Acta Crystallographica Section D
Biological Crystallography
ISSN 0907-4449

Introduction of a leucine half-zipper engenders multiple high-quality crystals of a recalcitrant tRNA synthetase

Min Guo, Ryan Shapiro, Paul Schimmel and Xiang-Lei Yang*

The Skaggs Institute for Chemical Biology and Department of Molecular Biology, The Scripps Research Institute, La Jolla, CA 92037, USA

Correspondence e-mail: xlyang@scripps.edu

Received 14 October 2009
Accepted 25 December 2009

Although *Escherichia coli* alanyl-tRNA synthetase was among the first tRNA synthetases to be sequenced and extensively studied by functional analysis, it has proved to be recalcitrant to crystallization. This challenge remained even for crystallization of the catalytic fragment. By mutationally introducing three stacked leucines onto the solvent-exposed side of an α-helix, an engineered catalytic fragment of the synthetase was obtained that yielded multiple high-quality crystals and cocrystals with different ligands. The engineered α-helix did not form a leucine zipper that interlocked with the same α-helix from another molecule. Instead, using the created hydrophobic spine, it interacted with other surfaces of the protein as a leucine half-zipper (LHZ) to enhance the crystal lattice interactions. The LHZ made crystal lattice contacts in all crystals of different space groups. These results illustrate the power of introducing an LHZ into helices to facilitate crystallization. The authors propose that the method can be unified with surface-entropy reduction and can be broadly used for protein-surface optimization in crystallization.

1. Introduction

Because of their essential role in establishing the rules of the genetic code in all life forms, the structures of aminoacyl-tRNA synthetases (AARSs) are of great interest. These enzymes catalyze the ligation of amino acids to cognate tRNAs (one enzyme for each amino acid) and thus establish the genetic code in protein biosynthesis. Based on the distinct architectures of the domains that catalyze aminoacylation, AARSs are divided into two classes (I and II; Arnez & Moras, 1997; Woese *et al.*, 2000; Kavran *et al.*, 2007). Alanyl-tRNA synthetase (AlaRS) is a class II synthetase. The N-terminal catalytic fragment/unit harbors the aminoacylation domain and a $tRNA^{Ala}$-recognition motif, while the C-terminal half encodes a domain for editing mischarged $tRNA^{Ala}$ and a C-Ala domain for binding to the L-shaped tRNA elbow (Naganuma *et al.*, 2009; Sokabe *et al.*, 2009; Guo, Chong, Beebe *et al.*, 2009). AlaRS from *Escherichia coli* was among the first synthetases to be cloned and sequenced (Putney, Melendez *et al.*, 1981). It has also been extensively characterized using biochemical and genetic approaches (Jasin *et al.*, 1983, 1984; Regan *et al.*, 1987; Clarke *et al.*, 1988; Park & Schimmel, 1988; Musier-Forsyth *et al.*, 1991; Beebe, Merriman *et al.*, 2003; Beebe, Ribas De Pouplana *et al.*, 2003). Because a single G·U base pair in the acceptor stem of $tRNA^{Ala}$ is critical for aminoacylation of $tRNA^{Ala}$ and because this base pair is well separated from the trinucleotide anticodon of the genetic code, the G·U base-pair-dependent aminoacylation of $tRNA^{Ala}$

was seen as evidence for an early primitive 'second genetic code' (Hou & Schimmel, 1988; Francklyn & Schimmel, 1989). The concept of the second genetic code was later extended by examples from other tRNA synthetases (de Duve, 1988; Perona *et al.*, 1989; Frugier *et al.*, 1992; Francklyn *et al.*, 1992; Martinis & Schimmel, 1992, 1995; Nureki *et al.*, 1994; Schimmel & Ribas de Pouplana, 1995; Saks & Sampson, 1996; Felden & Giegé, 1998; Schimmel & Alexander, 1998). In addition, *E. coli* AlaRS was found to regulate its own gene transcription (Putney & Schimmel, 1981) and thus provided an early example of the expanded functions of AARSs.

The extensive functional studies of AlaRS have long been hampered by a lack of structural information. An early study reported the crystallization of a small fragment of the protein but did not yield any structure (Frederick *et al.*, 1988). 16 years later, a catalytic fragment of the orthologous *Aquifex aeolicus* AlaRS was crystallized and its structure was solved (Swairjo *et al.*, 2004). This work was followed by successful crystallizations of fragments of other thermophilic orthologs (Naganuma *et al.*, 2009; Sokabe *et al.*, 2009). While these structures could be used to interpret the large number of functional data on the *E. coli* enzyme, the lack of structural information for *E. coli* AlaRS itself nonetheless introduced some ambiguity into the interpretation.

In addition to AlaRS, many other biologically important proteins which have been subjected to extensive functional analysis are a structural mystery. Structural genomics studies have shown that the majority of native proteins are recalcitrant to crystallization (Price *et al.*, 2009), possibly because they have evolved to avoid native-state aggregation *in vivo* and the evolved properties make proteins difficult to crystallize *in vitro* (Doye *et al.*, 2004). This notion supports the idea of modifying protein surfaces to facilitate crystallization. Considering that surface entropy is the dominant factor in crystallization (Price *et al.*, 2009) and that polar side chains (such as lysine, glutamate or glutamine) play a negative entropic role (Longenecker *et al.*, 2001; Mateja *et al.*, 2002; Derewenda, 2004), it is conceivable that surface-entropy reduction (SER) by replacing polar side chains with alanines may provide favorable protein–protein interfaces to facilitate crystallization (Derewenda, 2004; Goldschmidt *et al.*, 2007).

In addition to SER, engineered leucine zippers provide another strategy to promote protein–protein interaction. Leucine zippers, which are commonly found in natural protein–protein interactions, are helical motifs that have leucine residues at about every third or fourth position, so that the protruding isobutyl side chains are lined up on one side of the helix. This arrangement creates a hydrophobic spine that can interlock with the same motif from another molecule to form a coiled-coil interface. Therefore, the leucine-zipper strategy is to introduce multiple leucine substitutions on one side of a helix to create a hydrophobic surface to promote dimerization (Yamada *et al.*, 2007). The created interaction may not only facilitate crystallization but may also influence the crystal packing.

