interface of the SepRS dimer, which reinforces the stable architecture of transsulfursome.
ATP was observed in the active sites of SepRS. The binding manner of ATP was similar to that of other canonical class II aaRS, especially PheRS\textsuperscript{14}. The bent ATP mainly interacted with motif 2 of SepRS via four conserved residues Arg228, Glu230, His237 and Tyr240. Superposition of the structures of SepRS-SepCysE(N-helix) on Archaeoglobus fulgidus SepRS-tRNA\textsubscript{Cys}.Sep (PDB entry 2DU3) showed that the OH group of Sep was very close to the α-phosphate of ATP, as necessary to form the phosphoseryl-adenylate (Fig. 2d).

**Dynamic architecture of the transsulfursome.** Two complex structures solved in the present work were combined to show that SepCysE consisted of an N-terminal helix (N-helix), an N-terminal domain (NTD) and a C-terminal domain (CTD). These components were connected by two linkers, linker1 and linker2, each composed of about 10 residues (Fig. 3, Supplementary Fig. 8a). Each component of SepCysE bound to each of the three molecules, SepRS, SepCysS, and tRNA\textsubscript{Cys}. Both the N-helix and NTD bound as dimers to the dimeric molecules of SepRS or SepCysS, while the CTD worked as “two monomers” of which only one “monomer” bound to monomeric tRNA\textsubscript{Cys} (Fig. 3). The two intervening linkers (linker1: between N-helix and NTD, and linker2: between NTD and CTD) made SepCysE highly flexible, enabling the movement of the SepCysE(NTD)-SepCysS complex with respect to the SepRS-SepCysE(N-helix) complex and also allowing a large motion of the SepCysE(CTD) with bound tRNA\textsubscript{Cys} during tRNA transfer from SepRS to SepCysS (see below). The flexibility of SepCysE(NTD)-SepCysS relative to SepRS via linker1 was further confirmed by SEC-SAXS and EM analyses.

A small angle X-ray scattering (SAXS) experiment was performed to study the structure of the transsulfursome in solution. For simplicity, CTD-truncated SepCysE was used for this experiment. A plot of the intensity of the scattering data (I), as a function of scattering angle (Q) was compared with the transsulfursome model derived from the crystal structures. To take account of the structural variability of the transsulfursome due to flexible linker1, a number of models were generated by changing the distance and orientation of these two rigid structures, and the theoretical curves calculated from the models were compared with the observed plot. As shown in Fig. 4a, the observed plot matched well to a theoretical curve with χ² agreement of 1.24 for “a straight model” after adjusting the distance between the two rigid structures, SepRS-SepCysE(N-helix) and SepCysE(NTD)-SepCysS connected by linker1. Whereas the fitting was sensitive to the distance between the two rigid structures, it was rather insensitive to the tilt angles and even better fitted models were obtained when the angle was also varied (Fig. 4b). Thus, an ensemble distribution of the structures with various tilt angles by linker1 was expected in solution.

The structure of the transsulfursome complex was further characterized by negative stain electron microscopy and class averaging. The particles were adsorbed onto the EM grid in side orientation. Over 20,000 particles were selected from the EM micrographs and subjected to image alignment and classification. This analysis confirmed that the complex possessed an overall architecture, roughly resembling a “three sectional staff”. Tetrameric SepRS was positioned in the center of the complex and dimeric SepCysE-SepCysS at the opposite ends (Fig. 4c). The SepCysE-SepCysS domains did not occupy stable positions but instead flexed independently of the SepRS domain. Thus, the complex could adopt straight and/or curved conformations without changing its overall composition (Fig. 4d). The wild-type complex underwent a tilting-motion from a curved to an S-shaped conformation (Supplementary Fig. 8a, Supplementary Movie 1) and the distribution of the center-to-center distance between the SepCysE-SepCysS components at the opposite ends of the complex ranged from 170 to 290 Å (Fig. 4d). This suggests that a hinge exists at the site of linker1 connecting SepRS-SepCysE(N-helix) and SepCysE(NTD)-SepCysS, and allows movement of these domains relative to each other. Because of

