nature chemical biology

Discovery of proteinaceous N-modification in lysine biosynthesis of *Thermus thermophilus*

Akira Horie[1,7], Takeo Tomita[1], Asako Saiki[1,7], Hidetoshi Kono[2,3], Hikari Taka[4], Reiko Mineki[4], Tsutomu Fujimura[4], Chiharu Nishiyama[5], Tomohisa Kuzuyama[1] & Makoto Nishiyama[1,6]

Although the latter portion of lysine biosynthesis, the conversion of α-aminoadipate (AAA) to lysine, in *Thermus thermophilus* is similar to the latter portion of arginine biosynthesis, enzymes homologous to ArgA and ArgJ are absent from the lysine pathway. Because ArgA and ArgJ are known to modify the amino group of glutamate to avoid intramolecular cyclization of intermediates, their absence suggests that the pathway includes an alternative N-modification system. We reconstituted the conversion of AAA to lysine and found that the amino group of AAA is modified by attachment to the γ-carboxyl group of the C-terminal Glu54 of a small protein, LysW; that the side chain of AAA is converted to the lysyl side chain while still attached to LysW; and that lysine is subsequently liberated from the LysW-lysine fusion. The fact that biosynthetic enzymes recognize the acidic globular domain of LysW indicates that LysW acts as a carrier protein or protein scaffold for the biosynthetic enzymes. This study thus reveals the previously unknown function of a small protein in primary metabolism.

It was once thought that bacteria and plants biosynthesize lysine (**1**) via the diaminopimelate (**2**, DAP) pathway, whereas fungi biosynthesize lysine from α-aminoadipate (**3**, AAA)[1–3]. We previously found that, in fact, the bacterium *T. thermophilus* synthesizes lysine via AAA[4]. The enzymes involved in the first half of the lysine biosynthetic pathway in *T. thermophilus* are similar to those in the leucine (**4**) biosynthetic pathway or tricarboxylic acid cycle[5–9], as is the case in fungal lysine biosynthesis through AAA. However, the latter half of the pathway is totally different from the corresponding portion of fungal lysine biosynthesis and involves enzymes similar to those in the arginine (**5**) biosynthetic pathway[10,11] (**Fig. 1a**). Since our discovery of lysine biosynthesis through AAA, evidence has mounted that many microorganisms synthesize lysine by a similar pathway, indicating that this pathway is one of the origins of lysine biosynthesis[12].

In bacteria, the arginine biosynthetic pathway commonly begin with the modification of the α-amino group of glutamate (**6**) by an acetyl group. The reaction is catalyzed by *N*-acetylglutamate (**7**) synthase (ArgA)[13,14] and/or *N*-acetylornithine (**8**) acetyltransferase (ArgJ)[14,15], yielding *N*-acetylglutamate while avoiding an undesirable cyclization to form Δ^1-pyroline-2-carboxylate (**9**). *N*-Acetylglutamate, in turn, is phosphorylated by *N*-acetylglutamate kinase (ArgB), reduced by *N*-acetylglutamylphosphate (**10**) reductase (ArgC), transaminated by *N*-acetylornithine aminotransferase (ArgD) and deacetylated by *N*-acetylornithine deacetylase (ArgE) to produce ornithine (**11**). Ornithine is converted to arginine by a series of reactions catalyzed by ornithine carbamoyltransferase (ArgF), argininosuccinate synthase (ArgG) and argininosuccinate lyase (ArgH). Distinct amino acid sequence homology between the enzyme components of the lysine and arginine biosynthetic pathways has led to the presumption that lysine biosynthesis in *T. thermophilus* proceeds similarly to the synthesis of ornithine from glutamate. Specifically, LysZ, LysY, LysJ and LysK show substantial amino acid sequence identity to ArgB (29%), ArgC (39%), ArgD (41%) and ArgE (20%), respectively, and are thought to function similarly to these homologs. However, there are no paralogs of *argA* and *argJ* in the lysine biosynthetic enzyme gene clusters. Instead, the major gene cluster contains two additional genes, *lysX* and *lysW* (**Fig. 1b**), whose products show no discernible homology to ArgA or ArgJ.

LysX shares homology with the ATP-dependent carboxylate-amine/thiol ligase superfamily, which includes glutathione synthetase, the enzyme that ligates glycine (**12**), an amino donor, with γ-glutamylcysteine (**13**), an amino acceptor, to produce glutathione (**14**)[16]. The enzymes in this family are known to activate the carboxyl group of the substrate by phosphorylation and to catalyze the condensation of the carboxyl group to the amino group, forming a C-N bond. A gene homologous to *lysX* is included in putative lysine biosynthetic gene clusters in all microorganisms that have been proposed to biosynthesize lysine through a pathway similar to that in *T. thermophilus* (**Fig. 2**). Therefore, it is likely that LysX activates a carboxyl group through phosphorylation using ATP as the phosphate donor and ligates it with the amino group of AAA to avoid the formation of a cyclic compound, Δ^1-piperidine-2-carboxylate (**15**, Pip2C), during

[1]Biotechnology Research Center, The University of Tokyo, Bunkyo-ku, Tokyo, Japan. [2]Computational Biology Group, Quantum Beam Science Directorate, and [3]Quantum Bioinformatics Team, Center for Computational Science and e-System, Japan Atomic Energy Agency, Kizugawa, Kyoto, Japan. [4]Division of Biochemical Analysis, Central Laboratory of Medical Sciences, Juntendo University School of Medicine, Bunkyo-ku, Tokyo, Japan. [5]Atopy (Allergy) Research Center, Juntendo University School of Medicine, Bunkyo-ku, Tokyo, Japan. [6]RIKEN SPring-8 Center, Sayo-cho, Sayo-gun, Hyogo, Japan. [7]Present addresses: Central Laboratories for Frontier Technology, Kirin Holdings Co., Yokohama, Japan (A.H); Life Science Division, Yamaha Motor Co., Shizuoka, Japan (A.S.). Correspondence should be addressed to M.N. (umanis@mail.ecc.u-tokyo.ac.jp).