Using both SER and leucine-zipper surface-optimization strategies, we reinvestigated the crystallization of *E. coli* AlaRS. We found that a catalytic fragment of the enzyme with three introduced leucines embedded on a predicted α-helix yielded many high-quality crystals and cocrystals with different ligands. However, the engineered α-helix did not form a leucine zipper with the same α-helix from another molecule. Instead, it interacted with other surfaces of the protein as a leucine half-zipper (LHZ) to enhance the crystal lattice interactions. Because the LHZ was engineered in a location that did not disturb enzyme activity, high-resolution structural information obtained from such constructs can be used to interpret the large archive of functional information obtained in previous investigations.

2. Materials and methods

2.1. Preparation of engineered proteins

The plasmid for expressing the *E. coli* AlaRS ND–ED fragment (residues 1–701; Guo, Chong, Beebe *et al.*, 2009), containing the N-terminal catalytic domain, the tRNA-recognition motif, the editing domain and a C-terminal 6×His tag, was constructed through PCR amplification of the desired region using primers containing *Nde*I–*Xho*I sites and ligated into plasmid pET20b to generate pET20b-AlaRS$_{701}$. Using site-directed mutagenesis, three groups of SER mutations (N32A/D33A, D54A/K55A and K326A/E327A) and three groups of leucine mutations (H104L/Q108L/E112L, K332L/G335L/D339L and E392L/R397L/T400L) were introduced. These AlaRS$_{701}$ proteins were expressed in *E. coli* strain BL21 (DE3) grown in LB medium with 100 µg ml^{-1} ampicillin. Protein expression was induced at an OD_{600} of 0.6 with 0.1 m*M* IPTG at room temperature for 3 h. Cells were lysed using a French press in Ni–NTA binding buffer (20 m*M* Tris–HCl pH 8.0, 500 m*M* NaCl, 15 m*M* imidazole). After centrifugation at 150 000*g* for 30 min, the proteins were purified from the supernatant by Ni–NTA affinity chromatography. The supernatants were added to Ni–NTA beads (Qiagen, Hilden, Germany), washed with buffer containing 20 m*M* Tris–HCl pH 8.0, 500 m*M* NaCl and 25 m*M* imidazole and eluted with buffer containing 20 m*M* Tris–HCl pH 8.0, 500 m*M* NaCl and 250 m*M* imidazole. The eluted AlaRS$_{701}$ proteins had >95% purity as judged by SDS–PAGE.

A plasmid for expressing the highly crystallizable *E. coli* AlaRS ND fragment (residues 1–441) with mutations H104L/Q108L/E112L was created by introducing a stop codon into the mutant pET20b-EcAlaRS$_{701}$ plasmid after Phe441. The protein, named AlaRS$_{441\text{-}LZ}$, was expressed in *E. coli* in the same way as described above for AlaRS$_{701}$. The fragment no longer contains the 6×His tag and was purified using three consecutive chromatography columns (DEAE Sepharose Fast Flow, Q Sepharose High Performance and Phenol Sepharose High Performance columns; GE Healthcare, Pittsburgh, Pennsylvania, USA). A NaCl gradient from buffer *A* (25 m*M* Tris–HCl pH 8.0) to buffer *B* (25 m*M* Tris–HCl pH 8.0, 500 m*M* NaCl) was used to run the DEAE and the Q columns. A different salt gradient from buffer *C* [25 m*M* Tris–HCl pH 8.0, 1.0 *M* $(NH_4)_2SO_4$] to buffer *A* was used to run the Phenol

Sepharose column. After the three-step purification, the protein was >95% pure as judged by SDS–PAGE. The active-site mutant (G237A) protein of $AlaRS_{441\text{-}LZ}$ was prepared in the same way with a similar purity. All purified proteins were dialyzed against 5 m*M* Tris–HCl buffer pH 8.0, 50 m*M* NaCl and 1 m*M* *β*-mercaptoethanol and concentrated to 60–100 mg ml^{-1} prior to crystallization.

2.2. Crystallization

A total of six engineered $AlaRS_{701}$ proteins were subjected to initial crystallization screening. Proteins (40 mg ml^{-1} final concentration) were mixed with 2 m*M* 5′-*O*-[*N*-(L-alanyl)-sulfamoyl]adenosine (Ala-SA; an adenylate analog), 10 m*M* $MgCl_2$ and 10 m*M* *β*-mercaptoethanol. Crystallization screens were set up using a Mosquito robot (TTP LabTech, Royston, England) using the sitting-drop vapor-diffusion method in 96-well plates. Each well contained 70 µl reservoir solution and an initial drop consisting of 0.1 µl sample solution mixed with 0.1 µl reservoir solution. A total of 768 conditions were screened using commercial crystallization screening solutions including Crystal Screen, Crystal Screen 2, Index Screen, SaltRx (Hampton Research, Aliso Veijo, California, USA), PEGs Suite, PEGs II Suite, JCSG+ (Qiagen), Wizard I and II

