

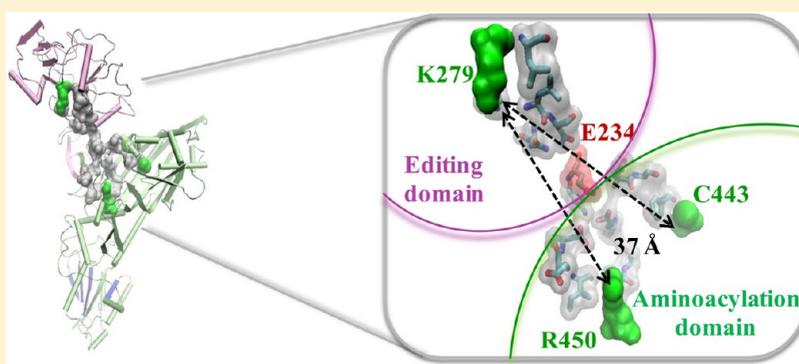
Multiple Pathways Promote Dynamical Coupling between Catalytic Domains in *Escherichia coli* Prolyl-tRNA Synthetase

James M. Johnson,[†] Brianne L. Sanford,[‡] Alexander M. Strom,[†] Stephanie N. Tadayon,[†] Brent P. Lehman,[†] Arrianna M. Zirbes,[†] Sudeep Bhattacharyya,^{*,†} Karin Musier-Forsyth,^{*,‡} and Sanchita Hati^{*,†}

[†]Department of Chemistry, University of Wisconsin–Eau Claire, Eau Claire, Wisconsin 54702, United States

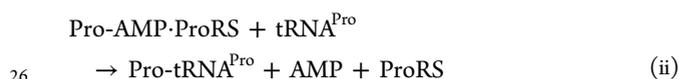
[‡]Department of Chemistry and Biochemistry, Center for RNA Biology, The Ohio State University, Columbus, Ohio 43210, United States

S Supporting Information



ABSTRACT: Aminoacyl-tRNA synthetases are multidomain enzymes that catalyze covalent attachment of amino acids to their cognate tRNA. Cross-talk between functional domains is a prerequisite for this process. In this study, we investigate the molecular mechanism of site-to-site communication in *Escherichia coli* prolyl-tRNA synthetase (Ec ProRS). Earlier studies have demonstrated that evolutionarily conserved and/or co-evolved residues that are engaged in correlated motion are critical for the propagation of functional conformational changes from one site to another in modular proteins. Here, molecular simulation and bioinformatics-based analysis were performed to identify dynamically coupled and evolutionarily constrained residues that form contiguous pathways of residue–residue interactions between the aminoacylation and editing domains of Ec ProRS. The results of this study suggest that multiple pathways exist between these two domains to maintain the dynamic coupling essential for enzyme function. Moreover, residues in these interaction networks are generally highly conserved. Site-directed changes of on-pathway residues have a significant impact on enzyme function and dynamics, suggesting that any perturbation along these pathways disrupts the native residue–residue interactions that are required for effective communication between the two functional domains. Free energy analysis revealed that communication between residues within a pathway and cross-talk between pathways are important for coordinating functions of different domains of Ec ProRS for efficient catalysis.

Class II prolyl-tRNA synthetases (ProRSs) catalyze covalent attachment of proline to tRNA^{Pro} in a two-step reaction:



ProRSs are modular proteins and are divided into two evolutionarily distinct groups based on sequence alignment and structural architecture.^{1,2} The “prokaryotic-like” ProRSs contain an insertion domain (INS) between motifs 2 and 3 of the catalytic domain, whereas “eukaryotic-like” ProRSs have C- and/or N-terminal extension domains.³ ProRSs from all three kingdoms of life have been shown to misactivate noncognate alanine and cysteine, resulting in mischarged tRNA^{Pro}.^{4,5} Many

aminoacyl-tRNA synthetases (AARSs) have evolved proof-reading capabilities to correct mistakes in amino acid activation (pretransfer editing) and charging (post-transfer editing).⁶ It has been shown that a domain (INS) of approximately 180 residues inserted within the catalytic core of prokaryotic-like ProRSs (Figure 1) is a post-transfer editing active site that hydrolyzes specifically mischarged Ala-tRNA^{Pro}.^{4,5,7,8} In contrast, Cys-tRNA^{Pro} is hydrolyzed by an INS homologue known as YbaK, which is encoded as a single-domain protein by many bacteria.^{9,10} Unlike prokaryotic-like ProRSs, most eukaryotic-like ProRSs do not possess the INS domain but in some cases encode free-standing editing-domain homologues. Some lower-

Received: January 21, 2013

Revised: May 13, 2013

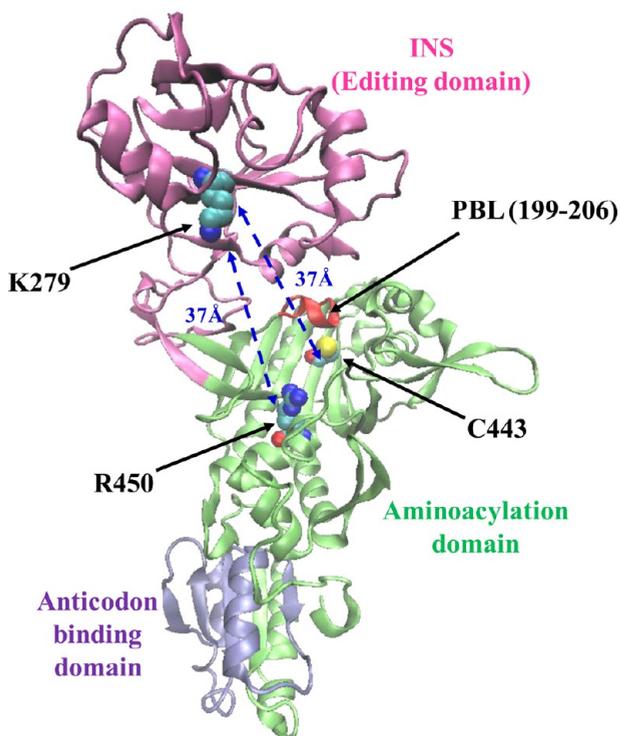


Figure 1. Ribbon representation of the three-dimensional structural model of the monomeric form of Ec ProRS. The homology model was derived from the X-ray crystal structure of Ef ProRS.¹⁵ The C_{α} - C_{α} distances between the starting residues (C443 and R450) and the end residue (K279) are shown.

47 eukaryotic ProRSs encode an N-terminal domain that displays a
48 low degree of homology with bacterial INS. This domain
49 exhibits post-transfer editing activity against Ala-tRNA^{Pro} in
50 *Plasmodium falciparum*¹¹ but is a defunct editing domain in
51 *Saccharomyces cerevisiae*.¹²

52 Bacterial ProRSs are modular enzymes, and efficient catalysis
53 and editing requires effective communication between distant
54 domains. *Escherichia coli* (Ec) ProRS, a representative member
55 of the prokaryotic-like group, contains three distinct domains
56 (Figure 1). The aminoacylation domain (motifs 1–3, consisting
57 of residues 64–81, 128–164, and 435–465, respectively)
58 catalyzes the activation of proline and the aminoacylation of
59 tRNA^{Pro}, as well as pretransfer editing;¹³ the anticodon binding
60 domain (residues 506–570) is critical for recognition of
61 cognate tRNA, and the INS (residues 224–407, located
62 between motifs 2 and 3 of the catalytic domain) is the post-
63 transfer editing active site.^{7,8} The aminoacylation domain and
64 the INS of Ec ProRS have been observed to depend on each
65 other in terms of their individual catalytic activities. For
66 example, mutation of a highly conserved aspartate (D350) in
67 INS to alanine resulted in reduced aminoacylation activity.⁷
68 Previous studies have also indicated that covalent connectivity
69 between domains is a prerequisite for efficient aminoacylation
70 and editing functions by ProRSs and other AARS systems.¹⁴
71 Deletion of the INS of Ec ProRS (Δ INS) has a severe impact
72 on catalysis; the amino acid activation efficiency of the deletion
73 variant was reduced 1200-fold.¹³ The addition of a separately
74 cloned and purified INS in *trans* failed to stimulate the amino
75 acid activation efficiency of the Δ INS construct (S. Hati and K.
76 Musier-Forsyth, unpublished data). The requirement of
77 covalent connectivity between domains for efficient function

of Ec ProRS suggests the existence of interdomain
78 communication in this enzyme. Moreover, structural studies
79 revealed that the catalytically important proline-binding loop
80 (PBL) of bacterial ProRSs undergoes a conformational change
81 from the “open” to “closed” state upon substrate binding.¹⁵ A
82 recent study showed that any perturbation in the surrounding
83 structural elements has a significant impact on PBL dynamics.¹⁶
84 The pathway by which substrate-induced conformational
85 change propagates from the activation center to the distant
86 protein segments that modulate PBL dynamics and conforma-
87 tional change is unknown.
88

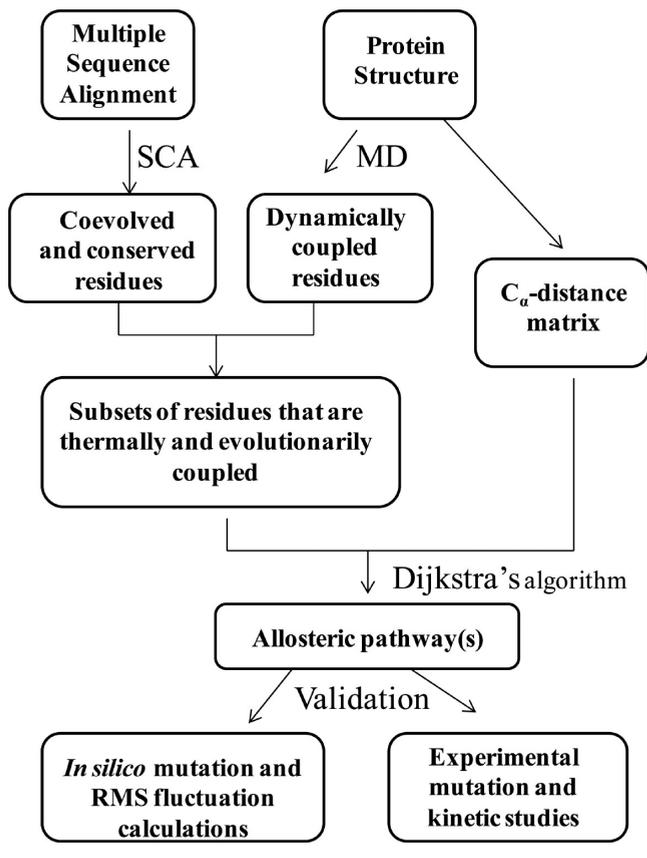
For multidomain proteins like Ec ProRS, domain–domain
89 communication is achieved by coupled-domain dynamics.^{17–20}
90 Recent MD simulation results revealed that the INS of Ec
91 ProRS is engaged in coupled motion with structural elements of
92 the catalytic domain.¹⁶ The collective dynamics of the PBL is
93 altered by the deletion of the INS or point mutation at the
94 INS–aminoacylation-domain junction.¹⁶ To understand the
95 molecular mechanism by which different structural elements of
96 Ec ProRS coordinate their function, it is important to identify
97 networks defined by key residue–residue interactions that
98 promote coupled-domain dynamics in this enzyme.
99

In this study, a bioinformatics-based computational method,
100 statistical thermal coupling analysis (STCA), was employed to
101 trace pathways of site-to-site communication in Ec ProRS.
102 Previous application of this method revealed that dynamically
103 coupled and evolutionarily constrained residues are important
104 for maintaining coupled-domain dynamics in *Thermus*
105 *thermophilus* leucyl-tRNA synthetase (Tt LeuRS).²⁰ Here
106 STCA was used to identify the residue–residue interaction
107 networks between the INS and aminoacylation domains of Ec
108 ProRS. In addition, mutational and kinetic studies, as well as
109 thermal fluctuation analyses, were performed to validate the
110 predicted networks of residue–residue interactions. In
111 summary, this study demonstrates that a modular protein like
112 Ec ProRS employs multiple pathways of residue–residue
113 interactions to communicate between distant functional sites.
114 Moreover, networks of these pathways involve residues that are
115 evolutionarily constrained and engaged in correlated motion.
116