Received 7 January; accepted 1 May; published online 20 July 2009; doi:10.1038/nchembio.198

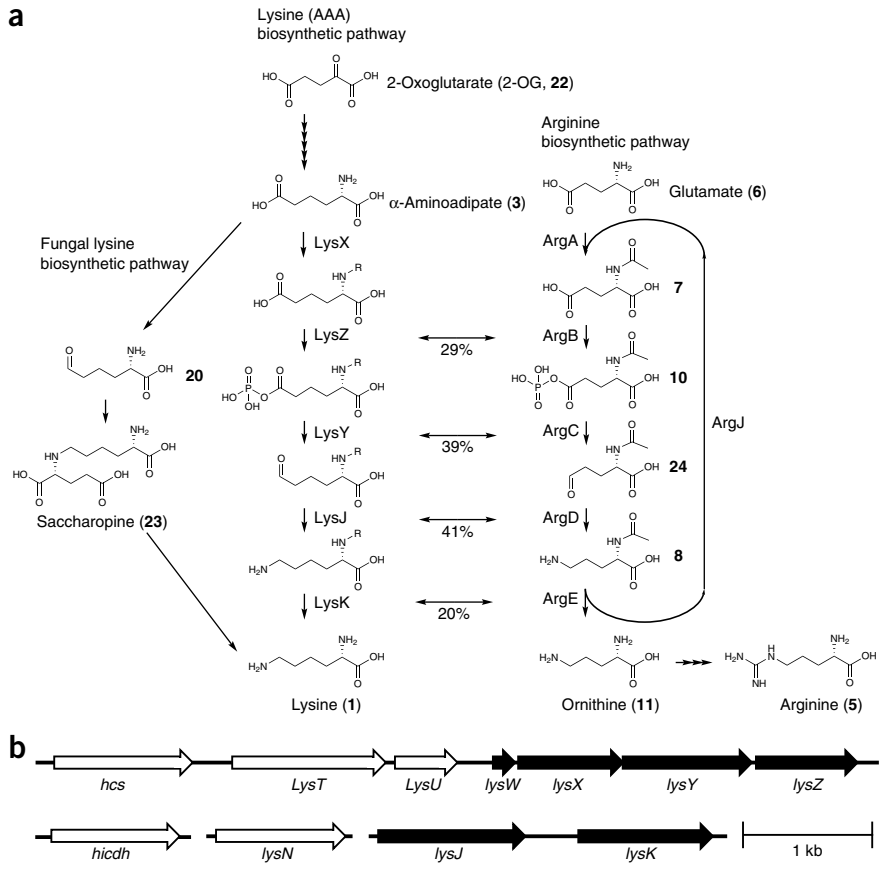

Figure 1 Lysine biosynthetic pathway. (**a**) Enzymes and intermediates in the bacterial and fungal lysine (**1**) biosynthetic pathways through AAA (**3**) and the related portion of the arginine (**5**) biosynthetic pathways. Identity in amino acid sequence between enzymes involved in the *T. thermophilus* AAA pathway and the corresponding homologs of arginine biosynthesis in *T. thermophilus* are shown under arrows. (**b**) Schematic representation of lysine biosynthetic gene clusters of *T. thermophilus*. The genes *lysW*, *lysX*, *lysY*, *lysZ*, *lysJ* and *lysK*, which are involved in the latter part of the AAA pathway, the conversion of AAA to lysine, are indicated by black arrows.

later reactions. Yet it is unclear what substrate serves as the amino acceptor and becomes an N-modification group. All our attempts to detect the activity of LysX that modifies the amino group of AAA using small chemical compounds, such as acetate (**16**), succinate (**17**), formate (**18**), glutamate and AAA, have so far failed.

LysX shows substantial amino acid sequence identity (28%) to *Escherichia coli* RimK, which is also a member of the ATP-dependent carboxylate-amine/thiol ligase superfamily and which adds 2–4 glutamate molecules to the C-terminal glutamate residue of the ribosomal protein S6 (ref. 17). On the basis of these observations, we speculated that LysX uses a C-terminal glutamate group of a small protein as the amino acceptor for AAA. What seemed to us the most likely candidate for such a small-protein amino acceptor for AAA is LysW, a 54-amino-acid acidic peptide with a conserved C-terminal sequence, EDWGE, encoded upstream of the *lysX* gene. All microorganisms that have been proposed to biosynthesize lysine through a pathway similar to the one in *T. thermophilus* also contain the *lysW* homolog in the putative lysine biosynthetic gene cluster (**Fig. 2**).

In this study, we have performed a successful synthesis of lysine from AAA and shown that first AAA is attached to the γ-carboxyl group of the C-terminal Glu54 of LysW and then its side chain is converted to the lysyl side chain in the LysW-conjugated form. Thus, LysW functions not only as an N-modification group for the amino group of AAA but also as a carrier protein or protein scaffold that is well recognized by the lysine biosynthetic enzymes.

RESULTS

Conversion of AAA to lysine

For *in vitro* conversion of AAA to lysine, we purified the LysW, LysX, LysY, LysZ, LysJ and LysK proteins from recombinant *Escherichia coli* cells harboring appropriate expression plasmids (**Supplementary Fig. 1**). After mixing all of these protein components and the putative cofactors of the enzymes and incubating them at 60 °C for 1 h, we detected a substantial amount of lysine (**Fig. 3a**). We did not observe lysine production when any one of the protein components was omitted from the reaction mixture (**Supplementary Fig. 2a–g**). Thus, the reaction mixture contained sufficient and necessary protein components for lysine synthesis.

In support of our hypothesis about LysW function, LysW of *T. thermophilus* is a 54-amino-acid acidic peptide with a calculated pI

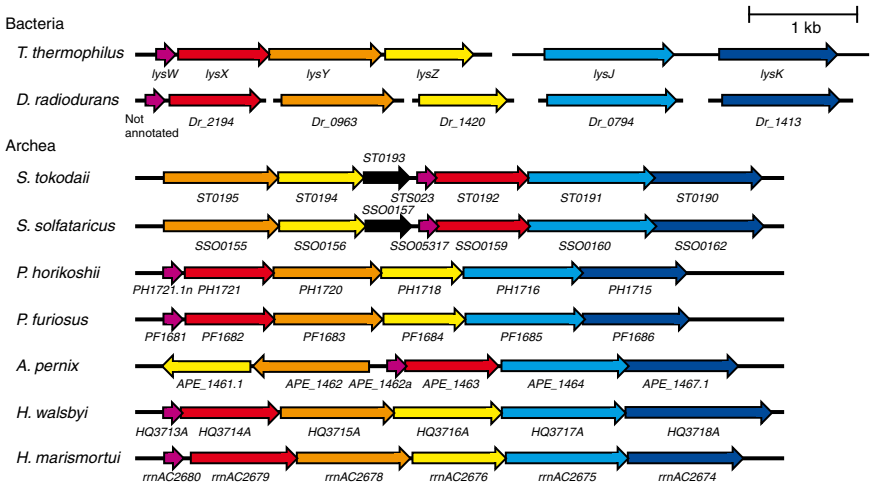

Figure 2 Gene clusters for lysine biosynthetic enzymes and carrier proteins in *T. thermophilus*, *D. radiodurans*, and several archaea. Homologs of *lysW*, *lysX*, *lysZ*, *lysY*, *lysJ* and *lysK* are shown as pink, red, yellow, orange, sky blue and dark blue arrows, respectively. Genome sequence data used are for *T. thermophilus*[28], *D. radiodurans*[29], *Sulfolobus tokodaii*[30], *Sulfolobus solfataricus*[31], *Pyrococcus horikoshii*[32], *Pyrococcus furiosus*[33], *Aeropyrum pernix*[34], *Haloquadratum walsbyi*[35] and *Haloarcula marismortui*[36].