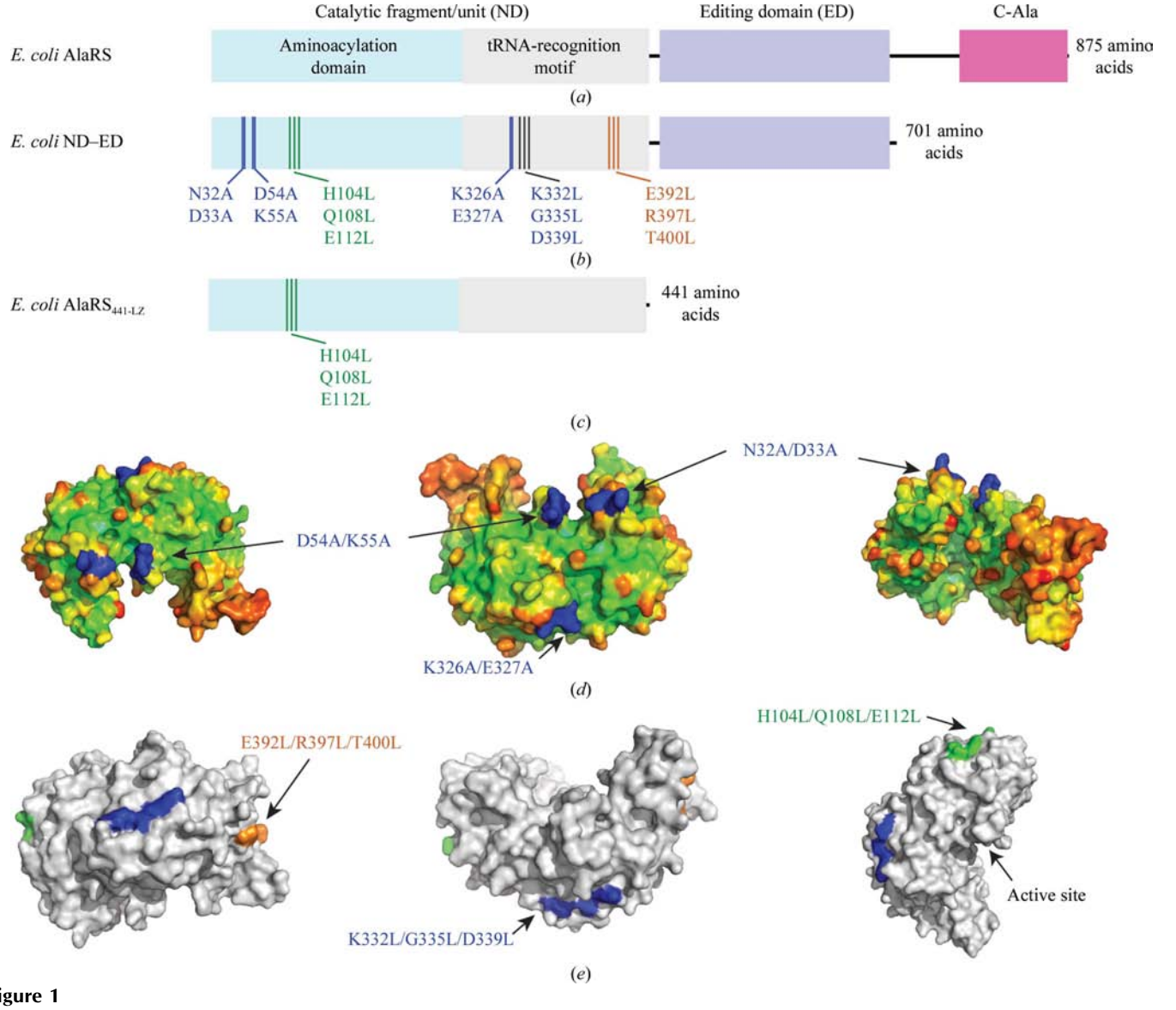

Figure 1
E. coli AlaRS and the fragments with mutagenesis sites designed for surface optimization. (*a*) Schematic representation of full-length *E. coli* AlaRS. The N-terminal catalytic fragment (ND; 1–442) includes the aminoacylation domain and a tRNA-recognition motif, which is followed by the editing domain (ED; 468–701) and the C-Ala domain (766–875). (*b*) The ND–ED fragment and the six groups of mutagenesis sites that were initially introduced. (*c*) The ND fragment containing one group of leucine mutations (*E. coli* $AlaRS_{441\text{-}LZ}$) that was robust for crystallization. (*d*) Surface locations of the three groups of mutations designed to reduce surface entropy. The surface of the *A. aeolicus* $AlaRS_{453}$ crystal structure (PDB code 1riq; Swairjo *et al.*, 2004) is colored according to the atomic *B* factors. The three mutational sites (N32A/D33A, D54A/K55A and K326A/E327A) on three high-mobility loop regions are colored in blue. (*e*) Surface locations of the three groups of leucine mutations designed according to the leucine-zipper strategy. Each group contains three leucine substitutions on the surface of the target helix. The group shown on the right (H104L/Q108L/E112L) formed the basis for the successful crystallizations of the ND domain described in this study.

Table 1
Selected crystal and X-ray diffraction data for the engineered *E. coli* $AlaRS_{441\text{-}LZ}$.

No.	Crystallization conditions	Resolution (Å)	Space group	Unit-cell parameters						Packing	Ligand
				a (Å)	*b* (Å)	*c* (Å)	α (°)	β (°)	γ (°)		
1	30% PEG 400, 0.1 *M* HEPES pH 7.8, 277 K	2.10	$P4_12_12$	114.75	114.75	125.37	90.0	90.0	90.0	Fig. 3(*a*)	Ala-SA
2	31% PEG 400, 0.1 *M* HEPES pH 7.9, 277 K	1.93	$P4_12_12$	114.88	115.04	125.57	90.0	90.0	90.0	Fig. 3(*a*)	Gly-SA
3	27% PEG 400, 0.1 *M* HEPES pH 7.5, 277 K	1.93	$P4_12_12$	114.86	114.86	126.08	90.0	90.0	90.0	Fig. 3(*a*)	Ser-SA
4	32% PEG 400, 0.1 *M* HEPES pH 7.8, 277 K	1.93	$P2_12_12_1$	114.88	115.04	125.57	90.0	90.0	90.0	Fig. 3(*a*)	G237A/Gly-SA
5	20% PEG 3350, 0.2 *M* NaCl, 0.1 *M* Tris–HCl pH 8.8, 296 K	2.85	$P2_12_12_1$	44.24	110.19	119.26	90.0	90.0	90.0	Fig. 3(*b*)	Ala-SA
6	20% PEG 400, 0.2 *M* Na_2SO_4, 277 K	2.11	$P2_12_12_1$	43.82	111.39	118.54	90.0	90.0	90.0	Fig. 3(*b*)	AMPPCP
7	20% PEG 3000, 0.1 *M* Tris–HCl pH 8.5, 277 K	2.27	$P2_12_12_1$	44.36	108.75	118.99	90.0	90.0	90.0	Fig. 3(*b*)	Ala/AMPPCP
8	30% PEG 3350, 0.2 *M* disodium tartrate, 0.1 *M* HEPES pH 7.4, 6 m*M* $CaCl_2$, 5 m*M* $MgCl_2$, 277 K	2.50	*P*1	67.12	81.21	103.97	64.2	87.0	73.4	Fig. 3(*c*)	Ala-SA
9	25% PEG 550 MME, 0.1 *M* MES pH 6.5, 296 K	2.80	$P2_12_12_1$	76.31	117.09	150.55	90.0	90.0	90.0	Fig. 3(*d*) Fig. 3(*e*)	G237A/Ser-SA† G237A/apo

