Supporting Information

Harms et al. 10.1073/pnas.1104808108

SI Text

Crystallization Conditions, Data Collection, and Refinement of Crystal Structures. Crystal structures of the Δ+PHS/I72R, Δ+PHS/A90R, and Δ+PHS/A109R variants were solved under cryogenic conditions. The Δ+PHS/I72R, Δ+PHS/A90R, and Δ+PHS/A109R variants were crystallized by the hanging drop vapor diffusion method. Three drops, each containing a 4 μL:4 μL mixture of protein and reservoir solution, were suspended over 1 mL of reservoir solution and equilibrated at 4 °C. Δ+PHS/I72R crystals were grown using a protein solution consisting of 7.1 mg/mL protein and 2 milliequivalents of thymidine-3′,5′-diphosphate (THP). The reservoir solution consisted of 25% (vol/vol) 2-methyl-2,4-pentanediol (MPD) and 25 mM potassium phosphate at pH 8.0. Δ+PHS/A90R crystals were grown using a protein solution consisting of 8.1 mg/mL protein, 3 millequivalents of $CaCl_2$, and 1 milliequivalents of THP. The reservoir solution consisted of 24% (vol/vol) MPD and 25 mM potassium phosphate at pH 8.0. THP was synthesized in our laboratory, as described previously (1). Δ+PHS/A109R crystals were grown using an 8.3 mg/mL protein solution with a reservoir solution consisting of 42% (vol/vol) MPD and 25 mM potassium phosphate at pH 9.0.

Diffraction data for each variant were collected at cryogenic temperatures (100 K) from single crystals suspended with mother liquor in a cryoloop, then flash frozen in liquid nitrogen. Diffraction data were collected on a Kappa ApexII diffractometer outfitted with a sealed copper tube, multilayer optics, and a CCD detector (Bruker/AXS). Reflections were indexed and integrated using ApexII software, and merged using XPREP. Structure determination and refinement were performed using the ccp4i interface to the CCP4 program suite (2, 3). Molecular replacement was done with PHASER (4), using the Δ+PHS/I92E structure (PDB ID code 1TQO) as a search model (5). For each search model, solvent atoms were removed, Glu-92 was truncated to Ala, and all B factors were set to 20 $Å^2$ prior to molecular replacement. In the case of Δ+PHS/I72R, Ile-72 was truncated to Ala. In the case of Δ+PHS/A109R, residues 42–43 and 50–53 were removed from the model. Iterative model building and refinement were performed using COOT (6) and refmac5 (7). Side chains for Ile-92, the substituted arginine (Arg), and the THP ligand (for Δ+PHS/I72R and Δ+PHS/A90R) were visible after the first round of refinement. Waters, phosphates, ions, and MPD were added in later rounds of refinement. Isotropic B factors were used for the Δ+PHS/I72R and Δ+PHS/A90R refinement, whereas anisotropic B factors were used for the Δ+PHS/A109R refinement.

1. Tener GM (1961) 2-Cyanoethyl phosphate and its use in the synthesis of phosphate esters1. *J Am Chem Soc* 83:159–168.
2. Collaborative Computational Project Number 4 (1994) The CCP4 suite: Programs for protein crystallography. *Acta Crystallogr D Biol Crystallogr* 50:760–763.
3. Potterton E, Briggs P, Turkenburg M, Dodson E (2003) A graphical user interface to the *CCP* 4 program suite. *Acta Crystallogr D Biol Crystallogr* 59:1131–1137.
4. McCoy AJ, et al. (2007) Phaser crystallographic software. *J Appl Crystallogr* 40:658–674.
5. Nguyen DM, Leila Reynald R, Gittis AG, Lattman EE (2004) X-ray and thermodynamic studies of Staphylococcal nuclease variants I92E and I92K: Insights into polarity of the protein interior. *J Mol Biol* 341:565–574.
6. Emsley P, Cowtan K (2004) Coot: Model-building tools for molecular graphics. *Acta Crystallogr D Biol Crystallogr* 60:2126–2132.
7. Murshudov GN, Vagin AA, Dodson EJ (1997) Refinement of macromolecular structures by the maximum-likelihood method. *Acta Crystallogr D Biol Crystallogr* 53:240–255.

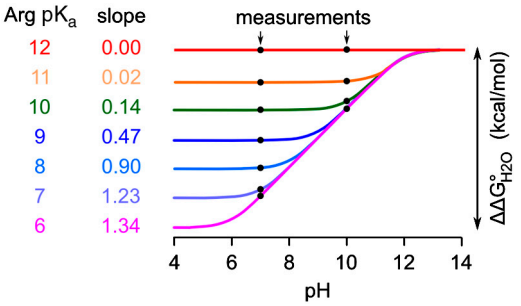

Fig. S1 Simulated curves of $\Delta\Delta G^{\circ}_{H_2O}$ vs. pH for a protein containing an Arg with a pK_a that varies between 6 (pink) and 12 (red). The linkage equation used for these simulations is $\Delta\Delta G(pH) = -RT \ln([1 + \exp(2.303(pK_a^D - pH))]/[1 + \exp(2.303(pK_a^N - pH))])$, where pK_a^N and pK_a^D are the pK_a values of Arg residues in the native and unfolded states of the protein, respectively. The value of pK_a^D was assumed to be 12. Experimental measurements were done at pH 7 and 10 (indicated with points). A slope of zero implies that the pK_a^N of the internal Arg > 10, a slope greater than 0 but less than 1.36, implies that the internal Arg has a pK_a^N between 7 and 10, and a slope of 1.36 implies that the internal Arg has a pK_a^N below 7.

