[9] H. Murakami, A. Ohta, H. Ashigai, H. Suga, *Nat. Methods* **2006**, *3*, 357.

[10] Y. Goto, T. Katoh, H. Suga, *Nat. Protoc.* **2011**; in press.

[11] a) Y. Goto, A. Ohta, Y. Sako, Y. Yamagishi, H. Murakami, H. Suga, *ACS Chem. Biol.* **2008**, *3*, 120; b) Y. Goto, H. Murakami, H. Suga, *RNA* **2008**, *14*, 1390; c) Y. Goto, H. Suga, *J. Am. Chem. Soc.* **2009**, *131*, 5040.

[12] a) T. Kawakami, S. Aimoto, *Chem. Lett.* **2007**, *36*, 76; b) T. Kawakami, S. Aimoto, *Tetrahedron Lett.* **2007**, *48*, 1903.

[13] T. Kawakami, A. Ohta, M. Ohuchi, H. Ashigai, H. Murakami, H. Suga, *Nat. Chem. Biol.* **2009**, *5*, 888.

[14] a) S. Luckett, R. S. Garcia, J. J. Barker, A. V. Konarev, P. R. Shewry, A. R. Clarke, R. L. Brady, *J. Mol. Biol.* **1999**, *290*, 525; b) M. L. Korsinczky, H. J. Schirra, D. J. Craik, *Curr. Protein Pept. Sci.* **2004**, *5*, 351.

– 492 –

Small-molecule inhibitor binding to an N-acyl-homoserine lactone synthase

Jiwoung Chung[a,1], Eunhye Goo[a,1], Sangheon Yu[a,1], Okhee Choi[a], Jeehyun Lee[a], Jinwoo Kim[b], Hongsup Kim[a], Jun Igarashi[c], Hiroaki Suga[c,d,e], Jae Sun Moon[f], Ingyu Hwang[a,2], and Sangkee Rhee[a,g,2]

[a]Department of Agricultural Biotechnology, Seoul National University, Seoul 151-921, Korea; [b]Department of Applied Biology, Gyeongsang National University, Jinju 660-701, Korea; [c]Japan Research Center for Advanced Science and Technology, University of Tokyo, Tokyo 153-8904, Japan; [d]Department of Chemistry, School of Science, University of Tokyo, Tokyo 113-0033, Japan; [e]Department of Molecular Medicine and Biopharmaceutical Sciences, Seoul National University, Seoul 151-742, Korea; [f]Plant Systems Engineering Center, Korea Research Institute of Bioscience and Biotechnology, Daejeon 305-333, Korea; and [g]Center for Fungal Pathogenesis, Seoul National University, Seoul 151-921, Korea

Quorum sensing (QS) controls certain behaviors of bacteria in response to population density. In Gram-negative bacteria, QS is often mediated by N-acyl-L-homoserine lactones (acyl-HSLs). Because QS influences the virulence of many pathogenic bacteria, synthetic inhibitors of acyl-HSL synthases might be useful therapeutically for controlling pathogens. However, rational design of a potent QS antagonist has been thwarted by the lack of information concerning the binding interactions between acyl-HSL synthases and their ligands. In the Gram-negative bacterium *Burkholderia glumae*, QS controls virulence, motility, and protein secretion and is mediated by the binding of N-octanoyl-L-HSL (C8-HSL) to its cognate receptor, TofR. C8-HSL is synthesized by the acyl-HSL synthase TofI. In this study, we characterized two previously unknown QS inhibitors identified in a focused library of acyl-HSL analogs. Our functional and X-ray crystal structure analyses show that the first inhibitor, J8-C8, binds to TofI, occupying the binding site for the acyl chain of the TofI cognate substrate, acylated acyl-carrier protein. Moreover, the reaction byproduct, 5′-methylthioadenosine, independently binds to the binding site for a second substrate, S-adenosyl-L-methionine. Closer inspection of the mode of J8-C8 binding to TofI provides a likely molecular basis for the various substrate specificities of acyl-HSL synthases. The second inhibitor, E9C-3oxoC6, competitively inhibits C8-HSL binding to TofR. Our analysis of the binding of an inhibitor and a reaction byproduct to an acyl-HSL synthase may facilitate the design of a new class of QS-inhibiting therapeutic agents.

Q uorum sensing (QS) is an intercellular signaling process that mediates certain behaviors of bacteria (including bioluminescence, biofilm formation, motility, and virulence factor production) in response to the bacterial cell population density (1–3). In Gram-negative bacteria, QS is often mediated by N-acyl-L-homoserine lactones (acyl-HSLs), which are synthesized by the LuxI family of acyl-HSL synthases from S-adenosyl-L-methionine (SAM) and acylated acyl-carrier protein (acyl-ACP), with the release of holo-ACP and 5′-methylthioadenosine (MTA) as byproducts (*SI Appendix*, Fig. S1A) (4, 5). Compounds of the acyl-HSL class share a homoserine lactone ring moiety, but the acyl chains conjugated to the ring via an amide bond vary in length, oxidation state at C3, and amount of saturation (*SI Appendix*, Fig. S1A). The recent finding that p-coumarate is an alternative substrate for acyl-ACP has extended the known range of possible acyl-HSL substrates (6). On the other hand, the acyl-HSL receptor is a transcriptional regulator that controls the expression of target genes in response to acyl-HSL binding (1–3).

Among the hundreds of genes regulated by QS, the most widely studied genes are those related to virulence; these genes are of particular interest because QS disruption is being investigated as a strategy for controlling virulent pathogens (7–9). QS inhibitors can act by suppressing acyl-HSL production, blocking QS receptors, or inactivating signal molecules (8–10). However, reports on the development of QS inhibitors have thus far focused only on acyl-HSL receptors (11, 12) or on enzymes involved in the SAM biosynthesis pathway (13).

We previously reported that the WT *Burkholderia glumae* strain BGR1 produces toxoflavin, a phytotoxin that acts as a key virulence factor in bacterial rice grain rot (14). In this bacterium, motility, protein secretion, and toxoflavin biosynthesis and transport are controlled by the QS signaling molecule N-octanoyl-L-HSL (C8-HSL). This molecule is synthesized by the acyl-HSL synthase TofI and mediates QS signaling by binding to its cognate receptor, TofR (14–16).

In the present study, we identified a TofI inhibitor, J8-C8, and a competitive inhibitor of C8-HSL binding to TofR, E9C-3oxoC6. We examined the TofI structure and the binding of J8-C8 and of MTA to TofI by X-ray crystal structure analysis of apo-TofI and a TofI/J8-C8/MTA ternary complex. These results provide an unprecedented structural view of the binding of an acyl-HSL synthase inhibitor and a reaction byproduct to an acyl-HSL synthase. Analysis of the complex suggests the molecular basis for the substrate specificity of acyl-HSL synthases toward acyl-ACPs and SAM.

Results

Identification of Two Distinct QS Inhibitors. To identify compounds that disrupt QS in *B. glumae*, we screened 55 compounds derived from previous "hits" in other Gram-negative bacterial QS systems (17). The compounds were screened by measuring the QS-mediated production of toxoflavin and C8-HSL by *B. glumae* BGR1 cells (*SI Appendix*, Fig. S2A). Two strong hits were identified—E9C-3oxoC6 and J8-C8—together with some minor hits (Fig. 1A and *SI Appendix*, Fig. S2). Both compounds diminished the production of C8-HSL and inhibited QS-dependent *ahpF* expression in BGR1 cells, but the effect of E9C-3oxoC6 was stronger (Fig. 1 B and C). In the mutant strain *B. glumae* BGS2 (*tofI::Ω*), E9C-3oxoC6 reduced toxoflavin production in the presence of 1 μM C8-HSL, but J8-C8 had no effect (Fig. 1D), suggesting that whereas QS inhibition by J8-C8 is mediated by TofI, E9C-3oxoC6 acts through a different mechanism. The IC_{50} values of J8-C8 and E9C-3oxoC6 for toxoflavin production were ~35 μM and 12 μM, respectively (*SI Appendix*, Fig. S3 A and B).
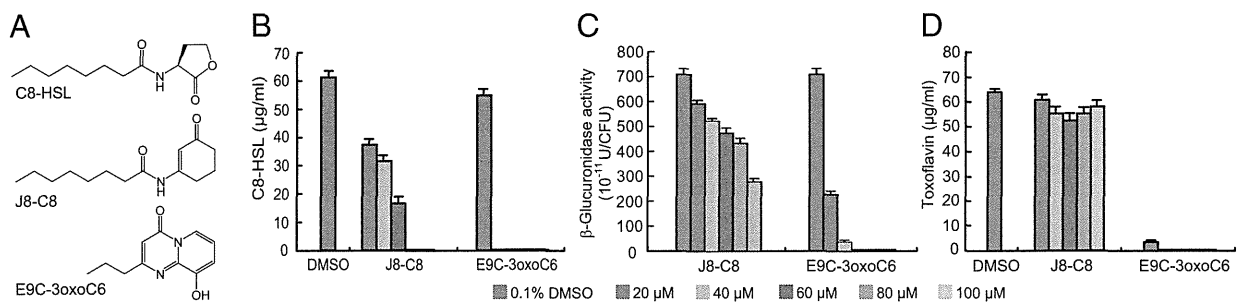
Fig. 1. Identification of small molecules that interfere with QS in *B. glumae*. (*A*) Chemical structures of C8-HSL, J8-C8, and E9C-3oxoC6. (*B*) Inhibition of C8-HSL production in *B. glumae* BGR1 cells. (*C*) Inhibition of *ahpF* expression by various concentrations of J8-C8 and E9C-3oxoC6 in BGR1 cells. (*D*) Inhibition of toxoflavin production by various concentrations of E9C-3oxoC6 in *B. glumae* BGS2 (*tofI*::Ω) cells in the presence of 1 μM C8-HSL. Error bars represent SDs from triplicate experiments.

J8-C8 Is a TofI Inhibitor. To assess our hypothesis that J8-C8 targets TofI, we measured the in vitro synthesis of C8-HSL by TofI in the presence of J8-C8 or E9C-3oxoC6. Whereas J8-C8 inhibited C8-HSL synthesis in a dose-dependent manner, E9C-3oxoC6 had no inhibitory effect on C8-HSL synthesis (Fig. 2*A*). When we then tested the effect of MTA on TofI activity, we found that 200 μM MTA inhibited TofI by ~30% and that J8-C8 and MTA had a synergistic effect, inhibiting TofI by ~70% when used in combination (Fig. 2*B*).

E9C-3oxoC6 Is a Competitive Inhibitor of C8-HSL Binding to TofR. To determine whether the target of E9C-3oxoC6 might be TofR, we conducted a series of competition experiments in which we measured the effect of E9C-3oxoC6 on toxoflavin production by BGS2 cells, which lack TofI. In the absence of E9C-3oxoC6, increasing the concentration of C8-HSL increased toxoflavin production in a C8-HSL concentration-dependent manner, with almost complete recovery of the WT level at 0.4 μM C8-HSL (Fig. 2*C*). Importantly, the addition of E9C-3oxoC6 at 12 μM (the IC$_{50}$ value) delayed the recovery of toxoflavin production, but this delay was fully reversed by 0.8 μM C8-HSL (Fig. 2*C*). These results suggest that E9C-3oxoC6 competitively inhibits C8-HSL binding to TofR.

Structure of apo-TofI(3MΔ). We determined the X-ray crystal structure of TofI in its apo form using a soluble, catalytically competent TofI mutant, TofI(3MΔ). This mutant harbors three single amino acid substitutions (F42M, I149M, and L152M) and two single amino acid deletions (ΔHis91 and ΔPro92) (*SI Appendix*, Fig. S1*B* and Table S1). The crystal structure of apo-

TofI(3MΔ) consists of eight β-strands and five α-helices folded into an α-β-α domain (Fig. 3*A* and *SI Appendix*, Fig. S4). Three highly disordered regions could not be modeled in the final structure of apo-TofI(3MΔ) (*SI Appendix*, Fig. S4). Structurally, these overall features in TofI are similar to those of the apo forms of two other acyl-HSL synthases, LasI (18) and EsaI (19), with an rmsd of 1.4 Å (164 Cα atoms) and 2.2 Å (139 Cα atoms), respectively (*SI Appendix*, Fig. S5). Notable differences in the *N*-terminal region were seen among the three structures.

As noted previously (18, 19), an apparent pocket exists on the concave side of the β-sheet; this pocket is located primarily between β4 and β5; is enclosed by three helices, α1, α3, and α4; and is oriented parallel to the β-sheet (Fig. 3*A*). In a structure of apo-TofI(3MΔ), five water molecules are clustered at the entrance to the pocket via direct or water-mediated hydrogen bonding to the side chain of Arg104 and the main-chain carbonyl oxygen of Phe105 (*SI Appendix*, Fig. S6*A*), leaving the inside of the pocket vacant.