† The apo form of G237A $AlaRS_{441\text{-}LZ}$ and its complex with Ser-SA were cocrystallized and one molecule of each was found in the asymmetric unit. The introduced LHZ formed different crystal lattice contacts in each molecule.

(Emerald BioSystems, Bainbridge Island, Washington, USA) at both 277 and 296 K. Only one of the six engineered $AlaRS_{701}$ proteins (H104L/Q108L/E112L $AlaRS_{701}$) produced crystals. The crystals were dissolved in 10 µl 0.1 *M* NaOH solution. The molecular weight of the crystallized protein was identified by MALDI–TOF mass spectrometry (Applied Biosystems, Foster City, California, USA) and indicated that a truncated form of the protein had crystallized. The subsequently produced H104L/Q108L/E112L $AlaRS_{441}$ proteins (named $AlaRS_{441\text{-}LZ}$ and G237A $AlaRS_{441\text{-}LZ}$) were subjected to the same high-throughput screenings as above. They were also cocrystallized with further ligands including Ala-SA, Gly-SA, Ser-SA, ATP, alanine and AMPPCP [adenylyl 5′-(β,γ-methylene)triphosphate]. Details of crystallization conditions and diffraction data are shown in Table 1. Methods for data collection, structure determination and refinement have been described elsewhere (Guo, Chong, Shapiro *et al.*, 2009).

3. Results

3.1. SER and leucine-zipper design

Our main goal was to use surface engineering to facilitate crystallization of the recalcitrant *E. coli* AlaRS. In addition, we hoped that the crystals obtained would not involve the active site in lattice contacts, which tends to interfere with ligand binding. In this regard, the *A. aeolicus* AlaRS catalytic fragment (residue 1–453) was previously crystallized in such a way that the active site and the potential tRNA-binding surface were extensively involved in crystal lattice contacts (Swairjo *et al.*, 2004). With this in mind, using the *A. aeolicus* AlaRS crystal structure as a guide, we focused on protein surfaces that are away from the active site and the tRNA-binding interface when searching for sites suitable for mutagenesis on the *E. coli* AlaRS.

E. coli AlaRS is an 875-amino-acid polypeptide that forms a tetramer in solution (Putney, Sauer *et al.*, 1981). The N-terminal catalytic unit (ND) contains 442 residues; it is linked to the editing domain (ED; residues 468–701), which is followed by a coiled-coil region and a C-Ala domain (Guo, Chong, Beebe *et al.*, 2009; Fig. 1*a*). The previously solved *A. aeolicus* AlaRS structure corresponds to the ND fragment (Swairjo *et al.*, 2004). In this structure, three loops on the surface showed the highest thermal mobility, as judged by atomic displacement parameters (or *B* factors). Two of them are in the aminoacylation domain and one is in the tRNA-recognition motif. These loops contain long polar residues in both *A. aeolicus* AlaRS and *E. coli* AlaRS (*e.g.* Lys, Glu, Asp or Asn). Using the SER strategy, we mutated specific polar residues to alanines in each of the three high-mobility regions (N32A/D33A, D54A/K55A and K326A/E327A; Figs. 1*b* and 1*d*). Interestingly, the web-based program *SERp* (Goldschmidt *et al.*, 2007) which is designed to predict SER sites also suggested K326A/E327A with the highest score (5.16) based on the primary sequence of *E. coli* AlaRS. Other sites suggested by the program [E201A/E202A (4.72), E143A/K144A/E145A (4.68) and E117A/K118A (4.46)] were either too close to the active site or were associated with low *B* factors in the *A. aeolicus* AlaRS structure and thus were not selected for our mutagenesis experiment.

A separate design was undertaken based on the strategy of introducing a leucine zipper into a surface helix (Yamada *et al.*, 2007). We found five helices that have extensive solvent-accessible surfaces and are away from the active site in the *A. aeolicus* AlaRS structure. Among these five helices, three were more stable based on their *B* factors and thus were selected as sites for introducing leucine mutations into the corresponding regions in *E. coli* AlaRS (Figs. 1*b* and 1*e*). Each group of leucine mutations contained three Leu substitutions at positions on the outer side of the targeted helix.