117 ■ MATERIALS AND METHODS

General Strategy. On the basis of reported experimental
118 and structural data, the starting and ending points of interaction
119 networks between the aminoacylation domain and the INS
120 were selected; these active sites are >30 Å apart (Figure 1).
121 The single cysteine residue (C443), which is important for amino
122 acid activation, was chosen as one of the starting points within
123 the amino acid activation site. The other starting residue is
124 R450, found to be important for substrate binding¹⁵ (Figure 1).
125 The end point was chosen to be K279 in the INS, which has
126 been shown to be required for the editing function⁸ and is 37 Å
127 from C443 and R450 (Figure 1). STCA was conducted in three
128 discrete steps (Scheme 1). First, statistical coupling analysis
129 (SCA)^{21,22} was performed to identify conserved and co-evolved
130 residues in the ProRS family. In a parallel study, the collective
131 motions of various domains of the protein were studied by
132 performing a long time scale molecular dynamics (MD)
133 simulation of Ec ProRS. In the second step, the evolutionary
134 dependence of the coupled-domain dynamics was explored. In
135 this step, results of the SCA and MD simulation were
136 computationally integrated (Scheme 1). This analysis identified
137 a subset of residues that are not only dynamically coupled but
138 also evolutionarily constrained. Next, distinct networks of
139

Scheme 1



The correlated motions between residue pairs of distant structural elements were studied by principal component analysis (PCA, also known as essential dynamics analysis) of collective motions,²⁹ as described previously.^{16,30} PCA is a procedure by which the low-frequency (high-amplitude) collective motions of a biomolecule, which are often more relevant for its functions, are extracted from a MD simulation trajectory.³¹ This method has been described in detail in our previous study with this enzyme.¹⁶ Briefly, a covariance matrix of the C_α atoms was generated using the simulated MD data set. The diagonalization of this covariance matrix produces eigenvectors and eigenvalues representing the direction and magnitude of the collective motion of the whole protein or protein segments of interest.

In this study, the last 20 ns of the MD simulation data was used to generate principal components of atomic (backbone C_α atoms) fluctuations using Carma.³² The first three principal components were used to perform PCA-based cluster analysis, which produced a new trajectory of conformations representing the predominant conformational fluctuations. Finally, these conformations were used to generate dynamic cross-correlation matrix C, in which the ijth element, C_{ij}, represents the cross-correlation coefficient between fluctuations of residues at sites i and j during the simulation:

$$C_{ij} = \frac{\langle (x_i - \langle x_i \rangle)(x_j - \langle x_j \rangle) \rangle}{\sigma_{x_i} \sigma_{x_j}} \quad (1)$$

The atomic (C_α) displacements of residues i and j are represented by x_i and x_j, respectively; the angular brackets represent ensemble averages, and σ_{x_i} and σ_{x_j} represent the standard deviations of these displacements.

The rms fluctuations of C_α atoms averaged over three replicate simulations were also obtained for the WT and mutants. In these calculations, the last 20 ns of MD simulation data was used, each comprising an ensemble of 200000 conformations. PCA was also conducted for the WT and mutant proteins to compare the effect of mutations on coupled dynamics. For each of these protein systems, an ensemble of 600000 conformations, obtained by combining three replicate trajectories, was used to perform PCA. The first three clusters representing the predominant conformational fluctuations were considered in this study. In addition, the collective dynamics of the catalytically important PBL of the WT and various mutants were also analyzed following the method described previously.^{16,30}

Statistical Coupling Analysis. SCA is based on the assumption that the coupling of two sites in a protein, whether for structural or functional reasons, should cause those two sites to co-evolve.^{21,22} SCA was conducted using the standard protocol, details of which have been published^{21,22,33} and are available at <http://www.hhmi.swmed.edu/Labs/rr/sca.html>. In this work, a multiple-sequence alignment of 492 protein sequences of the ProRS family was generated using BLAST.³⁴ Only ProRS sequences that are significantly identical in sequence (>67%) with Ec ProRS were included in this study. Conservation constant ΔC_i^{stat} and coupling constant ΔΔC_{ij}^{stat} were obtained using standard procedures described previously.²⁰

Statistical Thermal Coupling Analysis: Integration of Evolutionary and Dynamic Information. In our earlier study of Tt LeuRS, we observed that coupled-domain motions are facilitated by networks of thermally and evolutionarily

interacting residues between the two distant sites were identified from this subset of residues using Dijkstra's algorithm.²³ Finally, the role of these selected residues in site-to-site communication was probed experimentally by conducting site-directed mutagenesis and kinetic studies. In addition, *in silico* mutations were introduced, and their impact on protein dynamics was examined by comparing root-mean-square (rms) fluctuations of the WT and mutated variants.¹⁶

Proteins were visualized using VMD.²⁴ SCA was conducted using a MATLAB script obtained from the Ranganathan lab (http://systems.swmed.edu/rr_lab/sca.html). The MD plots and SCA plots were created and all data processing conducted using MATLAB R2006b (The MathWorks Inc., Natick, MA).

COMPUTATIONAL METHODS

Molecular Dynamics Simulations. MD simulations were conducted using the three-dimensional homology model of monomeric Ec ProRS (residues 1–567) (provided by S. Cusack). The model was generated using the crystal structure of *Enterococcus faecalis* ProRS,¹⁵ which is 48% identical to the Ec enzyme. Mutant proteins were built using the Mutator plugin of VMD.²⁴ Simulations were performed in water (TIP3P model²⁵) with substrate-free enzymes using the all-atom CHARMM22 force field²⁶ within the NAMD package.²⁷ All simulations were conducted with a 500 ps equilibration step followed by a 25 ns production MD run. The details of the MD simulation protocol were as described previously.¹⁶ To evaluate the statistical significance of the MD simulation analysis, three replicates were generated for each protein system, as described in the protein simulation studies by Roy and Laughton.²⁸

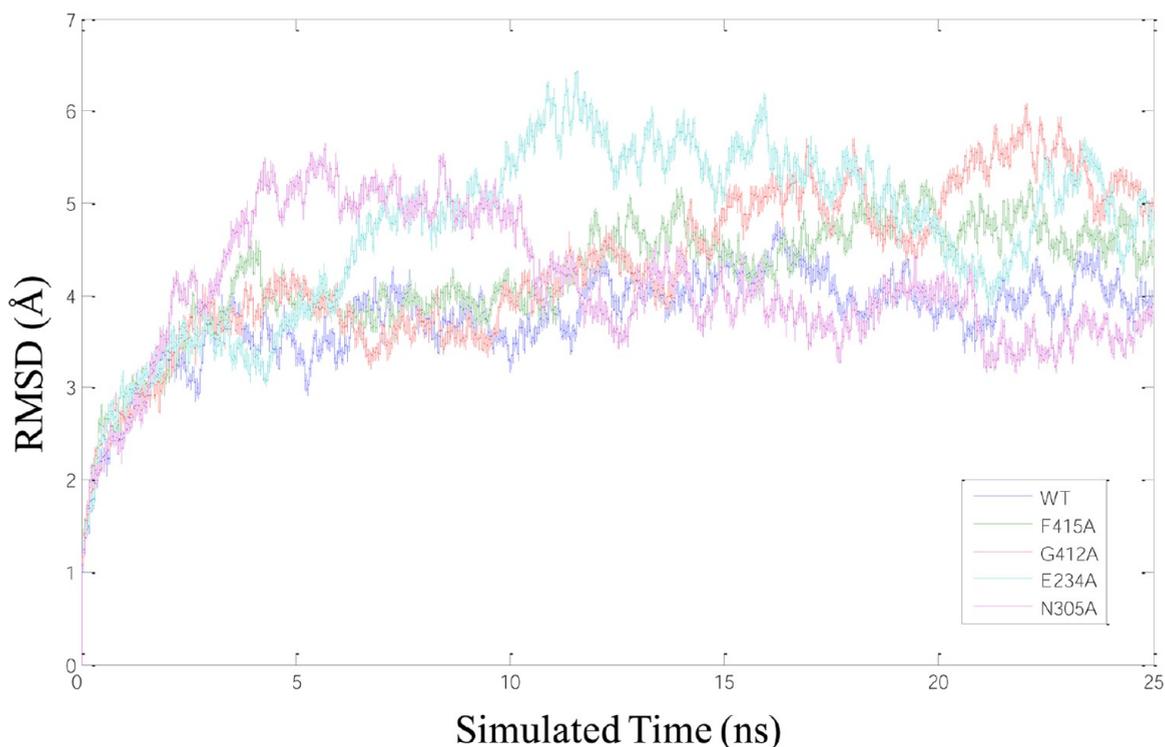


Figure 2. rmsd of the C_{α} atoms from their initial coordinate as a function of time for WT Ec ProRS and four variants. Calculations of rmsds were performed using 25 ns MD simulation data.

229 constrained residues.²⁰ To identify residue clusters that are
 230 important for maintaining coupled-domain dynamics, we
 231 extracted a subset of residues that are simultaneously coupled
 232 through evolution and dynamics (Scheme 1). The motional
 233 coupling information from MD was integrated with the
 234 evolutionary conservation and co-evolution data set obtained
 235 from the SCA study.

236 The conserved and co-evolved residues were treated
 237 separately. The conserved and dynamically coupled residues
 238 were chosen by selecting only those residues that exhibit
 239 significant conservation ($\Delta G_i^{\text{stat}} \geq 0.5$; ΔG_i^{stat} values range from
 240 0 to 1.0), as well as motional coupling [$C_{ij} \geq 0.8$; C_{ij} values
 241 range from -1.0 (anticorrelated motion) to 1.0 (correlated
 242 motion)] with each other. On the basis of recent studies of the
 243 significance of correlated motion in long-range communication,
 244 only positive correlations were considered.^{35,36} The value of C_{ij}
 245 was set to ≥ 0.8 to obtain residues that are engaged in strong
 246 correlated motions.

247 To select the co-evolved and dynamically coupled residues,
 248 the dynamic cross-correlation matrix (**C** matrix) was truncated
 249 by including only those columns that are present in the
 250 normalized SCA-derived **G** matrix. Next, the co-evolutionary
 251 dynamic coupling, **CDC**, matrix was created by multiplying
 252 each ij th element of the **G** matrix with the corresponding
 253 element of the truncated **C** matrix:

$$254 \quad \text{CDC}_{ij} = \Delta \Delta G_{ij}^{\text{stat}} \times C_{ij} \quad (2)$$

255 The **CDC** matrix therefore contains the covariance information
 256 of residue pairs that are co-evolved as well as dynamically
 257 coupled. In this study, the co-evolved and dynamically
 258 correlated residues are extracted by choosing only those
 259 residues for which $\text{CDC}_{ij} \geq 0.4$ ($\Delta \Delta G_{ij}^{\text{stat}} \geq 0.5$; $C_{ij} \geq 0.8$).