J8-C8 Binding Mode in the TofI(3MΔ)/J8-C8/MTA Ternary Complex. The crystal structure of the TofI(3MΔ)/J8-C8/MTA ternary complex shows the binding of J8-C8 and MTA to TofI(3MΔ) and the structural stabilization of the segment Gly32–Glu40, which is highly disordered in the apo form (Figs. 3*B* and 4*A* and *SI Appendix*, Fig. S6*C*). This segment forms part of the loop between α1 and β2 (*SI Appendix*, Fig. S4). The structure of TofI(3MΔ) in the ternary complex is essentially identical to that in the apo form (rmsd of 0.23 Å for 175 Cα atoms), except for the ordering of the α1–β2 loop residues.
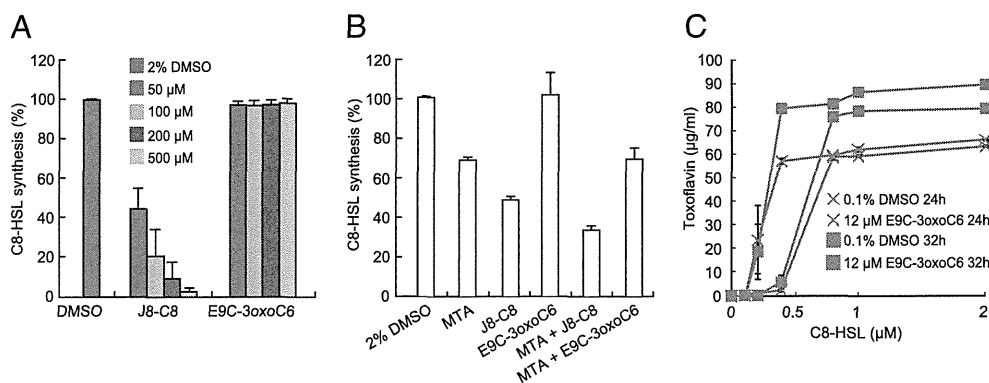

Fig. 2. Molecular mechanisms of J8-C8 and E9C-3oxoC6 inhibitory activity. (*A*) Inhibition of C8-HSL synthesis by J8-C8 in BGR1 cells. (*B*) Inhibition of C8-HSL synthesis by MTA and synergistic inhibitory effect of J8-C8. The amount of C8-HSL produced for each treatment is shown relative to that produced for DMSO treatment alone (100%). (*C*) Toxoflavin inhibition in BGS2 cells in the presence of 12 μM E9C-3oxoC6 is relieved by increasing amounts of C8-HSL. Levels of toxoflavin production were measured in the presence of 12 μM E9C-3oxoC6 or 0.1% DMSO at 24 h and 32 h after the addition of C8-HSL. Error bars represent SDs from triplicate experiments.
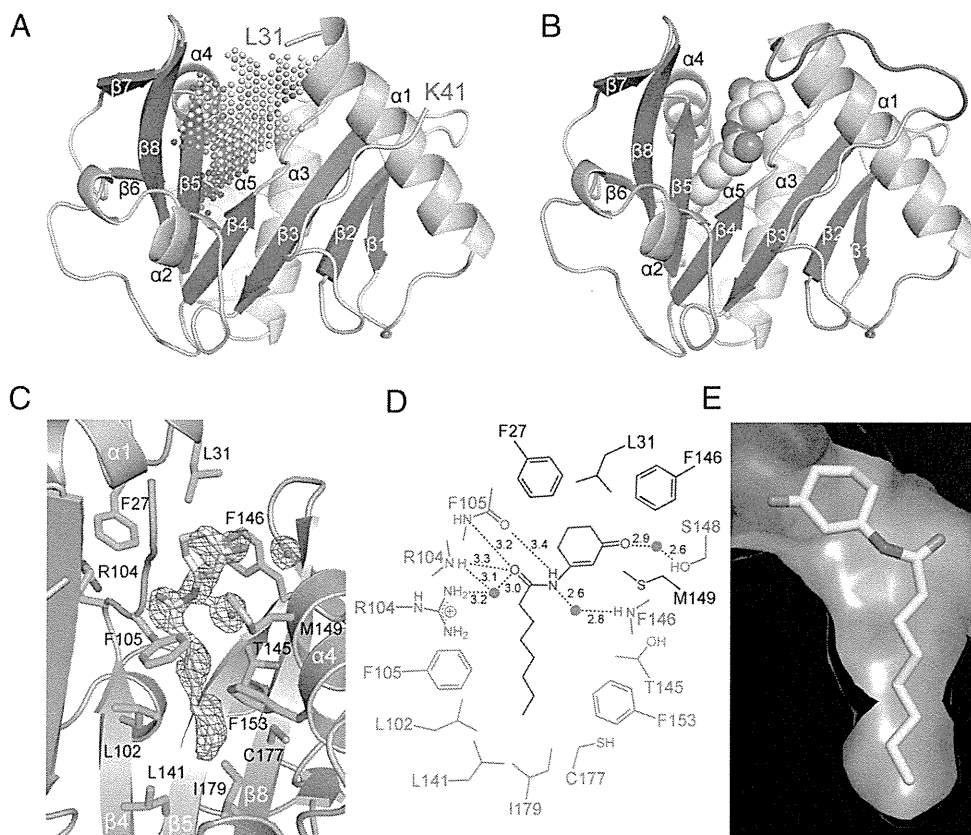
Fig. 3. Structure of TofI and the binding site of J8-C8. (*A*) The overall structure of apo-TofI(3MΔ) is shown with the putative pocket region, as defined by the program PocketPicker (23), indicated by dots. (*B*) The TofI(3MΔ)/J8-C8/MTA ternary complex is presented in an orientation identical to that in *A*. The bound J8-C8 is shown as a space-filling model, and the stabilized loop residues between α1 and β2 are shown in red. For clarity, a bound MTA in the ternary complex is not shown. (*C*) The J8-C8 binding site in the ternary complex is displayed with its neighboring residues and with the 2*F*ₒ–*F*𝒸 electron density map (contoured at 0.9σ) for J8-C8. (*D*) Schematic diagram of the interactions between TofI(3MΔ) and J8-C8 in the ternary complex. Enzyme residues involved in direct hydrogen bonds are indicated in red with interatomic distances (Å). Residues involved in van der Waals interactions within 4.0 Å of the acyl chain or ring moiety of J8-C8 are shown in green and black, respectively. A water molecule is indicated with a red circle. (*E*) The pocket-shaped binding site of J8-C8 in the ternary complex is shown as a surface model. This view is rotated by ∼180° from that shown in *C*.

In the ternary complex, J8-C8 is bound to the pocket observed in the apo-TofI(3MΔ) structure. The octanoyl chain is almost in an elongated conformation (Fig. 3*C* and *SI Appendix*, Fig. S6*B*) and is surrounded by hydrophobic residues of β4 (Leu102 and Phe105), β5 (Thr145 and Phe146), α4 [Met149 (Ile149 in the WT enzyme) and Phe153], and α1 (Phe27 and Leu31). The innermost part of the pocket is sealed off by a layer of residues that includes Leu141, Cys177, and Ile179. Compared with the apo structure, no significant positional changes occur in the main and side chains of residues in the vicinity of the pocket, indicating that the pocket remains in a highly static mode. Specifically, different conformations of the side chains, which are defined with the ordered electron density, were found in the pocket for Thr145 and Leu102 (*SI Appendix*, Fig. S6 *A* and *B*), but these minor changes, with displacement of ∼1.2 Å, had no significant effect on the size and shape of the pocket. The partially disordered ring moiety in J8-C8 (*SI Appendix*, Fig. S7) is likely stabilized by stacking interactions with Phe146 at a distance of 3.9 Å, as well as by van der Waals interactions with Phe27 and Leu31. In addition, the carbonyl oxygen on the ring faces toward the *N*-terminal end of helix α4 with the positive end of the helix dipole and is within hydrogen-bonding distance to a water molecule that forms a hydrogen bond at a distance of 2.6 Å to Ser148 (Fig. 3*D*). The amide bond of the inhibitor appears to make several hydrogen bonds with residues in the binding site. Specifically, the nitrogen and carbonyl oxygen in the amide bond are within hydrogen-bonding distance of the main chain atoms of

Phe105 (3.2–3.4 Å) and Arg104 (3.3 Å), and of water molecules (2.6–3.0 Å) that might form a hydrogen-bonding network with the nearby residues Arg104 and Phe146. J8-C8 is almost buried in the pocket, such that only ∼6% (33 Å²) of the possible surface area is exposed. The ring moiety of J8-C8 constitutes most of the exposed area, suggesting that the hydrophobicity of the octanoyl chain is a major driving force in stabilizing the binding of J8-C8 to TofI(3MΔ). The dissociation constants for J8-C8 with TofI variants in solution were in the range of 3.6–4.1 μM, as determined by fluorescence titration (*SI Appendix*, Fig. S8).

MTA Binding Mode in the TofI(3MΔ)/J8-C8/MTA Ternary Complex. Unlike the J8-C8 inhibitor bound to the concave side of the β-sheet, the reaction byproduct MTA is bound to the convex side of the β-sheet, which is surrounded by helix α1, a long loop between α1 and β2 (including the newly stabilized segment Gly32–Glu40), α2, and β7 (Fig. 4*A*). In the tertiary complex, MTA remains partially exposed to solvent, with a solvent-exposed surface area of 92 Å². Its binding site is separated from the amide bond of J8-C8 by ∼8.0 Å, with intervention by the bulky side chains of Phe146 and Trp33 (Fig. 4 *A* and *B*). The adenine moiety of MTA forms several hydrogen bonds at distances of 2.6–3.2 Å with the side chains of residues Asp45, Gln46, and Arg104, as well as hydrophobic interactions with Leu78, Val82, and Phe83 (Fig. 4 *C* and *D*). In addition to these interactions, the 5′-methylthioribose ring of MTA mediates a face-to-face stacking interaction with the side chain of Trp33 within 3.5–3.9 Å. It
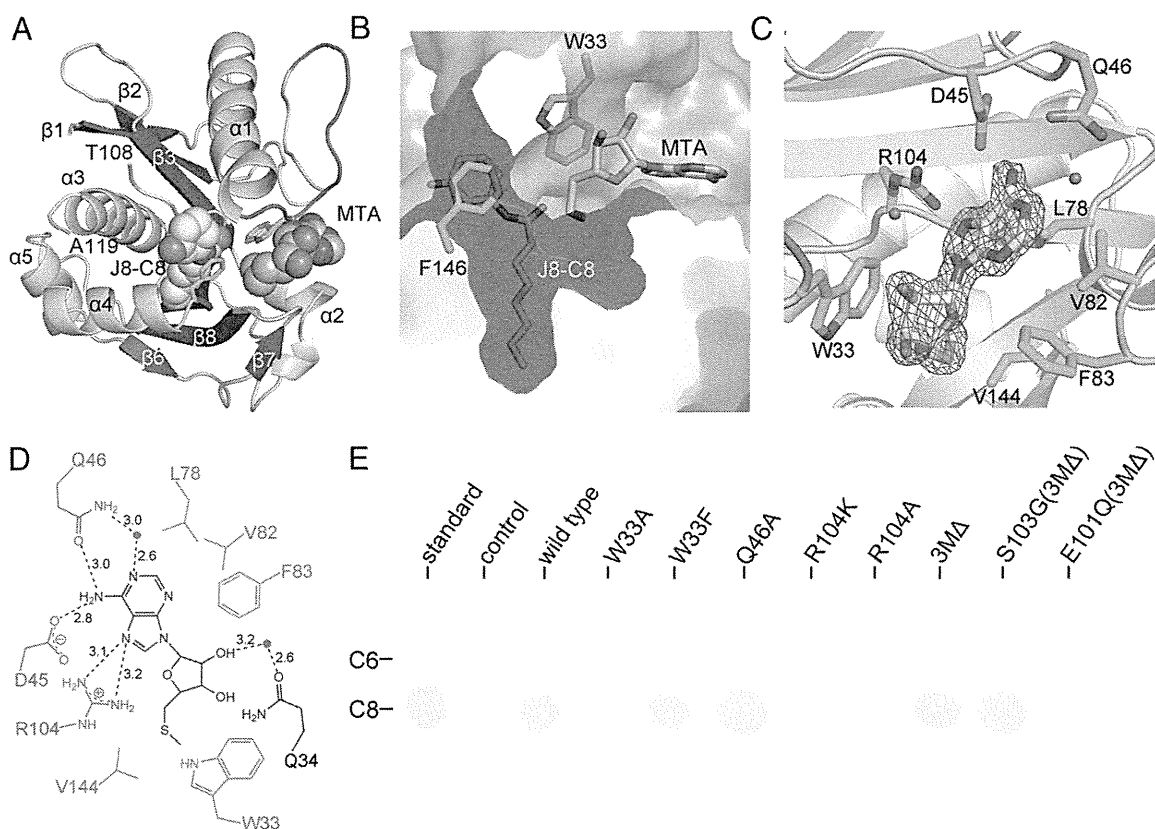
Chung et al.

Fig. 4. Structural and functional features of the ternary complex. (*A*) The overall structure of the ternary complex. J8-C8 and MTA are represented by space-filling models, and the stabilized loop residues between α1 and β2 are shown in red. (*B*) Close-up molecular surface representation of the J8-C8 and MTA binding sites separated by Phe146 and Trp33. (*C*) Close-up view of the MTA-binding region and the nearby interacting residues. The F_o–F_c electron density map for MTA is contoured at 3.5σ. (*D*) Schematic diagram of the interactions between bound MTA and residues in its binding site. Residues participating in hydrogen bonds (red) and in van der Waals interactions (green) are shown. (*E*) Detection of acyl-HSLs produced by various TofI enzymes. Unlike the other enzymes, the S103G and E101Q mutants contain additional mutations shown in TofI(3MΔ) to express the mutant protein as a soluble form. Details of the thin-layer chromatographic analysis are provided in *SI Appendix*, Fig. S1*B*.

forms a hydrophobic interaction with Val144, whereas the hydroxyl group at C2′ of the ribose ring forms a water-mediated hydrogen bond with Gln34.

Notably, residues Trp33, Asp45, and Arg104 are strictly conserved among members of the LuxI-type acyl-HSL synthase family (*SI Appendix*, Fig. S4). In a functional analysis, we found that the aromaticity of Trp33 is essential for C8-HSL production (Fig. 4*E*). Specifically, the W33F mutant, which maintains a stacking interaction with the MTA 5′-methylthioribose ring, was catalytically active. The R104K and R104A mutants were catalytically inactive, suggesting the functional significance of the arginine side chain in a hydrogen-bonding network (directly with the adenine moiety of MTA and via a water-mediated interaction with J8-C8) (Figs. 3*D* and 4*D*). These observations are consistent with a previous mutational analysis of an acyl-HSL synthase RhlI (20). As for Asp45, our mutant enzymes, such as D45N and D45A, were not expressed successfully as a soluble form, although a functional study of EsaI indicated that this conserved aspartate is crucial for activity (19). Given that the side chains of the MTA-interacting residues, with the exception of those in the Gly32–Glu40 loop, maintained identical orientations between the apo and ternary structures in this study, the stabilization of residues Gly32–Glu40, particularly Trp33, appears to be a major impetus for the binding of MTA.

Discussion

The binding of several QS inhibitors to acyl-HSL receptors has been characterized in detail (12), but their putative competition

with acyl-HSLs for binding to their cognate receptors has not been demonstrated previously. To the best of our knowledge, our competition experiments provide previously undescribed biochemical evidence of the competitive behavior of E9C-3oxoC6 toward TofR binding. This approach may be beneficial for validating the competition between other QS inhibitors and cognate acyl-HSLs for binding to proteins of the LuxR family.

One of the compounds identified in this study, J8-C8, is a structural analog of the TofI product C8-HSL and is the previously unidentified inhibitor of a LuxI-type acyl-HSL synthase. Thus far, acyl-HSL synthase structural information has been limited to the apo forms of LasI (18) and EsaI (19). Thus, our crystallographic analysis of a TofI/J8-C8/MTA ternary complex provides unprecedented molecular details of the binding interactions between an acyl-HSL inhibitor and an acyl-HSL synthase, and suggests the molecular basis for the selectivity of TofI for octanoyl-ACP from among the pool of acyl-ACPs presumed to be present during catalysis. Consistent with these functional features, the size of the pocket in TofI is highly specific for the octanoyl chain (Fig. 3*E*), and J8-C6, which differs from J8-C8 only in its hexanoyl chain, failed to disrupt QS in *B. glumae* in our experiments (*SI Appendix*, Fig. S2*A*), possibly because of its poor binding to TofI (*SI Appendix*, Fig. S8*D*). These structural considerations suggest that J8-C8 acts as a competitive inhibitor of the octanoyl-ACP substrate.