---

**Fig. 3** Architecture of transsulfursome. Left, transsulfursome model consisting of tetrameric SepRS (green) at the center, two dimers of SepCysS (blue) at both ends, and SepCysE (magenta, red, yellow) at the position connecting SepRS and SepCysS. Right, SepCysE consists of N-helix (magenta), NTD (yellow), and CTD (red) connected by linker1 (black) and linker2 (red). The three domains of SepCysE (N-helix, NTD, and CTD) each binds SepRS, SepCysS, and tRNA\textsubscript{Cys}, respectively as shown in the schematic representation at the bottom. Note that the orientation of SepCysE (CTD) is also arbitral due to the flexibility of linker2.

---
the flexibility of the complex and preferred orientation on the EM grid, we were not able to determine a low resolution 3D volume.

To support this notion, we truncated 10 residues of linker1 (transsulfursome (del_linker1)) and analyzed the resulting complex by negative stain EM and image classification. As expected, shortening of the flexible connector, resulted in reduction of the overall flexibility of the complex. Most of the class-averages, obtained from ~ 12,000 particles of this complex, appeared straight and the tilting-motion of SepCysE-SepCysS relative to SepRS was greatly reduced (Fig. 4d, Supplementary Fig. 8b, Supplementary Movie 2). The increased rigidity was also reflected in the narrower distribution of the distance between the SepCysE-SepCysS domains (230–260 Å, Fig. 4d). Thus, motion of SepCysE-SepCysS relative to SepRS is linker1 dependent.

Fig. 4 SAXS and EM analyses showing the domain flexibility of the transsulfursome. a, b SAXS analysis of the transsulfursome. Experimental scattering data (blue) fit well to a theoretical curve (red) calculated from the combined model from crystal structures. Fitting is not sensitive to the azimuth or tilt angles of the (SepCysE(NTD)-SepCysS) domains. a Best fitted “straight model” with center-to-center distance of end domains (SepCysE(NTD)-SepCysS) of about 240 Å. b Tilted models fit well as long as (SepCysE(NTD)-SepCysS) is within an appropriate distance from SepRS. Given here is one example of a well-fitted “bent model.” c, d Negative stain EM analysis of wild and mutant (del_linker1) transsulfursome. c Selected class-average of a wild type transsulfursome bent (above) and a typical mutant(del_linker1) straight conformation (bottom) (scale bars, 10 nm), with interpretative cartoons showing SepRS, SepCysS and SepCysE in green, blue and khaki, respectively. The cartoons were obtained after graphically mapping a low-pass filtered 3D model of each component into the respective class average. Rulers indicate the distribution range of the total distance between the SepCysE-SepCysS domains at opposite ends of the complex. d Center-to-Center distance distribution between the SepCysE-SepCysS domains measured for all class-averages of the wild type transsulfursome (above) and the mutant (del_linker1) (bottom) data set, respectively. Representative class averages are shown and marked by connecting lines. Each class average contains 60–140 single particles. Note that the two SepCysE-SepCysS domains tilt independent from each other and the complex can adopt thereby bent and S-shaped conformations. Typical micrographs and larger sets of class averages are shown in Supplementary Fig. 8.
Discussion