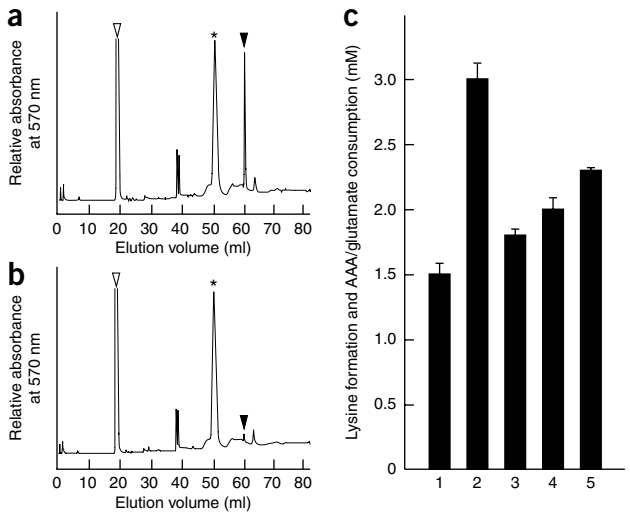

Figure 3 *In vitro* lysine synthesis. (**a**,**b**) Chromatograms of amino acid analyses of reaction mixtures. Proteins added were: (**a**) LysW, LysX, LysZ, LysY, LysJ and LysK; (**b**) LysW (E54A), LysX, LysZ, LysY, LysJ and LysK. AAA (**3**) and Lys (**1**) are indicated by open and closed arrowheads and ethanolamine by an asterisk. (**c**) Lysine formation and AAA and glutamate consumption in the *in vitro* reaction. The vertical axis shows the amounts of lysine, AAA or glutamate (**6**) either formed or consumed as positive values. The amounts, which were quantified with an amino acid analyzer, are shown as bars with s.e.m. ($n = 3$). 1, lysine formation in reaction mixture using all active protein components in the presence of 10 mM AAA (as in **a**); 2, AAA consumption in reaction mixture using all active protein components in the presence of 10 mM AAA (as in **a**); 3, lysine formation in reaction mixture using all active protein components in the presence of 10 mM each of AAA and glutamate; 4, AAA consumption in reaction mixture using all active protein components in the presence of 10 mM each of AAA and glutamate; 5, glutamate consumption in reaction mixture using all active protein components in the presence of 10 mM each of AAA and glutamate.

of 3.7, and *in silico* modeling suggests that it has a quite acidic surface (**Supplementary Fig. 3a**). The modeled structure of LysW shows that it is composed of two structural domains, the globular domain (residues 1–43) and C-terminal extension (residues 44–54); the latter includes five conserved residues, both of which have considerable acidic surface (**Supplementary Fig. 3a**). All of the enzymes involved in the conversion of AAA to lysine (LysX, LysZ, LysY, LysJ and LysK) have basic regions surrounding the active sites (**Supplementary Fig. 3b–f**), indicating that they are likely to interact with LysW. In fact, several residues in this region are actually conserved among all the putative enzymes involved in lysine biosynthesis in hyperthermophiles (**Supplementary Fig. 4a–e**). In contrast, the corresponding regions of their counterparts in arginine biosynthesis, or of homologs, have no such large, negatively charged surface (**Supplementary Fig. 3b–f**). Therefore, it is very likely that each enzyme in lysine biosynthesis specifically recognizes the globular domain of LysW through electrostatic interactions around these active sites. Another feature of LysW is its highly conserved five-residue C-terminal sequence, EDWGE. Among these five residues, we hypothesized that the amino group of AAA is attached to the C-terminal Glu54.

To test our hypothesis, we performed the lysine synthesis assay using a LysW mutant carrying an alanine in place of the conserved C-terminal Glu54 and found that no lysine was produced (**Fig. 3b**). Consistent with these results, a *T. thermophilus* mutant whose *lysW* gene had been replaced with the mutated gene encoding LysW-Glu54Ala (**Supplementary Fig. 5**) showed a lysine-auxotrophic phenotype (**Supplementary Fig. 6a**,**b**). These observations indicate that the Glu54 is very important for the function of LysW in lysine biosynthesis both *in vivo* and *in vitro*.

When we applied 10 mM of AAA to the reaction mixture with wild-type LysW, 1.5 ± 0.1 mM (all values are given as mean ± s.e.m.) of lysine was synthesized (**Fig. 3c**, lane 1) and 3.0 ± 0.2 mM of AAA was consumed (**Fig. 3c**, lane 2), indicating that two molecules of AAA were used to synthesize one molecule of lysine. This two-to-one stoichiometry can be explained by one molecule of AAA being used as a substrate to attach to the γ-carboxyl group of the C-terminal Glu54 of LysW, while the other is used by LysJ for the conversion of LysW-γ-AAA semialdehyde to LysW-γ-Lys, as will be described below. In most cases, aminotransferases catalyze the transfer of the amino group of glutamate to an amino acceptor. When we added AAA and glutamate simultaneously to the reaction mixture at 10 mM each for the enzymatic conversion of AAA to lysine, we observed no enhancement of lysine production (1.8 ± 0.1 mM; **Fig. 3c**, lane 3). However, we observed consumption largely equivalent to lysine production for both AAA (2.0 ± 0.05 mM) (**Fig. 3c**, lane 4) and glutamate (2.3 ± 0.03 mM) (**Fig. 3c**, lane 5). This result means that 1 mol of lysine was synthesized from 1 mol of AAA with consumption of 1 mol of glutamate, suggesting that LysJ used glutamate but not AAA for the aminotransferase reaction when both possible amino donors were present. This result indicates that glutamate is an intrinsic amino donor recognized more specifically than AAA by LysJ, although LysJ can also potentially use AAA as an amino donor.

Chemical structure determination of intermediates

Our results obtained using a Glu54Ala mutant of LysW suggested that the C-terminal glutamate group of LysW modifies the amino group of AAA in lysine biosynthesis. To confirm this hypothesis, we determined the chemical structures of the LysW derivatives in the reactions.