The TofI reaction scheme (*SI Appendix*, Fig. S1*A*) indicates that the MTA-binding site characterized in the TofI ternary complex is the binding site for SAM, a substrate common to all acyl-HSL

synthases. Many of the MTA-interacting residues are highly or strictly conserved in the acyl-HSL synthase family (*SI Appendix*, Fig. S4); thus, this common substrate can be accommodated by all members of this family. In the ternary complex, the binding sites for J8-C8 and MTA are ~8.0 Å apart and are separated primarily by Phe146 and Trp33 (Fig. 4*B*). Behind these residues, a channel leading toward the interior of the enzyme is large enough to accommodate the methionine moiety of SAM and to connect the inhibitor- and MTA-binding sites (*SI Appendix*, Fig. S10*A*). Thus, reactions for acylation and lactonization could occur in this region on binding of octanoyl-ACP and SAM.

A previous kinetic analysis of an acyl-HSL synthase, RhlI, suggested that the ionizing residue plays a part in the acylation reaction that leads to deprotonation of the α-amino group of SAM and subsequent nucleophilic attacks on the thioester bond of acyl-ACP (21). After acylation, lactonization appears to proceed via a direct attack of the carboxylate oxygen on the methylene carbon adjacent to the sulfonium ion within an *N*-acyl–SAM intermediate (21). In particular, various experiments consistently indicated that this lactonization step does not require a general acid-base catalysis (21). Thus, formation of acyl-HSL depends on the presence of a catalytic base. In previous structural and functional analyses of EsaI (19), a water molecule that can be activated by enzyme residue(s) in a hydrogen-bonding network has been suggested as the possible catalytic base in the acylation reaction. Catalytic features similar to this proposition were also characterized in the modeled structure for a TofI(3MΔ)/J8-C8/SAM complex (*SI Appendix*, Fig. S9). No obvious residues exist to act as a catalytic base near the amide bond of J8-C8, presumably mimicking the thioester bond of acyl-ACP, except for two water molecules. In fact, Glu101 and Ser103, which are conserved in the LuxI family of acyl-HSL synthases (*SI Appendix*, Fig. S4) (20), are the only possible candidates that could activate a water molecule via a hydrogen-bonding network, although the side chain of Glu101 appears to be in a more favorable orientation for the proposed role compared with that of Ser103, which faces away from a water molecule (*SI Appendix*, Fig. S9). Mutation of the corresponding glutamate residue, Glu97 in EsaI (19) and Glu101 in RhlI (20), nearly abolished enzyme activity. In EsaI, mutation of the serine residue corresponding to Ser103 in TofI also inactivated the enzyme (19); however, in our study, the S103G mutant enzyme, with additional mutations in TofI(3MΔ), retained enzyme activity almost comparable to that of TofI(3MΔ) (Fig. 4*E*). Our functional analysis of the E101Q mutant indicated that Glu101 is important for catalytic activity (Fig. 4*E*). Taken together, these structural and functional data support a water molecule as a catalytic base that is activated via a hydrogen-bonding network with Glu101. Nonetheless, we cannot rule out an alternative possibility, that conformational changes induced by the binding of SAM and/or acyl-ACP could properly juxtapose a possible catalytic base into the α-amino group of SAM. The observed perturbation of the p*K* value for the ionizing residue on binding of acyl-ACP suggests those conformational changes (21), which could not be identified in our ternary structure lacking the ACP moiety. In fact, different conformations of the *N*-terminal region, which corresponds to the MTA-binding site in TofI, were characterized among acyl-HSL synthases, even in the absence of ligand (*SI Appendix*, Fig. S5), implying that this particular region has a dynamic feature for catalysis (18). Further investigation is needed to resolve the mechanism of the acylation reaction.

Unlike the MTA-binding site, the inhibitor-binding pocket of acyl-HSL synthases exhibits sequence variations that alter its size and shape. The pocket-forming residues have relatively well-conserved hydrophobic features, but are not identical (*SI Appendix*, Fig. S4 and Table S2). Close inspection of the corresponding pocket regions in the apo forms of LasI (18) and EsaI (19) reveals sequence variations in and near the pocket area, as well as localized structural differences in α3 and α4, leading to different shapes and sizes of the pocket (*SI Appendix*, Fig. S10). These changes are consistent with the previously proposed tunnel- and pocket-shaped binding sites for the 3-oxo-C12 acyl chains in LasI and the 3-oxo-C6 acyl chains in EsaI (19). Thus, the specificity of acyl-HSL synthases for their cognate acyl-ACPs is likely achieved through sequence variations in the pocket.

Our structural and functional analysis showing that the inhibitor J8-C8 occupies the binding site for the acyl chain of acyl-ACP suggests the molecular basis for the acyl chain-length specificity of acyl-HSL synthases and the binding of MTA. Compounds mimicking J8-C8 with variations in acyl chain length would be candidates for narrow-range, rather than wide-spectrum, therapeutic agents. This structural information is expected to serve as a basis for fragment-based design of a new class of QS inhibitors that act against acyl-HSL synthases. In particular, we speculate that a compound containing both J8-C8 and MTA moieties would be a potent and broad-spectrum agent against various QS-dependent pathogens, including those affecting humans.

Materials and Methods

Acyl-HSL and Toxoflavin Production Assays. Assays for production of QS signaling molecules (22) and toxoflavin (14) were performed as described previously. Toxoflavin was kindly provided by Dr. Tomohisa Nagamatsu, Okayama University, Okayama, Japan.

Crystallization and Structure Determination. All TofI variants used in this study were constructed using the full-length *tofI* gene as a template for PCR carried out with sequence-specific or mutagenic primers (*SI Appendix*, Table S4). The resulting TofI variants were expressed in *Escherichia coli* strain BL21(DE3) pLysS (Novagen). For crystallization, the TofI(3MΔ) mutant was purified by ion-exchange and size-exclusion chromatography. Single-wavelength data for the apo-TofI(3MΔ) and its ternary complexes were obtained to 2.3 Å and 2.0 Å resolution, respectively, on beamlines 4A and 6C at the Pohang Accelerator Laboratory, Pohang, Korea. Structures were solved by molecular replacement and refined to final R_{work}/R_{free} values of 22.9/27.0% for the apo form and 20.0/23.1% for its ternary complex. For functional assays, enzymes with C-terminal His-tags were purified by immobilized metal-affinity chromatography.

Fluorescence Measurements. The binding of J8-C8 or J8-C6 to TofI variants was followed using steady-state fluorescence measurements at 25 °C. Ligand binding was measured by following the change in the intrinsic fluorescence of TofI(Δ) or TofI(3MΔ) as a function of ligand concentration. The excitation and emission wavelengths were 280 nm and 350 nm, respectively.

1. Fuqua C, Greenberg EP (2002) Listening in on bacteria: Acyl-homoserine lactone signaling. *Nat Rev Mol Cell Biol* 3:685–695.
2. Waters CM, Bassler BL (2005) Quorum sensing: Cell-to-cell communication in bacteria. *Annu Rev Cell Dev Biol* 21:319–346.
3. Parker CT, Sperandio V (2009) Cell-to-cell signaling during pathogenesis. *Cell Microbiol* 11:363–369.
4. Schaefer AL, Val DL, Hanzelka BL, Cronan JE, Jr., Greenberg EP (1996) Generation of cell-to-cell signals in quorum sensing: Acyl homoserine lactone synthase activity of a purified *Vibrio fischeri* LuxI protein. *Proc Natl Acad Sci USA* 93: 9505–9509.
5. Moré MI, et al. (1996) Enzymatic synthesis of a quorum-sensing autoinducer through use of defined substrates. *Science* 272:1655–1658.
6. Schaefer AL, et al. (2008) A new class of homoserine lactone quorum-sensing signals. *Nature* 454:595–599.
7. Rasko DA, Sperandio V (2010) Anti-virulence strategies to combat bacteria-mediated disease. *Nat Rev Drug Discov* 9:117–128.

– 497 –

8. Hentzer M, et al. (2003) Attenuation of *Pseudomonas aeruginosa* virulence by quorum-sensing inhibitors. *EMBO J* 22:3803–3815.
9. Smith RS, Iglewski BH (2003) *Pseudomonas aeruginosa* quorum sensing as a potential antimicrobial target. *J Clin Invest* 112:1460–1465.
10. Dong YH, et al. (2001) Quenching quorum-sensing–dependent bacterial infection by an *N*-acyl homoserine lactonase. *Nature* 411:813–817.
11. Rasmussen TB, Givskov M (2006) Quorum-sensing inhibitors: A bargain of effects. *Microbiology* 152:895–904.
12. Zou Y, Nair SK (2009) Molecular basis for the recognition of structurally distinct autoinducer mimics by the *Pseudomonas aeruginosa* LasR quorum-sensing signaling receptor. *Chem Biol* 16:961–970.
13. Gutierrez JA, et al. (2009) Transition state analogs of 5′-methylthioadenosine nucleosidase disrupt quorum sensing. *Nat Chem Biol* 5:251–257.
14. Kim J, et al. (2004) Quorum sensing and the LysR-type transcriptional activator ToxR regulate toxoflavin biosynthesis and transport in *Burkholderia glumae*. *Mol Microbiol* 54:921–934.
15. Kim J, et al. (2007) Regulation of polar flagellum genes is mediated by quorum sensing and FlhDC in *Burkholderia glumae*. *Mol Microbiol* 64:165–179.
16. Goo E, Kang Y, Kim H, Hwang I (2010) Proteomic analysis of quorum-sensing–dependent proteins in *Burkholderia glumae*. *J Proteome Res* 9:3184–3199.
17. Smith KM, Bu Y, Suga H (2003) Library screening for synthetic agonists and antagonists of a *Pseudomonas aeruginosa* autoinducer. *Chem Biol* 10:563–571.
18. Gould TA, Schweizer HP, Churchill ME (2004) Structure of the *Pseudomonas aeruginosa* acyl-homoserinelactone synthase LasI. *Mol Microbiol* 53:1135–1146.
19. Watson WT, Minogue TD, Val DL, von Bodman SB, Churchill ME (2002) Structural basis and specificity of acyl-homoserine lactone signal production in bacterial quorum sensing. *Mol Cell* 9:685–694.
20. Parsek MR, Schaefer AL, Greenberg EP (1997) Analysis of random and site-directed mutations in *rhlI*, a *Pseudomonas aeruginosa* gene encoding an acylhomoserine lactone synthase. *Mol Microbiol* 26:301–310.
21. Raychaudhuri A, Jerga A, Tipton PA (2005) Chemical mechanism and substrate specificity of RhlI, an acylhomoserine lactone synthase from *Pseudomonas aeruginosa*. *Biochemistry* 44:2974–2981.
22. Shaw PD, et al. (1997) Detecting and characterizing *N*-acyl-homoserine lactone signal molecules by thin-layer chromatography. *Proc Natl Acad Sci USA* 94:6036–6041.
23. Weisel M, Proschak E, Schneider G (2007) PocketPicker: Analysis of ligand binding-sites with shape descriptors. *Chem Cent J* 1:7, doi:10.1186/1752-153X-1-7.

12094 | www.pnas.org/cgi/doi/10.1073/pnas.1103165108   Chung et al.

– 498 –

Apolar surface area determines the efficiency of translocon-mediated membrane-protein integration into the endoplasmic reticulum

Karin Öjemalm[a], Takashi Higuchi[b], Yang Jiang[c], Ülo Langel[c], IngMarie Nilsson[a], Stephen H. White[d], Hiroaki Suga[b,1], and Gunnar von Heijne[a,e,1]

[a]Center for Biomembrane Research, Department of Biochemistry and Biophysics, Stockholm University, SE-106 91 Stockholm, Sweden; [b]Department of Chemistry, Graduate School of Science, The University of Tokyo, 113-0033 Tokyo, Japan; [c]Department of Neurochemistry, Stockholm University, SE-10691 Stockholm, Sweden; [d]Department of Physiology and Biophysics and the Center for Biomembrane Systems, University of California, Irvine, CA 92697-4560; and [e]Science for Life Laboratory Stockholm University, SE-171 77 Solna, Sweden

Integral membrane proteins are integrated cotranslationally into the membrane of the endoplasmic reticulum in a process mediated by the Sec61 translocon. Transmembrane α-helices in a translocating polypeptide chain gain access to the surrounding membrane through a lateral gate in the wall of the translocon channel [van den Berg B, et al. (2004) Nature 427:36–44; Zimmer J, et al. (2008) Nature 455:936–943; Egea PF, Stroud RM (2010) Proc Natl Acad Sci USA 107:17182–17187]. To clarify the nature of the membrane-integration process, we have measured the insertion efficiency into the endoplasmic reticulum membrane of model hydrophobic segments containing nonproteinogenic aliphatic and aromatic amino acids. We find that an amino acid's contribution to the apparent free energy of membrane-insertion is directly proportional to the nonpolar accessible surface area of its side chain, as expected for thermodynamic partitioning between aqueous and nonpolar phases. But unlike bulk-phase partitioning, characterized by a nonpolar solvation parameter of 23 cal/(mol $\cdot$ Å^2), the solvation parameter for transfer from translocon to bilayer is 6–10 cal/(mol $\cdot$ Å^2), pointing to important differences between translocon-guided partitioning and simple water-to-membrane partitioning. Our results provide compelling evidence for a thermodynamic partitioning model and insights into the physical properties of the translocon.

flexizyme | hydrophobicity | nonproteinogenic amino acid | transmembrane helix

In eukaryotic cells, membrane proteins destined for the plasma membrane and the various compartments along the endo- and exocytic pathways are synthesized by endoplasmic reticulum (ER)-bound ribosomes and cotranslationally integrated into the ER membrane in a process mediated by the Sec61 translocon complex; the homologous SecYEG translocon mediates membrane-protein integration into the inner membrane of prokaryotes (1, 2). Subsequent to membrane integration, membrane proteins fold and oligomerize in the ER and are then moved further along the secretory pathway by vesicular transport.

During the membrane-integration step, hydrophobic segments in the translocating nascent polypeptide chain exit the Sec61 translocon through a lateral gate and become embedded in the surrounding lipid bilayer (3–5). Cotranslational, translocon-mediated integration of transmembrane α-helices into the ER membrane sets the stage for all subsequent folding and oligomerization events and hence represents a critical step in the maturation of membrane proteins.