260 **Identification of Interaction Networks Using Dijkstra's**
 261 **Algorithm.** From the short-listed residues (Scheme 1),

residue–residue interaction networks between C443 or R450 262
 (aminoacylation domain) and K279 (INS) of Ec ProRS were 263
 identified using Dijkstra's algorithm.²³ In this method, each of 264
 the C_{α} atoms of the protein backbone represents a node. The 265
 connectivity between two adjacent nodes was described by a 266
 binary connection matrix **P** of inter-residue (C_{α} – C_{α}) contacts. 267
 The C_{α} – C_{α} distance matrix, **D**, was computed from the 268
 Cartesian coordinates of all C_{α} atoms of the protein. Based on a 269
 C_{α} – C_{α} cutoff distance D_{ij}^0 , P_{ij} is equal to 1 if $D_{ij} < D_{ij}^0$ and zero 270
 otherwise. The interaction networks (paths) between two 271
 functional sites were identified and listed in terms of a “cost”, 272
 which is equal to the sum of all C_{α} – C_{α} distances between 273
 adjacent residues in a given path. 274

275 ■ EXPERIMENTAL METHODS

Materials. All amino acids (Sigma) were of the highest 276
 quality (>99% pure) and used without any further purification. 277
 $[\gamma\text{-}^{32}\text{P}]\text{ATP}$ and $[\text{}^{32}\text{P}]\text{PP}_i$ were from Perkin-Elmer. Primers for 278
 site-directed mutagenesis and polymerase chain reaction were 279
 from Integrated DNA Technologies. 280

Enzyme Preparation. Overexpression and purification of 281
 histidine-tagged WT and mutant Ec ProRS were performed as 282
 described previously.^{37,38} Plasmids encoding D198A, E234A, 283
 H302A, N305A, G412A, F415A, H302A/G412A, N305A/ 284
 G412A, and E218A/N305A Ec ProRS were generated by site- 285
 directed mutagenesis of pCS-M1S³⁸ using primers listed in 286
 Table S1 of the Supporting Information. Results of mutagenesis 287
 were confirmed by DNA sequencing (Biotechnology Center, 288
 University of Wisconsin—Madison, Madison, WI). Protein 289
 expression was induced in Ec SG13009 (pREP4) competent 290
 cells with 0.1 mM isopropyl β -D-thiogalactoside for 4 h at 37 291
 $^{\circ}\text{C}$. Histidine-tagged proteins were purified using a Talon 292
 cobalt affinity resin, and the desired protein was eluted with 100 293
 mM imidazole. Protein concentrations were determined 294

295 initially by the Bio-Rad assay (Bio-Rad Laboratories) followed
296 by active site titration.³⁹

297 **RNA Preparation.** Ec tRNA^{Pro} was transcribed using T7
298 RNA polymerase from BstNI-linearized plasmid as described
299 previously⁴⁰ and purified by denaturing 12% polyacrylamide gel
300 electrophoresis.

301 **Enzyme Assays.** The ATP-PP_i exchange assay was
302 performed at 37 °C according to the published method.⁴¹
303 The concentrations of proline ranged from 0.05 to 2 mM. The
304 enzyme concentrations used were 10–40 nM for proline
305 activation. Kinetic parameters were determined from Line-
306 weaver-Burk plots and represent the average of at least three
307 determinations.

308 ATP hydrolysis reactions for monitoring pretransfer editing
309 were conducted as described previously¹³ using 500 mM
310 alanine and 4 μM ProRS.

311 Aminoacylation assays were performed at room temperature
312 in a reaction mixture containing 50 mM HEPES (pH 7.5), 20
313 mM KCl, 25 mM MgCl₂, 0.1 mg/mL bovine serum albumin, 20
314 mM 2-mercaptoethanol, and 4 mM ATP. Cognate amino-
315 acylation reaction mixtures also contained 23 μM [³H]proline,
316 0.5 μM tRNA^{Pro}, and 100 nM Ec ProRS. When charged
317 tRNA^{Pro} was prepared to be used in deacylation assays, the
318 same reaction mixture was used, and [¹⁴C]alanine was used to
319 acylate G1:C72/U70 tRNA^{Pro} (8 μM) by Ec alanyl-tRNA
320 synthetase (8 μM). After incubation for 1.5 h, 1% acetic acid
321 was used to quench the reaction, and the mischarged tRNA was
322 purified by repeated phenol/CHCl₃ extraction (5:1 solution,
323 pH 4.5), followed by ethanol precipitation.

324 Deacylation assays were conducted at room temperature in
325 reaction mixtures containing 50 mM HEPES (pH 7.5), 5 mM
326 MgCl₂, and 1 μM G1:C72/U70 [¹⁴C]Ala-tRNA^{Pro} and initiated
327 with 0.5 μM Ec ProRS. A buffer-only background curve was
328 also performed and subtracted from each curve.

329 ■ RESULTS

330 **MD Simulation.** For the substrate-free WT ProRS, the
331 root-mean-square-deviation (rmsd) of each frame of the 25 ns
332 MD simulation trajectory was computed relative to the initial
333 coordinates (Figure 2). The rmsd values of these frames
334 fluctuated with a mean value of 1.5–2.0 Å during the last 20 ns
335 of the simulations, and similar fluctuations in rmsd values were
336 obtained for all three replicate simulations. These variations are
337 consistent with previous studies of substrate-unbound
338 AARSs.^{42–44}

339 **Dynamic Cross-Correlation Analysis.** The dynamic
340 cross-correlation matrix (C) for the WT enzyme (Figure S1
341 of the Supporting Information) was generated using the first
342 three principal components. Analysis of the cross-correlation of
343 fluctuations of residues revealed both inter- and intradomain
344 dynamic correlation. The aminoacylation domain and the INS
345 are mainly engaged in anticorrelated motions; i.e., their
346 displacements are in opposite directions ($C_{ij} < 0$). However,
347 various structural elements within the aminoacylation domain
348 and the INS are engaged mainly in correlated motion ($C_{ij} > 0$).
349 The C matrix (Figure S1 of the Supporting Information) was
350 used to extract residues that are engaged in strong correlated
351 motion ($C_{ij} \geq 0.8$).

352 **Conserved and Co-Evolved Residues.** Earlier, we
353 showed that evolutionarily constrained residues are important
354 for maintaining coupled-domain dynamics in Tt LeuRS.²⁰ To
355 identify the conserved residues responsible for coupled-domain
356 dynamics between aminoacylation and editing active sites of Ec

ProRS, the ΔG_i^{stat} value of each residue, a quantitative measure
357 of the conservation of a residue at the *i*th position of the
358 sequence, was calculated using SCA.^{21,22,33} The ΔG_i^{stat} values
359 were obtained as a normalized one-dimensional vector (Figure
360 S2 of the Supporting Information). The number of conserved
361 residues in Ec ProRS obtained at different ΔG_i^{stat} cutoffs is
362 reported in Table 1. In addition, a 567 × 152 coupling matrix
363

Table 1. Numbers of Conserved and Co-Evolved Residues within Ec ProRS That Are Engaged in Correlated Motion^a

ΔG_i^{stat}	≥ 0.50	≥ 0.55	≥ 0.60	≥ 0.65	≥ 0.70
no. of conserved residues	207	182	151	104	61
CDC _{ij}	≥ 0.40	≥ 0.45	≥ 0.50	≥ 0.55	≥ 0.60
no. of co-evolved residues	96	69	48	28	18

^aThese residue numbers were determined by varying cutoff values for ΔG_i^{stat} (evolutionary conservation constant) and CDC_{ij} ($\text{CDC}_{ij} = \Delta \Delta G_{ij}^{\text{stat}} \times C_{ij}$; co-evolutionary dynamic coupling constant) while maintaining a correlation coefficient of residue–residue fluctuations (C_{ij}) of ≥ 0.8 .

364 containing the evolutionary coupling indices ($\Delta \Delta G_{ij}^{\text{stat}}$) of the
365 567 total residues for 152 perturbation sites^{21,22} was obtained
366 from SCA (Figure 3). It is evident that only a small fraction of
367 Ec ProRS residues have high $\Delta \Delta G_{ij}^{\text{stat}}$ values. The multiplication
368 of this 567 × 152 SCA matrix with the truncated 567 × 152
369 MD matrix yields the CDC matrix (eq 2). The numbers of
370 evolutionarily and dynamically coupled residues determined for
371 various CDC_{ij} cutoffs are listed in Table 1. These pools of
372 conserved and co-evolved residues were further used to identify
373 long-range interaction networks described below.

374 **Interaction Networks across Domains.** To map the
375 interaction networks between the aminoacylation domain
376 (C443/R450) and the INS (K279), Dijkstra's algorithm²³
377 was applied to the pool of dynamically and evolutionarily
378 (conserved and/or co-evolved) coupled residues (Scheme 1),
379 as described in the methods sections. Because the average van
380 der Waals radius of an amino acid is ~3 Å and the acceptable
381 distance for the noncovalent interactions is 2.0–3.0 Å, a strong
382 noncovalent interaction will be prevalent when two C_α atoms
383 are separated by a distance of 8.0–9.0 Å. Therefore, in this
384 study, distance cutoff D_{ij}^0 for the potential interactions between
385 the neighboring C_α atoms was varied between 8.0 and 9.0 Å.

386 Several residue–residue interaction networks were identified
387 between the aminoacylation and INS domains. These
388 contiguous interaction networks were obtained by using various
389 cutoff values for ΔG_i^{stat} , C_{ij} , and D_{ij}^0 as listed in Table 2 while
390 maintaining a CDC_{ij} value of ≥ 0.4 ($\text{CDC}_{ij} = \Delta \Delta G_{ij}^{\text{stat}} \times C_{ij}$;
391 $\Delta \Delta G_{ij}^{\text{stat}} \geq 0.5$; $C_{ij} \geq 0.8$). The four probable contiguous
392 pathways (paths I–IV), identified on the basis of the distance
393 between the starting and ending residues (cost), as well as the
394 degree of dynamic correlation between networking residues
395 (average correlation coefficient value), are listed in Table 2 and
396 shown in Figure 4. These predicted paths (interaction
397 networks) are dominated by polar residues. Moreover, close
398 scrutiny of residues in the predicted pathways revealed that
399 conserved residues are dominant over co-evolved residues.

400 To evaluate the proposed communication pathways, we
401 conducted mutational studies on some of the conserved
402 residues (Table 2, underlined) that are present in more than
403 one of the predicted pathways. The impact of altering pathway
404 residues on amino acid activation, tRNA aminoacylation, and
405 editing reactions was investigated.

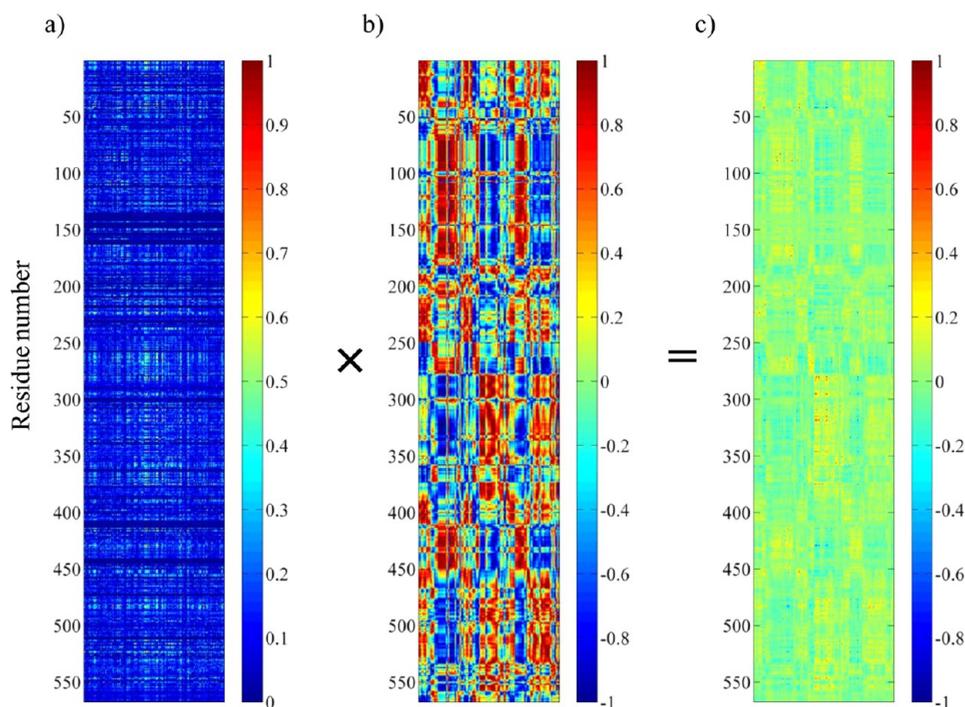


Figure 3. Generation of the co-evolutionary dynamic coupling (CDC) matrix from the SCA coupling matrix and truncated dynamic cross-correlation matrix. (a) Unclustered SCA coupling matrix from SCA of the ProRS family. The X-axis represents the 152 perturbation sites, while the Y-axis corresponds to residues 1–567 of Ec ProRS. The color gradient, as indicated in the color bar, is as follows: blue squares for the lowest and red squares for the highest statistical coupling energies, $\Delta\Delta G_{ij}^{\text{stat}}$. (b) Truncated MD cross-correlation matrix generated by taking only those columns of perturbation sites (residues) that are present in the SCA matrix. The cross-correlation values range from 1 (correlated) to -1 (anticorrelated). (c) The SCA–MD plot obtained as the CDC matrix by multiplying individual elements of the SCA matrix with the corresponding elements of the truncated MD matrix. Values range from 1.0 (co-evolved and thermally correlated) to -1.0 (co-evolved and thermally anticorrelated).