The ternary complex SepCysE-SepCysS-tRNACys solved in the present work presents the structure of which the 3′-end moiety of tRNACys is at the active site for the second reaction, while SepRS-tRNACys complex solved previously shows the structure for the first reaction13. The distance between these two active sites in the transsulfursome is not fixed due to the conformational change caused by the flexible linker1. However, model building shows that the distance is no less than 90 Å. Therefore, tRNACys must move a long distance between these two sites during two-step reactions. Furthermore, as both SepRS and SepCysS access to the same side of tRNACys, it is not possible for SepRS and SepCysS to bind to the tRNACys simultaneously and to make a ternary complex. In contrast, as SepCysE(CTD) accesses to tRNACys from opposite side, it binds to the tRNACys without disturbing the binding of the SepRS or SepCysS to tRNACys. These considerations suggest a scenario for this consecutive reactions where SepCysE(CTD) carries tRNACys from the first active site to the second active site of SepRS while SepCysS(N-helix) remains fixed to SepRS. The model building shows such complex is possible by extending the two linkers of SepCysE. Two flexible linkers of SepCysE allow the movement of tRNACys over this long distance. The discriminator base of U73 and possibly the Sep moiety of tRNACys are recognized by SepCysS, followed by conversion to the canonical Cys-tRNACys. This scenario is consistent with the experiment that SepCysE enhances the binding of SepRS and tRNACys 11.

Fig. 5 Proposed mechanism of the tRNA-dependent cysteine biosynthesis by transsulfursome. a tRNA channeling for consecutive reactions of the transsulfursome. Only upper half of the transsulfursome is drawn here. (Left) The first reaction state (SepRS active form). This model was built by superposing the SepCysE(CTD)-tRNACys complex (present work) with the SepRS-tRNA complex (PDB:2DU3). The stretching of linker1 and linker2 allows the SepCysE(CTD)-tRNACys complex to reach the binding site on SepRS, while anchoring SepCysE(N-helix) to its binding site on SepRS. The anticodon of tRNACys is recognized by the site close to the SepCysE(N-helix) binding site. (Middle) The intermediate state. The mischarged Sep-tRNACys moves from the first reaction site on SepRS to the second reaction site on SepCysS without being released from bound SepCysE(CTD). Two flexible linkers of SepCysE allow the movement of Sep-tRNACys over this long distance. (Right) The second reaction state (SepCysS active form). The discriminator base of U73 and possibly the Sep moiety of Sep-tRNACys are recognized by SepCysS, followed by conversion to the canonical Cys-tRNACys. The anticodon of tRNACys may not be recognized by SepRS in this step. b A schematic drawing of the tRNA-dependent cysteine biosynthesis by transsulfursome. N-helix (purple) and NTD (orange) of SepCysE bridge the SepRS (green) and SepCysS (blue) to form ternary complex (transsulfursome) with a molar ratio of 4:4:4. SepCysE (CTD) (red) shuttles tRNACys between two distinct active sites on transsulfursome for sequential reactions catalyzed by SepRS and SepCysS. Each of the two enzymes recognizes tRNACys separately in the two steps. The released Cys-tRNACys is delivered to ribosome by EF-1α to finally incorporate a cysteine into proteins. Two tRNA substrates are bound to the opposite positions on transsulfursome. The recognition sites of tRNACys by SepRS and SepCysS are marked with black and yellow asterisks, respectively.
SepRS when CCA terminus binds to the second reaction site (Fig. 5a, right). In this model, two linkers are critically important to satisfy the flexibility required for this long-distance swing of tRNA Cys. Therefore, we examined this model by making two SepCysE mutants of which each of two linkers were truncated. The importance of the linkers was studied by an in vivo genetic study. The ΔSepRSAΔSepCysE double deletion mutant strain is a cysteine auxotroph, which can be complemented by expressing wild-type SepRS and SepCysE from a shuttle vector. However, expression of wild-type SepRS together with either SepCysE(del-linker1) or SepCysE(del-linker2) in ΔSepRSAΔSepCysE still required cysteine for growth (Fig. 6). This result indicates that both linkers are essential for the function of the transsulfursome.