In all, three groups of SER mutations (N32A/D33A, D54A/K55A, K326A/E327A) and three groups of leucine mutations (H104L/Q108L/E112L, K332L/G335L/D339L, E392L/R397L/T400L) were designed (Figs. 1*b*, 1*d* and 1*e*). In addition to the considerations stated above, none of the six sites of mutagenesis were involved in crystal lattice interactions in the

A. aeolicus AlaRS crystals. The intention was to introduce new sites for potential lattice interactions.

3.2. One group of leucine mutations made the N-terminal catalytic fragment robust for crystallization

Each of the six groups of mutations was initially introduced into a construct that contained the N-terminal catalytic unit and the editing domain (ND–ED) of *E. coli* AlaRS, $AlaRS_{701}$ (Fig. 1*b*). The ND–ED contains all the sites needed for specific recognition of the cognate amino acid and tRNA substrate and is catalytically active for tRNA aminoacylation and for editing the mischarged tRNA (Jasin *et al.*, 1983; Beebe *et al.*, 2008). All six $AlaRS_{701}$ mutants were purified and subjected to high-throughput screening of 768 conditions at both 277 and 296 K. Of the six mutants, only that containing leucine mutations H104L/Q108L/E112L produced crystals. The H104L/Q108L/E112L $AlaRS_{701}$ construct gave crystals that diffracted to 3.3 Å resolution, but the structure determined by molecular replacement (using the *A. aeolicus* AlaRS structure as the initial search model) showed that this crystal contained only the ND domain. Further analysis of the crystals by MALDI–TOF mass spectrometry confirmed that the ED region was cleaved off during crystallization; the cleavage site was identified to be Phe441, which is at the end of the ND domain.

To remove the protease contamination that cleaved off the ED, we prepared another batch of H104L/Q108L/E112L $AlaRS_{701}$ protein with protease inhibitors added at all steps of purification. Interestingly, when cleavage was successfully inhibited, the intact H104L/Q108L/E112L $AlaRS_{701}$ could not be crystallized using the same screening protocol. This result suggested that removal of the ED was necessary for crystallization. This observation is in line with our finding that the ND region of *E. coli* AlaRS does not directly interact with the ED and that the two domains are flexible with respect to each other (Guo, Chong, Beebe *et al.*, 2009).

According to the protease cleavage site identified by mass spectrometry, we next made a new construct of AlaRS ND with the same leucine mutations H104L/Q108L/E112L and without a His tag ($AlaRS_{441\text{-}LZ}$; Fig. 1*c*). After purification, the protein was subjected to high-throughput crystallization screening. Crystals were readily obtained. To exemplify the extremely high tendency of $AlaRS_{441\text{-}LZ}$ to crystallize, more than 80 of the 96 conditions of the PEG Suite screen yielded crystals. Moreover, 26 different conditions at room temperature and 29 conditions at 277 K yielded large-sized crystals that were suitable for data collection. Importantly, this ND fragment of *E. coli* AlaRS is the minimal fragment that is catalytically active for amino-acid activation (Putney, Royal *et al.*, 1981; Jasin *et al.*, 1983; Regan *et al.*, 1987). Previous efforts to crystallize a similar fragment (residues 1–461) of *E. coli* AlaRS were largely unsuccessful and did not yield any structure (Frederick *et al.*, 1988). Evidently, the leucine mutations (H104L/Q108L/E112L) greatly facilitated the crystallization of the ND fragment of *E. coli* AlaRS. In addition, cocrystals of $AlaRS_{441\text{-}LZ}$ (and its active-site mutant G237A $AlaRS_{441\text{-}LZ}$) with ligands such as Ala-SA, Gly-SA, Ser-SA, ATP, alanine and AMPPCP were also readily obtained.

3.3. Leucine mutations form leucine half-zippers that are involved in crystal lattice interaction

More than ten data sets were collected from $AlaRS_{441\text{-}LZ}$ or G237A $AlaRS_{441\text{-}LZ}$ crystals containing different ligands and grown from different crystallization conditions (Table 1). These crystals exhibited five different lattices (with four different space groups), with resolutions extending in some instances to 1.93 Å (Table 1). The crystal structures of *E. coli* $AlaRS_{441\text{-}LZ}$ from multiple data sets were readily solved by molecular replacement. As expected, the introduced leucine residues (Leu104/Leu108/Leu112) lined up on the outer side of the target helix (Lys103–Thr115) in the catalytic domain (Fig. 2). Without exception, they formed crystal lattice contacts in all crystals (Figs. 3*a*–3*e*), explaining the high success rate of $AlaRS_{441\text{-}LZ}$ in crystallization. Interestingly, the engineered helix did not form a leucine zipper with the same helix from a symmetry-related molecule as expected. Instead, the introduced leucines acted as a leucine half-zipper (LHZ) and made hydrophobic contacts with other parts of the neighboring molecule in all crystals. In most cases, the LHZ interacted with an α-helix (Glu421–Ser439) on the C-terminal end of a neighboring molecule (Figs. 3*a*–3*d* and 3*f*). In one case, the LHZ interacted with a loop region (Pro30–Asp33) that was ~20 Å away from the active site in a neighboring molecule (Figs. 3*e* and 3*g*). In all cases, the engineered LHZ appeared to be the key to promoting crystal packing by directly forming crystal lattice contacts.

Most importantly, because the designed LHZ was located opposite to the active site (Fig. 2), crystal packing through the LHZ did not affect the active site

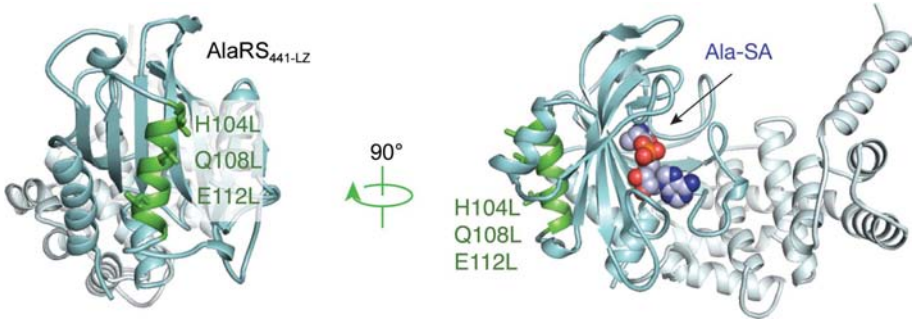

Figure 2
The crystal structure of *E. coli* $AlaRS_{441\text{-}LZ}$ in complex with the Ala-SA ligand. The helix with the leucine mutations (Lys103–Thr115, green) is located in the aminoacylation domain and opposite to the active site (shown by the bound Ala-SA). The aminoacylation domain is colored cyan and the following tRNA-recognition motif is colored light blue.

