Methods

General. Taq polymerase and restriction enzymes were obtained from New England Biolabs. Primers were purchased from Integrated DNA Technologies. All other chemicals were purchased from Sigma unless otherwise noted. The *E. coli* strain, origami, was obtained from Novagen, and the isogenic parental strains, AD494 and DHB4, were gifts from Professor Jon Beckwith (Harvard University).

DnaE constructs and expression. The cysteine point mutants described in this work were generated by PCR using template DNA encoding an artificially fused DnaE intein (Ming Xu, NEB) with short N- and C-extein residues. Primers introduced the desired substitution and generated a 5' XhoI site and 3' PstI site which were used to insert the amplified fragment in frame between the CFP and YFP coding sequences in the plasmid, C-I_a-Y[1]. FRET-active C-I_a-Y precursor was expressed from the arabinose-inducible promoter in pBAD33[2], as described previously[1].

For high level protein expression and crystallography the fused DnaE intein with the N-extein amino acid sequence FCPG, and the C-extein sequence, FN, was PCR amplified, digested with BamHI and XhoI, and cloned into pet-45b (Novagen). The resulting plasmid, His_6-^{CPGC}DnaE-spl, was transformed into origami-DE3, grown at 37 °C in LB broth (10 g/l bacto-tryptone, 5 g/l bacto-yeast extract, 10 g/l NaCl) with ampicillin (100 μg/ml) and kanamycin (10 μg/ml) to A_{600} = 0.6, followed by induction with 0.3 mM IPTG (Gold Biotechnology) at 25 °C for 4 h. Cells were harvested by centrifugation and the resulting pellets were lysed with BPER-II. The cleared lysate was combined with an equal volume of 2x Ni^{2+}-NTA buffer (0.1 M $NaPO_4$, 0.6 M NaCl, 0.05 M imidazole, pH 8) and the precursor was purified at 4 °C using either

5 ml Ni-NTA-Sepharose 6 Fast Flow column (GE Healthcare) or a 1 ml Ni-NTA spin column (Qiagen). Eluted His_6-^{CPGC}DnaE-spl precursor was treated with enterokinase (Genscript) to remove the N-terminal His tag, and further purified by size-exclusion chromatography.

DnaE intein activity measurements. Activity of DnaE intein variants during bacterial expression was assessed by in-gel YFP fluorescence. C-I_a-Y precursors were expressed in 5 ml cultures by addition of arabinose to 0.4% as described previously[1]. Total soluble protein was extracted from cell pellets with 300 µl of B-PER-II (Pierce/Thermo Scientific), centrifuged at 14,000 x*g* for 10 min at 4 °C, and separated without pre-boiling by nonreducing SDS/PAGE. Gels were scanned for YFP fluorescence using a Typhoon 9400 imager (GE Healthcare) with the excitation laser at 488 nm and emission filter at 580 nm. The relative abundance of precursor (C-I_a-Y) to cleaved product (I_a-Y) was determined by fluorescence intensity using the ImageQuant software (GE Healthcare).

Rates of DTT-induced N-extein cleavage of intein precursors were determined using an *in vitro* FRET-based assay described previously[1]. Reaction solutions were prepared in the wells of a Flat-bottom black, polystyrene, 96-well assay plates (Corning/Costar) by combining 10 µL of cleared bacterial lysate containing soluble precursor and 80 µl of assay buffer (20 mM Tris pH 8.0, 10 mM EDTA). N-extein cleavage was initiated by adding an additional 10 µl of assay buffer containing DTT (0.2 M), or hydroxylamine (1.0 M), or hydroxylamine (1.0 M) with TCEP (0.05 M). A catalytically inactive intein precursor, $CI_{aa}Y$, with C1A and N159A mutations was prepared identically and served as a control for these experiments[1]. Reaction solutions were excited at 400 nm with emission measured at 485 nm and at 540 nm using a Biotek Synergy-2, 96-well fluorescence spectrophotometer. Fluorescence values at 540 nm were divided by the

fluorescence values at 485 nm, providing the FRET ratio. Rate constants (k_{obs}) were calculated from FRET ratios determined at selected intervals over 8 h by fitting the data to a single or double exponential decay using Excel Solver (Microsoft Inc.).

Crystallography. Purified ^{CGPC}DnaE-splc was concentrated to approximately 6 mg/ml in a buffer containing 20 mM Tris HCl, pH 8.0 and 200 mM NaCl. Crystals were obtained by hanging drop vapor diffusion methods at room temperature, using a precipitation buffer containing 54% ammonium sulfate, 2% PEG400, 100 mM HEPES, pH 7.0. Prior to data collection, all crystals were transferred to a cryosolvent consisting of the precipitation buffer complemented with 20% (v/v) glycerol. Crystals were flash cooled in the cold nitrogen gas stream (100K). All X-ray diffraction data were collected using an in-house R-axis IV++ (Rigaku) detector and processed using CrystalClear (Rigaku). Data collection statistics are summarized in Table 1. The structure was determined by molecular replacement[3] with the structure of the spliced intein (PDB, 1ZD7) used as the search model. The structure was refined using CNS[4] combined with model building using COOT[5]. The final refinement statistics are summarized in Table 1. All but one residue (Gly-1) are within the preferred (98.1%) or allowed (1.2%) regions of the Ramachandran plot, as defined in the program MOLPROBITY[6].