Table 2. Probable Pathways of Communication between the Aminoacylation and INS Domains in Ec ProRS As Identified Using STCA^a

pathway	parameters	residue networks	cost (Å)
C443 → K279			
I (8 residues)	$C_{ij} \geq 0.8$ $\Delta G_i^{\text{stat}} = 0.5$ $\text{CDC}_{ij} = 0.4$ $D_{ij}^{\circ} = 8.0$	C443 → <u>F415</u> → T199 → <u>D198</u> → <u>E234</u> → N389 → <u>H302</u> → K279	43.9 (0.62, 0.51, 60)
II (9 residues)	$C_{ij} \geq 0.8$ $\Delta G_i^{\text{stat}} = 0.6$ $\text{CDC}_{ij} = 0.4$ $D_{ij}^{\circ} = 8.0$	C443 → <u>F415</u> → T199 → <u>D198</u> → <u>E234</u> → <u>N305</u> → L304 → L281 → K279	46.6 (0.66, 0.59, 0.62)
R450 → K279			
III (10 residues)	$C_{ij} \geq 0.8$ $\Delta G_i^{\text{stat}} = 0.6$ $\text{CDC}_{ij} = 0.4$ $D_{ij}^{\circ} = 8.0$	R450 → V411 → <u>G412</u> → E209 → <u>D198</u> → <u>E234</u> → <u>N305</u> → L304 → L281 → K279	51.5 (0.47, 0.60, 0.63)
IV (10 residues)	$C_{ij} \geq 0.8$ $\Delta G_i^{\text{stat}} = 0.5$ $\text{CDC}_{ij} = 0.4$ $D_{ij}^{\circ} = 8.0$	R450 → I409 → <u>E218</u> → D219 → N232 → <u>E234</u> → <u>N305</u> → L304 → L281 → K279	53.3 (0.74, 0.65, 0.64)

^aAbbreviations: C_{ij} , correlation coefficient of residue–residue fluctuations; ΔG_i^{stat} , evolutionary conservation constant; $\Delta\Delta G_{ij}^{\text{stat}}$, evolutionary coupling constant; $\text{CDC}_{ij} = \Delta\Delta G_{ij}^{\text{stat}} \times C_{ij}$, co-evolutionary dynamic coupling constant; D_{ij}° , distance cutoff. Numbers in parentheses in column 4 represent average C_{ij} values between adjacent residues in a given path obtained from dynamic cross-correlation matrix **C** for the three replica simulations. Residues shown in bold represent the terminal residues. All residues except those shown in italics are conserved residues; residues shown in italics represent the co-evolved residues. The underlined residues were chosen for mutational studies.

406 **Amino Acid Activation.** First, the WT enzyme and six
407 single mutants [F415A, G412A, and D198A (aminoacylation
408 domain); N305A, H302A, and E234A (INS)] were examined

for their role in amino acid activation. Using the ATP–PP_i exchange
409 reaction, we found that mutation of some of these 410
411 residues to alanine had a considerable impact on proline 412

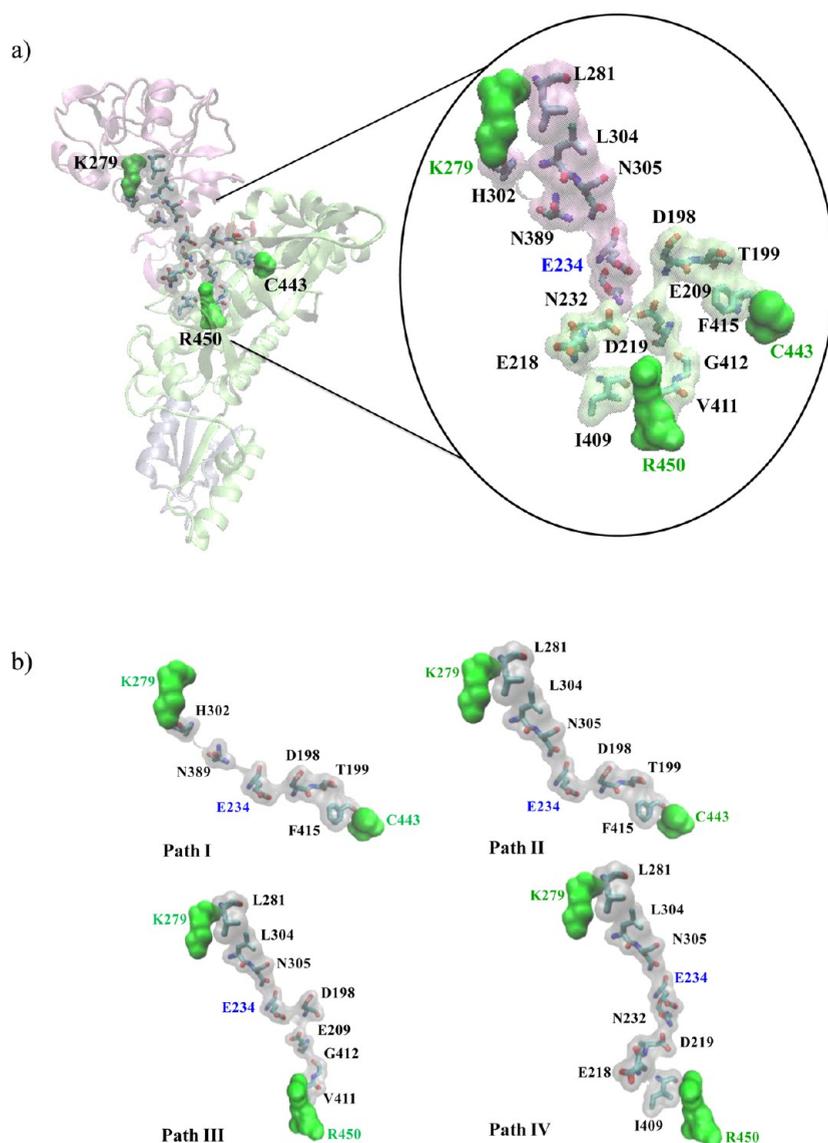


Figure 4. (a) Representation of residue–residue interaction networks between the substrate binding domain (R450/C443) and the INS (K279) identified in this study. Terminal residues in the pathways are shown in green space-filling surface representation. (b) Paths I–IV are shown in space-filling surface representation.

Table 3. Kinetic Parameters for Amino Acid Activation by WT and Mutant Variants of Ec ProRS^a

Ec ProRS	k_{cat} (s ⁻¹)	K_{M} (mM)	$k_{\text{cat}}/K_{\text{M}}$ (mM ⁻¹ s ⁻¹)	$k_{\text{cat}}/K_{\text{M}}$ (relative)	fold decrease	$\Delta\Delta G$ (kcal/mol)
WT	12.6 ± 4.9	0.18 ± 0.03	71	1.0	–	–
D198A	6.98 ± 0.37	0.33 ± 0.01	21	0.30	3.4	0.75
E218A ^b	4.4 ± 2.3	3.40 ± 0.68	1.3	0.02	55	2.5
E234A	6.7 ± 1.9	1.03 ± 0.25	6.5	0.09	11	1.5
H302A	7.3 ± 1.9	0.22 ± 0.04	33	0.46	2.1	0.46
N305A	0.61 ± 0.18	0.45 ± 0.18	1.4	2.0 × 10 ⁻²	51	2.4
G412A	12.8 ± 0.57	0.300 ± 0.003	43.0	0.61	1.6	0.29
F415A	0.131 ± 0.010	0.76 ± 0.29	0.17	2.4 × 10 ⁻³	400	3.7
H302A/G412A	10.7 ± 0.80	0.62 ± 0.26	17	0.24	4.2	0.88
N305A/G412A	ND	ND	ND	–	–	–
E218A/N305A	ND	ND	ND	–	–	–

^aResults are the average of three trials with the standard deviation indicated. $\Delta\Delta G$ was calculated according to the equation $\Delta\Delta G = -RT \ln(\text{fold decrease in } k_{\text{cat}}/K_{\text{M}})$, where R is the gas constant, 1.986 cal K⁻¹ mol⁻¹, and T is 310 K. ND indicates not detectable under the experimental conditions used. ^bData from ref 16.

412 activation (Table 3). Alanine substitution of F415 of the
413 catalytic domain, which undergoes a conformational change

upon prolyl-adenylate binding,¹⁵ resulted in an ~400-fold
decrease in $k_{\text{cat}}/K_{\text{M}}$. Mutation of residues in the INS domain 415

416 also had a significant impact on amino acid activation. For
 417 example, significantly lower levels of amino acid activation
 418 efficiency were observed for N305A (~50-fold reduced) and
 419 E234A (~10-fold reduced) variants. To examine if residues in
 420 the predicted pathways are indeed coupled, amino acid
 421 activation efficiencies of double mutants were also evaluated.
 422 The H302A/G412A variant, obtained by alanine substitution of
 423 residues that belong to the two separate pathways (H302, path
 424 I, and G412, path III), exhibited a >4-fold decreased proline
 425 activation efficiency (Table 3); the H302A and G412A single
 426 mutants exhibited a 2.1- and 1.6-fold reductions in amino acid
 427 activation efficiency, respectively. The free energy analysis
 428 [$\Delta\Delta G = -RT \ln(\text{fold decrease in } k_{\text{cat}}/K_M)$, where R is the gas
 429 constant, $1.986 \text{ cal K}^{-1} \text{ mol}^{-1}$, and T is 310 K]⁴² showed that
 430 the free energy change of the double mutant (0.88 kcal/mol) is
 431 somewhat greater than the sum of the free energy change of the
 432 H302A (0.46 kcal/mol) and G412A (0.29 kcal/mol) mutants
 433 (Table 3). This difference in free energy change implies weak
 434 coupling between H302 and G412.⁴⁵ Proline activation was not
 435 detected for the other two double mutants tested, N305A/
 436 G412A (path III) and E218A/N305A (path IV). The lack of
 437 activity of the path III variant suggests strong coupling between
 438 N305A and G412A because the effect of the double mutant
 439 ($\Delta\Delta G > 3.7$ based on the level of detection of this assay) is
 440 greater than expected on the basis of the single mutant effects
 441 (2.4 and 0.29 kcal/mol for N305A and G412A, respectively). In
 442 a recent study, we showed that the proline activation efficiency
 443 was reduced by ~50-fold in the E218A variant.¹⁶ Thus, the lack
 444 of activity of the path IV double mutants is not surprising. No
 445 conclusions can be drawn regarding energetic coupling between
 446 E218A and N305A because the combined $\Delta\Delta G$ (5 kcal/mol)
 447 for the two single mutants is beyond the detection limit of this
 448 assay and no activity was detected for the double mutant.

449 **Aminoacylation.** The aminoacylation activity of single and
 450 double mutants was examined by steady-state kinetic assays
 451 under conditions where synthesis of Pro-tRNA^{Pro} was linear
 452 with time and the initial rate of reaction was proportional to
 453 k_{cat}/K_M . The aminoacylation activity of F415A and N305A
 454 variants was nearly abolished with rates 70-fold slower than the
 455 WT rate (Figure 5a). The overall aminoacylation activity was
 456 also slightly reduced for the H302A and E234A variants (2-
 457 fold) and the D198A (5.5-fold) variant (Figure 5a,b).
 458 Interestingly, mutation of G412 to alanine resulted in
 459 significantly weakened aminoacylation capability for both the
 460 single (G412A, 7-fold) and double (H302A/G412A, 5.5-fold)
 461 mutants (Figure 5b), although an only ~2–4-fold reduction in
 462 proline activation efficiency was observed for these variants,
 463 which were weakly coupled (Table 3). When $\Delta\Delta G$ is calculated
 464 for the aminoacylation reaction (Table 4), strong coupling is
 465 observed between H302 and G412 because the $\Delta\Delta G$ for
 466 double mutant H302A/G412A (1.0 kcal/mol) is significantly
 467 smaller (subadditive) than the sum of the two single mutants
 468 H302A and G412A ($\Delta\Delta G = 0.4 \text{ kcal/mol} + 1.2 \text{ kcal/mol} = 1.6$
 469 kcal/mol). Aminoacylation activity was not detected for double
 470 mutants N305A/G412A and E218A/N305A (Figure 5). On
 471 the basis of the detection limit of the aminoacylation assays, no
 472 conclusions can be drawn regarding the coupling of these
 473 residues. Taken together, these observations suggest that
 474 mutation of putative pathway residues has distinct impacts on
 475 the two steps of the tRNA aminoacylation reaction.