The multiple domain architecture of SepCysE confers flexibility to the transsulfursome, which enables SepCysE-SepCysS moiety to bend to SepRS to facilitate the access of tRNA acceptor arm to active site of SepRS. However, this bending keeps SepRS out of reach for another half of the dimeric moiety of SepCysE-SepCysS. That is, the dimeric SepCysE-SepCysS can convey only one tRNA at a time (Fig. 5b). In all, the transsulfursome can afford to catalyze two tRNAs at a time although it possesses four sets of active sites. This is compatible with our result that one transsulfursome only binds two tRNAs (Supplementary Fig. 9) and also to the previous results that the tetrameric SepRS binds two tRNAs and that only two of the four chemically equivalent subunits are active. It is possible that the left and the right active sites of the transsulfursome are alternately used for reaction. As a reference, the alternate mechanism is believed to be adopted by HisRS.

The highly interdigitated structure of SepRS may facilitate the alternate catalysis. On the basis of these results, we propose the scheme of the tRNA-dependent cysteine biosynthesis by the indirect pathway (Fig. 5b), where multidomain SepCysE binds SepRS and SepCysS to assemble into a functional unit (transsulfursome) and mediates the transfer of tRNA substrate between the active sites of SepRS and SepCysS via flexible linkers. The released final product Cys-tRNACys is delivered to ribosome by EF-1α to incorporate a cysteine into proteins.

Among the indirect pathways known today, the system for Gln/Asn-tRNA Gln/Asn synthesis is best studied. The complex named transamidosome is responsible for this synthesis. The transamidosome studied here is very different from this system, although both systems are involved in the indirect aminoclaylation. First, formation of the transamidosome is tRNA-independent and could be stable throughout the two reactions, whereas transamidosome does not form a stable complex. This is reasonable because dynamic dissociation of the complex is important for transamidosome, as non-discriminating GluRS/AspRS show dual-specificity to tRNA Glu/tRNA Asp and tRNA Gln/tRNA Asn, and the second enzyme amidotransferase (GacB) is also employed in the two kinds of transamidosomes. By contrast, the transsulfursome is specialized for synthesizing Cys-tRNACys, and stable formation of the transsulfursome is beneficial for the efficient function of this complex in a high-temperature habitat. Second, unlike the case of the transamidosome, both SepRS and SepCysS have access to the same site of the major groove of tRNA Cys to recognize the discriminator base U73. This suggests that the two enzymes evolved independently and only at a later stage they were combined together to form a system for Cys-tRNACys synthesis. This scenario implies that it is difficult to form a ternary complex such that the CCA terminus of tRNA is close to both active sites of the two enzymes and that the system requires a large movement of tRNA rather than just flipping the acceptor arm of tRNA between the two active sites, as in the case of transamidosome. Third, whereas in the transamidosome the first enzyme (Glu/AspRS) appears to continuously recognize the anticodon of the cognate tRNA throughout the entire two-step reaction, our model of the transsulfursome suggests that the anticodon is recognized only at the first step. In fact, it is unlikely that SepRS could specifically recognize tRNA Cys in both steps because tRNA Cys needs to be accessed from completely different directions at each step. The recognition of U73 of tRNA Cys by SepCysS also implies that the anticodon recognition by SepRS is redundant in the second step, although we cannot rule out the possibility that tRNA Cys maintains some kind of interaction with SepRS to safeguard the transfer of tRNA Cys.

The present structure analysis revealed how SepCysE assembles two enzymes (SepRS and SepCysS) and a substrate (tRNA) together to create a mechanism of aminoclaylation of tRNA in the indirect pathway. However, the indirect pathway for aminoclaylation of tRNA is associated with the risk that the mischarged intermediate is used for protein synthesis. Therefore, the direct pathway is employed in most extant organisms. Indeed, in some archaeal lineages, CysRS was horizontally transferred and replaced the indirect mechanism. Nonetheless, the ancestral indirect pathway for Cys-tRNACys synthesis has been preserved in methanogenic archaea, although the reason for this is unclear. One possible explanation is the energetic advantage by the direct link of amino acid synthesis and protein synthesis in methanogens where cysteine in the proteome is in a large demand. Whatever the reason for this would be, the structure of transsulfursome reflects the most primordial feature of a multiprotein complex each of whose components have acquired individual functions. SepCysE, which is made up of three different binding
domains, enhances the two enzymes together. By binding the substrate tRNA, it further increases the reaction efficiency while reducing the risk that the mischarged intermediate is released. This complex may not only have facilitated the addition of Cys to the genetic code but also contributed to the evolution of other multi-protein complexes. The architecture of the transsulfursome has the potential to guide protein engineering for site-specific incorporation of the unnatural amino acid O-phosphoserine (Sep) into the proteome.15, 36