We incubated AAA and LysW in three different reaction mixtures containing either (i) LysX; (ii) LysX, LysZ, LysY and LysJ; or (iii) all five protein components, LysX, LysZ, LysY, LysJ and LysK, and then examined the change in the mobility of LysW on 12% Tricine–sodium dodecyl sulfate PAGE (Tricine-SDS-PAGE). When we subjected reaction mixture (ii) to gel electrophoresis, we observed a new, slowly migrating protein band with an apparent molecular mass of 6 kDa (**Fig. 4a**, lane 4), whereas on the gel for reaction mixture (i), no band showed any mobility shift (**Fig. 4a**, lane 3). When AAA and LysW were incubated with reaction mixture (iii)—that is, when LysK was added to mixture (ii)—the shifted band disappeared (**Fig. 4a**, lane 5). We extracted the bands corresponding to these LysW derivatives in (**Fig. 4a**, lanes 3 and 4) from the gel and analyzed their C-terminal sequences by MALDI-TOF-MS and MS/MS analyses after trypsin treatment. Because LysW contains two arginine residues at positions 13 and 41 (**Fig. 4b**), trypsin treatment generated three tryptic peptides, consisting of residues 1–13, 14–41 and 42–54. Of the two bands in reaction mixture (ii), the fast-migrating band possessed the normal C-terminal 13 amino acid residues of LysW, NH_2-LEPAPEEAEDWGE-COOH, and had a molecular mass of 1,471.9 Da (**Supplementary Fig. 7**). In contrast, the corresponding C-terminal tryptic peptide from the slowly migrating band had a molecular mass of 1,599.7 Da (**Fig. 4c**), which corresponded to the mass of the C-terminal 13-amino-acid peptide with one lysine residue attached. MS/MS analysis revealed that the lysine was indeed attached to the C-terminal Glu54 of LysW (**Fig. 4d**). We next performed a similar analysis for the LysW protein in reaction mixture (i) that showed no apparent mobility shift on Tricine-SDS-PAGE (**Fig. 4a**, lane 3). Trypsin treatment gave two C-terminal fragments with normal molecular mass of 1,471.9 Da (minor component) and an increased molecular mass of 1,614.7 Da

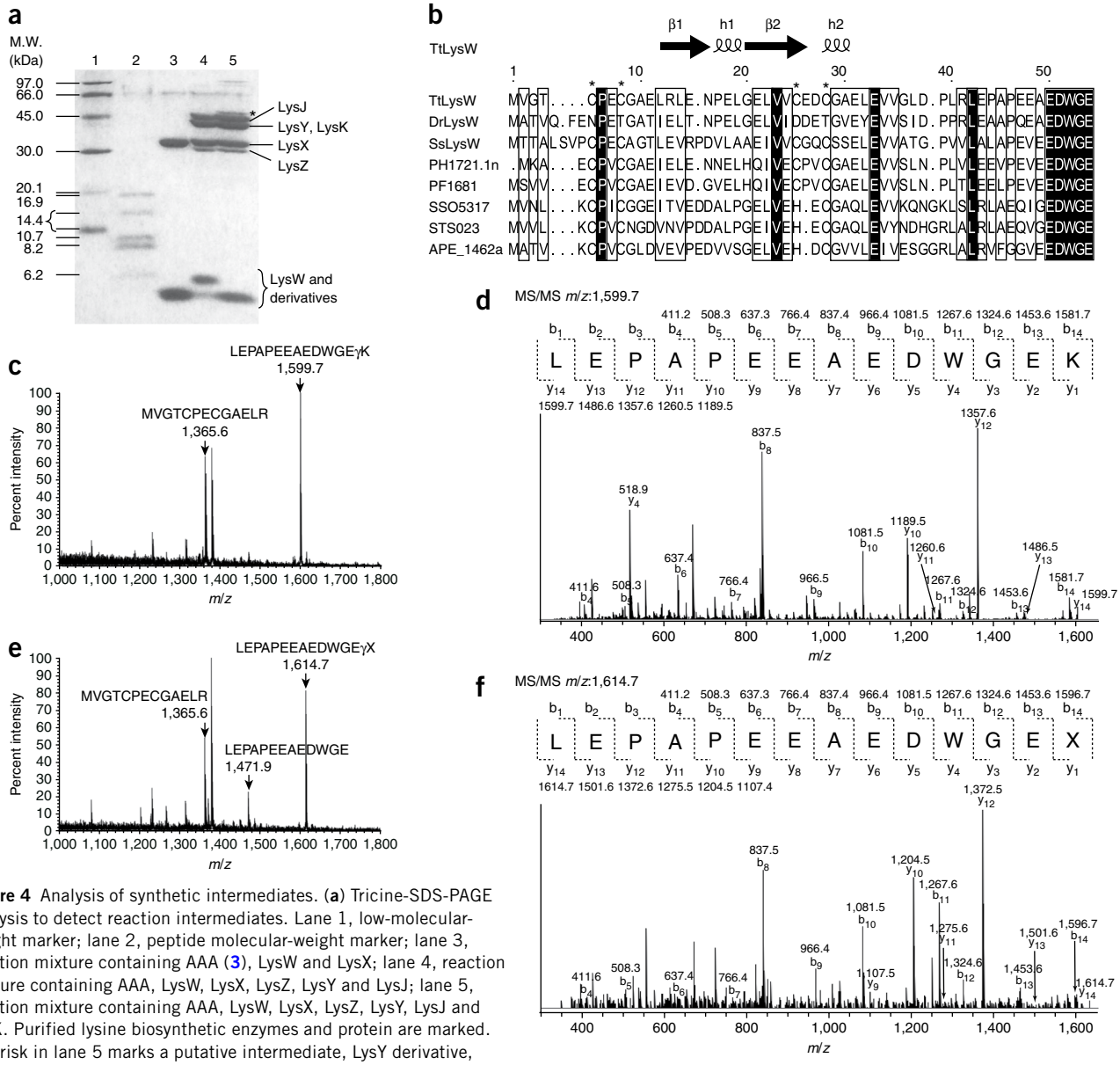

Figure 4 Analysis of synthetic intermediates. (**a**) Tricine-SDS-PAGE analysis to detect reaction intermediates. Lane 1, low-molecular-weight marker; lane 2, peptide molecular-weight marker; lane 3, reaction mixture containing AAA (3), LysW and LysX; lane 4, reaction mixture containing AAA, LysW, LysX, LysZ, LysY and LysJ; lane 5, reaction mixture containing AAA, LysW, LysX, LysZ, LysY, LysJ and LysK. Purified lysine biosynthetic enzymes and protein are marked. Asterisk in lane 5 marks a putative intermediate, LysY derivative, in which LysW-AAA semialdehyde is covalently attached to the active site cysteine. LysW is seen as the smallest clear bands in lanes 3, 4 and 5. A band shift can be seen immediately above the LysW band in lane 4. (**b**) Amino acid sequence alignment of LysW homologs. Four cysteine residues putatively responsible for metal bindings are shown by asterisks. TtLysW, DrLysW, SsLysW, PH1721.1n, PF1681, SSO5317, STS023 and APE_1462a indicate LysW from *T. thermophilus* and the LysW homologs of *D. radiodurans*, *S. sahachiroi*, *P. horikoshii*, *S. solfataricus*, *S. tokodaii* and *A. pernix*, respectively. (**c**) MALDI-TOF mass spectrum of the trypsin-treated LysW derivative after LysX, LysZ, LysY and LysJ reactions, which was extracted from the band in lane 4 of **b**. The N-terminal tryptic fragment, MVGTCPECGAER, is also shown. (**d**) MS/MS spectrum of the ion at *m/z* 1,599.7 in **c**. (**e**) MALDI-TOF mass spectrum of the trypsin-treated LysW derivative after LysX reactions, which was extracted from the band in lane 3 of **b**. AAA residue is indicated by 'X'. (**f**) MS/MS spectrum of the ion at *m/z* 1,614.7 in **e**. AAA residue is indicated by 'X'. For **d**,**f**, the amino acid sequence of the precursor ion is shown in the upper part of the figure. Theoretical molecular weights of the fragments are given above and below the sequence. Detected molecular masses of the fragments are indicated in the MS/MS spectrum.