(Fig. 3). Moreover, kinetic analyses showed that $AlaRS_{441\text{-}LZ}$ is enzymatically similar to wild-type $AlaRS_{441}$. These combined features make $AlaRS_{441\text{-}LZ}$ suitable for studying the conformational changes associated with ligand binding (Guo, Chong, Shapiro *et al.*, 2009).

4. Discussion

Using rational design, we generated six groups of surface-engineered mutations on the ND–ED fragment of *E. coli* AlaRS (Fig. 1*b*) and found that one group (H104L/Q108L/E112L) could turn the ND fragment into a protein that was robust for crystallization ($AlaRS_{441\text{-}LZ}$; Fig. 1*c*). With this group of leucine mutations we obtained a series of cocrystals of $AlaRS_{441\text{-}LZ}$ with different ligands (Table 1). The inability to crystallize the longer ND–ED construct with the same LHZ indicates that conformational flexibility was a substantial barrier to crystallization which could not be overcome by surface optimization alone. Consistent with our observations, statistical analysis of large-scale protein crystallizations also demonstrated that surface entropy and backbone flexibility both anticorrelate strongly with successful structural determination (Price *et al.*, 2009). Therefore, for flexible proteins that are recalcitrant to crystallization we recommend that a method to minimize the flexibility (*e.g.* by adding a specific binding partner) should be used together with surface optimization.

Interestingly, although the leucine mutations (H104L/Q108L/E112L) were designed to form a self-interacting zipper-like contact as proposed initially (Yamada *et al.*, 2007), it turned out that the leucine side chains packed with other sites, forming an interface that only involved the LHZ from one protein molecule (Figs. 3*f* and 3*g*). Therefore, these leucines created a hydrophobic surface that tends to interact broadly rather than simply self-adhering. The effect is to create like a 'sticky patch'. On the other hand, leucine has one of the lowest side-chain entropy values among the 20 amino acids (Cieślik & Derewenda, 2009), suggesting that the

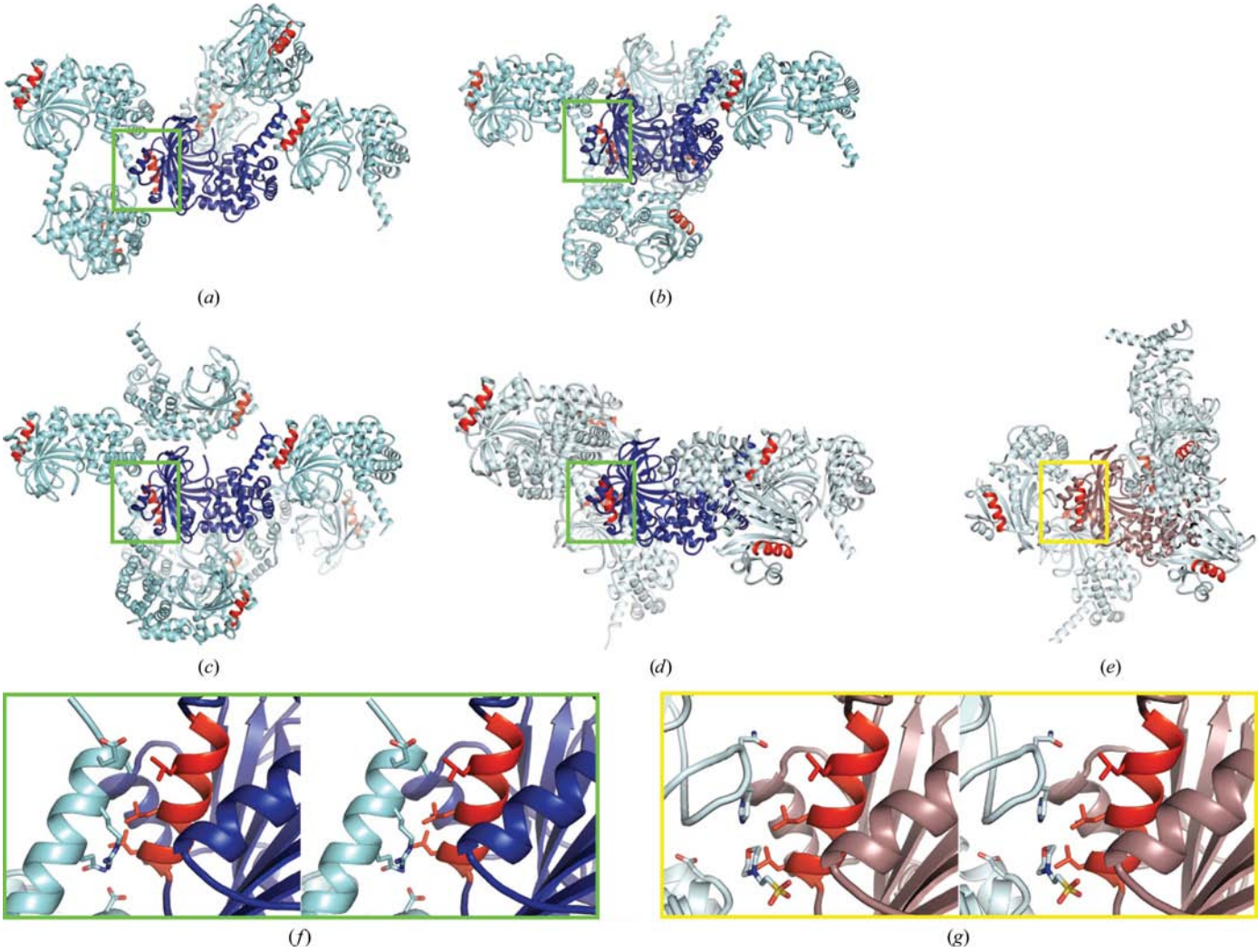

Figure 3
The leucine mutations form crystal contacts as leucine half-zippers (red) in all five different crystal lattices of *E. coli* $AlaRS_{441\text{-}LZ}$. In most cases (*a*–*d*), the introduced leucines interact with the C-terminal helix (Glu421–Ser439) of the neighboring molecule. Only in one crystal packing (*e*) do the introduced leucines interact with a surface-loop region (Pro30–Asp33) in the aminoacylation domain. (*f*, *g*) Close-up stereoviews of the interactions with the introduced LHZ. The details of the crystals are listed in Table 1.