MoaA constructs and expression. The full-length MoaA intein was amplified from *Pyrococcus abyssi* genomic DNA (generous gift of Dr. Kenneth V. Mills). The amplified fragment included six native N-extein residues, NLNCWY, and three native C-extein residues, CFF, followed by a His_6 tag to facilitate precursor purification. A control MoaA construct was also amplified that carried a single C-3A mutation in the N-extein. Both fragments were cloned in-frame to the C-

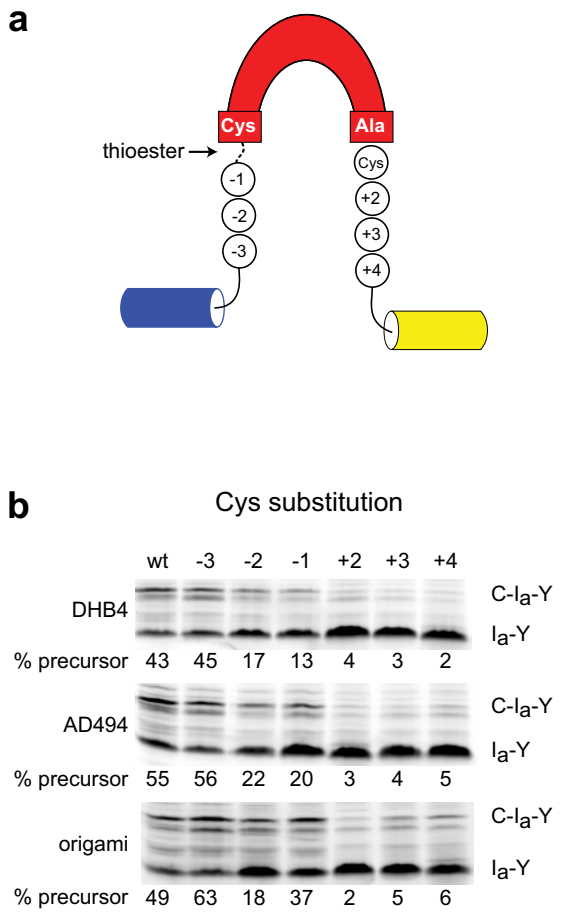
terminus of CFP in pBAD33. For precursor expression,, the resulting plasmids, CFP-^{CWYC}MoaA- $His_{6,}$ and CFP- ^{AWYC}MoaA- $His_{6,}$ were transformed into *E. coli*, grown at 37 °C to A_{600} = 0.8-1, then treated with arabinose at 0.4% followed by incubation for 3 h at 15 °C. Protein extraction followed the procedure described above for the DnaE intein.

MoaA intein activity measurements. Redox sensitivity of the CFP-^{CWYC}MoaA-His_6 precursor was determined by in-gel CFP fluorescence, following *in vitro* incubation at 22 ˚C and at 55 ˚C in the presence and absence of TCEP (0.017 M). Reactions were initiated immediately after elution from a Ni^{2+}-NTA column. Aliquots of the reactions were removed at 0, 2, 4 and 24 h, mixed with nonreducing SDS-PAGE loading buffer, and frozen at -80 ˚C. Thawed samples were separated without pre-boiling by SDS-PAGE, followed by fluorescence scanning using the Typhoon 9400 imager with excitation laser at 457 nm and emission filter at 526 nm. The relative abundance of precursor (CFP^{CWYC}MoaA-His_6) to splicing product (CFP-His_6) was determined by fluorescence intensity, as above. The experiments utilized a MoaA intein mutant (K57R) that arose spontaneously during culture. This mutation appeared to promote higher expression without compromising intein activity or redox sensitivity, as determined by repeat experiments with the wild-type intein.

For precursor accumulation studies, CFP-^{CWYC}MoaA-His_6 and CFP-^{AWYC}MoaA-$His_{6\ were}$ expressed in DHB4, AD494, and origami, extracted with B-PER-II, and analyzed by in-gel CFP fluorescence, as above. The experiments with CFP-^{CWYC}MoaA-His_6 and CFP-^{AWYC}MoaA-His_6 precursors were preformed in triplicate.

1. Amitai, G., Callahan, B.P., Stanger, M.J., Belfort, G. & Belfort, M. Modulation of intein activity by its neighboring extein substrates. *Proc. Natl. Acad. Sci. USA* **106**, 11005-11010 (2009).
2. Guzman, L.M., Belin, D., Carson, M.J. & Beckwith, J. Tight regulation, modulation, and high-level expression by vectors containing the arabinose PBAD promoter. *J. Bacteriol.* **177**, 4121-30 (1995).
3. McCoy, A.J. et al. Phaser crystallographic software. *J. Appl. Crystallogr.* **40**, 658-674 (2007).
4. Brunger, A.T. et al. Crystallography & NMR system: A new software suite for macromolecular structure determination. *Acta. Crystallogr. D Biol. Crystallogr.* **54**, 905-21 (1998).
5. Emsley, P. & Cowtan, K. Coot: model-building tools for molecular graphics. *Acta. Crystallogr. D Biol. Crystallogr.* **60**, 2126-32 (2004).
6. Laskowski, R.A., McArthur, M.W., Moss, D.S. & Thornton, J.M. PROCHECK: a program to check the stereochemical quality of protein structures. *J. Appl. Crystallogr.* **26**, 282-291 (1993).