Methods
Expression and purification of transsulfursome. The pET-15b vector encoding M. jannaschii SepRS was transformed into E. coli strain B834(DE3) plus pRARE2 (Novagen). The pET-15b vector encoding SepCysE together with pCDF Duet-1 derived plasmid encoding SepCysE were cotransformed into B834(DE3) plus pRARE2. Overexpression of the proteins was induced by 0.5 mM IPTG at 25 °C overnight. For purification of the proteins, cells expressing SepCysE-SepCysS and SepRS were separately sonicated and the extracts were heat-treated at 75 °C for 30 min. The SepCysE-SepCysS was first purified with a HisTrap HP column (GE Healthcare) and eluted with a linear gradient of 20–500 mM imidazole in buffer containing 50 mM HEPES-NaOH (pH 7.5), 300 mM NaCl, and 5 mM MgCl2. The collected eluate was then mixed with purified SepRS before loading onto an additional HisTrap HP column (GE Healthcare). The eluted fractions were pooled and purified using a Hitrap Heparin HP column and eluted with a linear gradient of 0.3–1 M of NaCl in the buffer containing 50 mM HEPES-NaOH (pH 7.5) and 5 mM MgCl2. The protein fractions were then loaded onto a HiLoad 6/10 Superdex 200 prep grade column (GE Healthcare) equilibrated with buffer containing 20 mM HEPES-KOH pH 8.0, 150 mM KCl, 5 mM MgCl2, and 10% glycerol. The 10% denaturing purification of the transsulfursome follows the same procedure as described above. The SepCysE-SepCysS complex was purified using the same protocol but without addition of SepRS.36 Mutagenesis was performed using a QuikChange mutagenesis kit (Agilent Technologies). The information priors used in mutagenesis are summarized in Supplementary Table 2.

Preparation of M. jannaschii tRNA<sub>Cys</sub>. The M. jannaschii tRNA<sub>Cys</sub> was transcribed in vitro by T7 polymerase.18 PCR was performed to synthesize the template DNA for tRNA<sub>Cys</sub> transcription. The sequence of the DNA oligonucleotides used in PCR were as follows: Forward primer 5′-TTAGGCTGACACCTATAGCCGG GTAGCTTGAGGAGGACGACT-3′; Middle primer 5′-CTAGGAC GGCGACTGAGATCCGGCCTTAGCCTGGT-3′; Reverse primer 5′-TGGACGGCGGG CCCTTGTACCAACGGCAG-3′. T7 polymerase is used for transcription. After transcription at 37 °C for 5 h, reaction was terminated and product was loaded onto DEAE column. The tRNA<sub>Cys</sub> was eluted by linear increase of 0.15–2 M NaCl gradients in buffer containing 50 mM MES (pH 6.5), 150 mM NaCl, and 0.2 M EDTA. The tRNA<sub>Cys</sub> was further purified by 10% denaturing Urea-PAGE and stored at 30 °C.