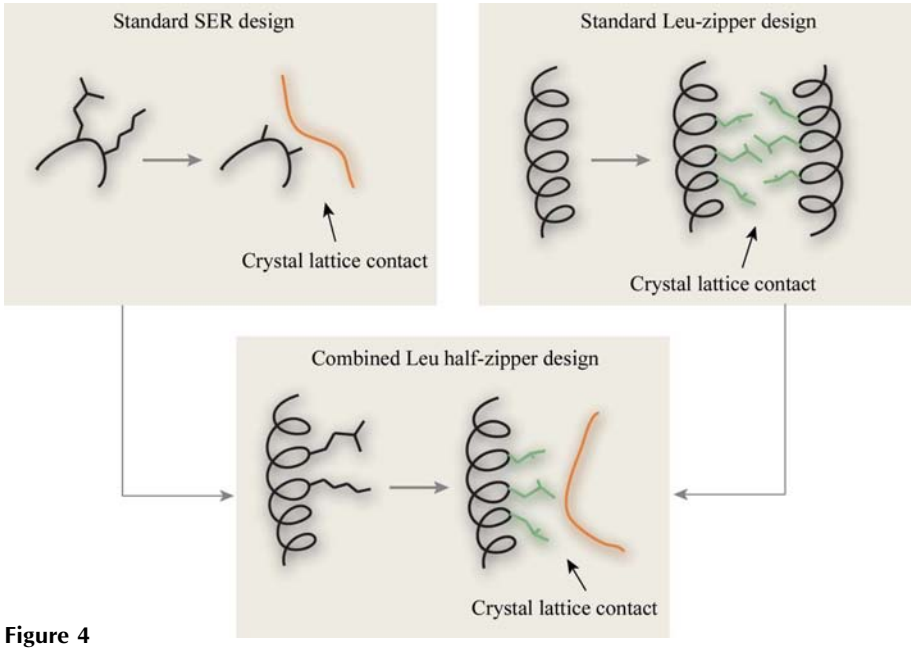

Figure 4
Schematic illustrations of the standard SER and leucine-zipper designs and the concept of LHZ as a unified surface-optimization strategy to facilitate protein crystallization.

replacement of most amino acids with leucine would result in surface-entropy reduction. For these reasons, the LHZ may promote crystallization not only by creating a hydrophobic surface to enhance protein–protein lattice interactions but also by lowering the surface entropy to facilitate crystal packing.

Alanine has the lowest side-chain entropy and therefore is the standard choice in SER design (Derewenda, 2004). However, when an alanine residue is introduced into an α-helix, the main-chain atoms participate in intramolecular hydrogen bonds and the short side chain provides only a limited surface for making intermolecular crystal lattice contacts. Therefore, a principle used in the SER-prediction program (Goldschmidt *et al.*, 2007) is to suggest alanine substitutions in loop regions. In contrast, when placed onto the surface of a helix, leucine, with its long hydrophobic side chain and low entropy, could work better than alanine to form crystal lattice contacts. Not surprisingly, leucine has the second highest propensity for involvement in crystal lattice interfaces (Cieślik & Derewenda, 2009).

Because α-helices are the most abundant secondary structure in proteins and many of them are found to be solvent-exposed, the strategy of improving crystallization by engineering the surface Lys, Glu or Gln residues on a helix to leucines can be applied broadly to many proteins. Whereas the initially proposed leucine-zipper strategy expected these leucine substitutions to form a self-interacting surface to facilitate crystallization (Yamada *et al.*, 2007), we showed here that the LHZ has a broader effect. The targeted leucine substitutions of Lys, Glu or Gln on the surface of a helix can reduce entropy and simultaneously provide additional surface area for making lattice contacts (Fig. 4). This strategy complements the well established SER strategy, which focuses on replacing flexible residues with alanines in loop regions. Thus, the LHZ strategy connects the standard SER and leucine-zipper designs into a unified surface-optimization concept for facilitating protein crystallization. The high success rate of crystallizations in our case supports the validity of this concept. The crystals of the engineered proteins studied here diffracted to the highest resolution of all AlaRS crystals from different species (Swairjo *et al.*, 2004; Swairjo & Schimmel, 2005; Naganuma *et al.*, 2009; Sokabe *et al.*, 2009).

We thank Dr Todd Yeates and the reviewers for their instructive comments which improved the clarity of the manuscript. X-ray diffraction data were collected on Stanford Synchrotron Radiation Laboratory (SSRL) beamlines 7-1, 9-1 and 11-1. The SSRL Structural Molecular Biology Program is supported by the Department of Energy, Office of Biological and Environmental Research and by the National Institutes of Health, National Center for Research Resources, Biomedical Technology Program and the National Institute of General Medical Sciences. This work was supported by grant GM 15539 from the National Institutes of Health and by a fellowship from the National Foundation for Cancer Research.