In previous studies, we have provided quantitative data on the propensities of the 20 natural amino acids to promote the integration of transmembrane helices into the ER membrane and have shown that they depend both on hydrophobicity and on position within the helix (3, 6). Although the partitioning of transmembrane helices between the Sec61 translocon and the lipid membrane bears strong similarities to partitioning of solutes between water and lipid membranes, translocon-to-bilayer partitioning may not be equivalent to water-to-bilayer partitioning (7). Insights into the differences between the two partitioning processes might be revealed if the physicochemical properties of the translocon could be probed chemically. However, given the somewhat idiosyncratic collection of proteinogenic amino acids used in nature, it has hitherto not been possible to vary side-chain chemistry in the systematic fashion required to unravel fully the physicochemical basis for translocon-mediated membrane partitioning. In order to probe the membrane-integration mechanism in greater detail, we have taken advantage of a suppressor tRNA-based technique to introduce nonproteinogenic aliphatic and aromatic amino acids into a model hydrophobic segment and measure their apparent free energies of membrane insertion. The results show not only that translocon/membrane partitioning is quantitatively different from simple water/membrane partitioning, but also reveal a physicochemical asymmetry between the cytoplasmic and lumenal ends of a transmembrane helix, which is likely related to the structure of the translocon.

Results

Experimental Approach. The basic approach is illustrated in Fig. 1. A suppressor tRNA (tRNA$_{sup}$) is charged with the desired nonproteinogenic amino acid using the Flexizyme system (8, 9) and is then added to an in vitro translation system programmed with an mRNA encoding an engineered version of the well-characterized membrane-protein leader peptidase (Lep). The Lep construct has two N-terminal transmembrane helices (TM1, TM2) and contains a hydrophobic test segment (H segment) flanked by two canonical Asn-X-Thr acceptors sites for N-linked glycosylation (G1, G2). A UAG stop codon that can be recognized by the charged tRNA$_{sup}$ serves to position the nonproteinogenic amino acid in the H segment.

When the translation reaction is carried out in the presence of ER-derived dog pancreas rough microsomes (RMs), TM1 and TM2 insert into the ER membrane as shown (10). If the H segment is recognized as a transmembrane segment by the translocon and inserted into the membrane, only the G1 acceptor site is
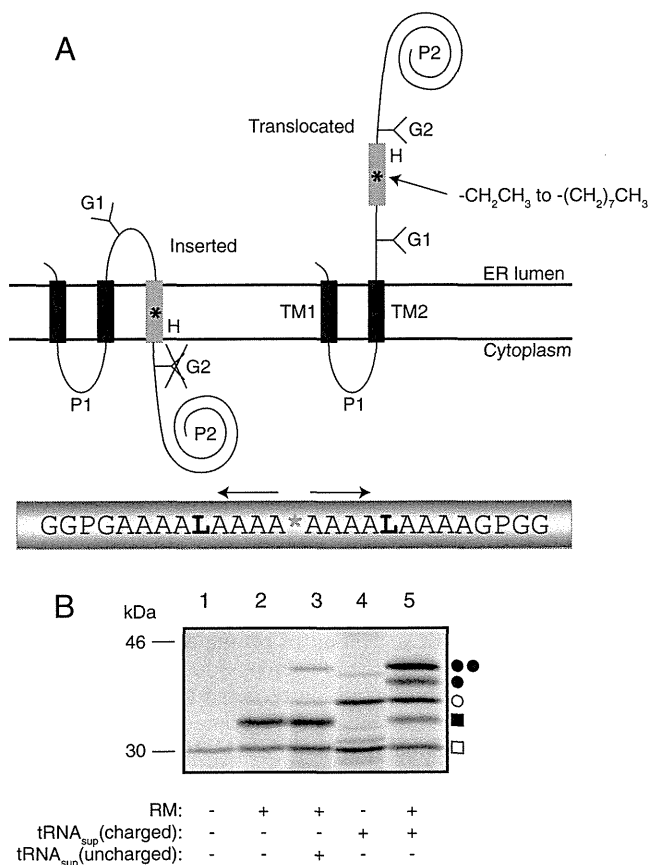
A

Fig. 1. (A) Model H segments are introduced into the Lep host protein between two engineered acceptor sites for N-linked glycosylation (G1, G2). A typical H segment is shown, including a nonproteinogenic amino acid (*) and GGPG...GPGG flanking regions. The fractions of membrane-inserted and noninserted H segment are determined by quantitation of radiolabeled singly and doubly glycosylated forms of the protein. (B) In vitro translation of a Lep construct with the H segment AAAALAAAAAAAAALAAAX (X = Me-Trp) in the absence and presence of charged or uncharged tRNA$_{sup}$ and dog pancreas RMs. □, Truncated Lep protein resulting from termination at the UAG stop codon in the H segment; ■, truncated Lep protein glycosylated on the G1 site; ○, full-length, nonglycosylated Lep protein; ●, full-length Lep protein glycosylated on the G1 site; ●●, full-length Lep protein glycosylated on the G1 and G2 sites. The degree of suppression in the absence of added tRNA$_{sup}$ is negligible (lane 2), and is very low (4%) with added uncharged tRNA$_{sup}$ (lane 3). Suppression is >50% in the presence of charged tRNA$_{sup}$ (lanes 4 and 5). The amounts of sample loaded in −RM and +RM lanes were adjusted to give roughly equal signals.

accessible to the lumenal oligosaccharyl transferase enzyme and receives a glycan moiety; if, in contrast, the H segment is translocated across the membrane, both G1 and G2 become glycosylated (11). The apparent free energy of membrane insertion of a given H segment is calculated as $\Delta G_{app} = -RT \ln(f_1/f_2)$, where R is the gas constant, T the absolute temperature (298 K), f_1 the amount of singly glycosylated molecules, and f_2 the amount of doubly glycosylated molecules (6). The H segments analyzed here have the composition GGPG-[1X,nL,(18-n) A]-GPGG, where X is the nonproteinogenic amino acid and n is chosen such that, for any given X, -1 kcal/mol $\leq \Delta G_{app} \leq +$ 1 kcal/mol. The GGPG...GPGG flanks are included in order to break any regular secondary structure and thereby insulate the H segment from the influences of the surrounding sequence and to prevent it from shifting in position along the membrane normal (6).

Membrane-Insertion Characteristics of Nonpolar Side Chains. Because hydrophobicity is central to transmembrane-helix integration

into the ER membrane, we first asked how nonpolar surface area correlates with propensity for membrane insertion. The most straightforward way to determine the relation between nonpolar surface area and ΔG_{app} is to choose amino acids with linear alkyl side chains for X. We therefore scanned amino acids with linear side chains containing two to eight carbons along an H segment with $n = 2$ leucines. The results are shown in Fig. 2A and listed in Table S1. Three important conclusions can be drawn: (i) ΔG_{app} varies in a regular fashion with the length of the side chain; (ii) the difference in ΔG_{app} between the terminal and central positions of the side chain increases with the length of the side chain; and (iii) for the longer side chains, there is a noticeable asymmetry in the curves when one compares positions close to the N-terminal end with those close to the C-terminal end of the H segment (especially evident comparing positions 6 and 14, Fig. 2 A and D). Cyclic aliphatic and nonpolar aromatic side chains yield similar results, Figs. 2B and 3A.

As seen in Fig. 2C, there is a strict correlation between the increase in side-chain accessible surface area (ASA) and ΔG_{app} that holds for aliphatic side chains with at least up to three times the ASA of the largest natural aliphatic amino acids. For the middle position in the H segment, ΔG_{app} decreases by approximately 10 cal/(mol · Å^2), whereas the decrease is smaller when the side-chain is closer to the ends of the H segment, Fig. 2D. A similar relation between ASA and ΔG_{app} holds for the nonpolar aromatic side chains, but the decrease in ΔG_{app} is only about 7 cal/(mol · Å^2) for the middle position of the H segment in this case, Fig. 2C.

Except for the slight asymmetry in the curves, these results are those expected if membrane insertion of the H segment is driven by thermodynamic partitioning into a lipid bilayer environment. First, similar to our results, the partitioning free energy of nonpolar compounds between aqueous buffer and a bilayer-mimetic solvent is proportional to ASA and can be characterized by the atomic solvation parameter σ. For partitioning between buffer and nonpolar solvents such as n-octanol, σ is approximately 23–25 cal/(mol · Å^2) for aliphatic side chains and approximately 16 cal/(mol · Å^2) for nonpolar aromatic side chains (12, 13). The corresponding values for translocon/bilayer partitioning in the middle of the membrane, approximately 10 and 7 cal/(mol · Å^2), are 2.5-fold smaller, but the value for nonpolar aromatic side chains is approximately 70% of the value for aliphatic side chains in both cases. Second, large aliphatic compounds are expected to partition preferentially into the center of the bilayer (14), as has been shown experimentally for n-hexane (15), and we see the same trend in our data, Fig. 2D.

Membrane-Insertion Characteristics of Polar Aromatic Side Chains. We also tested a selection of polar aromatic side chains, some containing non-carbon atoms, Fig. 3. As found previously, Tyr and Trp behave differently compared to non-polar aromatic residues such as Phe in that they promote membrane insertion significantly better when located near the ends of the H segment than when in the middle (3, 6); the same behavior is now seen for the aniline side chain but not for methylated tyrosine, methylated tryptophan, or the benzothiophene side chain. This observation points to the hydrogen-bonding ability of the Tyr, Trp, and aniline side chains (and the lack thereof in the Phe, methyl-Tyr, methyl-Trp, and benzothiophene side chains) as a critical distinguishing factor. A study of the interactions of Trp analogs with and without H-bonding ability showed that the "aromaticity" of Trp was the dominant cause of the preferential partitioning of Trp into lipid bilayer interfaces (16). Our results indicate that H bonding becomes important when polar aromatic residues are inserted in locations below the membrane-water interface.

Transmembrane Asymmetry. Finally, what could be the cause of the slight transbilayer asymmetry in the ΔG_{app} curves in Fig. 2A?
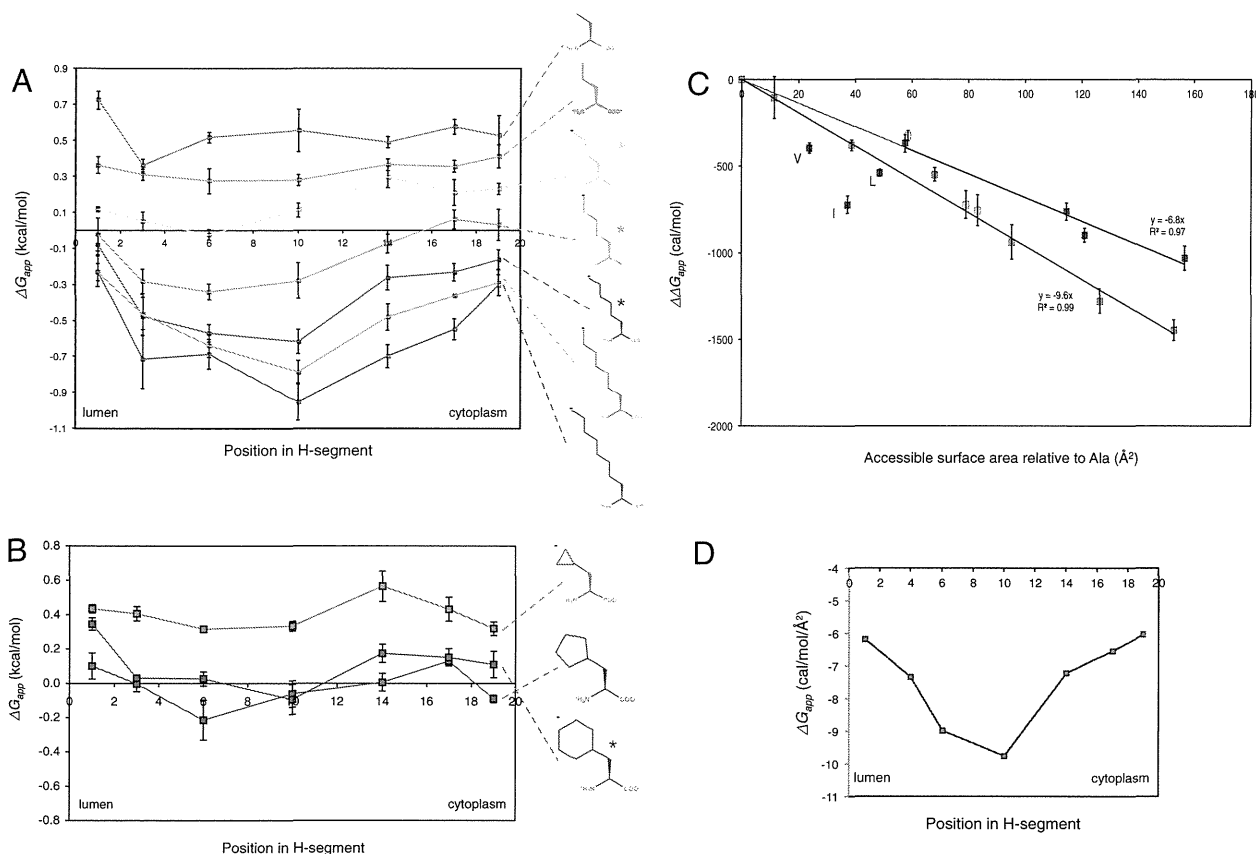
Fig. 2. (*A*) Apparent free energy of insertion (ΔG_{app}) for 19-residue-long H segments carrying a single nonproteinogenic amino acid with a linear alkyl side chain (shown on the right) in the indicated positions. The H segments all have the sequence AAAALAAAAAAAAALAAAA with the Ala in the indicated position replaced by the nonproteinogenic amino acid. The lumenal, N-terminal end of the H segment is on the left and the cytoplasmic end on the right. Error bars show standard deviations; * indicates side chains for which the ΔG_{app} value for position 6 is significantly smaller than the value for position 14 (two-sided *t* test, $p < 0.01$), i.e., that have a significant asymmetry in the ΔG_{app} profile. (*B*) ΔG_{app} for cyclic alkyl side chains. The H segments all have the sequence AAAA-LAAAAAAAAALAAAA with the Ala in the indicated position replaced by the nonproteinogenic amino acid. (*C*) ΔG_{app} decreases in proportion to the increase in ASA of the nonproteinogenic amino acid. $\Delta\Delta G_{app}$ is measured relative to the AAAALAAAAAAAAALAAAA H segment and the nonproteinogenic amino acid is in the middle position (position 10) in the H segment. The change in ASA is calculated relative to a model AAAALAAAAAAAAALAAAA α-helix. Red data points are for linear alkyl side chains, yellow for cyclic alkyl side chains (cyclopropyl, cyclopentyl, cyclohexyl), green for aromatic side chains lacking polar groups (Phe, 1-naphthyl, 2-naphthyl, biphenyl), and blue for the branched natural amino acids Leu, Ile, and Val. The average change in ΔG_{app} is −9.6 cal/(mol · Å²) for the linear alkyl side chains and −6.8 cal/(mol · Å²) for the nonpolar aromatic side chains. (*D*) Average change in ΔG_{app} per square angstrom (Å²) of accessible surface area as a function of position in the H segment (calculated from the data in *A* using linear regression as in *C*).