476 **Pretransfer Editing.** The pretransfer editing reaction was
 477 studied by monitoring enhanced hydrolysis of ATP in the
 478 presence of alanine¹³ relative to that observed in the absence of

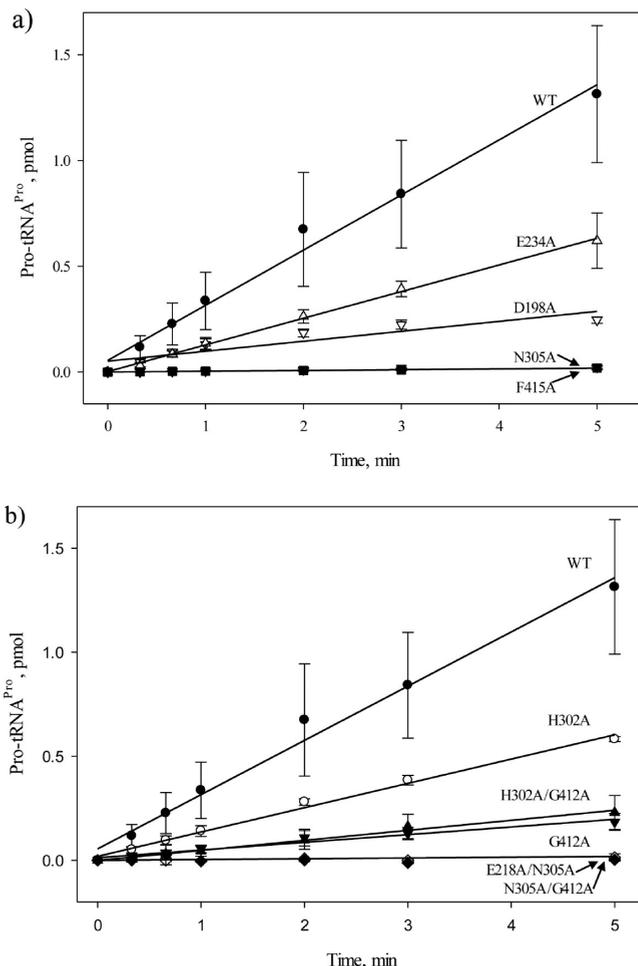


Figure 5. Initial rates of aminoacylation of tRNA^{Pro} with proline by WT and single- and double-point variants of Ec ProRS. For the sake of clarity, the results are presented in two panels (a and b). The assays were performed at room temperature with 0.5 μM tRNA^{Pro} and 100 nM Ec ProRS. Linear fits of the data are shown.

Table 4. Kinetic Parameters for Aminoacylation by WT and Mutant Variants of Ec ProRS^a

Ec ProRS	k_{cat}/K_M (relative)	fold decrease	$\Delta\Delta G$ (kcal/mol)
WT	1	1	–
D198A	0.18	5.5	1.0
E218A ^b	0.71	1.4	0.20
E234A	0.48	2.1	0.43
H302A	0.45	2.2	0.46
N305A	0.014	70	2.5
G412A	0.14	7.1	1.2
F415A	0.014	70	2.5
H302A/G412A	0.18	5.5	1.0
N305A/G412A	ND	–	–
E218A/N305A	ND	–	–

^aAminoacylation assays were performed under conditions where the initial rates were proportional to RNA concentrations. This indicated that $V_0/[S]$ was an accurate reflection of k_{cat}/K_M . The relative k_{cat}/K_M values were normalized and set to 1 for WT. $\Delta\Delta G$ was calculated according to the equation $\Delta\Delta G = -RT \ln(\text{fold decrease in } k_{\text{cat}}/K_M)$, where R is the gas constant, $1.986 \text{ cal K}^{-1} \text{ mol}^{-1}$, and T is 298 K. ND indicates not detectable under the experimental conditions used. ^bData from ref 16.

479 the amino acid. Reduced pretransfer editing activity was
 480 observed for all mutants except G412A, which demonstrated
 481 levels of pretransfer editing similar to that of WT ProRS
 482 (Figure 6). This result may be due to the variant's ability to
 483 activate alanine slightly more efficiently than the WT enzyme
 484 (data not shown).

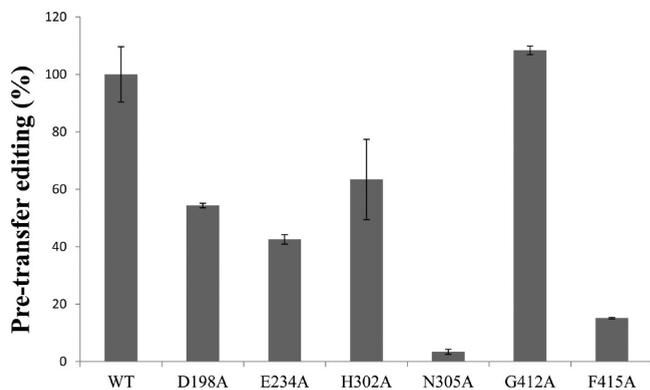


Figure 6. Relative alanine pretransfer editing activity of WT and mutant variants of Ec ProRS. The assay was performed at 37 °C using 4 μM ProRS and 500 mM alanine. Results are reported as percent activity relative to WT, which was set to 100%.

485 **Post-Transfer Editing.** The impact of mutations on Ala-
 486 tRNA^{Pro} hydrolysis was also studied (Figure 7). A significant

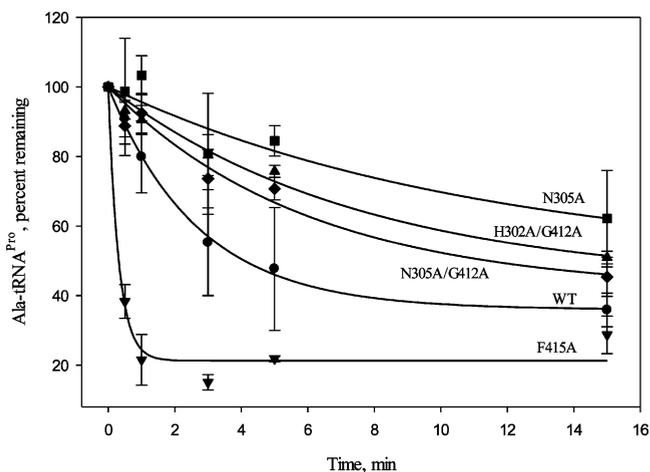


Figure 7. Deacylation of Ala-tRNA^{Pro} by WT and mutant variants of Ec ProRS. The assays were performed at room temperature with 1 μM G1:C72/U70 [¹⁴C]Ala-tRNA^{Pro} and 0.5 μM Ec ProRS.

487 reduction in post-transfer activity was observed for the N305A
 488 mutant (~4-fold reduced relative to that of WT). This residue
 489 resides at the interface between the INS and activation
 490 domains, far from the editing active site. Most of the other
 491 mutant ProRS variants tested (D198A, E234A, H302A, G412A,
 492 and E218A/N305A) displayed editing activity that was similar
 493 to that of WT ProRS (data not shown). The H302A/G412A
 494 and N305A/G412A double mutants exhibited small (~3- and
 495 ~2-fold, respectively) decreases in activity. Interestingly, the
 496 F415A mutant, which was severely defective in proline
 497 activation and aminoacylation, hydrolyzed mischarged Ala-
 498 tRNA^{Pro} at a rate nearly 9-fold greater than that of the WT
 499 enzyme (Figure 7).

Protein Flexibility and PBL Dynamics. To examine if the
 500 alteration in enzyme activity observed for the ProRS variants
 501 correlates with a change in protein dynamics caused by
 502 mutation of on-pathway residues, we analyzed the backbone
 503 flexibility and collective dynamics of these systems. This was
 504 accomplished by using MD simulations of the four selective
 505 substrate-free mutants, namely, E234A, N305A, G412A, and
 506 F415A, for which significant changes in catalysis were observed.
 507 In each case, three replicates of 25 ns MD simulations were
 508 conducted. To assess if each system has reached an equilibrated
 509 state, the rmsd of each frame of the MD simulation trajectory
 510 was computed from their respective initial coordinates for all
 511 four mutants (Figure 2). The rmsd values were observed to
 512 fluctuate within 2.0 Å during the production period (final 20
 513 ns) of the simulations, indicating stability for each system. Also,
 514 the quality of the simulations was assessed by computing the
 515 rms fluctuation of each amino acid from the time-averaged
 516 structure using the last 20 ns of MD simulation data. The rms
 517 fluctuations for each replica, as well as the replica-averaged
 518 fluctuations of the WT and the four ProRS variants, are
 519 reported in Figure 8. Analysis of these rms fluctuation data

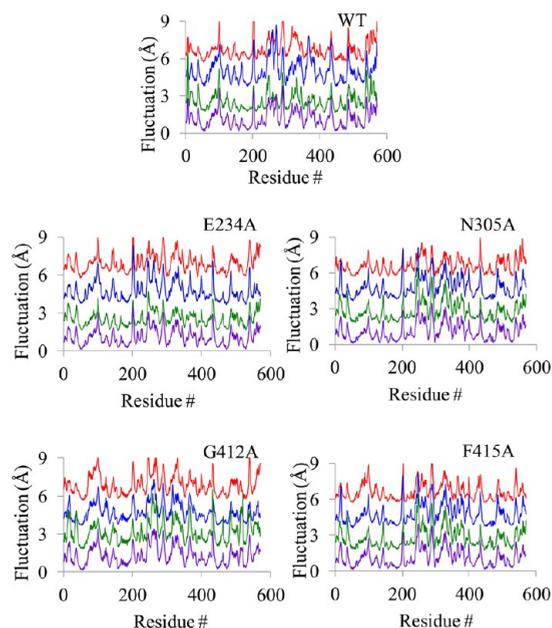


Figure 8. Root-mean-square fluctuations of individual amino acids for WT ProRS and four variants. In each stacked plot, the rms fluctuations of C_α atoms calculated from the time-averaged structures over the last 20 ns of MD trajectories are shown. Fluctuations for the three replicates are separated by 2 Å and color-coded for the sake of clarity: red for replica 1, blue for replica 2, and green for replica 3. In each case, the bottom plot (purple) represents the replica-averaged rms fluctuations. The calculated propagated uncertainties are 0.3 Å for WT, 0.2 Å for E234A, 0.2 Å for N305A, 0.3 Å for G412A, and 0.3 Å for F415A.

indicates that the backbone flexibilities are quite reproducible
 521 for each of these protein systems, with only a propagated
 522 uncertainty of 0.2–0.3 Å for the three replica simulations
 523 (Figure 8). These results indicate that all simulations have
 524 reached equilibrated states.