Supplemental Figure 1



Supporting Figure 1. Activity of Cys point mutants in *E. coli.* (a) FRET-based intein reporter construct with the fused *Ssp* DnaE intein (red) inserted between cyan and yellow fluorescent proteins. Precursor (C-I_a-Y) is FRET active, whereas cleavage products are not. N- and C-extein residues targeted for cysteine scanning mutagenesis are circled. (b). Cys-3 variant confers redox-responsive intein activity in *E. coli*. Activity of the Cys point mutants and of the wild-type control were determined from the fraction of unreacted precursor (C-I_a-Y) remaining 7 h after induction of protein expression. The concentration of unreacted precursor (C-I_a-Y) and cleaved product (I_a-Y) were quantified by using densitometry (excitation 488 nm, emission 580 nm). Numbers below the gel images indicate the fraction of unreacted precursor (C-I_a-Y) present in the sample, calculated as the fluorescence intensity of C-I_a-Y divided by the summed intensities of C-I_a-Y and I_a-Y. These data are representative of independent experiments carried out at least twice.

Table 1 Data collection and refinement statistics

Data collection	
Space group	$P2_12_12_1$
Cell dimensions	
a, *b*, *c* (Å)	40.22,53.74,67.96
α, β, γ (°)	90.,90.,90.
Resolution (Å)	29.1-1.55(1.60-1.55)*
R_{sym} or R_{merge}	0.060(0.445)
$I / \sigma I$	11.0(2.0)
Completeness (%)	99.7(98.1)
Redundancy	3.73(3.39)
Refinement	
Resolution (Å)	1.55
No. reflections	21996
R_{work} / R_{free}	0.220/0.260
No. atoms	
Protein	1293
Ligand/ion	10
Water	209
B-factors (Å^2)	
Protein	16.1
Ligand/ion	30.1
Water	32.0
R.m.s. deviations	
Bond lengths (Å)	0.005
Bond angles (°)	1.3

*Values in parentheses are for highest-resolution shell.

Table 2 Data collection, phasing and refinement statistics (MIR)

	Crystal 1 name	Crystal 2 name
Data collection		
Space group		
Cell dimensions		
a, *b*, *c* (Å)		
α, β, γ (°)		
Resolution (Å)	##(high res shell) *	
R_{sym} or R_{merge}	##(high res shell)	
$I / \sigma I$	##(high res shell)	
Completeness (%)	##(high res shell)	
Redundancy	##(high res shell)	
Refinement		
Resolution (Å)		
No. reflections		
R_{work} / R_{free}		
No. atoms		
Protein		
Ligand/ion		
Water		
B-factors		
Protein		
Ligand/ion		
Water		
R.m.s deviations		
Bond lengths (Å)		
Bond angles (°)		

* Number of xtals for each structure should be noted in footnote. *Values in parentheses are for highest-resolution shell.

[AU: Equations defining various R values are standard and hence are no longer defined in the footnotes.]
[AU: Phasing data should be reported in the methods section]
[AU: Ramachandran statistics should be in methods section at the end of the refinement sub-section.]
[AU: Wavelength of data collection, temperature, beamline should all be in methods section.]

Table 3 Data collection, phasing and refinement statistics for MAD (SeMet) structures

	Native	Crystal 1 name			Crystal 2 name	
Data collection						
Space group		common #			common #	
Cell dimensions						
a, *b*, *c* (Å)		common #			common #	
α, β, γ (°)		common #			common #	
	Peak	*Inflection*	*Remote*	*Peak*	*Inflection*	*Remote*
Wavelength	#	#	#	#	#	#
Resolution (Å)	#	#	#	#	#	#
R_{sym} or R_{merge}	#	#	#	#	#	#
$I / \sigma I$	#	#	#	#	#	#
Completeness (%)	#	#	#	#	#	#
Redundancy	#	#	#	#	#	#
Refinement						
Resolution (Å)		common #			common #	
No. reflections						
R_{work} / R_{free}						
No. atoms						
Protein						
Ligand/ion						
Water						
B-factors						
Protein						
Ligand/ion						
Water						
R.m.s deviations						
Bond lengths (Å)						
Bond angles (°)						

* Number of xtals for each structure should be noted in footnote. *Values in parentheses are for highest-resolution shell.

[AU: Equations defining various R values are standard and hence are no longer defined in the footnotes.]
[AU: Phasing data should be reported in the methods section]
[AU: Ramachandran statistics should be in methods section at the end of the refinement sub-section.]
[AU: Wavelength of data collection, temperature, beamline should all be in methods section.]