SUPPORTING INFORMATION

Jun Lu[†], Ruiying Wu[‡], Joshua N. Adkins[§], Andrzej Joachimiak[‡], and J. N. Mark Glover[†*]

SUPPLEMENTARY EXPERIMENTAL PROCEDURES

GC/MS-EI (Gas chromatography/mass spectrometry-electron ionization) analysis. Organic solvent tetrachloroethylene was used to extract hydrophobic ligands from 2 mg of purified $TraJ_F^{11-130}$ and same volume of its buffer control. After concentration, the sample was injected for GC-MS-EI analysis at the Mass Spectrometry Facility, Department of Biochemistry, University of Alberta. Results are searched against NIST/EPA/NIH Mass Spectral Library for the ligand identity.

Purification of $TraJ_F^{11-130}$ for GCMS-EI analysis of potential $TraJ_F$ ligands. To identify the physiological ligand potentially present in the hydrophobic pocket at the dimeric interface of the F plasmid TraJ, His-tagged $TraJ_F^{11-130}$ was purified from 5 liters of cells containing pJLJ5629 without IPTG induction such that the protein is expressed at low levels to allow high occupancy of the potential physiological ligand. The protein was further purified similarly as described in Experimental Procedures except that no DTT was present throughout the purification.

Limited proteolysis of His_6-$TraJ_F^{11-130}$. Four aliquots of 10 μg of His_6-$TraJ_F^{11-130}$ purified from *E. coli* BE280 ($TnaA^-$) cells in the absence of reducing agents were diluted in buffer (50 mM Tris-HCl, 300 mM NaCl, pH 7.5) to a final concentration of 1 μg/μL. Two of the aliquots also contain a certain additive at the following final concentrations: DTT at 3 mM or indole at 3 mM. After incubation at 25°C for a week, 3 aliquots were mixed with 0.2 μg of

trypsin and was further incubated at 25°C for 2 hours and then mixed with 1 μL of PMSF (10 mg/mL). One aliquot without additives was used as undigested control. Samples were separated by SDS-PAGE. The molecular weight of the tryptic fragments were determined by matrix-assisted laser desorption/ionisation time-of-flight (MALDI-TOF) mass spectrometry. Mass spectrometry was performed at Institute of Biomolecular Design, University of Alberta.

***In vitro* cross-linking assays.** Four aliquots of 10 μg of His_6-$TraJ_F^{11-130}$ purified from *E. coli* BE280 ($TnaA^-$) cells in the absence of reducing agents were diluted in buffer (50 mM sodium phosphate, 300 mM NaCl, pH 7.5) to a final concentration of 1 μg/μL. Two aliquot also contains a certain additive at the following final concentrations: DTT at 3 mM or indole at 3 mM. After incubation at 25°C for a week, 3 aliquot were mixed with BS^3 (Bis[sulfosuccinimidyl] suberate) (Pierce) to a final concentrations of 300 μM. One aliquot without additives was used as the negative control. After 30 minutes of incubation at 25°C, 1 μL of 1M Tris-HCL (pH 7.5) was added to each aliquot, followed by another 5 minutes of incubation at room temperature. Proteins were separated by SDS-PAGE and visualized by Coomassie blue staining.

SUPPLEMENTARY RESULTS

Characterizing potential ligands at the dimeric interface of $TraJ_F$. In the electron density map of both SeMet-$TraJ_F^{1-140}$ and $TraJ_F^{11-130}$, there is a strong density unaccounted for in an internal pocket formed at the dimeric interface. The higher resolution data from the $TraJ_F^{11-130}$ crystals revealed electron density suggesting a six-member ring structure with substituents at two adjacent positions in the ring (Figure S2). Calculation of an anomalous difference map from the same native data set revealed two strong adjacent anomalous difference peaks in the ring. GC-MS-EI analysis of a purified $TraJ_F^{11-130}$ sample indicated the presence of DTD (Dithiane diol, oxidized DTT) in the protein solution. DTD indeed fits the density, with the two sulphur atoms positioned on the anomalous difference peaks.