We asked whether the asymmetry correlates with the N-to-C-terminal polarity of the H segment or with the orientation of the H segment relative to the membrane. To invert the membrane orientation of H segments containing hexyl and octyl side chains in different positions, we inserted an additional GGPG-[7L,12A]-GPGG transmembrane segment between TM2 and the H segment, Fig. 4. Comparing Figs. 4 and 2*A*, there is a general trend that the membrane-insertion efficiency is higher when the alkyl side chain is in the lumenal half of the H segment (i.e., the asymmetry in the ΔG_{app} curves correlates with the membrane orientation of the H segment), suggesting that it reflects an asymmetry in either the ER membrane or the Sec61 translocon. Although data are scarce, the ER is thought not to have a strong asymmetry in lipid composition between the two leaflets (17, 18). Therefore, it is more likely that the origin of the asymmetry is to be sought in the translocon. Although we cannot know the exact nature of the translocon asymmetry until a high-resolution structure of a functionally open translocon becomes available, the recent structure of a translocon with a partially open lateral gate (5) reveals the nascent-chain conduit as a deep canyon lined mainly by apolar residues and with scattered polar groups projecting from the canyon walls, Fig. 5. It does not seem unreasonable that the complex shape and physicochemical hetero-

geneity of the translocon channel might underlie the observed asymmetry in the ΔG_{app} curves.

Discussion

Our results provide compelling evidence for the thermodynamic partitioning model of translocon-mediated integration of transmembrane helices into the ER membrane (3, 6, 19–21). In its simplest version, this model pictures the transmembrane helix as equilibrating between the translocon channel and the surrounding membrane. As judged from the available X-ray structures of prokaryotic homologs of the Sec61 translocon (4, 5, 22, 23), the channel is quite narrow and lined by a mixture of polar and apolar amino acids, providing an environment that is less polar than aqueous buffer. Likewise, the ER membrane, which, like all biological membranes, has a high protein content, may offer an environment that is less apolar than a pure lipid bilayer (24). Indeed, the solvation parameters for partitioning of aliphatic and nonpolar aromatic side chains obtained here [approximately 10 and 7 cal/(mol · Å²)] are a factor 2.5 smaller than found in classical solute transfer experiments (13, 25–28), suggesting that simple water-to-lipid partitioning measurements do not capture the full complexity of translocon-to-membrane partitioning (7). The asymmetry in the ΔG_{app} curves, Figs. 2 and 4, and effects on the hydrophobicity threshold for H segment insertion caused

Fig. 4. ΔG_{app} for N_{cyt}-C_{lum} orientated linear hexyl and octyl side chains. The GGPG...GPGG flanked H segments are inserted C terminal to an added third transmembrane helix (GGPGALAALALAALAALALAALAGPGG) in the Lep host protein and acceptor sites for N-linked glycosylation are engineered into the construct, as shown in the cartoon. If the H segment does not insert into the membrane, only the G1' site will receive a glycan; if it does insert, the G3' and G4' sites will also be modified. The H segments all have the composition [X,18A] and have the nonproteinogenic amino acid X in the indicated position. The cytoplasmic, N-terminal end of the H segment is on the left and the lumenal on the right. The ΔG_{app} value for position 3 is significantly larger than the value for position 17 (two-sided t test, $p < 0.01$) for the octyl side chain (indicated by *).

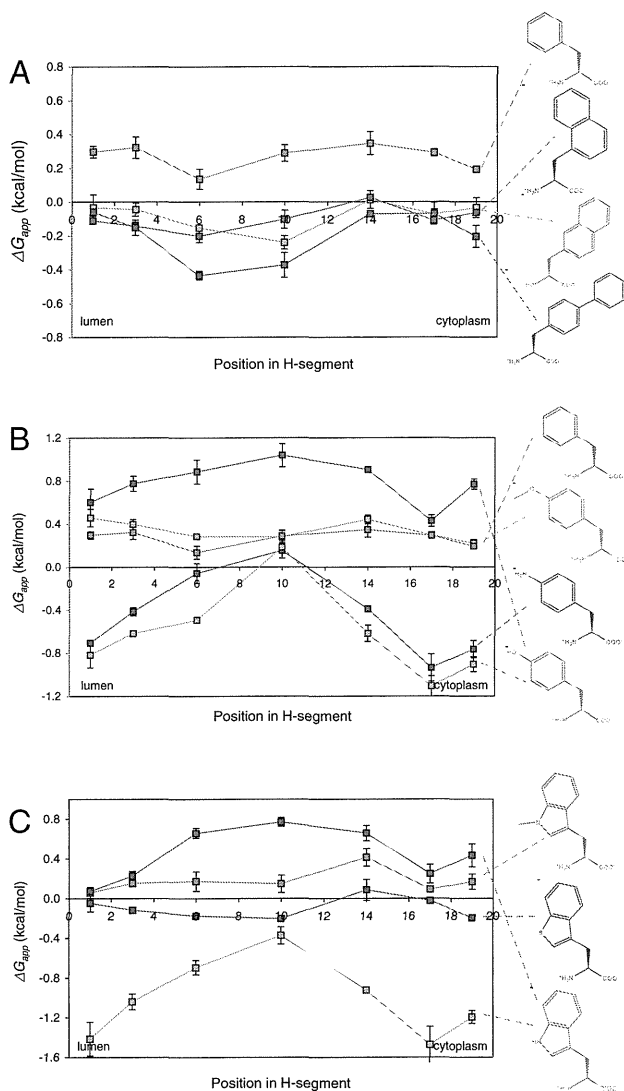**Fig. 3.** ΔG_{app} for different aromatic amino acids. The H segments used in A have the sequence AAAALAAAAAAAAALAAAA, with the Ala in the indicated position replaced by the nonproteinogenic amino acid. In B, the H segments used for Phe, Me-Tyr, and Tyr (green trace) also have this sequence, whereas the sequence used for Tyr (light-blue trace) and the aniline side chain is AAAALALAAAAALALAAAA. In C, AAAALAAAAAAAAALAAAA is used for Trp (green trace), Me-Trp and the benzothiophene side chain, and AAAALALAAAAALALAAAA is used for Trp (light-blue trace).

by mutations in the Sec61 translocon (29) support the notion that our measurements report on partitioning between the translocon channel and the surrounding lipid, although we cannot completely rule out more complicated models where the H segment can also explore the membrane-water interface region in the vicinity of the translocon during the membrane-insertion step.

Little is known about the energetics of the passage of nascent membrane-protein chains through the Sec61 translocon. Is the chain pushed steadily through the translocon by the ribosome? Or does the chain diffuse through the translocon aided by accessory proteins, such as BiP, which give directionality via a Brownian ratchet mechanism (30, 31)? If the chain is driven steadily through the translocon by the ribosome, Schow et al. (7) have suggested that there are only two independent equilibria that need be considered: one between translocon and bilayer (ΔG_{tbi}) and one between water and bilayer (ΔG_{wbi}). The first equilibrium process determines whether a transmembrane helix enters the membrane and the second determines if it stays there. Schow et al. (7) further estimated that $\Delta G_{wbi} \approx 2.6\Delta G_{tbi}$, and
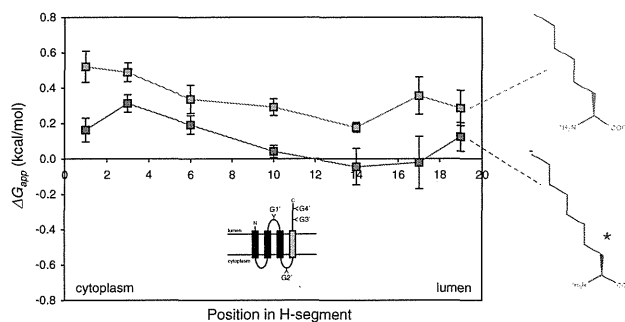
suggested that the discrepancy could be explained by assuming that the translocon/bilayer solvation parameter $\sigma_{tbi} = \sigma_{wbi}/2.6 = 8.8$ cal/(mol · Å^2). This value is remarkably close to the value for σ_{tbi} (approximately 10 cal/(mol · Å^2)) reported here. Although the agreement may be fortuitous, it does suggest a possible route to understanding quantitatively the physical chemistry of translocon-to-membrane partitioning.

Materials and Methods

Enzymes and Chemicals. All enzymes were purchased from Fermentas, except Phusion DNA polymerase from Finnzyme and SP6 RNA polymerase from Promega. The QuikChange™ Site-Directed Mutagenesis kit and deoxyribonucleotides were from Stratagene, and the Megashortscript™ T7 kit was from Ambion Inc. The plasmid pGEM1, the rabbit reticulocyte lysate system, and the RNasin were from Promega. Oligonucleotides were from Eurofins MWG Operon. All chemicals were from Sigma-Aldrich, except DMSO from J.T. Baker Chemicals, ethanol from Kemetyl, and [35S]-methionine from PerkinElmer.

DNA Manipulations. For cloning, a modified version of the *Escherichia coli lepB* gene in a pGEM1 vector was used (6). It harbors an engineered test segment (H segment) in the coding region of the P2 domain between a Spe1 cleavage site in codons 226–227 and a Kpn1 cleavage site in codon 253 (WT *lepB* codon positions) as well as two glycosylation acceptor sites for N-linked glycosylation at codons 96–98 (G1: Asn-Ser-Thr) and codons 258–260 (G2: Asn-Ala-Thr). The sequences of the H segment's flanking regions are QET-KENGIRLSETSGGPG-(H segment)-GPGGVPGQQNATWIVPP (Spe1 and Kpn1 cleavage sites underlined). The introduction of amber stop codons (TAG) into the termini of the H-segment encoding sequence was done by site-directed mutagenesis using Pfu Turbo polymerase. To introduce amber stop codons in or near the center of the H segment, double-stranded oligonucleotides encoding the H segment with the amber stop (including also GGPG/GPGG) and flanked by N-terminal Spe1 and C-terminal Kpn1 sticky ends were first generated by annealing of two pairs of complementary oligonucleotides with overlapping overhangs (each 18–45 nucleotides long), followed by annealing of the pairs via the complementary overhangs and cloning into the *lepB* gene between the Spe1 and Kpn1 cleavage sites (6).

For Lep constructs with a N_{cyt}-C_{lum} orientated H segment, a large part of the P2 domain was replaced by the corresponding part of the P2 domain of a Lep construct where all positive charges after codon 179 until 30 codons upstream of the end of the protein were replaced by alanines, and a new H segment with composition GGPG-ALAALALAALAALALAALA-GPGG was introduced between Apa1 and Mfe1 cleavage sites, located in codons −71 and −45 relative to the first codon of the original H segment. The construct further has two additional glycosylation acceptor sites for N-linked glycosylation, one (G2': Asn-Ala-Thr) 16–18 codons downstream of the new H segment and the other (G4': Asn-Ser-Thr) 7–9 codons downstream of G3' (the G3' site is the same as the G2 site in the original construct with
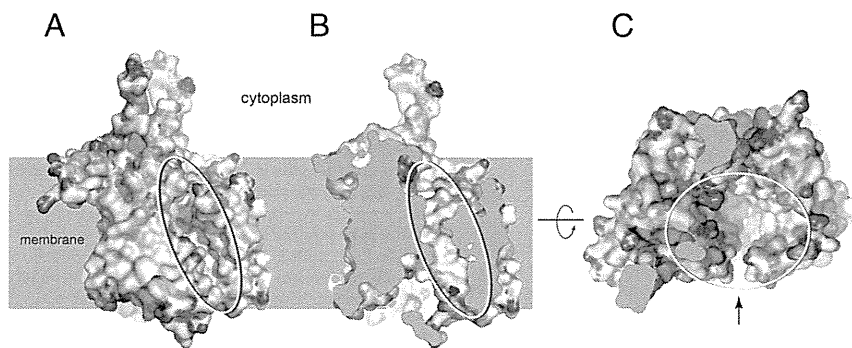
Fig. 5. Different views of the nascent-chain conduit in *Pyrococcus furiosus* SecYEβ (5) whose structure reveals a partially open ("primed") translocon. (*A*) View along the membrane plane, showing the partially open lateral gate (yellow oval). (*B*) Same view as in *A* but with the front of the molecule removed to show the rear wall of the channel. (*C*) View of the upper parts of the channel (yellow oval) seen from the cytoplasmic end. The lateral gate is indicated by the arrow. Carbon atoms are displayed in white, oxygen atoms in red, nitrogen atoms in blue, and sulfur atoms in yellow.

N_{out}-C_{in} orientated H segment). The gene fragment of the replacing P2 domain was generated by PCR amplification and inserted between a Xho1 cleavage site (*lepB* codons 180 and 181) and a Sma1 cleavage site directly downstream of the 3′ end of the *lepB* gene.

Preparation of Flexible tRNA Aminoacylation Ribozyme [Dinitrobenzyl Flexizyme (dFx); Enhanced Flexizyme (eFx)] and Microbacteriophage L5 (ML)-Derived tRNA$^{Asn}_{cta}$ [ML-tRNA$^{Asn}_{cta}$ (tRNA$_{sup}$)]. Preparations were done using the same protocol. First, double-stranded DNA templates encoding the RNA species and an N-terminal T7 promotor sequence (8, 32) were generated by PCR extension of annealed overlapping oligonuclotides. DNA templates were then amplified by PCR using primers complementary to both ends of the templates, followed by phenol/chloroform extraction and ethanol precipitation. The DNA was used in a second step for transcription by T7 polymerase using the Ambion Megashortscript™ T7 kit, and the RNA product was isopropanol precipitated and purified over 12% denaturing PAGE. After cutting out the RNA band, RNA was eluted for 2 h in 0.3 M NaCl, ethanol precipitated, and dissolved in dH$_2$O (70–250 μM final concentration).

Materials for the Synthesis and Characterization of Amino Acid Derivatives. All experiments dealing with air- and moisture-sensitive compounds were conducted under an atmosphere of dry argon. For TLC analysis, Merck precoated plates (silica gel 60 F$_{254}$, Art 5715, 0.25 mm) were used. For flash column chromatography, silica gel 60 (Merck Art 7734, 70–230 mesh) was used. For silica gel preparative TLC preparation, Merck precoated plates (silica gel 60 F$_{254}$, Art 5744, 0.5 mm) were used. Cyanomethylester (CME) and 3,5-dinitrobenzylester derivatives were prepared using a previously described procedure (8, 33). For general procedures of synthesis and characterization of compounds, see *SI Text*.