(major component) (**Fig. 4e**). The molecular mass of the larger C-terminal tryptic peptide corresponded to the mass of the C-terminal 13-amino-acid peptide plus one AAA molecule. MS/MS analysis revealed that the AAA was attached to the C-terminal Glu54 of LysW (**Fig. 4f**).

The analysis revealed that AAA-to-lysine conversion proceeds with AAA and intermediates fused to the C-terminal Glu54 of LysW. We tried to determine the chemical structures of the LysW derivatives modified by LysX plus LysZ and by LysX, LysY and LysZ. However, we did not detect either of the two putative LysW derivatives, containing (i) the AAA moiety phosphorylated by LysZ and (ii) the AAA semialdehyde subsequently formed by LysY, in our analysis, probably because of the instability of these intermediates.

Purified LysW is red in color and produces four peaks at 345, 373, 496 and 570 nm in the absorption spectrum (**Supplementary Fig. 8a**), indicating that Fe ion is bound[18] to LysW, possibly at the four conserved cysteine residues (Cys5, Cys8, Cys25 and Cys28; **Fig. 4b**). The circular dichroism spectrum indicated that LysW has a specific structure even at a high temperature of 95 °C (**Supplementary Fig. 8b**).

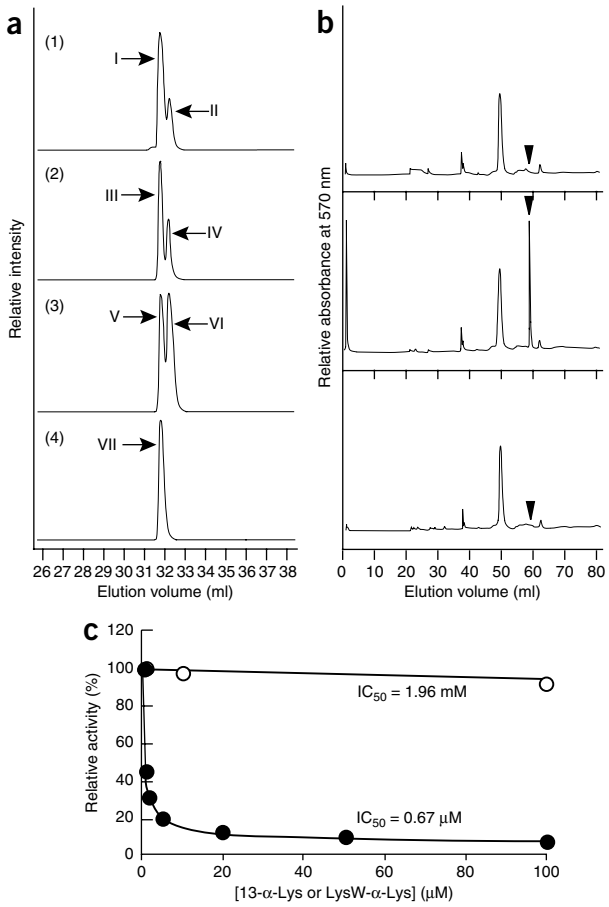

Figure 5 γ-linkage-specific reactions of lysine biosynthetic enzymes. (**a**) HPLC chromatograms of trypsin-treated fragments of reaction intermediates and synthetic oligopeptides. (1) Co-injection of synthetic peptides, 13-γ-Lys (I) and 13-α-Lys (II); (2) co-injection of tryptic C-terminal fragment of LysW generated by a series of reactions by LysX, LysZ, LysY and LysJ (III) with 13-α-Lys. (IV); (3) co-injection of tryptic C-terminal fragment of LysW derivative (V), the same as III in (2), with tryptic C-terminal fragment of recombinant LysW-α-Lys (VI); (4) co-injection of tryptic C-terminal fragment of LysW derivative, the same as III in (2), with 13-γ-Lys (VII). (**b**) Chromatograms of amino acid analysis for LysK reactions. Substrates: top, 13-α-Lys; middle, 13-γ-Lys; bottom, recombinant LysW-α-Lys. Arrows indicate positions of lysine. (**c**) Effect of LysW-α-Lys and 13-α-Lys addition on LysK activity for 13-γ-Lys. Closed and open circles show LysK activity for 13-γ-Lys in the presence of LysW-α-Lys and 13-α-Lys, respectively.

AAA is attached to γ-carboxyl group of Glu54 of LysW

Glutamate possesses two carboxyl groups, at the α and γ positions, and in a C-terminal glutamate residue both are available for reaction. To determine which carboxyl group of the C-terminal Glu54 of LysW accepts the amino group of AAA in the lysine biosynthetic system, we prepared two synthetic oligopeptides: one possessing the C-terminal 13-amino-acid peptide fused with one lysine molecule by α-linkage at C terminus (13-α-Lys) and the other with an identical peptide sequence but a γ-linked lysine residue (13-γ-Lys). We also prepared C-terminal tryptic fragments from the LysW derivative that showed slow mobility on the gel (**Fig. 4a**, lane 4) and from LysW-α-Lys, a genetically generated LysW derivative containing one additional α-linked lysine residue at the C terminus. When we subjected 13-α-Lys and 13-γ-Lys to HPLC, 13-γ-Lys eluted from the column faster than 13-α-Lys (**Fig. 5a**). In contrast, the C-terminal tryptic fragment of the lysine-attached LysW derivative generated in reaction mixture (ii) (discussed in the previous section) eluted separately from 13-α-Lys or the C-terminal tryptic fragment of recombinant LysW-α-Lys but at a volume identical to that for 13-γ-Lys (**Fig. 5a**). These results indicate that AAA is attached to the γ-carboxyl group of Glu54 of LysW. We hereafter referred to the intermediates in reaction mixture (ii) as LysW-γ-Lys.

To further confirm the γ-bond linkage specificity of the enzymes involved in lysine biosynthesis, we performed an enzyme assay using LysK. First we assessed whether the lysine-releasing activity of LysK is specific to γ-linkage. We used three compounds as substrates: (i) LysW-α-Lys, (ii) 13-α-Lys and (iii) 13-γ-Lys. Incubation of these possible substrates with LysK resulted in the release of lysine from 13-γ-lysine (**Fig. 5b**, middle) but not from 13-α-lysine (**Fig. 5b**, top) or LysW-α-Lys (**Fig. 5b**, bottom). This result indicates that the bond at the C terminus of the intermediate has γ-peptide linkage and that LysK shows strict γ-linkage specificity. This result is consistent with the role of LysK, which is to liberate lysine from LysW-γ-Lys in the last step of lysine biosynthesis.