Crystalllographic analysis of SepCysE-SepCysS-S<sub>i</sub> tRNACys. SepCysE-SepCysS and tRNA<sub>Cys</sub> were mixed at a molar ratio of 1:1.2, with a protein concentration of 16 μM. T7 promoter is underline. After transformation, the two enzymes together. By binding the substrate tRNA, it further increases the reaction efficiency while reducing the risk that the mischarged intermediate is released. This complex may not only have facilitated the addition of Cys to the genetic code but also contributed to the evolution of other multi-protein complexes. The architecture of the transsulfursome has the potential to guide protein engineering for site-specific incorporation of the unnatural amino acid O-phosphoserine (Sep) into the proteome.15, 36

Structure determination and refinement of transsulfursome. The structure was solved by the molecular replacement method using AutoMR of the Phenix package. The structures of M. jannaschii SepRS (PDB entry 2DUT), M. jannaschii SepCysE-SepCysS-SepCysE(NTD) (PDB entry 3WKR), and E. coli tRNA<sub>Cys</sub> (PDB entry 1B23) were used as search models. However, only SepRS was found by this search. In the Fo-Fc map subsequently created, a pair of N-terminal helices of SepCysE was obtained. A truncated model of SepCysE was found in the active site of SepRS. Structural refinement was performed using phenix.refine with secondary structure restraints, and final R<sub>work</sub> and R<sub>free</sub> factors converged to 18.5% and 20.9%, respectively. Statistical information of the structure refinement is summarized in Table 1.

Isothermal titration calorimetry experiment. Isothermal titration calorimetry (ITC) experiment was performed to investigate the interaction of proteins and tRNA. Protein and tRNA were dialyzed to buffer containing 20 mM HEPES-KOH pH 8.0, 150 mM KCl, 5 mM MgCl<sub>2</sub>, and 10% glycerol. The protein samples were loaded to the cell of Nano ITC (TA Instruments) and 25 aliquots of 2 μL of tRNA were injected into the cell with a time interval of 180 s under a stirring speed of 250 rpm at room temperature. The data were analyzed using the program NanoAnalyze. The control (buffer to buffer, to tRNA to buffer) was also performed to normalize the experiment.

Gel filtration assay. The binding affinity of SepCysE-SepCysS and tRNA mutants was analyzed by gel filtration. The 50 μM protein was incubated with 30 μM tRNA at room temperature for 1 h in the solution containing 20 mM HEPES-KOH pH 8.0, 150 mM KCl, 5 mM MgCl<sub>2</sub>, and 5% glycerol. The sample was loaded onto a Superdex 200 10/300 GL column (GE Healthcare) and the elution profile was monitored at 254 and 280 nm. To investigate the binding between SepRS and SepCysE, 100 μM SepRS variants were incubated with 400 μM SepCysE variants and loaded onto a Superdex 200 10/300 GL column equilibrated with binding buffer containing 20 mM HEPES-NaOH pH 7.5, 300 mM NaCl, and 10% glycerol.

Electrophoretic mobility shift assay. Electrophoretic mobility shift assay (EMSA) was performed to analyze the binding ratio of proteins and tRNA. For transsulfursome and SepCysE-SepCysS, 10 μM tRNA was mixed with stepwise increases of proteins in the binding buffer containing 20 mM HEPES-KOH pH 8.0, 150 mM KCl, 5 mM MgCl<sub>2</sub>, and 10% glycerol and incubated at room temperature for 30 min. To characterize SepCysE(CTD) as a universal RNA-binding domain, 10 μM tRNA<sub>Cys</sub>, tRNA<sub>Sec</sub>, tRNA<sub>Phe</sub> and tRNA<sub>His</sub> were separately incubated with 80 μM SepCysE(CTD). The samples were loaded onto a 10% non-denaturing polyacrylamide gel and electrophoresis was performed at a constant voltage of 100 V. The gel was separated with ethidium bromide and Coomassie brilliant blue.
mask including the complete complex. The final class-averages contained 24–167 members, showed improved features for all components and were used to measure the center-to-center distance between the SepCysE-SepCysS components at the opposite ends of the complex, as shown in Fig. 4c, d. A sequence of selected class-averages in random order was used to create a movie (Supplementary Movie 1 and 2).