Next, we probed the impact of these mutations on the
 526 collective dynamics. For each system, PCA was conducted
 527 using the combined 60 ns trajectory (last 20 ns of three
 528 replicates) and the first three clusters representing the
 529 predominant collective dynamics were extracted. The rms 530

531 fluctuations of C_{α} atoms were computed from their respective
 532 average structures, normalized, and averaged over the three
 533 clusters. The impact of mutation of on-pathway residues on the
 534 protein flexibility was examined by computing the difference of
 535 these cluster-averaged rms fluctuations between the WT and a
 536 specific mutant (Figure 9). Analysis of the results indicates that

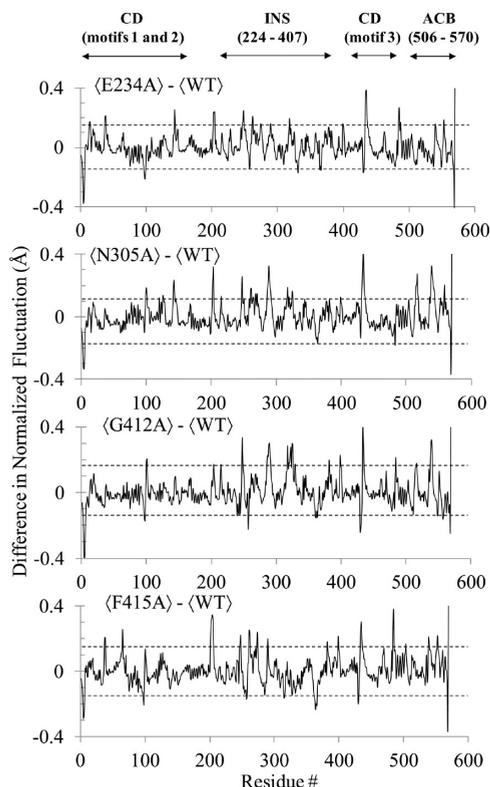


Figure 9. Changes in the normalized rms fluctuations of C_{α} atoms observed in the collective dynamics of the four mutants with respect to WT Ec ProRS. The angular bracket indicates that the rms fluctuations are averaged over the first three clusters representing the predominant collective protein motions. The propagated uncertainties for each of these plots are within 0.15 Å and are shown with two parallel dotted lines. Abbreviations: CD, catalytic (aminoacylation) domain; INS, insertion domain; ACB, anticodon binding domain.

537 the flexibility of the protein backbone was altered by varied
 538 extents because of the alanine substitutions at these sites. The
 539 overall flexibility of the catalytic-domain residues, including the
 540 PBL, was found to be considerably impacted by the alanine
 541 substitution of N305 (Figure 9), which is located in the editing
 542 domain and has been observed to have significantly reduced
 543 proline activation and aminoacylation activities. The fluctua-
 544 tions of the editing-domain residues were also impacted in all
 545 four mutants, with the most significant changes observed for
 546 N305A, G412A, and F415A. A noticeable alteration in editing
 547 activity was also observed for these mutants, indicating that
 548 coupled thermal motions are important for Ec ProRS function.

549 The PBL is critical for substrate binding and catalysis.
 550 Therefore, to examine the impact of these point mutations on
 551 the collective dynamics of the PBL, we conducted PCA of WT
 552 Ec ProRS and the same four Ec ProRS variants (F415A,
 553 G412A, E234A, and N305A) using the last 20 ns of MD
 554 simulation data for the C_{α} atoms of the PBL. The root-mean-
 555 square projections (rmsps)¹⁶ of the PBL for the WT and
 556 mutant proteins are shown in Figure 10. In this study, the rmsp

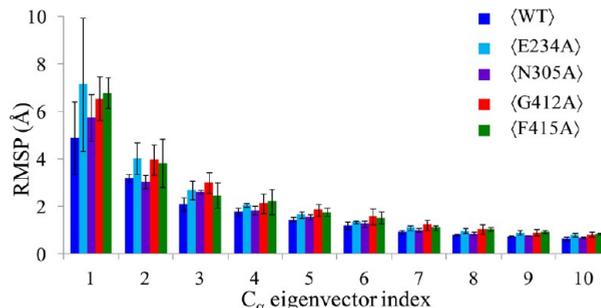


Figure 10. Analysis of root-mean-square projections (rmsp) of the PBL C_{α} atoms for WT ProRS and the four variant proteins. The angular bracket indicates that the rmsp values are averaged over the three replica simulations. These average rmsps and standard deviations for eigenvectors 1–10 are shown.

(averaged over three replicate simulations) represents the
 557 collective displacement of the loop along principal components
 558 1–10 (i.e., PC1–PC10, respectively), which are C_{α} eigenvec-
 559 tors. The PCA shows that each mutation had an impact on the
 560 collective dynamics of the PBL. The displacement of the PBL
 561 along PC1 is significantly altered for all the mutant proteins
 562 compared to the WT enzyme. However, the error in PBL
 563 displacement along PC1 for E234A is large, making it difficult
 564 to ascertain the precise effect of this mutation on the PBL
 565 dynamics, and further studies are required. Noticeable changes
 566 in PBL displacement were also observed for PC2 and PC3.
 567 Therefore, this analysis indicates that mutation of the pathway
 568 residues could impact PBL dynamics and potentially alter
 569 substrate binding. 570

DISCUSSION

Protein Dynamics and Site-to-Site Communication. A
 572 protein’s internal dynamics is critical for many important
 573 biochemical processes ranging from catalysis to allostery.
 574 Various studies have demonstrated that coupling of dynamics
 575 between domains is a prerequisite to the coordination of
 576 biological events occurring in distant sites.^{17–20,46–48} It has
 577 been reported that correlated backbone motion provides
 578 pathways for transfer of structural and dynamics information.³⁶
 579 Alternatively, long-range intraprotein communication can be
 580 transmitted by correlated side chain fluctuations.³⁵ These
 581 studies suggest that both local interactions and global dynamics
 582 are critical for long-range communication. Moreover, evolu-
 583 tionary studies of myosin motor protein and other proteins
 584 systems, including G protein-coupled receptors and hemoglo-
 585 bin, have demonstrated that conserved residues are key
 586 contributors to allosteric communication.^{21,22,49} Recently,
 587 nuclear magnetic resonance experiments have provided
 588 information about the dynamic processes through which the
 589 KIX domain of the CREB binding protein communicates
 590 allosteric information.⁵⁰ This study also revealed that the
 591 information is transmitted through an evolutionarily conserved
 592 network of residues. Similarly, we showed that coupled-domain
 593 motions are mediated by conserved and evolutionarily coupled
 594 residues in Tt LeuRS.²⁰ Recent experimental^{42,51} and computa-
 595 tional^{20,44,52} studies of other AARS systems have also revealed
 596 the existence of residue–residue interaction networks that
 597 promote interdomain communication. 598

An earlier study using principal component analysis of the
 599 trajectory of MD simulations of prokaryotic-like ProRSs
 600 demonstrated that the INS is engaged in coupled motion 601

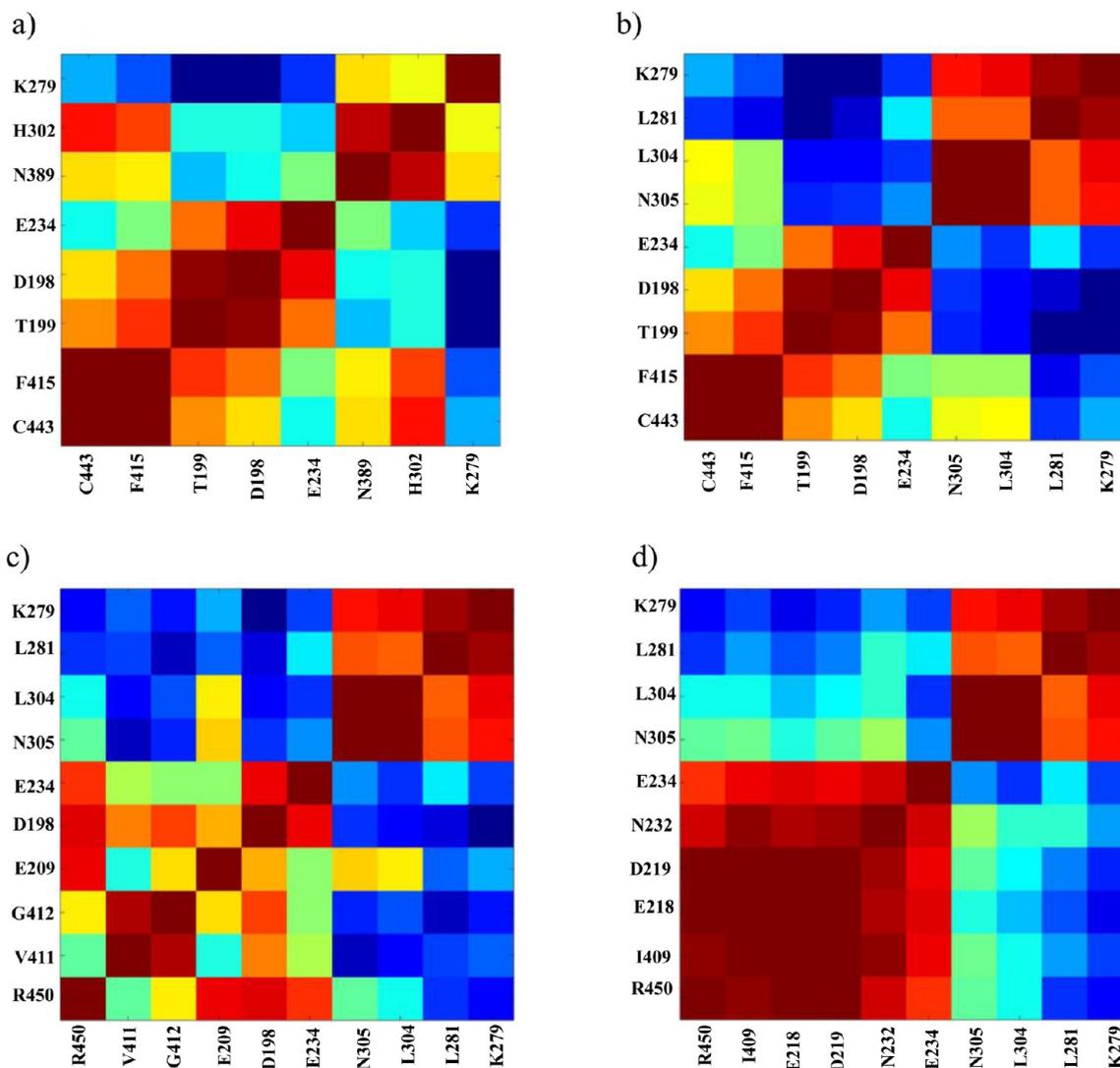


Figure 11. Coupling of thermal motions between residues in various pathways. Cross-correlations in fluctuation between residue pairs in (a) path I, (b) path II, (c) path III, and (d) path IV, extracted from the C matrix of the MD simulation. A strongly correlated motion between residues is colored red, whereas strongly anticorrelated motions are colored blue.

602 with various structural elements of the aminoacylation domain, 603 including the PBL.¹⁶ The collective dynamics of the PBL was 604 found to be altered by the deletion of the INS or point 605 mutation at the INS–aminoacylation-domain junction (*vide* 606 *infra*).¹⁶ Because the PBL undergoes a large-scale conformational 607 transition upon substrate binding,¹⁵ we sought to 608 understand the molecular mechanism by which the PBL 609 dynamics is modulated by distant protein segments. The STCA 610 method was specifically applied to identify pre-existing 611 residue–residue interactions between the INS and amino- 612 acylation domains of Ec ProRS that could modulate PBL 613 dynamics. Four probable pathways of communication between 614 the aminoacylation domain (C443/R450) and INS (K279) 615 (Table 2 and Figure 4) were identified. Residues involved in 616 these pathways were found to be predominantly conserved and 617 polar (Table 2). In addition, the majority of pathway residues 618 are noncatalytic, suggesting that these residues are conserved in 619 the ProRS family to maintain functional dynamics.

620 **Enzymatic Function and Mutation.** The impact of the 621 alanine substitution of pathway residues on amino acid 622 activation is significant (Table 3). Earlier studies showed that

623 mutation of two separate pathway residues, E218 (path IV) and 624 C443 (paths I and II), resulted in activation efficiencies 625 significantly lower than those of WT ProRS.^{16,38} Interestingly, 626 mutation of residues E234 and N305, which are located farther 627 from the aminoacylation active site, also impacts amino acid 628 activation. For example, a 50-fold decrease in amino acid 629 activation efficiency was observed for the N305A variant 630 compared to that of the WT enzyme. In general, the mutations 631 that resulted in decreased amino acid activation efficiency also 632 resulted in similar reductions in aminoacylation efficiency, as 633 well as in pre- and post-transfer editing activities, which 634 strongly suggests that the predicted pathways of interdomain 635 communication are important for substrate binding and 636 catalysis. In one case, a catalytic-domain mutant (F415A) 637 severely defective in amino acid activation, aminoacylation, and 638 pretransfer editing, which are all believed to occur in the 639 aminoacylation active site, was more active than WT ProRS in 640 post-transfer editing, an activity that resides in the INS domain. 641 This suggests that F415 is coupled to the INS. Moreover, 642 mutations of residues at the INS–aminoacylation-domain 643 interface (N305A and E234A) have a greater impact on overall

644 enzyme function, consistent with interdomain communication
645 occurring between the INS and aminoacylation domain
646 through their interface region.