The presence of DTD within the pocket at the $TraJ_F$ dimer interface suggested that a similar small molecule ligand might bind this pocket to modulate dimerization, providing a potential mechanism for the regulation of TraJ activity and conjugation in response to a small molecule signal. We reasoned that the native ligand might have been lost and exchanged for DTD during purification. In addition, overexpression of $TraJ_F^{11-130}$ might have exceeded the amounts of native ligand available in the cells. Therefore, to isolate the $TraJ_F$ PAS domain with any bound natural ligands, we purified $TraJ_F^{11-130}$ from cells expressing the protein at close to physiological levels (to allow higher occupancy of potential physiological ligand) and in the absence of DTT (see Supplementary Experimental Procedures). GC/MS-EI analysis of this sample revealed the presence of a small molecule most likely to be indole based on its fragmentation pattern (Figure S3). In *E. coli*, indole is derived from tryptophan catalysed by tryptophanase (TnaA) and serves as a common intercellular signal [1, 2]. To test a role for indole in $TraJ_F$ function *in vivo*, we compared levels of TraJ-dependent F plasmid P_Y activity in wild type

or *tnaA E. coli* strains (Figure S4A). $TraJ_F$ activated P_Y to comparable levels in both the wild type ($TnaA^+$) and in two distinct $TnaA^-$ strains, indicating that indole is dispensable for the *in vivo* function of $TraJ_F$. To identify other potential ligands, we further purified $TraJ_F^{11\text{-}130}$ from a non-indole-producing strain in the absence of DTT and analysed the purified protein with GC-MS-EI. No ligands were identified in the analysis of three independent preparations of the protein.

$TraJ_F^{11\text{-}130}$ purified in the absence of DTT and indole elutes at the same volume by size exclusion chromatography as protein purified with DTT and indole, suggesting that indole and DTT might not affect $TraJ_F^{11\text{-}130}$ dimerization. We further used biochemical assays to test for the potential effects of small molecule ligands of $TraJ_F$ PAS dimer stability. We assayed the *in vitro* structural stability of $TraJ_F^{11\text{-}130}$ with and without DTD or indole through limited trypsin analysis and chemical crosslinking (Figure S4B and S4C). Aliquots of $TraJ_F^{11\text{-}130}$ purified in the absence of DTT and indole were mixed with DTT or indole, and after equilibration at room temperature, were subjected to trypsin digestion or BS^3 chemical crosslinking as described in Experimental Procedures. $TraJ_F^{11\text{-}130}$ with and without DTT (or indole) have similar patterns of limited trypsin digestion (Figure S4B), resulting in a predominant major tryptic fragment, $TraJ_F^{38\text{-}130}$, agreeing with our previous analysis [3]. BS^3 chemical crosslinking also showed that neither indole nor DTT appreciably affected $TraJ_F^{11\text{-}130}$ dimerization efficiency (Figure S4C). Therefore, DTT and indole do not appear to be required for proper folding and dimerization of $TraJ_F^{11\text{-}130}$ in solution.

SUPPLEMENTARY REFERENCES

[1] Burns, R. O., and Demoss, R. D. (1962) Properties of tryptophanase from *Escherichia coli*, *Biochim Biophys Acta 65*, 233-244.

[2] Lee, J. H., and Lee, J. (2010) Indole as an intercellular signal in microbial communities, *FEMS Microbiol Rev 34*, 426-444.

[3] Lu, J., Peng, Y., Arutyunov, D., Frost, L. S., and Glover, J. N. (2012) Error-prone PCR mutagenesis reveals functional domains of a bacterial transcriptional activator, TraJ, *J Bacteriol 194*, 3670-3677.

[4] Jobling, M. G., and Holmes, R. K. (1990) Construction of vectors with the p15a replicon, kanamycin resistance, inducible lacZ alpha and pUC18 or pUC19 multiple cloning sites, *Nucleic Acids Res 18*, 5315-5316.

[5] Stols, L., Gu, M., Dieckman, L., Raffen, R., Collart, F. R., and Donnelly, M. I. (2002) A new vector for high-throughput, ligation-independent cloning encoding a tobacco etch virus protease cleavage site, *Protein Expr Purif 25*, 8-15.

[6] Tabor, S., and Richardson, C. C. (1985) A bacteriophage T7 RNA polymerase/promoter system for controlled exclusive expression of specific genes, *Proc Natl Acad Sci U S A 82*, 1074-1078.

[7] Sugino, Y., and Hirota, Y. (1962) Conjugal fertility associated with resistance factor R in *Escherichia coli*, *J Bacteriol 84*, 902-910.