References

Arnez, J. G. & Moras, D. (1997). *Trends. Biochem. Sci.* **22**, 211–216.
Beebe, K., Merriman, E. & Schimmel, P. (2003). *J. Biol. Chem.* **278**, 45056–45061.
Beebe, K., Mock, M., Merriman, E. & Schimmel, P. (2008). *Nature (London)*, **451**, 90–93.
Beebe, K., Ribas De Pouplana, L. & Schimmel, P. (2003). *EMBO J.* **22**, 668–675.
Cieślik, M. & Derewenda, Z. S. (2009). *Acta Cryst.* D**65**, 500–509.
Clarke, N. D., Lien, D. C. & Schimmel, P. (1988). *Science*, **240**, 521–523.
Derewenda, Z. S. (2004). *Structure*, **12**, 529–535.
Doye, J. P., Louis, A. A. & Vendruscolo, M. (2004). *Phys. Biol.* **1**, 9–13.
Duve, C. de (1988). *Nature (London)*, **333**, 117–118.
Felden, B. & Giegé, R. (1998). *Proc. Natl Acad. Sci. USA*, **95**, 10431–10436.
Francklyn, C. & Schimmel, P. (1989). *Nature (London)*, **337**, 478–481.
Francklyn, C., Shi, J. P. & Schimmel, P. (1992). *Science*, **255**, 1121–1125.
Frederick, C. A., Wang, A. H., Rich, A., Regan, L. & Schimmel, P. (1988). *J. Mol. Biol.* **203**, 521–522.
Frugier, M., Florentz, C. & Giegé, R. (1992). *Proc. Natl Acad. Sci. USA*, **89**, 3990–3994.
Goldschmidt, L., Cooper, D. R., Derewenda, Z. S. & Eisenberg, D. (2007). *Protein Sci.* **16**, 1569–1576.
Guo, M., Chong, Y. E., Beebe, K., Shapiro, R., Yang, X. L. & Schimmel, P. (2009). *Science*, **325**, 744–747.
Guo, M., Chong, Y. E., Shapiro, R., Beebe, K., Yang, X. L. & Schimmel, P. (2009). *Nature (London)*, **462**, 808–812.
Hou, Y. M. & Schimmel, P. (1988). *Nature (London)*, **333**, 140–145.

Jasin, M., Regan, L. & Schimmel, P. (1983). *Nature (London)*, **306**, 441–447.
Jasin, M., Regan, L. & Schimmel, P. (1984). *Cell*, **36**, 1089–1095.
Kavran, J. M., Gundllapalli, S., O'Donoghue, P., Englert, M., Söll, D. & Steitz, T. A. (2007). *Proc. Natl Acad. Sci. USA*, **104**, 11268–11273.
Longenecker, K. L., Garrard, S. M., Sheffield, P. J. & Derewenda, Z. S. (2001). *Acta Cryst.* D**57**, 679–688.
Martinis, S. A. & Schimmel, P. (1992). *Proc. Natl Acad. Sci. USA*, **89**, 65–69.
Martinis, S. A. & Schimmel, P. (1995). *tRNA: Structure, Biosynthesis and Function*, edited by D. Söll & U. L. RajBhandary, pp. 349–370. Washington, DC: American Society for Microbiology.
Mateja, A., Devedjiev, Y., Krowarsch, D., Longenecker, K., Dauter, Z., Otlewski, J. & Derewenda, Z. S. (2002). *Acta Cryst.* D**58**, 1983–1991.
Musier-Forsyth, K., Usman, N., Scaringe, S., Doudna, J., Green, R. & Schimmel, P. (1991). *Science*, **253**, 784–786.
Naganuma, M., Sekine, S., Fukunaga, R. & Yokoyama, S. (2009). *Proc. Natl Acad. Sci. USA*, **106**, 8489–8494.
Nureki, O., Niimi, T., Muramatsu, T., Kanno, H., Kohno, T., Florentz, C., Giegé, R. & Yokoyama, S. (1994). *J. Mol. Biol.* **236**, 710–724.
Park, S. J. & Schimmel, P. (1988). *J. Biol. Chem.* **263**, 16527–16530.
Perona, J. J., Swanson, R. N., Rould, M. A., Steitz, T. A. & Söll, D. (1989). *Science*, **246**, 1152–1154.
Price, W. N. II *et al.* (2009). *Nature Biotechnol.* **27**, 51–57.
Putney, S. D., Melendez, D. L. & Schimmel, P. R. (1981). *J. Biol. Chem.* **256**, 205–211.
Putney, S. D., Royal, N. J., Neuman de Vegvar, H., Herlihy, W. C., Biemann, K. & Schimmel, P. (1981). *Science*, **213**, 1497–1501.
Putney, S. D., Sauer, R. T. & Schimmel, P. R. (1981). *J. Biol. Chem.* **256**, 198–204.
Putney, S. D. & Schimmel, P. R. (1981). *Nature (London)*, **291**, 632–635.
Regan, L., Bowie, J. & Schimmel, P. (1987). *Science*, **235**, 1651–1653.
Saks, M. E. & Sampson, J. R. (1996). *EMBO J.* **15**, 2843–2849.
Schimmel, P. & Alexander, R. (1998). *Proc. Natl Acad. Sci. USA*, **95**, 10351–10353.
Schimmel, P. & Ribas de Pouplana, L. (1995). *Cell*, **81**, 983–986.
Sokabe, M., Ose, T., Nakamura, A., Tokunaga, K., Nureki, O., Yao, M. & Tanaka, I. (2009). *Proc. Natl Acad. Sci. USA*, **106**, 11028–11033.
Swairjo, M. A., Otero, F. J., Yang, X. L., Lovato, M. A., Skene, R. J., McRee, D. E., Ribas de Pouplana, L. & Schimmel, P. (2004). *Mol. Cell*, **13**, 829–841.
Swairjo, M. A. & Schimmel, P. R. (2005). *Proc. Natl Acad. Sci. USA*, **102**, 988–993.
Woese, C. R., Olsen, G. J., Ibba, M. & Söll, D. (2000). *Microbiol. Mol. Biol. Rev.* **64**, 202–236.
Yamada, H. *et al.* (2007). *Protein Sci.* **16**, 1389–1397.