647 Amino acid activation results for the N305A/G412A double
648 mutant ($\Delta\Delta G > 3.7$, based on the level of detection) also
649 support coupling between pathway residues even though they
650 are located in distant domains. This mutant is totally inactive in
651 amino acid activation, an effect that is more severe than
652 predicted on the basis of the single mutations ($\Delta\Delta G = 2.5$
653 kcal/mol + 0.31 kcal/mol = 2.81 kcal/mol). Moreover, the
654 aminoacylation activity of this double mutant was completely
655 abolished. Note that this double mutant is active in hydrolyzing
656 Ala-tRNA^{Pro}; only a 2-fold reduction in post-transfer editing
657 activity was observed. This study also revealed cross-talk
658 between the predicted pathways. Free energy analysis for the
659 H302A (path I)/G412A (path III) double mutant for the
660 aminoacylation reaction indicates subadditivity suggesting
661 coupling between H302 and G412, which implies communi-
662 cation between pathways. Close scrutiny of the predicted
663 pathways showed that they are intertwined through residue
664 E234, a component of all four of the predicted pathways. This
665 observation suggests that E234, which is located in a bottleneck
666 position of the β -strand connecting the catalytic and INS
667 domains, may serve as a key hinge in site-to-site communi-
668 cation.

669 **Coupled Dynamics between the Two Functional**
670 **Sites.** If the identified residue networks facilitate coupled-
671 domain dynamics, then a mutation along these predicted paths
672 is expected to have an observable impact on the protein
673 dynamics. This hypothesis was supported by the rms
674 fluctuation analysis of the WT and mutant ProRS variants,
675 which demonstrates that alanine substitution of on-pathway
676 residues has a considerable impact on the flexibility of various
677 domains of Ec ProRS (Figure 9). In addition, the collective
678 dynamics of the PBL residues were also impacted by these
679 mutations (Figure 10). Thus, noncatalytic residues present in
680 the communication pathways are indeed important for
681 propagating substrate-induced conformational transitions from
682 the active site pocket to the catalytically important PBL. Taken
683 together, the *in silico* mutational study supports the role of the
684 predicted interaction networks in maintaining distant coupled-
685 domain dynamics.

686 In addition, analysis of cross-correlation coefficients between
687 adjacent pairs of residues in each pathway revealed an
688 interesting pattern of dynamic correlations (Figure 11).
689 Adjacent residues in a pathway are primarily engaged in
690 correlated motion; however, two distinct clusters of residues are
691 observed. Residues within each cluster are engaged in strong
692 correlated motion about the hinge residue (E234) between the
693 INS and the aminoacylation domain. This observation is
694 significant as it portrays a unique feature of coupled-domain
695 dynamics in Ec ProRS; although correlated motions exist within
696 each cluster, the coupling of these motions ultimately results in
697 anticorrelated dynamics between the two domains (Figure S1
698 of the Supporting Information). The existence of anticorrelated
699 motion between the INS and the aminoacylation domain
700 apparently modulates the PBL conformational transition from
701 the open to closed state that is required for substrate binding
702 and for protecting the Pro-AMP from hydrolysis.

703 Proteins are dynamical in nature; mutation of any residue
704 will have some impact on protein dynamics and may affect its
705 function. Although mutation of catalytic residues is expected to
706 impact an enzyme's function, in this study, we found that site-

directed changes of noncatalytic residues that are part of key
communication pathways also affect enzyme function. Most of
these residues in the predicted pathways not only impact the
function of the domain in which they are located but also affect
the distant domain with which they are dynamically coupled.
For example, alanine substitution of F415 of the aminoacylation
domain has a significant impact on the editing reaction.
Similarly, the N305A mutation (INS-domain residue) exhibited
significantly reduced efficiencies in proline activation and
aminoacylation of tRNA^{Pro}. On the other hand, residues that
do not belong to the predicted communication pathways are
expected to have little impact on the function of a distant
domain. Previous experimental results showed that alanine
substitution of several residues of the INS domain (*viz.*, T257,
H369, and D386) resulted in an only minor change in
aminoacylation activity (≤ 2.5 -fold decrease) and proline
activation remained unaffected by these mutations.⁷ These
three residues are not members of the predicted pathways.
Analysis of the dynamic cross-correlation matrix indeed
revealed low average correlation coefficients (-0.15 to
 -0.29) of these residues with respect to those of residues in
the predicted pathways. In summary, coupled motions between
INS and the aminoacylation domain are maintained by some
defined residue-residue interaction networks and play
important roles in both aminoacylation and editing functions.

CONCLUSIONS

There are two proposed models⁵³ for long-range site-to-site
communication: the "induced-fit" model in which a substrate-
induced conformational change is propagated through a single
pathway of residue-residue interactions and the "population-
shift" model according to which a perturbation at a distant site
alters the conformational equilibrium through multiple
"preexisting" pathways of residue-residue interactions. Sepa-
rate experimental and computational studies have suggested
that allosteric signal propagation is mediated by a network of
coupled residues and could be regulated by enthalpic
(conformational) and/or entropic (dynamic) changes.⁵³⁻⁵⁷ In
this study, an effort was made to identify the preexisting
network of residue-residue interactions that could facilitate
substrate-induced conformational changes from one site to
another in Ec ProRS through correlated motion.

It is well established that the structural and dynamic
information of a protein is encoded in its primary sequence.
Therefore, the substrate-free structure can be used to extract
information regarding dynamically coupled residues. Moreover,
amino acid residues that are important for structure, function,
and dynamics should be conserved or have co-evolved resulting
in a functionally more efficient enzyme. Therefore, integration
of evolutionary and dynamic information is essential for
identifying residues important for maintaining coupled
dynamics and, thereby, function of proteins. The STCA
method takes into account both dynamic and evolutionary
information in predicting preexisting pathways of interdomain
communication in multidomain proteins like Ec ProRS.

Results from the STCA analysis suggest that Ec ProRS
employs multiple pathways of communication between the INS
and aminoacylation domains to modulate the catalytically
important PBL dynamics. A significant number of residues
participating in interdomain communication are evolutionarily
conserved. Altering these conserved residues has a considerable
impact on enzyme function. Moreover, *in silico* mutation
followed by principal component analyses suggested that

769 residues on the predicted pathways of communication alter
770 protein dynamics. Taken together, this study has identified
771 residues that are likely to be involved in maintaining the
772 coupled dynamics between functional domains of Ec ProRS.
773 Moreover, this study supports the notion that to facilitate long-
774 range site-to-site communication, multidomain proteins like Ec
775 ProRS use parallel paths of residue–residue interactions.
776 Additional mutational studies are underway to explore the
777 relative importance of these pathways.

778 ■ ASSOCIATED CONTENT

779 ● Supporting Information

780 Dynamic cross-correlation matrix of WT Ec ProRS obtained
781 from the PCA-based cluster analysis, a normalized plot of
782 ΔG_i^{stat} versus residue number obtained from the statistical
783 coupling analysis, and a table containing the list of primers used
784 in this study. This material is available free of charge via the
785 Internet at <http://pubs.acs.org>.

786 ■ AUTHOR INFORMATION

787 Corresponding Author

788 *S.B.: e-mail, bhattas@uwec.edu; phone, (715) 836-2278; fax,
789 (715) 836-4979. K.M.-F.: e-mail, musier@chemistry.ohio-state.edu;
790 phone, (614) 292-2021; fax, (614) 688-5402. S.H.: e-mail,
791 hatis@uwec.edu; phone, (715) 836-3850; fax, (715) 836-4979.

792 Funding

793 This work was supported in part by XSEDE [Grants
794 CHE110018 (S.B.) and MCB110173 (S.H.)], the National
795 Institutes of Health [Grants GM049928 (K.M.-F.) and
796 GM085779 (S.H.)], and the Office of Research and Sponsored
797 Programs of the University of Wisconsin—Eau Claire.

798 Notes

799 The authors declare no competing financial interest.

800 ■ ABBREVIATIONS

801 AARS, aminoacyl-tRNA synthetase; Ec, *E. coli*; INS, insertion
802 domain; MD, molecular dynamics; PBL, proline-binding loop;
803 PCA, principal component analysis; Pro-AMP, prolyl-adeny-
804 late; ProRS, prolyl-tRNA synthetase; rms, root-mean-square;
805 rmsd, root-mean-square deviation; rmsp, root-mean-square
806 projection; SCA, statistical coupling analysis; STCA, statistical
807 thermal coupling analysis; Tt LeuRS, *T. thermophilus* leucyl-
808 tRNA synthetase; WT, wild-type.

809 ■ REFERENCES

810 (1) Cusack, S., Yaremchuk, A., Krikiviy, I., and Tukalo, M. (1998)
811 tRNA(Pro) anticodon recognition by *Thermus thermophilus* prolyl-
812 tRNA synthetase. *Structure* 6, 101–108.
813 (2) Stehlin, C., Burke, B., Yang, F., Liu, H., Shiba, K., and Musier-
814 Forsyth, K. (1998) Species-specific differences in the operational RNA
815 code for aminoacylation of tRNA^{Pro}. *Biochemistry* 37, 8605–8613.
816 (3) Musier-Forsyth, K., Burke, B., and Cusack, S. (2005) *Prolyl-tRNA*
817 *synthetases*, Landes Biosciences, Georgetown, TX.
818 (4) Ahel, I., Stathopoulos, C., Ambrogelly, A., Sauerwald, A.,
819 Toogood, H., Hartsch, T., and Soll, D. (2002) Cysteine activation is
820 an inherent in vitro property of prolyl-tRNA synthetases. *J. Biol. Chem.*
821 277, 34743–34748.
822 (5) Beuning, P. J., and Musier-Forsyth, K. (2000) Hydrolytic editing
823 by a class II aminoacyl-tRNA synthetase. *Proc. Natl. Acad. Sci. U.S.A.*
824 97, 8916–8920.
825 (6) Mascarenhas, A., Martinis, S. A., An, S., Rosen, A. E., and Musier-
826 Forsyth, K. (2009) Fidelity Mechanisms of the Aminoacyl-tRNA
827 Synthetases. In *Protein Engineering*, pp 153–200, Springer-Verlag, New
828 York.