Table S1. Plasmids and oligonucleotides

Plasmid and oligonucleotide	Description and references
pIZ2023	pSLT *traJ* cloned in pGEM-T easy vector; a gift from Prof. Josep Casadesus, Universidad de Sevilla
pJLac101-P_Y	P_Y-*lacZ* fusion cloned in pJLac101 [3]
pJLJ001	The F plasmid *traJ* cloned in pK184 [3]
pJLJ002	F-like plasmid pSLT *traJ* cloned in pK184; this work
pJLJ003	F-like plasmid R100 *traJ* cloned in pK184; this work
pJLJ004	A pSLT-F plasmid hybrid *traJ* ($TraJ_{pSLT}^{1-125}$:$TraJ_F^{121-226}$) cloned in pK184; this work
pJLJ005	An R100-F plasmid hybrid *traJ* ($TraJ_{R100}^{1-120}$:$TraJ_F^{121-226}$) cloned pK184; this work
pJLJ2123	F plasmid $traJ^{1-140}$ cloned in pT7-7; this work
pJLJ2829	F plasmid $traJ^{11-130}$ cloned in pT7-7; this work
pJLJ5629	F plasmid $traJ^{11-130}$ cloned in pK184 such that the $TraJ^{11-130}$ is constitutively expressed at close to physiological levels of $TraJ_F$ [3]; this work
pK184	P15a replicon-base cloning vector, Km^R [4]
pMCSG7	Bacterial expression ligation independent cloning vector with T7 promoter (pET21 derivative), adds N-terminal His tag and TEV protease site; Amp^R; ColE1 pBR322-type bacterial origin of replication [5]
pT7-7	Protein expression vector; pMB1 replicon; Amp^R [6]
pJSLT128	$TraJ_{pSLT}^{1-128}$ cloned in pMCSG7; this work
R100	An F-like conjugative plasmid [7]
JLU308	TAT GGA TCC CTT CTG GTT ACC ACT TAT GTT TGC AG
JLU321	TAGAATTCACCATCACCATCACCATGAGAACCTGTACTTCCAAGGAT ATCCGATGGATCGTATTCAACAAAAACATG
JLU323	ATG GAT CCC TAT TAT CTG ATA TCA TCA CTG ACA ATC G
JLU328	TAGAATTCACCATCACCATCACCATGAGAACCTGTACTTCCAAGGAC AAATAGATCTGCTGGAAAATCTGACGG
JLU329	ATGGATCCTATTAATTATAACTGTCTTTATGACGAACATGAGCAGC
JLU330B	ATGAATTCGAGGAGGTTCCTATGTATCCGACAGATCCTAGACAACTT AATACTGAACGTC
JLU331B	TTTGGATCCTTTATTGCCCGGAGCAGGCC

JLU346	GCTGCTCATGTTCGTCATAAAGACAG
JLU347	CTGTCTTTATGACGAACATGAGCAGCGCAGAAAAAAAGCCATAAAA CATTAACCACCTC
JLU349	ATAGAATTCGTGAGGAGGTTCCTATGCAAATAGATCTGGCGGAAAA TCTGACGGCAGTTATTCAGG
JLU350	CGCTGACTGGACAGATATTGTCCAGAATCTATTCTG
JLU351	CAGAATAGATTCTGGACAATATCTGTCCAGTCAGCGCTTAATGGACA CAGAAATATTATTCTGTGG
JLU352	CGCCCACAGAATAATATTTCTGTGTCCATTAAGAAATGAC
JLU353	GTCATTTCTTAATGGACACAGAAATATTATTCTGTGGGCGTTTTATGA TGCTGCTCATGTTCGTCATAAAGACAG
JLU356	ATAGAATTCGTGAGGAGGTTCCTATGCACCATCACCATCACCATCAA ATAGATCTGCTGGAAAATCTGACGGC
JLU390	ATAGAATTCTTTGAGGAGGTTCCTATGTGTCCGACGGATCGTAGAG
JLU391	CTGTCTTTATGACGAACATGAGCAGCTAAATTTATCTGCCACAACAC AATTATCTCATTATTATCGAATG
JLU394	TATGGATCCTCACTTAACACAATAAAATTCACGAATAATGTCTATCA ATG
TraJA	TACTTCCAATCCAATGCCATGTGTCCGACGGATCGTAGAGAAA
TraJB	TTATCCACTTCCAATGTTAATGATCTGGAAATAAATTTATCTGCCACA ACACA