We next examined the effect of 13-α-Lys or LysW-α-Lys on the release of lysine from 13-γ-Lys by LysK. Only slight inhibition (IC_{50} of 1.96 mM) resulted from the addition of 13-α-Lys to the reaction mixture at a concentration of 100 μM. However, LysW-α-Lys strongly inhibited the reaction, with an IC_{50} of 0.67 μM (**Fig. 5c**). This result indicates that LysK recognizes not only the scissile γ-linked C-N bond but also regions other than the C-terminal amino acid residues of LysW.

Characterization of LysX

We examined the activity of LysX for various substrates in the reaction. In our proposed catalytic mechanism, LysX activates the γ-carboxyl group of Glu54 of LysW by phosphorylation, with concomitant conversion of ATP to ADP, and then the amino group of AAA attacks the phosphorylated carboxyl group of LysW to form LysW-γ-AAA, with concomitant release of inorganic phosphate and ADP. To detect the activity of LysX, we monitored the released inorganic phosphate in the reaction mixture. Incubating 1.5 mM each of wild-type LysW, AAA and ATP with LysX at 60 °C for 1 h resulted in the release of 1.5 mM phosphate. In contrast, when we used 1.5 mM of either LysW (E54A) or APE_1462a—a LysW homolog in *Aeropyrum pernix* with 35% identity in the amino acid sequence with a conserved EDWGE sequence at the C terminus—for the reaction instead of wild-type LysW, we observed only negligible phosphate release (**Supplementary Fig. 9a**). This result further indicates that LysX recognizes not only the C-terminal glutamate residue of LysW but also other portions—possibly the globular domain—of LysW specifically. We also determined that LysX discriminates AAA from glutamate in the reaction (**Supplementary Fig. 9b**).

We next determined the kinetic parameters of LysX for LysW, AAA and ATP (**Supplementary Table 1**). It might be considered reasonable that LysX would have similar apparent K_m values (K_m^{app}) for LysW, AAA and ATP; however, this turned out not to be the case. The K_m^{app} value for LysW (0.078 mM) is markedly lower than those for the other substrates, which were 2.1 mM for AAA and 2.6 mM for ATP. This low K_m^{app} value may ensure efficient recognition of LysW by LysX because the cellular concentration of LysW would be expected to be lower than those of small compounds such as glutamate, AAA and ATP.

DISCUSSION

Microorganisms have developed various systems for masking the amino group of intermediates in different metabolic pathways. In arginine biosynthesis, cyclization into an unproductive five-membered ring intermediate, Δ^1-pyrroline-2-carboxylate, is

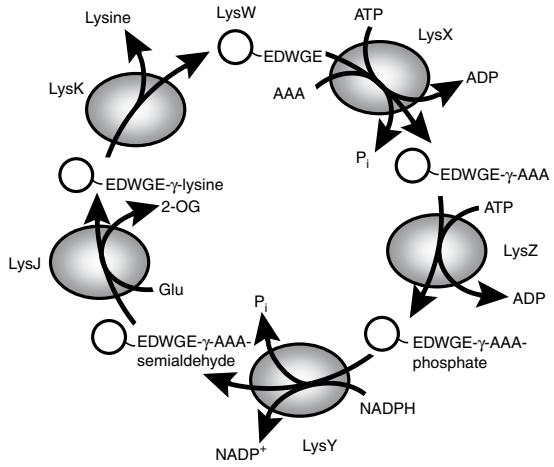

Figure 6 Proposed LysW-mediated mechanism of lysine biosynthesis. The conversion of AAA (**3**) to lysine (**1**) in the *T. thermophilus* lysine biosynthetic pathway is illustrated schematically. Enzymes LysX, LysZ, LysY, LysJ and LysK are shown as spheres, and LysW is shown in a white circle with five C-terminal conserved residues. Cofactors and products of reactions are also shown.

prevented by acetylation of the amino group of glutamate by ArgA or ArgJ[13–15]. Masking of the amino group to prevent intramolecular cyclization is also seen in the catabolic pathway of putrescine (**19**)[19]. Lysine biosynthesis in *T. thermophilus* includes an unstable intermediate, AAA semialdehyde (**20**), which readily forms a six-membered ring compound, Pip2C, by intramolecular cyclization and must prevent efficient lysine biosynthesis. Here we determined that *T. thermophilus* follows the approach of proteinaceous modification but not the system of using a small chemical compound, such as the acetyl group, for N-modification. Therefore, it seems plausible that the LysW modification system must provide some advantage compared to the small-compound protection systems in *T. thermophilus*. All other microorganisms that synthesize lysine through a pathway similar to that in *T. thermophilus* are hyperthermophilic archaea and contain LysW homologs with four conserved cysteine residues. The only exceptions are *Deinococcus* spp., which are related to *Thermus* and with it form the *Deinococcus-Thermus* phylum. *D. radiodurans* and *D. geothermalis*, whose genome sequences have been determined, are a mesophile and a moderate thermophile, respectively. In the LysW homologs of *Deinococcus*, four conserved cysteine residues at positions 5, 8, 25 and 28 are replaced by asparagine, threonine, aspartic acid and threonine, respectively; therefore, it is unlikely that deinococcal LysW homologs bind Fe ions. We assume that the extreme thermostability of LysW confers an advantage by providing efficient protection of the amino group of AAA at elevated temperature. We also think that LysW might prevent the intramolecular cyclization of biosynthetic intermediates more efficiently than small protecting compounds.

The electrostatic potentials of LysW and lysine biosynthetic enzymes, calculated using the crystal structures of the enzymes or simulated models, are indicative of electrostatic docking between LysW and these enzymes. The structural feature suggests that LysW is designed to be transferred successively among a set of enzymes, functioning as a 'carrier protein' or 'protein scaffold', to ensure the efficient synthesis of lysine (**Fig. 6**). Electrostatic interactions between the enzymes and LysW are thought to be desirable for an efficient reaction at elevated temperatures because these interactions function well under conditions in which thermophiles grow. To date, acyl carrier protein (ACP)[20] and peptidyl carrier protein (PCP)[21] are known as proteinaceous carriers of intermediates in fatty acid biosynthesis and peptide biosynthesis, respectively. Both are classified in the same family and function through a common mechanism whereby specific serine residues are modified with the phosphopantetheinyl group, and the acyl chain or peptidyl chain is attached and elongated in phosphopantetheinyl-conjugated form. In fatty acid synthetase (FAS), the elongating acyl group is attached to the SH group of the phosphopantetheinyl group of holo-ACP that is formed by phosphopantetheinyl transferase from apo-ACP and the phosphopantetheinyl group of acetyl-CoA. Holo-ACP plays a central role in the entire synthetic process by shuttling between the catalytic centers of FAS complex[22]. Notably, the crystal structure of FAS, which was recently determined, revealed specific binding through electrostatic interactions between ACP and these enzymes[23,24].