Acylation of ML-tRNA$^{Asn}_{cta}$. Flexizyme (dFx in all cases, except for aromatic amino acids where eFx was used) and ML-tRNA$^{Asn}_{cta}$ (tRNA$_{sup}$), each 250 μM in 71 mM Hepes-K buffer, pH 7.5 (total volume 7 μL), were heated at 95 °C for 2 min and cooled to room temperature over 5 min. One microliter MgCl$_2$ (200 mM for dFx and 3 M for eFx reactions) and 2 μL of amino acid 3,5-dinitrobenzyl ester (25 mM, nonaromatic amino acids) or amino acid CME (25 mM, aromatic amino acids) were added. The reaction was carried out for 2 h on ice and then stopped with 40 μM AcONa (0.3 M, pH 5.2). RNA was ethanol precipitated twice, once with a 0.1 M ethanolic AcONa solution (pH 5.2) and once with 95% ethanol. More detailed information can be found elsewhere (8).

Expression in Vitro and Quantification of Membrane-Insertion Efficiency. All constructs were transcribed for 60 min at 37 °C using a standard SP6 polymerase transcription protocol (34). Resulting mRNA was translated for 80 min at 30 °C in rabbit reticulocyte lysate (40–60 pg mRNA per μL translation mix) in the presence of [^{35}S]-Met (370 pCi/μL translation mix; 1 Ci = 37 GBq), dog pancreas RMs (50 pL/μL translation mix), amino acid mix (each 75 μM/μL translation mix), amino acid-tRNA$_{sup}$ (3.5 pmol/μL translation mix), and RNasin (3 units per μL translation mix). Translation products were analyzed by SDS-PAGE. Gels were visualized on a Fuji FLA-3000 PhosphoImager using the Image Reader 8.1j software and quantified using ImageGauge V 3.45 and the Qtiplot 0.9.3-rc2 softwares (35). The degree of membrane integration of each H segment was calculated as an apparent equilibrium constant between the membrane-integrated and nonintegrated forms: $K_{app} = f_1/f_2$, where f_1 is the fraction of singly and f_2 the fraction of doubly glycosylated Lep molecules, and the results were then converted to apparent free energies, $\Delta G_{app} = -RT \ln K_{app}$. All reported ΔG_{app} values are averages of three or four independent measurements. The degree of suppression of the UAG stop codon obtained in the in vitro translation system varied between 50% and 75% for the different aminoacylated tRNA$_{sup}$ species, and was between 2% and 6% (all positions shown in Fig. 2 were tested) when an uncharged tRNA$_{sup}$ was present in the translation mix.

Accessible Surface Area Calculations. Model helices of the H segments without flanking GGPG/GPGG sequences were generated using MacPyMol and, if needed, the side chains of the nonnatural amino acids manually modeled at the appropriate position in the helix. The accessible surface area was then calculated using Naccess version 2.1.1 (copyright S. Hubbard and J. Thornton 1992–1996) with a rolling probe size of 1.4 Å.

1. Rapoport TA (2007) Protein translocation across the eukaryotic endoplasmic reticulum and bacterial plasma membranes. *Nature* 450:663–669.

2. White SH, von Heijne G (2008) How translocons select transmembrane helices. *Annu Rev Biophys* 37:23–42.

3. Hessa T, et al. (2007) Molecular code for transmembrane-helix recognition by the Sec61 translocon. *Nature* 450:1026–1030.

4. van den Berg B, et al. (2004) X-ray structure of a protein-conducting channel. *Nature* 427:36–44.

5. Egea PF, Stroud RM (2010) Lateral opening of a translocon upon entry of protein suggests the mechanism of insertion into membranes. *Proc Natl Acad Sci USA* 107:17182–17187.

6. Hessa T, et al. (2005) Recognition of transmembrane helices by the endoplasmic reticulum translocon. *Nature* 433:377–381.

7. Schow EV, et al. (2011) Arginine in membranes: The connection between molecular dynamics simulations and translocon-mediated insertion experiments. *J Membr Biol* 239:35–48.

8. Murakami H, Ohta A, Ashigai H, Suga H (2006) A highly flexible tRNA acylation method for non-natural polypeptide synthesis. *Nat Methods* 3:357–359.

9. Xiao H, Murakami H, Suga H, Ferre-D'Amare AR (2008) Structural basis of specific tRNA aminoacylation by a small in vitro selected ribozyme. *Nature* 454:358–361.

10. Johansson M, Nilsson I, von Heijne G (1993) Positively charged amino acids placed next to a signal sequence block protein translocation more efficiently in *Escherichia coli* than in mammalian microsomes. *Mol Gen Genet* 239:251–256.

11. Sääf A, Wallin E, von Heijne G (1998) Stop-transfer function of pseudo-random amino acid segments during translocation across prokaryotic and eukaryotic membranes. *Eur J Biochem* 251:821–829.

12. Wimley WC, Creamer TP, White SH (1996) Solvation energies of amino acid sidechains and backbone in a family of host-guest pentapeptides. *Biochemistry* 35:5109–5124.

Öjemalm et al.

13. Karplus PA (1997) Hydrophobicity regained. *Protein Sci* 6:1302–1307.
14. Marqusee JA, Dill KA (1986) Solute partitioning into chain molecule interphases: Monolayers, bilayer membranes, and micelles. *J Chem Phys* 85:434–444.
15. White SH, King GI, Cain JE (1981) Location of hexane in lipid bilayers determined by neutron diffraction. *Nature* 290:161–163.
16. Yau WM, Wimley WC, Gawrisch K, White SH (1998) The preference of tryptophan for membrane interfaces. *Biochemistry* 37:14713–14718.
17. Buton X, Morrot G, Fellmann P, Seigneuret M (1996) Ultrafast glycerophospholipid-selective transbilayer motion mediated by a protein in the endoplasmic reticulum membrane. *J Biol Chem* 271:6651–6657.
18. van Meer G, Voelker DR, Feigenson GW (2008) Membrane lipids: Where they are and how they behave. *Nat Rev Mol Cell Biol* 9:112–124.
19. Heinrich S, Mothes W, Brunner J, Rapoport T (2000) The Sec61p complex mediates the integration of a membrane protein by allowing lipid partitioning of the transmembrane domain. *Cell* 102:233–244.
20. Gumbart J, Chipot C, Schulten K (2011) Free-energy cost for translocon-assisted insertion of membrane proteins. *Proc Natl Acad Sci USA* 108:3596–3601.
21. Rychkova A, Vicatos S, Warshel A (2010) On the energetics of translocon-assisted insertion of charged transmembrane helices into membranes. *Proc Natl Acad Sci USA* 107:17598–17603.
22. Zimmer J, Nam Y, Rapoport TA (2008) Structure of a complex of the ATPase SecA and the protein-translocation channel. *Nature* 455:936–943.
23. Tsukazaki T, et al. (2008) Conformational transition of Sec machinery inferred from bacterial SecYE structures. *Nature* 455:988–991.
24. Johansson AC, Lindahl E (2009) The role of lipid composition for insertion and stabilization of amino acids in membranes. *J Chem Phys* 130:185101.
25. Chothia C (1976) The nature of the accessible and buried surfaces in proteins. *J Mol Biol* 105:1–12.
26. Eisenberg D, McLachlan AD (1986) Solvation energy in protein folding. *Nature* 319:199–203.
27. Hermann RB (1977) Use of solvent cavity area and number of packed solvent molecules around a solute in regard to hydrocarbon solubilities and hydrophobic interactions. *Proc Natl Acad Sci USA* 74:4144–4145.
28. Reynolds JA, Gilbert DB, Tanford C (1974) Empirical correlation between hydrophobic free energy and aqueous cavity surface area. *Proc Natl Acad Sci USA* 71:2925–2927.
29. Junne T, Kocik L, Spiess M (2010) The hydrophobic core of the Sec61 translocon defines the hydrophobicity threshold for membrane integration. *Mol Biol Cell* 21:1662–1670.
30. Simon SM, Peskin CS, Oster GF (1992) What drives the translocation of proteins? *Proc Natl Acad Sci USA* 89:3770–3774.
31. Liebermeister W, Rapoport TA, Heinrich R (2001) Ratcheting in post-translational protein translocation: A mathematical model. *J Mol Biol* 305:643–656.
32. Ohta A, Murakami H, Higashimura E, Suga H (2007) Synthesis of polyester by means of genetic code reprogramming. *Chem Biol* 14:1315–1322.
33. Saito H, Kourouklis D, Suga H (2001) An in vitro evolved precursor tRNA with aminoacylation activity. *EMBO J* 20:1797–1806.
34. Nilsson I, Johnson AE, von Heijne G (2003) How hydrophobic is alanine? *J Biol Chem* 278:29389–29393.
35. Hedin LE, et al. (2010) Membrane insertion of marginally hydrophobic transmembrane helices depends on sequence context. *J Mol Biol* 396:221–229.

ACCOUNTS

of chemical research

Flexizymes: Their Evolutionary History and the Origin of Catalytic Function

JUMPEI MORIMOTO,[†,‡] YUUKI HAYASHI,[†,§]
KAZUHIRO IWASAKI,[†,‡] AND HIROAKI SUGA[*,†,‡,§]

†Department of Chemistry, School of Science, ‡Department of Chemistry and
Biotechnology, School of Engineering, and §Research Center for Advanced
Science and Technology, The University of Tokyo, Tokyo, Japan

CONSPECTUS

Transfer RNA (tRNA) is an essential component of the cell's translation apparatus. These RNA strands contain the anticodon for a given amino acid, and when "charged" with that amino acid are termed aminoacyl-tRNA. Aminoacylation, which occurs exclusively at one of the 3'-terminal hydroxyl groups of tRNA, is catalyzed by a family of enzymes called aminoacyl-tRNA synthetases (ARSs). In a primitive translation system, before the advent of sophisticated protein-based enzymes, this chemical event could conceivably have been catalyzed solely by RNA enzymes. Given the evolutionary implications, our group attempted in vitro selection of artificial ARS-like ribozymes, successfully uncovering a functional ribozyme (r24) from an RNA pool of random sequences attached to the 5'-leader region of tRNA. This ribozyme preferentially charges aromatic amino acids (such as phenylalanine) activated with cyanomethyl ester (CME) onto specific kinds of tRNA.

During the course of our studies, we became interested in developing a versatile, rather than a specific, aminoacylation catalyst. Such a ribozyme could facilitate the preparation of intentionally misacylated tRNAs and thus serve a convenient tool for manipulating the genetic code. On the basis of biochemical studies of r24, we constructed a truncated version of r24 (r24mini) that was 57 nucleotides long. This r24mini was then further shortened to 45 nucleotides. This ribozyme could charge various tRNAs through very simple three-base-pair interactions between the ribozyme's 3'-end and the tRNA's 3'-end. We termed this ribozyme a "flexizyme" (Fx3 for this particular construct) owing to its flexibility in addressing tRNAs.

To devise an even more flexible tool for tRNA acylation, we attempted to eliminate the amino acid specificity from Fx3. This attempt yielded an Fx3 variant, termed dFx, which accepts amino acid substrates having 3,5-dinitrobenzyl ester instead of CME as a leaving group. Similar selection attempts with the original phenylalanine-CME and a substrate activated by (2-aminoethyl)amidocarboxybenzyl thioester yielded the variants eFx and aFx (e and a denote enhanced and amino, respectively). In this Account, we describe the history and development of these flexizymes and their appropriate substrates, which provide a versatile and easy-to-use tRNA acylation system. Their use permits the synthesis of a wide array of acyl-tRNAs charged with artificial amino and hydroxy acids.

In parallel to these efforts, we initiated a crystallization study of Fx3 covalently conjugated to a microhelix RNA, which is an analogue of tRNA. The X-ray crystal structure, solved as a co-complex with phenylalanine ethyl ester and U1A-binding protein, revealed the structural basis of this enzyme. Most importantly, many biochemical observations were consistent with the crystal structure. Along with the predicted three regular-helix regions, however, the flexizyme has a unique irregular helix that was unexpected. This irregular helix constitutes a recognition pocket for the aromatic ring of the amino acid side chain and precisely brings the carbonyl group to the 3'-hydroxyl group of the tRNA 3'-end. This study has clearly defined the molecular interactions between Fx3, tRNA, and the amino acid substrate, revealing the fundamental basis of this unique catalytic system.

History of Aminoacylation Ribozymes

In the "modern world" translation system, aminoacylation of the 3'-terminus of tRNA is catalyzed by a family of protein enzymes, aminoacyl-tRNA synthetases (ARSs). On the other hand, the present evidence from the crystal structural as well as biochemical studies of ribosome have revealed that

ribosome's catalytic center consists of only RNA, that is, ribosome is a ribozyme. This suggests that a primitive translation catalytic system, including not only ribosome but also ARSs, could have consisted of entirely RNA molecules. However, naturally occurring ARS ribozymes are yet unknown and thus the above hypothesis is not fully supported by the available knowledge from nature. Because sophisticated protein enzymes such as ARSs could not be evolved before the advent of the translation system, it is critical to see if RNA molecules are able to catalyze aminoacylation.

In vitro selection (or SELEX) is a powerful technique that aims at isolating functional RNAs from a pool of random sequences of RNA. In fact, in the last nearly two decades, we have witnessed the discovery of many artificial ribozymes capable of catalyzing various chemical reactions in vitro even though such naturally occurring ribozymes are unknown in the modern life. Among them, a few research teams have successfully isolated ribozymes capable of charging certain amino acids onto RNA. One of the earlier pioneer works was reported by Yarus et al. where they isolated artificial ribozymes that catalyze self-aminoacylation of their own CCG 3'-terminal 2'/3'-OH with Phe-AMP or Tyr-AMP[1,2] as an aminoacylation donor. In 1998, Famulok and Jenne also reported an artificial rybozyme that catalyzes self-aminoacylation with N-biotin-Phe-AMP, but the site of aminoacylation is not 3'-terminal 2'/3'-OH but an internal 2'-OH.[3] Both classes of ribozymes were unable to aminoacylate onto the 3'-end of tRNA or tRNA-like CCA, and therefore, a major question remained unanswered whether ribozymes are able to charge amino acids onto the specific site of 3'-terminal CCA end of tRNA or tRNA molecules.