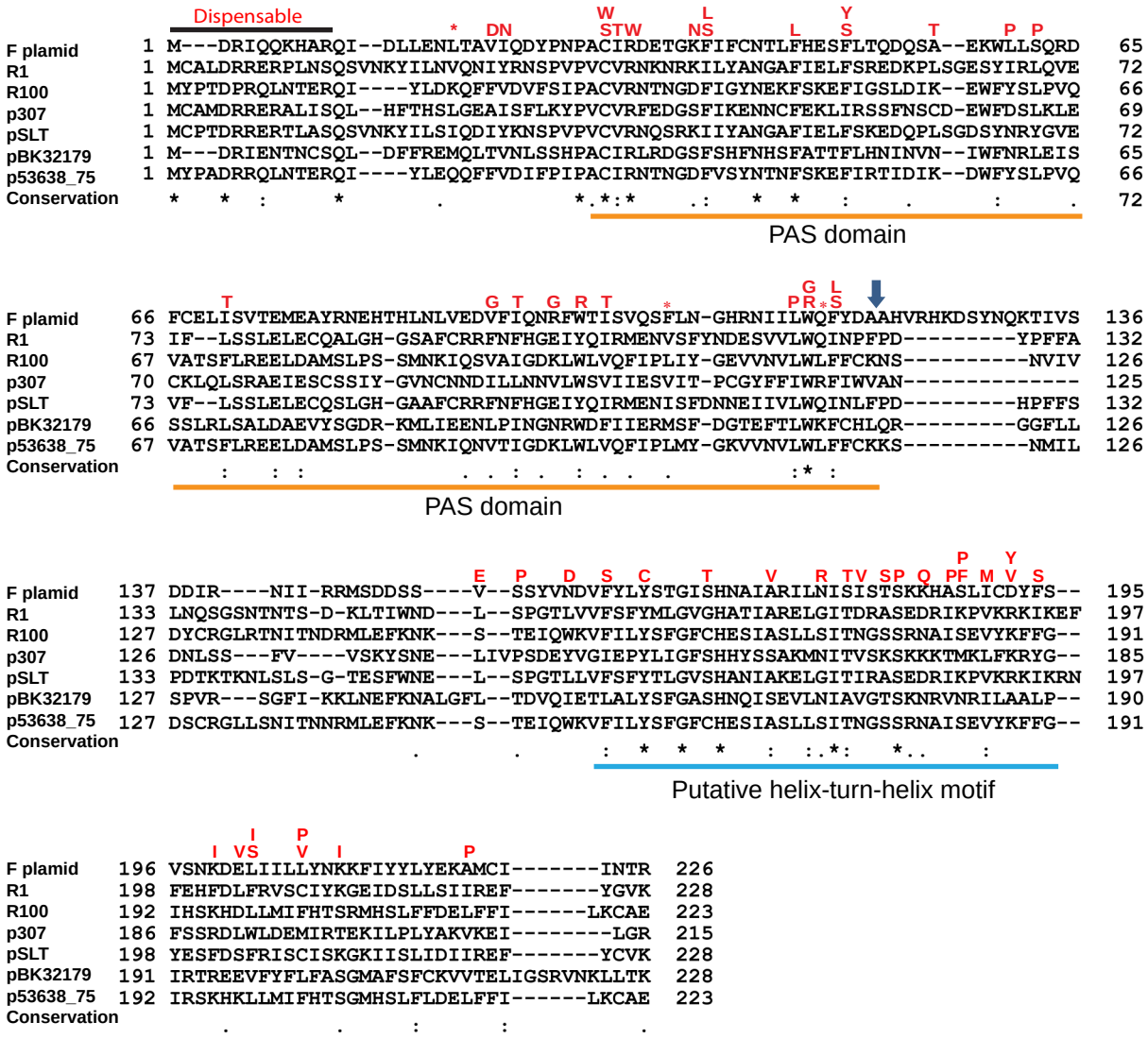

Figure S1. Sequence alignment of representative variants of the F-family TraJ. The labels of sequences indicate the corresponding plasmid sources of the TraJ variants, including the F plasmid (GenBank: BAA97942.1), R1 plasmid (P05837), R100 (P13972), P307 (GenBank: AAA25523.1) from *E. coli*, pSLT (GenBank: CAB40570.1) from *Salmonella enterica subsp. enterica serovar* Typhimurium, pBK32179 (YP_007366589.1) from *Klebsiella pneumoniae*, and p53638_75 (YP_001919411.1) from *E. coli* 53638. The alignment was done through T-Coffee (Tree-based Consistency Objective Function For AlignmEnt Evaluation). The Amino acid substitutions in $TraJ_F$, each of which affects the *in vivo* function as previously described are shown in red above the corresponding residues. Deletion of the first 10 residues of $TraJ_F$ did not affect function, designated as "dispensable". "*" indicate sites of further mutational analysis as shown in Table 2. The downward arrow indicates the boundary for PAS domain swapping among F, R100, and pSLT TraJ as shown in Table 3.