Most recently, biosynthesis of azinomycin B (**21**), a secondary metabolite with potent antitumor activity produced by *Streptomyces sahachiroi*, was analyzed[25]. Azinomycin B contains unusual building blocks derived from amino acids. Based on the sequence information for the biosynthetic gene cluster, it was proposed that the building blocks derived from amino acids are synthesized similarly to those of the AAA-to-lysine conversion in lysine biosynthesis shown here, because the cluster contains genes homologous to *lysX*, *lysY*, *lysZ* and *lysK*. We found a putative open reading frame, in the region between *aziC7* and *aziC4*, encoding a protein of 59 amino acids that showed 46% amino acid sequence identity to LysW, including the four conserved cysteine residues and the C-terminal EDWGE sequence (**Fig. 4b**), although the open reading frame is not annotated as a gene in the report. This observation suggests that a similar LysX-LysW system is used for secondary as well as primary metabolism. This study not only sets the stage for future structural analysis to uncover more details about protein-protein interaction but also elucidates one possible role of small proteins that are annotated as proteins with unknown cellular functions.

METHODS

Plasmids. Construction of expression vectors is described in **Supplementary Methods** using the oligonucleotides listed in **Supplementary Table 2**.

Protein purification. Procedure for protein purification is described in **Supplementary Methods**.

***In vitro* lysine production assay.** For cell-free synthesis of lysine from AAA, a reaction mixture containing 10 μM $ZnSO_4$, 10 μM $MgCl_2$, 10 μM $CoCl_2$, 5 mM pyridoxal phosphate (PLP), 10 mM NADPH, 100 mM Tris-HCl (pH 8.0) and 10 mM AAA was prepared. After the addition of proteins (50 μg LysW, 50 μg LysX, 13 μg LysZ, 37 μg LysY, 33 μg LysJ and 17 μg LysK) and 20 mM ATP, the reaction mixture was incubated at 60 °C for 1 h. Similar reactions containing an equal amount of LysW mutant E54A in place of the wild-type protein, or lacking a single enzyme from the basic contents, were also performed. The reaction mixture was mixed with an equal volume of 6% trichloroacetic acid (TCA) and centrifuged at 20,630*g* for 10 min. The amino acid content of the supernatant was measured with an L-8500A high-speed amino acid analyzer (Hitachi). Three measurements were performed for each sample and results are shown with standard errors (s.e.m.).

Amino acid sequence alignment. Alignment was done with ClustalW[26] and alignment with secondary structures of the modeled LysW structure was drawn based on ESPript[27].

***In vitro* synthesis and determination of chemical structures of the intermediates.** To synthesize the final intermediate, LysW-γ-Lys, in the biosynthetic pathway, LysK and the putative cofactor, $CoCl_2$, were removed from the

reaction mixture shown above. For synthesis of the first intermediate, LysW-γ-AAA, from AAA, LysK, LysJ, LysY and LysZ and the putative cofactors for corresponding enzymes ($CoCl_2$, PLP and NADPH) were removed from the reaction mixture. After the reaction, the reaction mixtures were directly subjected to 12% Tricine-SDS-PAGE to separate LysW- and LysW-fused intermediates. The chemical structures of the intermediates were determined by in-gel digestion followed by LC-MS/MS analysis and MALDI-TOF analysis, as described in detail in the **Supplementary Methods**.

Assay of LysK activity for synthetic oligopeptides. For the LysK activity assay, two synthetic oligopeptides and recombinant LysW-α-Lys were prepared as described in **Supplementary Methods**. Each oligopeptide contains the C-terminal 13 residues of LysW fused with one lysine residue at its C-terminal glutamine residue. One oligopeptide had the C-terminal lysine in α-peptide bond linkage whereas the other had the lysine residue in γ-peptide bond linkage. The synthetic oligopeptides or recombinant LysW-α-Lys of 1 mM were previously mixed in a 100-μl reaction mixture containing 0.2 mM $CoSO_4$ and 100 mM Tris-HCl (pH 8.0). The reaction was initiated by the addition of 0.2 μg LysK to the mixtures. The reaction was carried out at 60 °C for 1 h. After the proteins were removed by precipitation through the TCA treatment described above, the amount of released lysine in the mixture was measured with a L-8500A amino acid analyzer (Hitachi).

To examine the inhibitory effect of LysW-α-Lys or 13-α-Lys on the LysK reaction using 13-γ-Lys as the substrate, LysW-α-Lys or 13-α-Lys was added to the reaction mixture at concentrations ranging from 1 to 100 μM LysW-α-Lys and from 1 to 10,000 μM 13-α-Lys, respectively. IC_{50} values for LysK activity were calculated with SigmaPlot 10 (Systat Software).

Note: Supplementary information and chemical compound information is available on the Nature Chemical Biology website.

ACKNOWLEDGMENTS

We thank H. Satsu for technical help with amino acid analysis. This work was supported in part by a grant-in-aid for scientific research from the Ministry of Education, Culture, Sports, Science, and Technology of Japan, from the Nagase Science and Technology Foundation, from the Asahi Glass Foundation and from the Charitable Trust Araki Medical and Biochemistry Memorial Research Promotion Fund.

AUTHOR CONTRIBUTIONS

Research planning and supervision were by T.T., T.K. and M.N.; biochemical experiments were by A.H.; gene knockout and replacement of *T. thermophilus* were by A.S.; LC-MS/MS and MALDI-TOF MS were by H.T., R.M., T.F. and C.N.; *in silico* modeling was by H.K. and T.T.; and manuscript writing was by A.H., T.T. and M.N.

Published online at http://www.nature.com/naturechemicalbiology/.
Reprints and permissions information is available online at http://npg.nature.com/reprintsandpermissions/.