Meanwhile, Szostak and Lohse reported an acyl-transferase ribozyme (ATRib) capable of transferring N-biotin-Phe (biotin is a selectable tag using streptavidin resin) from the 3'-end of a short RNA to its own 5'-hydroxyl group.[4] Because this ribozyme was originally aimed at mimicking the function of ribosome as a peptidyl transfer catalyst, the acyl-donor RNA was designed to be 5'-CAACCA-3' as a mimic of tRNA's 3'-end and the 3'-end of ATRib has an internal guide sequence complementary to this RNA sequence. Using the microreversibility of acyl-transfer reaction, Suga et al. turned ATRib into two new ribozymes that are able to charge amino acids onto the 3'-terminus of tRNAs. A ribozyme, reported in 2000 and referred to as AD02,[5] has a 70 nucleotide (nt) accessory domain at the 3'-terminus of ATRib and catalyzes two steps of reactions; in the first step, AD02's accessory domain catalyzes self-aminoacylation of N-biotin-Gln assisted

by cyanomethyl ester (CME) onto the 5'-OH group, and in the second step the 5'-acyl group is transferred to the 3'-end of tRNA upon binding to the ATRib internal guide sequence. Although the catalytic ability of AD02 is very modest where only 4% of tRNA is aminoacylated, this ribozyme showed remarkable specificity toward Gln side chain over Met (0.047), Leu (0.004), Phe (<0.001), and Val (<0.001) (the values in parenthese are relative specificity constants).[6] This work represents the first example of a tRNA aminoacylation catalyst consisting of only a RNA scaffold. The second ribozyme, reported in 2002 and referred to as BC28,[7] also has an accessory domain at the 3'-terminus of ATRib and a loop, referred to as anti-anticodon (AC) loop, embedded in this domain, and recognizes a specific AC loop sequence of tRNA by forming six base pairs. BC28 accepts various kinds of amino acids from the 3'-end of 5'-AACCA-3' (note that one complementary base less than the original donor substrate) to its own 5'-terminal OH group such as ATRib, and also transfers back to the 3'-end of a specific tRNA designated by the interaction between the anti-AC loop and tRNA's AC loop, yielding up to 17% aminoacyl-tRNA. Interestingly, a base mutation(s) in the anti-AC loop is able to reprogram the specificity of BC28 that aminoacylates a desired tRNA with a complete match of six base pairs over other tRNAs containing a mispair(s). Although these ribozymes consisting of the ATRib scaffold endorse some of unique characteristic functions as ARSs, their modest efficiencies of tRNA aminoacylation due to their complex mechanisms involving the equilibrium shift of acyl-transfer chemistry dismissed further development of these ribozymes.

In Vitro Evolution of Flexizyme

In prokaryotes, precursor tRNAs, in which 5'-leader accessory sequences are attached to the 5'-end of the body of tRNA, are transcribed and processed by ribonucleoprotein enzymes, RNase P, to yield mature tRNAs onto which amino acids are charged by cognate ARSs (Figure 1A). The catalytic component of RNase P is known to be RNA (referred to as M1 RNA), representing one of the molecular fossils of the RNA world. One can hypothesize that even though the 5'-leader sequences have no function in the "modern world", they could have catalytic ability for self-aminoacylation of the 3'-terminus of tRNA sequences and later be removed by a M1 RNA-like ribozyme in the "RNA world" (Figure 1B). This hypothesis is also able to give a simple explanation on how such 5'-leader ribozymes could specify cognate tRNAs, that is, they are covalently linked to cognate tRNAs for self-aminoacylation, and thus the specificity is set in the respective precursor tRNAs. We thus decided to set an experiment
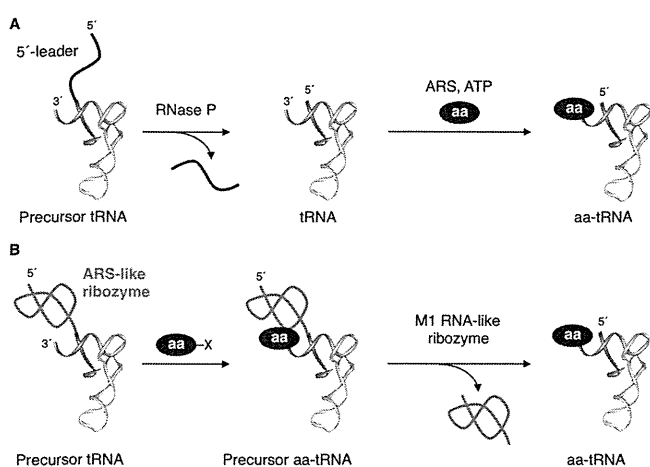
FIGURE 1. Comparison of (A) the current aa-tRNA maturation pathway and (B) a possible early aa-tRNA maturation pathway. aa denotes amino acid. (A) 5′-leader sequences of precursor tRNAs are removed by a RNA−protein complex, called RNase P, and then the 3′-terminal OH group is aminoacylated by protein ARSs. (B) ARS-like ribozymes could have existed in 5′-leader sequences of precursor tRNAs. They might have catalyzed aminoacylation of their own 3′-terminal OH and then be removed by M1-like ribozyme to yield mature aminoacyl-tRNAs. Note that all the components other than amino acid are composed of RNA molecules in this possible primitive pathway.

to discover such a catalytic precursor tRNA by means of in vitro selection.

Our previous selection outcome of AD02, in which the 3′-region of accessory domain was able to charge Gln onto the 5′-OH group of ATRib using the corresponding CME donor, particularly encouraged us to design a new selection scheme for the discovery of catalytic precursor tRNAs. We thus constructed a precursor tRNA library bearing random RNA sequences (20-nt 5′-primer region followed by 70-nt random) at the 5′-leader region of a chosen tRNA (Figure 2A) and performed in vitro selection to isolate RNA species capable of self-aminoacylating in the presence of N-biotin-L-phenylalanine cyanomethyl ester (N-biotin-Phe-CME) (Figure 3A). This selection campaign successfully yielded a single family of ribozyme, referred to as pre-24, that specifically charged N-biotin-Phe onto the 3′-terminal OH group(s) of the tRNA region.[8] Moreover, its 90-nt catalytic domain, referred to as r24, could be truncated down to 57-nt r24mini (Figure 2B) without any loss of catalytic activity.[9]

Biochemical studies on r24 revealed several intriguing features: r24 and r24mini (1) have an internal guide sequence, GGU, in L3 that forms complementary base pairs with the $\underline{A}^{73}\underline{CCA}^{76}$ (pairing bases are underlined);[9] (2) could be disconnected from tRNA, becoming a trans-acting ribozyme;[8] (3) could use not only N-biotin-Phe-CME but also Phe-CME (Figure 4A), Phe-AMP, and Phe-thioester;[8] (4) could

selectively charge Phe onto the 3′-OH group, not 2′-OH, at the tRNA 3′-terminus;[10] and (5) of which critical bases in unpaired joining regions (J1/2 and J2/3) for Phe binding were defined by chemical mappings.[9]

Although the specificity toward tRNA and Phe achieved in r24mini are essential features as an ARS-like ribozyme in the evolutionary point of view, during the course of our studies we had become more interested in developing a versatile aminoacylation catalyst rather than a specific one. We envisioned that such a ribozyme should allow us to prepare misacylated tRNAs at our will and thus serve as a convenient tool for manipulating the genetic code in translation. With such an application in mind, we decided to turn our effort to the evolution of more versatile catalysts. Based on r24mini, we constructed a doped RNA pool where bases in P3 and L3 were randomized (Figure 2C). Bases composing J2/3 and $U^{40}U^{41}$ in L3 (underlined bases in Figure 2B, C) and 5′-GGU-3′ in L3 were kept as the original sequences because their importance in the recognition of amino acid substrate and tRNA, respectively, was supported by biochemical studies.[9] With this RNA pool, we performed in vitro selection of active species. The outcome of active sequences showed that the bases in the top strand of P3 were completely conserved with the original bases whereas those beyond GGU in L3 had no sequence similarity in active species. These results suggested that the nonconserved bases are unnecessary for catalytic activity. To this end, r24mini was further truncated to a 45-nt construct (Figure 2D). This new shorter construct exhibited better activity in terms of the yield of Phe-tRNA product in trans, exhibiting k_{cat} of 0.15 min^{-1} and K_M to tRNA of 5 μM with an ability of multiple turnovers up to 16 times. Moreover, it was able to charge a variety of aromatic amino acids, including Phe analogues, onto tRNAs with A, G, and U at position 73. We thus referred to this construct of ribozyme as flexizyme3[11] (Fx3; 3 stands for the number of base pair interactions with the tRNA 3′-end).

Although this prototype Fx3 acquired versatility toward tRNAs, it was not flexible toward amino acid substrates; it was capable of charging only aromatic amino acids such as Phe and their derivatives onto tRNA. It should be noted that Fx3 retained all properties described for r24mini. As a result, we hypothesized that the critical recognition element could be embedded in the aromatic side chain (benzyl group) of the amino acid substrate. To overcome the limitation of substrate recognition of Fx3, we decided to perform in vitro evolution of Fx3 variants that are able to accept more diverse kinds of side chains in the substrate.[12] To increase the chance of success, we set two designs in the selection
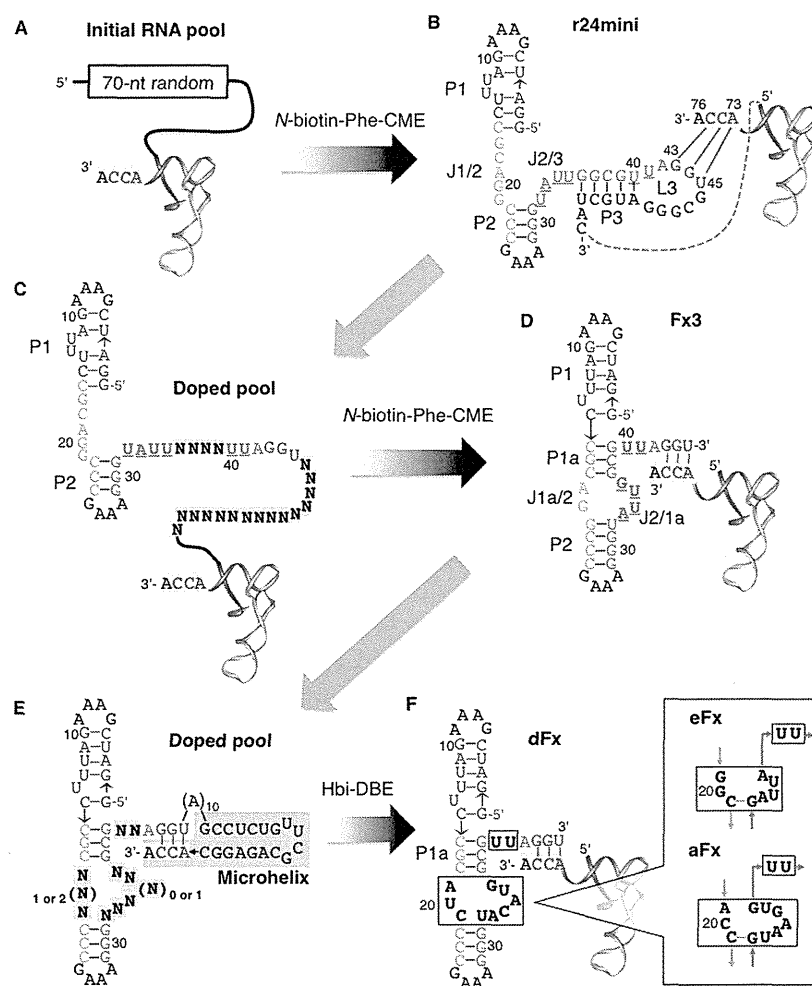
FIGURE 2. Diagram summarizing the evolutionary history of flexizyme. Black arrows indicate in vitro selection and sequence optimization steps. Amino acid substrate used for selection is written above the arrows. Gray arrows indicate the construction of doped pool (partially randomized RNA sequences). The secondary structures of (A), (B), and (C) are determined by biochemical studies. Those of (D), (E), and (F) are described based on the crystal structural analysis. In (B), (C), and (D), underlined bases are critical for binding sites of amino acid substrate determined by biochemical data.[9] In (C) and (E), randomized bases are highlighted as bold characters on a yellow background. In (F), sequences derived from randomized positions are highlighted as bold characters in a rectangle box. (A) Initial RNA pool for in vitro selection of ARS-ribozymes. (B) First generation of the artificial ARS-ribozyme. (C) Doped RNA pools constructed based on the sequence of r24mini. (D) Second generation ARS-ribozyme, that is flexible toward tRNA, referred to as Fx3. (E) Doped RNA pool constructed based on the sequence of Fx3. Bases of microhelix RNA are highlighted by a gray background. (F) Third generation ARS-ribozymes. The entire sequences of dFx is shown and sequences derived from randomized positions of eFx and aFx are in boxes on the right.
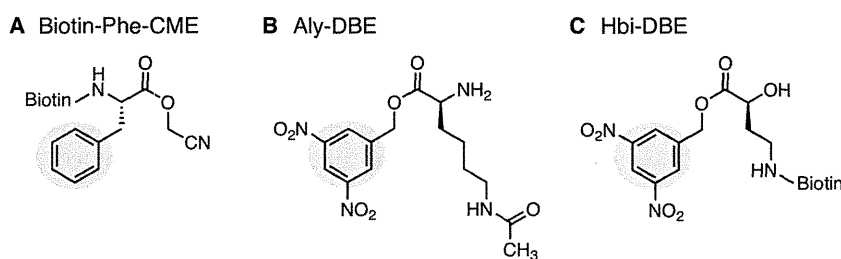

FIGURE 3. Amino acid substrates used in the evolutional attempts described in this Account. Aromatic rings accommodated in flexizymes are highlighted with yellow circles. Leaving groups are highlighted by gray boxes. (A) *N*-Biotin-phehylalanine cyanomethyl ester (Biotin-Phe-CME). (B) *ε*-(*N*-acetyl)lysine 3,5-dinitrobenzyl ester (Aly-DBE). (C) *δ*-(*N*-Biotinyl)Amino-*α*-(*S*)-hydroxybutanoic acid 3,5-dinitrobenzyl ester (Hbi-DBE).

strategy: (1) In order to alter Fx3's recognition element from the side chain to a generic group of substrate, the benzyl group was embedded in the leaving group; and (2) in order to alter the amino acid substrate recognition of Fx3, random

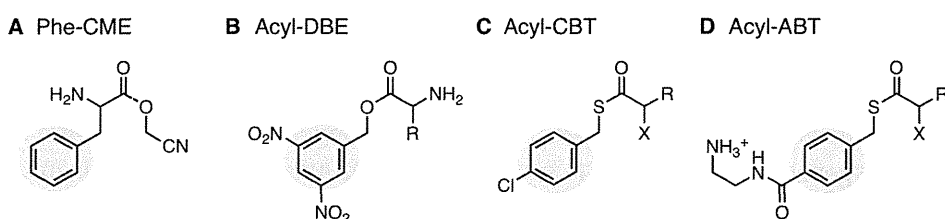**A** Phe-CME   **B** Acyl-DBE   **C** Acyl-CBT   **D** Acyl-ABT

FIGURE 4. Generic structure of amino acid substrate for flexizymes. Aromatic ring accommodated in flexizymes is highlighted with a yellow circle. Leaving group is shown in a gray rectangle. (A) Phenylalanine cyanomethyl ester (Phe-CME), suitable substrate for r24mini, Fx3, and eFx. (B) Amino acid dinitrobenzyl ester (Acyl-DBE), suitable substrate for dFx. (C) Amino acid 4-chlorobenzyl thioester (Acyl-CBT), suitable substrate for eFx. (D) Amino acid (2-aminoethyl)amidocarboxybenzyl thioester (Acyl-ABT), suitable substrate for aFx.

sequences were introduced to putative bases interacting with the aromatic side chain. To see if the strategy of (1) could be valuable, we synthesized *ε*-*N*-acetyl-lysine 3,5-dinitrobenzyl ester (Aly-DBE) (Figure 3B) where the cyanomethyl ester group was replaced with 3,5-dinitrobenzyl ester and tested it using Fx3. Although the yield of aminoacyl-tRNA was poor (~3%), Aly-DBE was accepted as an amino acid substrate of Fx3 in contrast to Aly-CME that was an inert substrate of Fx3.