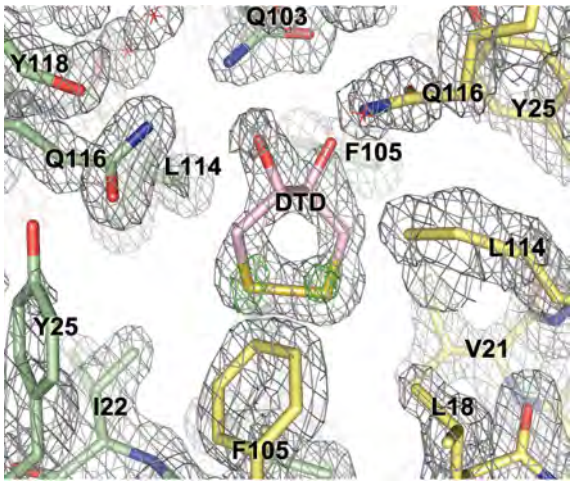

Figure S2. Electron density maps for the region surrounding the DTD molecule at the dimerization interface. In grey is the $2F_O$-F_C electron density map, calculated to 1.55 Å resolution and contoured at 1.5 σ, and in green is an anomalous difference electron density map, calculated to 1.55 Å resolution and contoured at 3.0 σ. The two protomers are colored in green and yellow, respectively, with the DTD in magenta with the sulphur atoms coloured yellow. Water molecules are shown in red "+".