(7) Wong, F. C., Beuning, P. J., Nagan, M., Shiba, K., and Musier- 829
Forsyth, K. (2002) Functional role of the prokaryotic proline-tRNA 830
synthetase insertion domain in amino acid editing. *Biochemistry* 41, 831
7108–7115. 832

(8) Wong, F. C., Beuning, P. J., Silvers, C., and Musier-Forsyth, K. 833
(2003) An isolated class II aminoacyl-tRNA synthetase insertion 834
domain is functional in amino acid editing. *J. Biol. Chem.* 278, 52857– 835
52864. 836

(9) An, S., and Musier-Forsyth, K. (2004) Trans-editing of Cys- 837
tRNA^{Pro} by *Haemophilus influenzae* YbaK protein. *J. Biol. Chem.* 279, 838
42359–42362. 839

(10) An, S., and Musier-Forsyth, K. (2005) Cys-tRNA(Pro) editing 840
by *Haemophilus influenzae* YbaK via a novel synthetase.YbaK.tRNA 841
ternary complex. *J. Biol. Chem.* 280, 34465–34472. 842

(11) Ahel, I., Korencic, D., Ibba, M., and Soll, D. (2003) Trans- 843
editing of mischarged tRNAs. *Proc. Natl. Acad. Sci. U.S.A.* 100, 15422– 844
15427. 845

(12) SternJohn, J., Hati, S., Siliciano, P. G., and Musier-Forsyth, K. 846
(2007) Restoring species-specific posttransfer editing activity to a 847
synthetase with a defunct editing domain. *Proc. Natl. Acad. Sci. U.S.A.* 848
104, 2127–2132. 849

(13) Hati, S., Zivovogel, B., Sternjohn, J., Wong, F. C., Nagan, M. C., 850
Rosen, A. E., Siliciano, P. G., Chihade, J. W., and Musier-Forsyth, K. 851
(2006) Pre-transfer editing by class II prolyl-tRNA synthetase: Role of 852
aminoacylation active site in “selective release” of noncognate amino 853
acids. *J. Biol. Chem.* 281, 27862–27872. 854

(14) Zhang, C. M., and Hou, Y. M. (2005) Domain-domain 855
communication for tRNA aminoacylation: The importance of covalent 856
connectivity. *Biochemistry* 44, 7240–7249. 857

(15) Crepin, T., Yaremchuk, A., Tukalo, M., and Cusack, S. (2006) 858
Structures of two bacterial prolyl-tRNA synthetases with and without a 859
cis-editing domain. *Structure* 14, 1511–1525. 860

(16) Sanford, B., Cao, B. V., Johnson, J. M., Zimmerman, K., Strom, 861
A. M., Mueller, R. M., Bhattacharyya, S., Musier-Forsyth, K., and Hati, 862
S. (2012) Role of coupled-dynamics in the catalytic activity of 863
prokaryotic-like prolyl-tRNA synthetases. *Biochemistry* 51, 2146–2156. 864

(17) Bu, Z., Biehler, R., Monkenbusch, M., Richter, D., and Callaway, 865
D. J. (2005) Coupled protein domain motion in Taq polymerase 866
revealed by neutron spin-echo spectroscopy. *Proc. Natl. Acad. Sci.* 867
U.S.A. 102, 17646–17651. 868

(18) Yu, H., Ma, L., Yang, Y., and Cui, Q. (2007) Mechanochemical 869
coupling in the myosin motor domain. II. Analysis of critical residues. 870
PLoS Comput. Biol. 3, e23. 871

(19) Chennubhotla, C., Yang, Z., and Bahar, I. (2008) Coupling 872
between global dynamics and signal transduction pathways: A 873
mechanism of allostery for chaperonin GroEL. *Mol. BioSyst.* 4, 287– 874
292. 875

(20) Weimer, K. M., Shane, B. L., Brunetto, M., Bhattacharyya, S., 876
and Hati, S. (2009) Evolutionary basis for the coupled-domain 877
motions in *Thermus thermophilus* leucyl-tRNA synthetase. *J. Biol.* 878
Chem. 284, 10088–10099. 879

(21) Lockless, S. W., and Ranganathan, R. (1999) Evolutionarily 880
conserved pathways of energetic connectivity in protein families. 881
Science 286, 295–299. 882

(22) Suel, G. M., Lockless, S. W., Wall, M. A., and Ranganathan, R. 883
(2003) Evolutionarily conserved networks of residues mediate 884
allosteric communication in proteins. *Nat. Struct. Biol.* 10, 59–69. 885

(23) Dijkstra, E. W. (1959) A note on two problems in connexion 886
with graphs. *Numerische Mathematik* 1, 269–271. 887

(24) Humphrey, W., Dalke, A., and Schulten, K. (1996) VMD: Visual 888
molecular dynamics. *J. Mol. Graphics* 14, 27–38. 889

(25) Jorgensen, W. L., Chandrasekhar, J., Madura, J. D., Impey, R. 890
W., and Klein, M. L. (1983) Comparison of simple potential functions 891
for simulating liquid water. *J. Chem. Phys.* 79, 926. 892

(26) MacKerell, A. D. J., Bashford, D., Bellott, M., Dunbrack, R. L. J., 893
Evanseck, J. D., Field, M. J., Fischer, S., Gao, J., Gou, J., Ha, S., Joseph- 894
McCarthy, D., Kuchnir, L., Kuczera, K., Lau, F. T. K., Mattos, C., 895
Michnick, S., Ngo, T., Nguyen, D. T., Prodhom, B., Reiher, W. E. I., 896
Roux, B., Schelenkrich, M., Smith, J. C., Stote, R., Straub, J., Watanbe, 897

- 898 M., Wiórkiewicz-Kuczera, J., Yin, D., and Karplus, M. (1998) Allatom
899 empirical potential for molecular modeling and dynamics studies of
900 proteins. *J. Phys. Chem. B* 102, 3586.
- 901 (27) Phillips, J. C., Braun, R., Wang, W., Gumbart, J., Tajkhorshid, E.,
902 Villa, E., Chipot, C., Skeel, R. D., Kale, L., and Schulten, K. (2005)
903 Scalable molecular dynamics with NAMD. *J. Comput. Chem.* 26, 1781–
904 1802.
- 905 (28) Roy, J., and Lughton, C. A. (2010) Long-timescale molecular-
906 dynamics simulations of the major urinary protein provide atomistic
907 interpretations of the unusual thermodynamics of ligand binding.
908 *Biophys. J.* 99, 218–226.
- 909 (29) van Aalten, D. M., Amadei, A., Linssen, A. B., Eijssink, V. G.,
910 Vriend, G., and Berendsen, H. J. (1995) The essential dynamics of
911 thermolysin: Confirmation of the hinge-bending motion and
912 comparison of simulations in vacuum and water. *Proteins* 22, 45–54.
- 913 (30) Mueller, R. M., North, M. A., Yang, C., Hati, S., and
914 Bhattacharyya, S. (2011) Interplay of flavin's redox states and protein
915 dynamics: An insight from QM/MM simulations of dihydronicotina-
916 mide riboside quinone oxidoreductase 2. *J. Phys. Chem. B* 115, 3632–
917 3641.
- 918 (31) Amadei, A., Linssen, A. B., and Berendsen, H. J. (1993)
919 Essential dynamics of proteins. *Proteins* 17, 412–425.
- 920 (32) Glykos, N. M. (2006) Software news and updates. Carma: A
921 molecular dynamics analysis program. *J. Comput. Chem.* 27, 1765–
922 1768.
- 923 (33) Hatley, M. E., Lockless, S. W., Gibson, S. K., Gilman, A. G., and
924 Ranganathan, R. (2003) Allosteric determinants in guanine nucleotide-
925 binding proteins. *Proc. Natl. Acad. Sci. U.S.A.* 100, 14445–14450.
- 926 (34) Altschul, S. F., Madden, T. L., Schaffer, A. A., Zhang, J., Zhang,
927 Z., Miller, W., and Lipman, D. J. (1997) Gapped BLAST and PSI-
928 BLAST: A new generation of protein database search programs.
929 *Nucleic Acids Res.* 25, 3389–3402.
- 930 (35) Dubay, K. H., Bothma, J. P., and Geissler, P. L. (2011) Long-
931 range intra-protein communication can be transmitted by correlated
932 side-chain fluctuations alone. *PLoS Comput. Biol.* 7, e1002168.
- 933 (36) Fenwick, R. B., Esteban-Martin, S., and Salvatella, X. (2011)
934 Understanding biomolecular motion, recognition, and allostery by use
935 of conformational ensembles. *Eur. Biophys. J.* 40, 1339–1355.
- 936 (37) Burke, B., Lipman, R. S., Shiba, K., Musier-Forsyth, K., and Hou,
937 Y. M. (2001) Divergent adaptation of tRNA recognition by
938 *Methanococcus jannaschii* prolyl-tRNA synthetase. *J. Biol. Chem.* 276,
939 20286–20291.
- 940 (38) Stehlin, C., Heacock, D. H., II, Liu, H., and Musier-Forsyth, K.
941 (1997) Chemical modification and site-directed mutagenesis of the
942 single cysteine in motif 3 of class II *Escherichia coli* prolyl-tRNA
943 synthetase. *Biochemistry* 36, 2932–2938.
- 944 (39) Fersht, A. R., Ashford, J. S., Bruton, C. J., Jakes, R., Koch, G. L.,
945 and Hartley, B. S. (1975) Active site titration and aminoacyl adenylate
946 binding stoichiometry of aminoacyl-tRNA synthetases. *Biochemistry* 14,
947 1–4.
- 948 (40) Liu, H., and Musier-Forsyth, K. (1994) *Escherichia coli* proline
949 tRNA synthetase is sensitive to changes in the core region of
950 tRNA(Pro). *Biochemistry* 33, 12708–12714.
- 951 (41) Heacock, D., Forsyth, C. J., Shiba, K., and Musier-Forsyth, K.
952 (1996) *Bioorg. Chem.* 24, 273–289.
- 953 (42) Ghosh, A., Sakaguchi, R., Liu, C., Vishveshwara, S., and Hou, Y.
954 M. (2011) Allosteric communication in cysteinyl tRNA synthetase: A
955 network of direct and indirect readout. *J. Biol. Chem.* 286, 37721–
956 37731.
- 957 (43) Budiman, M. E., Knaggs, M. H., Fetrow, J. S., and Alexander, R.
958 W. (2007) Using molecular dynamics to map interaction networks in
959 an aminoacyl-tRNA synthetase. *Proteins* 68, 670–689.
- 960 (44) Ghosh, A., and Vishveshwara, S. (2007) A study of
961 communication pathways in methionyl-tRNA synthetase by molecular
962 dynamics simulations and structure network analysis. *Proc. Natl. Acad.*
963 *Sci. U.S.A.* 104, 15711–15716.
- 964 (45) Horovitz, A. (1996) Double-mutant cycles: A powerful tool for
965 analyzing protein structure and function. *Folding Des.* 1, R121–R126.
- (46) Daily, M. D., and Gray, J. J. (2009) Allosteric communication 966
occurs via networks of tertiary and quaternary motions in proteins. 967
PLoS Comput. Biol. 5, e1000293. 968
- (47) Fidelak, J., Ferrer, S., Oberlin, M., Moras, D., Dejaegere, A., and 969
Stote, R. H. (2010) Dynamic correlation networks in human 970
peroxisome proliferator-activated receptor- γ nuclear receptor protein. 971
Eur. Biophys. J. 39, 1503–1512. 972
- (48) Zheng, W., Liao, J. C., Brooks, B. R., and Doniach, S. (2007) 973
Toward the mechanism of dynamical couplings and translocation in 974
hepatitis C virus NS3 helicase using elastic network model. *Proteins* 67, 975
886–896. 976
- (49) Tang, S., Liao, J. C., Dunn, A. R., Altman, R. B., Spudich, J. A., 977
and Schmidt, J. P. (2007) Predicting allosteric communication in 978
myosin via a pathway of conserved residues. *J. Mol. Biol.* 373, 1361– 979
1373. 980
- (50) Bruschweiler, S., Schanda, P., Kloiber, K., Brutscher, B., 981
Kontaxis, G., Konrat, R., and Tollinger, M. (2009) Direct observation 982
of the dynamic process underlying allosteric signal transmission. *J. Am.* 983
Chem. Soc. 131, 3063–3068. 984
- (51) Rodriguez-Hernandez, A., and Perona, J. J. (2011) Heat maps 985
for intramolecular communication in an RNP enzyme encoding 986
glutamine. *Structure* 19, 386–396. 987
- (52) Bhattacharyya, M., and Vishveshwara, S. (2011) Probing the 988
allosteric mechanism in pyrrolysyl-tRNA synthetase using energy- 989
weighted network formalism. *Biochemistry* 50, 6225–6236. 990
- (53) del Sol, A., Tsai, C. J., Ma, B., and Nussinov, R. (2009) The 991
origin of allosteric functional modulation: Multiple pre-existing 992
pathways. *Structure* 17, 1042–1050. 993
- (54) Gunasekaran, K., Ma, B., and Nussinov, R. (2004) Is allostery an 994
intrinsic property of all dynamic proteins? *Proteins* 57, 433–443. 995
- (55) Tsai, C. J., del Sol, A., and Nussinov, R. (2008) Allostery: 996
Absence of a change in shape does not imply that allostery is not at 997
play. *J. Mol. Biol.* 378, 1–11. 998
- (56) Popovych, N., Sun, S., Ebricht, R. H., and Kalodimos, C. G. 999
(2006) Dynamically driven protein allostery. *Nat. Struct. Mol. Biol.* 13, 1000
831–838. 1001
- (57) Daily, M. D., and Gray, J. J. (2007) Local motions in a 1002
benchmark of allosteric proteins. *Proteins* 67, 385–399. 1003