This result motivated us to perform in vitro evolution of flexizyme mutants capable of more efficiently using acyl-DBE. We thus constructed a library of Fx3—microhelix RNA conjugate in which the bases involved in amino acid recognition of Fx3 were randomized (Figure 2E) and performed in vitro selection of flexizyme mutants capable of charging microhelix with *δ*-(*N*-biotinyl)amino-*α*-(*S*)-hydroxybutanoic acid DBE (Hbi-DBE) (Figure 3C). This experiment turned Fx3 to a new family of ribozymes capable of catalyzing amino-acylation onto not only microhelix RNA but also tRNAs in cis as well as trans with much higher efficiencies. One of such a 46-nt ribozyme, referred to as dinitro-flexizyme (dFx, Figure 2F), was found to be tolerant with alteration of a variety of side chains, chirality, *N*-acyl/alkyl substitutions, and non-*α*-amino acid structures as far as they are esterified with DBE, for example, artificial *α*-L-amino acids, *α*-D-amino acids,[13] *α*-*N*-acyl amino acids,[14] *α*-*N*-alkyl amino acids,[15,16] *β*-amino acids (unpublished data), and *α*-hydroxy acids[17] (Figure 5A–F). In parallel to the above selection, we also performed a selection to improve Fx3 catalytic activity using *N*-biotin-Phe-CME. This selection yielded enhanced flexizyme (eFx) (Figure 2F) that shows higher catalytic activity toward various aromatic amino acid CMEs. Coincidentally, we later found that eFx surprisingly accepts substrates activated by chlorobenzyl thioester (CBT) (Figure 4C). Because the benzyl group is embedded in the leaving group of the substrate, eFx turns out to be also a versatile catalyst that accepts those with nonproteinogenic side chains.
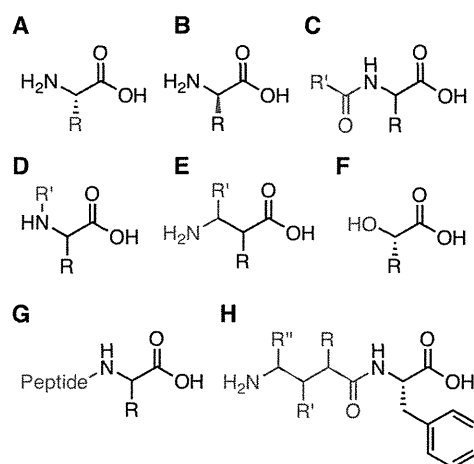

FIGURE 5. Representative examples of nonproteinogenic and artificial amino acids that have been tested for flexizymes. Nonstandard structures are highlighted in red. (A) L-Amino acid with nonproteinogenic side chain. (B) D-Amino acid. (C) *α*-*N*-acyl-amino acid. (D) *α*-*N*-Alkyl-amino acid. (E) *β*-Amino acid. (F) *α*-Hydroxy acid. (G) *α*-*N*-Peptidyl-amino acid. (H) *γ*-Aminoacyl-amino acid.

Having these two flexizymes, dFx and eFx, enables us to prepare a wide variety of acyl-tRNAs. The dFx/acyl-DBE system is our generic system for tRNA aminoacylation, since (1) synthesis of acyl-DBEs is simple, and (2) dFx is active against most acyl-DBEs dissolved in a generic reaction buffer containing some portions of DMSO if necessary. On the other hand, some acyl-DBEs could be poor or insufficient substrates of dFx, for instance, due to their steric hindrance of side chain in the combination of *N*-alkyl modification. Then, the eFx/acyl-CBT system is a preferable choice to increase the substrate reactivity, resulting in higher yields of acyl-tRNAs.

More recently, we have added another variant of flexizymes, referred to as aFx ("a" stands for amino) (Figure 2F), that uses amino acids that are activated by (2-aminoethyl)-amidocarboxybenzyl thioester (ABT) (Figure 4D). This leaving group has a primary amino group that is protonated in the generic reaction buffer at pH 7.5–8.5 to give the

ammonium salt.[18] Even though dFx and eFx function with many substrates activated with DBE and CBT, due to intrinsic hydrophobicity of the leaving groups we have encountered cases where the substrate activated by these leaving groups are poorly soluble in such a generic reaction buffer. In such cases, we first attempt to use DMSO as a cosolvent to dissolve the substrate since dFx and eFx still function up to 40% and 30% DMSO, respectively. However, DMSO occasionally fails to rescue the solubility of the substrate. In this case, the aFx/aminoacyl-ABT system often overcomes such an insolubility problem since the ABT group is able to increase water-solubility of substrate.

The choice of flexizymes in the combination of cognate leaving groups on the substrate gives us opportunities to test many different acyl-substrates; thus far, more than 300 kinds of nonstandard structures of amino and hydroxy acids (Figure 5A—F) have been charged onto tRNAs and tested for the incorporation into polypeptide chain using in vitro translation systems in our laboratory. Moreover, we have found that the flexizyme system also charges polypeptides up to pentapeptides, including those containing α-D-amino acids, β-amino acids, and γ-amino acids (Figure 5G and H), onto tRNA. These peptidyl-tRNAs derived from initiator tRNA were used to initiate the translation to yield peptides containing N-terminal unique peptides.[19,20]

A virtue of the flexizyme system is the simplicity of procedures. All we need are just a few steps of chemical synthesis of acyl-donor substrate and a few hours of its incubation with a tRNA in the presence of an appropriate flexizyme. Virtually any kind of amino and hydroxy acid could be used in the combination with any tRNAs with a variety of anticodon and body sequences. This flexibility has facilitated the genetic code reprogramming by integration with custom-made cell-free translation systems, recently referred to as flexible in-vitro translation (FIT) systems.[21] Because the primary focus in this Account is on flexizymes, we discuss this system elsewhere.

An Overview of Flexizyme Structure Determined by Crystal Structural Studies

The most intriguing feature of flexizymes is their substrate flexibilities toward amino acids bearing the benzyl (or aromatic) groups residing in the side chain or leaving group and tRNA albeit their small size consisting of only 45 or 46 nt in total length. Earlier biochemical data performed on r24 were fully accountable by the three-dimensional structure of Fx3 generated by the X-ray crystal diffraction data. Therefore, we here concentrate our discussion based on the crystal structure of Fx3.

After nearly 5 years of collaborations with Ferré-D'Amaré's group, we were able to successfully solve the crystal structure of Fx3 tethered to a microhelix RNA, a tRNA acceptor stem analogue.[22] A cocrystallization technique pioneered by Ferré-D'Amaré[23] was utilized in this study, where a U1A-binding loop sequence inserted into the Fx3's functionally dispensable L1 loop was cocrystallized with U1A protein. To define the active site of Fx3, the complex was also cocrystallized with an inhibitor, L-phenylalanine ethyl ester (Phe-EE), instead of active L-Phe-CME substrate (Figure 6B). In its crystal structure, bases are numbered differently from those in the original Fx3 construct, and the following discussion utilizes the numbers indicated in Figure 6A. In this section and the following two sections, A, U, G, or C represents a base in flexizyme and tA, tU, tG, or tC represents a base in tRNA.

Earlier biochemical studies[8] have indicated that flexizyme is composed of two helixes (P1 and P2) and its 3'-terminal three bases form an additional helix (P3) with tRNA. This prediction has turned out to be mostly true in the crystal structure of Fx3 albeit base pairs in P1 stem predicted by biochemical study are slightly different from those observed in the crystal structure. Significantly, two additional "unexpected" short helices, P1a and J1a/2-J2/1a, are newly found in the crystal structure (Figure 6A), showing that four of these helixes (P1, P1a, J1a-J2/1a, and P2) form a large main helical stack. Nucleotides of 52—54 near the 3'-end of flexizyme form a unique hairpin-shaped turn (J1a/3), helping the neighboring P3 stem to direct away perpendicularly from the main helical stack. There are three magnesium ions binding to flexizyme in the crystal structure. This is consistent with the conclusion from the biochemical data showing that there are at least two magnesium ion binding sites.[24] One of the magnesium ions exists in the irregular helix, likely participating in the formation of a catalytic core; and the other two exist in P1 stem, likely stabilizing its structure.

Structural Basis of Amino Acid Recognition

As described above, each flexizyme acquired its versatility through the independent courses of evolution experiments. Although they have different properties in terms of acyl-donor substrate specificities, all flexizymes have the common structural scaffold. Thus, we here focus on discussions of the catalytic domains on the basis of an available crystal structure of Fx3.

Fx3 accepts various phenylalanine derivatives, indicating its critical recognition element for amino acids is embedded in the benzyl (aromatic) side chain. The crystal structure
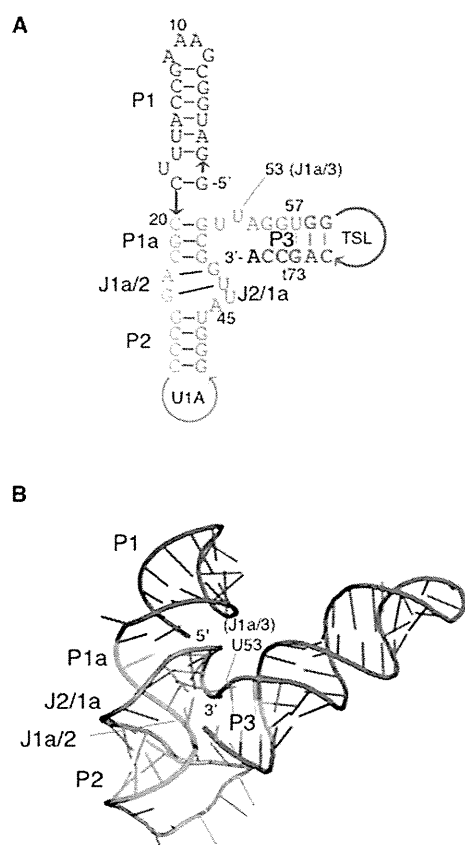
A

B

FIGURE 6. Secondary and tertiary structures of Fx3. G1−C7 and C19−G12 compose P1 stem. C20−C22 and G51−G49 compose P1a stem. G26−C28 and U43−G41 compose P2 stem. G55-U57 and C75−G73 compose P3 stem. A23−G25, U44−G48, and U52−A54 compose junctions J1a/2, J2/1a, and J1a/3, respectively. Sequences of flexizyme are drawn using three different colors, and sequences of tRNA are colored in red to facilitate comparison of the secondary structure and the tertiary structure. The U1A loop was inserted to the original Fx3 construct to enhance crystallization of the complex. The microhelix sequences other than 5′- and 3′-sequences were represented simply as TSL (T-stem loop). (A) The secondary structure determined based on the crystal structural analysis. (B) The 3D structure of Fx3 complexed with a microhelix RNA. This figure was generated by the data of accession code 3CUL from the Protein Data Bank using The PyMol Molecular Graphics System.

bound to Phe-EE shows that the aforementioned irregular helix composed of J1a/2 (A23−G25) and J2/1a (U44−G48) provides the site for the benzyl side chain (Figure 7A, B). In more details, three non-Watson−Crick base pairs (A23•G48, G24•U47, and G25•U44) and intervening two bases (A45 and U46) constitute the irregular helix, where A23•G48 expands the minor groove, making the neighboring G24•U47 pair more stretched than the ordinary G•U. This stretched G24•U47 and G25•U44 pairs together with the adjoining two unpaired bases, A45 and U46, forming the amino acid binding pocket. A hydrated magnesium binds to the major groove of this irregular helix and stabilizes the
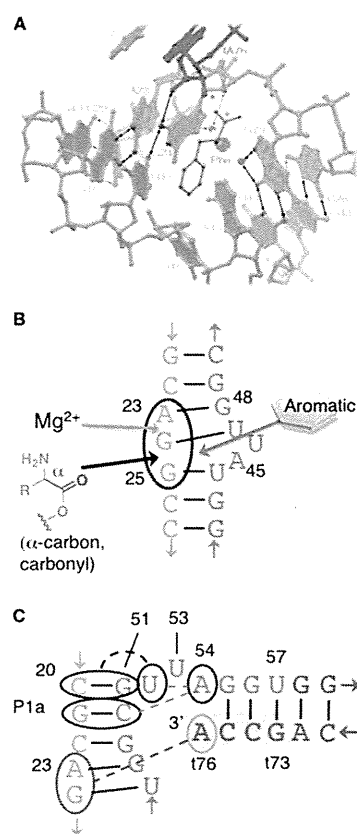
A

B

C

FIGURE 7. Amino acid and tRNA recognition elements of Fx3. Colors of bases are the same as those in Figure 6. Base-pairing interactions are indicated by solid lines. Bases and/or base pairs interacting with each other by noncanonical interactions via hydrogen bonding or stacking are circled by solid lines and indicated by dashed lines. (A) The 3D structure of flexizyme's active site. Adapted with permission from ref 22. Copyright 2008 Nature. (B) Amino acid recognition elements. Binding sites of the aromatic ring, α-carbon, and carbonyl group of Phe-EE are indicated by arrows. (C) tRNA recognition elements.

binding pocket. The phenyl group of Phe-EE is accommodated in the pocket formed by G24•U47 and U46 and interacting with O6 of G24 by its partial positive charge at the center of the phenyl ring. The α-carbon and carbonyl group of Phe-EE are accommodated in the cleft between G24 and G25. On the other hand, no electron density was observed for the ethyl group of ester moiety, but there is a space opening to the outer space and directing away from the flexizyme. This observation agrees with the fact that Fx3 accepts a relatively bulky leaving group, such as AMP, suggesting that the leaving group of the activated amino acid does not interact with the active pocket. Moreover, the α-amino group points away from the active pocket of flexizyme and thus is not precisely recognized by the flexizyme, either.

From these observations, we can speculate the mechanism how other advanced flexizymes such as dFx accept acyl-donor substrates irrespective of side chain or α-position