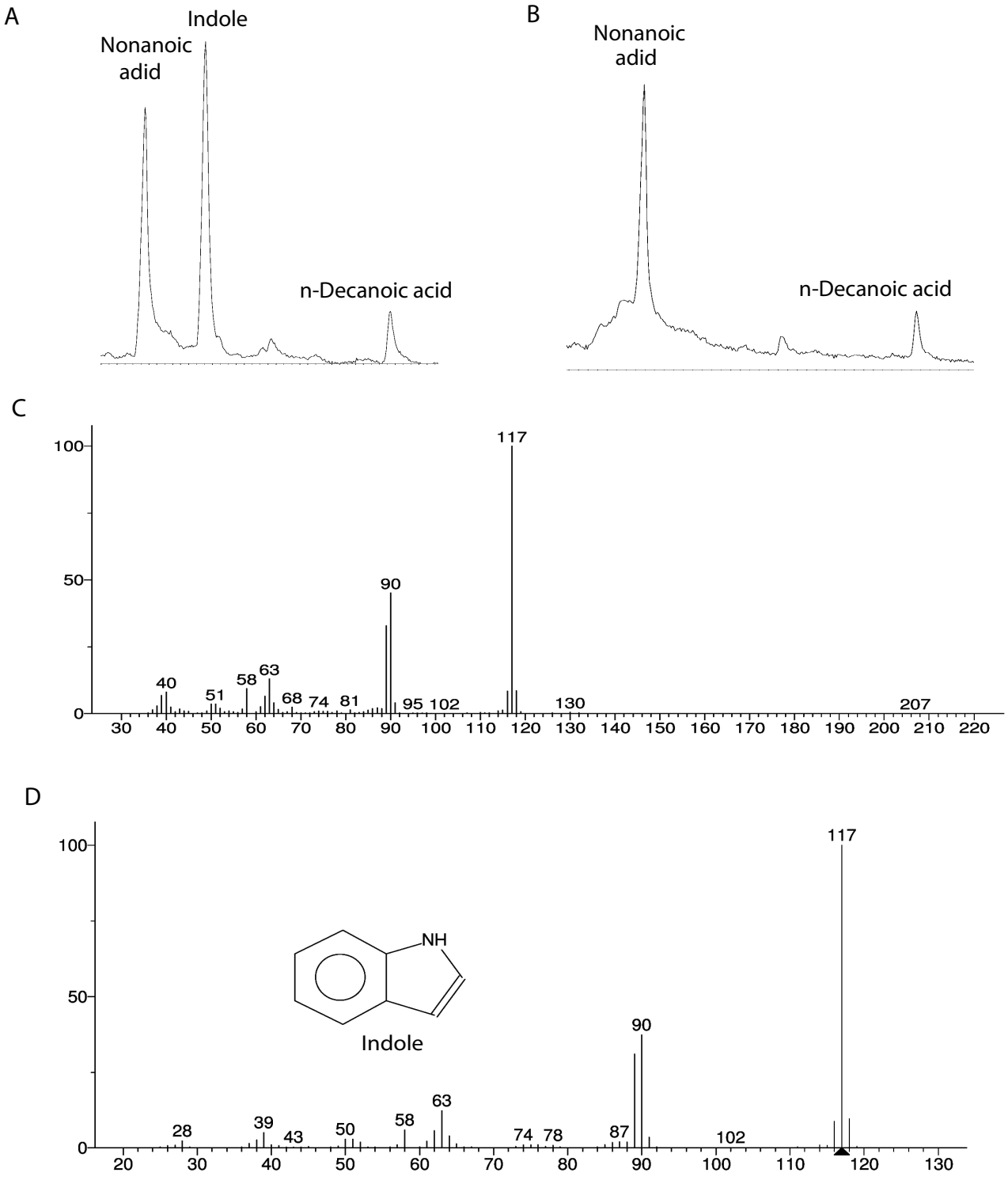

Figure S3. Presence of indole in $TraJ_F^{11\text{-}130}$ purified in absence of DTT. (A) Total ion current chromatogram of the extract from the $TraJ_F^{11\text{-}130}$ sample. (B) Total ion current chromatogram of the extract from the buffer control. (C) Experimental mass spectrum of indole in the extract from the $TraJ_F^{11\text{-}130}$ sample. (D) Standard mass spectrum of indole in the NIST08 library.

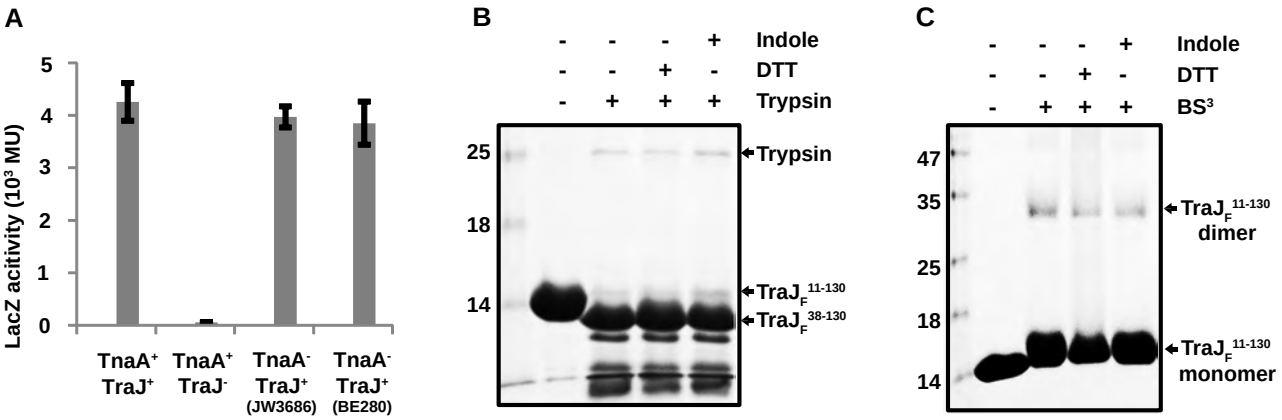

Figure S4. Potential $TraJ_F$ small molecule ligands do not modulate the stability of the PAS dimer or the activity of $TraJ_F$ *in vivo*. (A) Tryptophanase (TnaA) does not regulate $TraJ_F$ activity in *E. coli*. $TraJ_F$ activity was measured by a transcriptional reporter gene assay using pJLAC101-P_Y that harbours an F plasmid P_Y-*lacZ* fusion gene. LacZ activity was assayed in an *E. coli* $TnaA^+$ (BW25113; indole-producing) strain or two different $TnaA^-$ strains (JW3686 and BE280, non-indole- producing). $TraJ_F$ was expressed *in trans* from pJLJ001. A pJLJ001 derivative with a frameshifted *traJ* was used as the $TraJ^-$ control. (B) Effect of indole or DTT on limited trypsin digestion of $TraJ_F^{11\text{-}130}$. Numbers on the left side of the figure indicate molecular weight markers (in kDa) (first lane). (C) Effect of indole or DTT on *in vitro* BS3 chemical cross-linking of $TraJ_F^{11\text{-}130}$.