**Mutagenesis of Methanococcus maripaludis.** The replacement of the ssCys gene (mmp1240, encoding SepCysS) with a pyrC resistance (pase) cassette was generated in the markerless ΔsepCysS strain by homologous recombination. We confirmed the genotype of the ΔsepRS/ΔsepCysE-SepCysS-pac strain (S700) by Southern hybridization. For expression of SepRS and SepCysS, the two genes, mmp0688 and mmp1240, were inserted into the vector pMEV2 under the control of the hmtA17 and sla promoters, respectively. The sepCysE-SepCysS strain (S702) was then obtained by transformation of S700 with pMEV2-mmp0688-mmp1240. The SepCys(Δ-RF) variant was constructed with the plasmid pMEV2-mmp0688-mmp1240 using the QuikChange mutagenesis kit (Agilent Technologies) and then transformed into strain S700. The mutations of SepCysE (MMPI217), including the variants of SepCysE(del-linker1) and SepCysE(del-linker2), were constructed by using the plasmid pMEV2-mmp0688-mmp1240 and then transformed into the ΔsepRS/ΔsepCysE-SepCysS-pac strain (S761).11

We grew. *M. maripaludis* at 37°C in 28-ml aluminum sealed tubes with 5 ml of McFaa medium (minimal medium + 0.4 M sodium formate + 10 mM sodium acetate + 1 mM L-Ala) buffered with 0.2 M glycolline (pH 8.0) and reduced with 3 mM dithiothreitol11. L-Cys (3 mM), pyruronic (2.5 μg ml⁻¹), and neomycin (0.5 mg ml⁻¹ in plates and 1 mg ml⁻¹ in broth) were added as needed. Antibiotics were omitted when comparing the growth of the WT and mutants. As the sulfur source, 3 mM sodium sulfide was added before inoculation.

**Data availability.** Coordinates and structure factors have been deposited in the Protein Data Bank under accession code PDB 5X6B (SepCysE-SepCysS-tRNACys) complex) and PDB 5X6C (SepCysS complex). Other data are available from the corresponding author upon reasonable request.

Received: 24 March 2017 Accepted: 26 September 2017 Published online: 15 November 2017

**References**

1. Schimmel, P. Aminoacyl-tRNA synthetases: general scheme of structure-function relationships in the polyptides and recognition of transfer RNA. Annu. Rev. Biochem. 56, 125–158 (1987).

2. Ibba, M. & Soll, D. Aminoacyl-tRNA synthesis. Annu. Rev. Biochem. 69, 617–650 (2000).

3. Bult, C. J., White, O., Olsen, G. J. & Zhou, L. Complete genome sequence of the archaeon *Methanococcus maripaludis*. Science 273, 1058–1060 (1996).

4. Ibba, M., Becker, H. D., Stathopoulos, C., Tumbula, D. L. & Söll, D. The homotetrameric phosphoseryl-tRNA synthetase from Methanosarcina mazei exhibits half-of-the-sites activity. Proc. Natl Acad. Sci. USA 105, 11238–11237 (2008).

5. Becker, H. D. & Stern, D. Thermus thermophilus: a role in the evolution of RNA-dependent amino acid amidation pathways. Proc. Natl Acad. Sci. USA 95, 12832–12837 (1998).

6. Nakamura, A., Yao, M., Chimnaronk, S., Sakai, N. & Tanaka, I. Ammonia channel couples glutaminase with transamidase reactions in GtCAB. Science 312, 1954–1958 (2006).

7. Heberle, B., Moreira, F., Moverare, T., Zdarsky, M. & Franklyn, C. S. Asymmetric amino acid activation by class II histidyl-tRNA synthetase from Escherichia coli. J. Bacteriol. 184, 20753–20762 (2009).

8. Bailly, M., Blaise, M., Lorber, B., Becker, H. D. & Stern, D. The transamidosome: a dynamic ribonucleoprotein particle dedicated to prokaryotic tRNA-dependent asparagine biosynthesis. Mol. Cell 28, 228–239 (2007).

9. Rumpfia, T., Sheppard, K. & Soll, D. The archaeal transamidosome for RNA-dependent glutamine biosynthesis. Nucleic Acids Res. 38, 5774–5783 (2010).