1. Broquist, H.P. Lysine biosynthesis (yeast). *Methods Enzymol.* **17**, 112–129 (1971).
2. Buchli, R. *et al.* Cloning and functional expression of a soluble form of kynurenine/α-aminoadipate aminotransferase from rat kidney. *J. Biol. Chem.* **270**, 29330–29335 (1995).
3. Vogel, H.J. Distribution of lysine pathway among fungi: evolutionary implications. *Am. Nat.* **98**, 446–455 (1964).
4. Kobashi, N., Nishiyama, M. & Tanokura, M. Aspartate kinase-independent lysine synthesis in an extremely thermophilic bacterium, *Thermus thermophilus*: lysine is synthesized via α-aminoadipic acid, not via diaminopimeric acid. *J. Bacteriol.* **181**, 1713–1718 (1999).
5. Nishida, H. *et al.* A prokaryotic gene cluster involved in synthesis of lysine through the amino adipate pathway: a key to the evolution of amino acid biosynthesis. *Genome Res.* **9**, 1175–1183 (1999).
6. Wulandari, A.P. *et al.* Characterization of bacterial homocitrate synthase involved in lysine biosynthesis. *FEBS Lett.* **522**, 35–40 (2002).
7. Miyazaki, J., Kobashi, N., Nishiyama, M. & Yamane, H. Characterization of homoisocitrate dehydrogenase involved in lysine biosynthesis of an extremely thermophilic bacterium, *Thermus thermophilus* HB27, and evolutionary implication of β-decarboxylating dehydrogenase. *J. Biol. Chem.* **278**, 1864–1871 (2003).
8. Jia, Y., Tomita, T., Yamauchi, K., Nishiyama, M. & Palmer, D.R.J. Kinetics and product analysis of the reaction catalyzed by recombinant homoaconitase from *Thermus thermophilus*. *Biochem. J.* **396**, 479–485 (2006).
9. Miyazaki, T., Miyazaki, J., Yamane, H. & Nishiyama, M. α-Aminoadipate aminotransferase from an extremely thermophilic bacterium, *Thermus thermophilus*. *Microbiology* **150**, 2327–2334 (2004).
10. Miyazaki, J., Kobashi, N., Nishiyama, M. & Yamane, H. Functional and evolutionary relationship between arginine biosynthesis and prokaryotic lysine biosynthesis through α-aminoadipate. *J. Bacteriol.* **183**, 5067–5073 (2001).
11. Miyazaki, J., Kobashi, N., Fujii, T., Nishiyama, M. & Yamane, H. Characterization of a *lysK* gene as an *argE* homolog in *Thermus thermophilus* HB27. *FEBS Lett.* **512**, 269–274 (2002).
12. Xu, Y., Labedan, B. & Glansdorff, N. Surprising arginine biosynthesis: a reappraisal of the enzymology and evolution of the pathway in microorganisms. *Microbiol. Mol. Biol. Rev.* **71**, 36–47 (2007).
13. Leisinger, T. & Haas, D. *N*-Acetylglutamate synthase of *Escherichia coli* regulation of synthesis and activity by arginine. *J. Biol. Chem.* **250**, 1690–1693 (1975).
14. Baetens, M., Legrain, C., Boyen, A. & Glansdorff, N. Genes and enzymes of the acetyl cycle of arginine biosynthesis in the extreme thermophilic bacterium *Thermus thermophilus* HB27. *Microbiology* **144**, 479–492 (1998).
15. Martin, P.R. & Mulks, M.H. Sequence analysis and complementation studies of the *argJ* gene encoding ornithine acetyltransferase from *Neisseria gonorrhoeae*. *J. Bacteriol.* **174**, 2694–2701 (1992).
16. Snoke, J.E. Isolation and properties of yeast glutathione synthetase. *J. Biol. Chem.* **213**, 813–824 (1955).
17. Kang, W.K., Icho, T., Isono, S., Kitakawa, M. & Isono, K. Characterization of the gene *rimK* responsible for the addition of glutamic acid residues to the C-terminus of ribosomal protein S6 in *Escherichia coli* K12. *Mol. Gen. Genet.* **217**, 281–288 (1989).
18. Iwasaki, T. *et al.* Rational design of a mononuclear metal site into the archaeal Rieske-type protein scaffold. *J. Biol. Chem.* **280**, 9129–9134 (2005).
19. Kurihara, S. *et al.* A novel putrescine utilization pathway involves γ-glutamylated intermediates of *Escherichia coli* K-12. *J. Biol. Chem.* **280**, 4602–4608 (2005).
20. Majerus, P.W., Alberts, A.W. & Vagelos, P.R. The acyl carrier protein of fatty acid synthesis: purification, physical properties, and substrate binding site. *Proc. Natl. Acad. Sci. USA* **51**, 1231–1238 (1964).
21. Stachelhaus, T., Huser, A. & Marahiel, M.A. Biochemical characterization of peptidyl carrier protein (PCP), the thiolation domain of multifunctional peptide synthetases. *Chem. Biol.* **3**, 913–921 (1996).
22. Lynen, F. On the structure of fatty acid synthetase of yeast. *Eur. J. Biochem.* **112**, 431–442 (1980).
23. Lomakin, I.B., Xiong, Y. & Steitz, T.A. The crystal structure of yeast fatty acid synthase, a cellular machine with eight active sites working together. *Cell* **129**, 319–332 (2007).
24. Jenni, S. *et al.* Structure of fungal fatty acid synthase and implications for iterative substrate shuttling. *Science* **316**, 254–261 (2007).
25. Zhao, Q. *et al.* Characterization of the azinomycin B biosynthetic gene cluster revealing a different iterative type I polyketide synthase for naphthoate biosynthesis. *Chem. Biol.* **15**, 693–705 (2008).
26. Thompson, J.D., Higgins, D.G. & Gibson, T.J. CLUSTALW: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res.* **22**, 4673–4680 (1994).
27. Gouet, P., Courcelle, E., Stuart, D.I. & Metoz, F. ESPript: analysis of multiple sequence alignments in PostScript. *Bioinformatics* **15**, 305–308 (1999).
28. Henne, A. *et al.* The genome sequence of the extreme thermophile *Thermus thermophilus*. *Nat. Biotechnol.* **22**, 547–553 (2004).
29. White, O. *et al.* Genome sequence of the radioresistant bacterium *Deinococcus radiodurans* R1. *Science* **286**, 1571–1577 (1999).
30. Kawarabayasi, Y. *et al.* Complete genome sequence of an aerobic thermoacidophilic crenarchaeon, *Sulfolobus tokodaii* strain7. *DNA Res.* **8**, 123–140 (2001).
31. She, Q. *et al.* The complete genome of the crenarchaeon *Sulfolobus solfataricus* P2. *Proc. Natl. Acad. Sci. USA* **98**, 7835–7840 (2001).
32. Kawarabayasi, Y. *et al.* Complete sequence and gene organization of the genome of a hyper-thermophilic archaebacterium, *Pyrococcus horikoshii* OT3. *DNA Res.* **5**, 55–76 (1998).
33. Robb, F.T. *et al.* Genomic sequence of hyperthermophile, *Pyrococcus furiosus*: implications for physiology and enzymology. *Methods Enzymol.* **330**, 134–157 (2001).
34. Kawarabayasi, Y. *et al.* Complete genome sequence of an aerobic hyper-thermophilic crenarchaeon, *Aeropyrum pernix* K1. *DNA Res.* **6**, 83–101, 145–152 (1999).
35. Bolhuis, H. *et al.* The genome of the square archaeon *Haloquadratum walsbyi*: life at the limits of water activity. *BMC Genomics* **7**, 169 (2006).
36. Baliga, N.S. *et al.* Genome sequence of *Haloarcula marismortui*: a halophilic archaeon from the Dead Sea. *Genome Res.* **14**, 2221–2234 (2004).