10. Lapointe, J., Duplain, L. & Proulx, M. A single glutamyl-tRNA synthetase aminoacylates tRNAGlu and tRNA(Gln) in *Bacillus subtilis* and efficiently miscleavages *Escherichia coli* tRNA1Gln in vitro. J. Bacteriol. 165, 88–93 (1986).

11. Becker, H. D. & Stern, D. The antiquity of biogeochemical processes in the archaeal domain of life. Proc. Natl Acad. Sci. USA 107, 11515–11518 (2010).

12. Yamashita, K., Zhou, Y., Tanaka, I. & Oya, M. New model-fitting and model-completion programs for automated iterative nucleic acid refinement. Acta Crystallogr. Sect. D Biol. Crystallogr. 67, 768–776 (2011).

13. Donoghue, P., Schnitzbauer, M. & Söll, D. Model building for analysis of small-angle x-ray scattering of biological macromolecules from atomic coordinates. Acta Crystallogr. Sect. D Biol. Crystallogr. 69, 768–776 (2013).

14. Adame, P. D. et al. PHENIX: a comprehensive Python-based system for macromolecular structure solution. Acta Crystallogr. Sect. D Biol. Crystallogr. 66, 213–221 (2010).

15. Shimizu, N. et al. Software development for analysis of small-angle x-ray scattering data. In AIP Conference Proceedings 1741, 050017 (AIP Publishing, 2016).

16. Svergun, D., Barberato, C. & Koch, M. H. J. CRYOS—a program to evaluate X-ray solution scattering of biological macromolecules from atomic coordinates. J. Appl. Crystallogr. 28, 768–775 (1995).

17. Moriya, T. et al. High-resolution single particle analysis from electron cryo-microscopy images using SPHIRE. J. Vis. Exp. 16, e55448–e55448 (2017).

18. Tang, G. et al. EMAN2: an extensible image processing suite for electron microscopy. J. Struct. Biol. 157, 38–46 (2007).
Acknowledgements
The synchrotron radiation experiments were performed with the approval of SPring-8 (Proposal No. 2015B1024, 2016A2724) and Photon Factory (Proposal No. 2014G080, 2015R-74, 2016G027, and 2016G141). We thank Ms. Xiaomei Sun for her help in ITC experiment, and Dr. Shinya Saijo for his help in SEC-SAXS experiment and the beamline staff of SPring-8 and Photon Factory for their assistance with data collection. This work was supported by Grant-in-Aid for Scientific Research (B) (15H04334 to I.T., 17H03637 to M.Y.) from the Ministry of Education, Culture, Sports, Science and Technology, Japan, and U.S. National Science Foundation (MCB-1632941 to Y.L. and MCB-1410102 to W.B.W.). S.R. thanks the Max Planck Society and the European Council under the European Union Seventh Framework Programme (FP7/2007–2013) (Grant No. 615984) for supporting this work. M.C. was supported by the International Graduate Program (IGP) ‘Training Program for Global Leaders in Life Science’.

Author contributions
M.C., A.N. and M.Y. designed the experiments. M.C. performed biochemical and crystallographic experiments. M.C., K.K. and M.Y. contributed to crystal structure analyses. M.C. and Y.K. performed in vitro assays. Y.L., F.L. and W.B.W. contributed to in vivo assays. M.C., N.S., A.S. and M.Y. collected and analyzed SAXS data. Y.T., P.L., C.G. and S.R. contributed to EM experiments and analyses. M.C., I.T. and M.Y. wrote the manuscript with help from all authors.

Additional information
Supplementary Information accompanies this paper at doi:10.1038/s41467-017-01543-y.
Competing interests: The authors declare no competing financial interests.

Reprints and permission information is available online at http://npg.nature.com/reprintsandpermissions/

Publisher’s note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article’s Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article’s Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit http://creativecommons.org/licenses/by/4.0/.

© The Author(s) 2017