Table S1. Thermodynamic parameters for GdnHCl-induced unfolding of Arg-containing variants

	pH 7						pH 10					
Variant	C_m, M*		$\Delta G^\circ_{H_2O}$, kcal mol^{-1}		m value, kcal mol^{-1} M^{-1}		C_m, M		$\Delta G^\circ_{H_2O}$, kcal mol^{-1}		m value, kcal mol^{-1} M^{-1}	
Δ+PHS	2.4	(0.0)	11.8	(0.1)	4.8	(0.1)	2.2	(0.0)	10.4	(0.1)	4.7	(0.1)
G20R	1.7	(0.0)	8.6	(0.1)	5.2	(0.1)	1.5	(0.0)	8.0	(0.1)	5.4	(0.1)
V23R	0.7	(0.0)	4.3	(0.1)	6.0	(0.1)	0.6	(0.1)	3.7	(0.1)	6.5	(0.1)
L25R	0.6	(0.0)	3.0	(0.1)	4.9	(0.1)	0.6	(0.1)	2.4	(0.1)	6.3	(0.1)
F34R	0.5	(0.0)	3.3	(0.1)	6.9	(0.1)	0.8	(0.1)	2.5	(0.1)	6.7	(0.1)
L36R	0.5	(0.0)	2.8	(0.1)	5.8	(0.1)	0.4	(0.1)	2.5	(0.1)	7.1	(0.1)
L37R	2.0	(0.0)	8.4	(0.1)	4.2	(0.1)	1.6	(0.0)	7.7	(0.1)	4.7	(0.1)
L38R	2.3	(0.0)	10.6	(0.1)	4.6	(0.1)	1.8	(0.0)	9.3	(0.1)	5.0	(0.1)
V39R	0.8	(0.0)	4.6	(0.1)	5.6	(0.1)	0.5	(0.1)	3.5	(0.1)	5.5	(0.1)
T41R	1.9	(0.0)	9.2	(0.1)	4.7	(0.1)	1.6	(0.0)	8.7	(0.1)	5.5	(0.1)
A58R	1.4	(0.0)	7.9	(0.1)	5.6	(0.1)	1.2	(0.0)	6.8	(0.1)	5.7	(0.1)
T62R	1.2	(0.0)	6.6	(0.1)	5.5	(0.1)	0.9	(0.0)	5.0	(0.1)	5.4	(0.1)
V66R	0.7	(0.0)	2.8	(0.1)	4.2	(0.1)	0.4	(0.1)	2.1	(0.1)	5.5	(0.1)
I72R	1.0	(0.0)	5.6	(0.1)	5.8	(0.1)	0.8	(0.0)	4.6	(0.1)	6.0	(0.1)
V74R	1.0	(0.1)	5.3	(0.1)	5.4	(0.1)	0.9	(0.0)	4.4	(0.1)	4.9	(0.1)
A90R	1.0	(0.0)	5.5	(0.1)	5.5	(0.1)	0.9	(0.0)	4.9	(0.1)	5.5	(0.1)
Y91R	1.1	(0.0)	5.4	(0.1)	5.0	(0.1)	0.8	(0.0)	4.7	(0.1)	5.5	(0.1)
I92R	—		—		—							
V99R†	0.3	(0.1)	1.5	(0.1)	5.7	(0.1)	0.2	(0.1)	0.7	(0.1)	6.1	(0.1)
N100R†	0.5	(0.0)	2.4	(0.1)	4.8	(0.1)	0.3	(0.1)	1.3	(0.1)	4.3	(0.1)
L103R	0.8	(0.0)	5.2	(0.1)	6.3	(0.1)	0.6	(0.0)	4.1	(0.1)	6.3	(0.1)
V104R	—		—		—							
A109R	1.5	(0.0)	8.1	(0.1)	5.6	(0.1)	1.2	(0.0)	7.0	(0.1)	5.6	(0.1)
N118R	2.0	(0.0)	10.0	(0.1)	5.0	(0.1)	1.8	(0.0)	9.2	(0.1)	5.1	(0.1)
L125R	0.5	(0.1)	2.3	(0.1)	4.8	(0.1)	0.4	(0.1)	0.5	(0.2)	4.0	(0.2)
A132R	1.0	(0.0)	4.8	(0.1)	4.7	(0.1)	0.7	(0.0)	4.5	(0.2)	4.8	(0.2)

*Values in parentheses are parameter fit errors.
†Native baseline value fixed for pH 10 fit.

Table S2. Data collection and refinement parameters for crystal structures

	Δ+PHS/I72R	Δ+PHS/A90R	Δ+PHS/A109R
Data collection			
Wavelength, Å	1.5418	1.5418	1.5418
Resolution, Å	46.6–2.0 (2.1–2.0)	38.3–2.15 (2.25–2.15)	34.3–1.9 (2.0–1.9)
Unique reflections	10,315 (1,358)	7,790 (980)	12,416 (846)
Completeness	0.997 (0.976)	1.000 (1.000)	0.997 (0.984)
Redundancy	20.8 (9.4)	15.6 (9.5)	12.1 (5.9)
Average$I/\sigma(I)$	42.5 (9.8)	36.5 (7.1)	41.5 (8.5)
R_{sym}	0.06 (0.21)	0.05 (0.26)	0.04 (0.19)
Wilson B, Å^2	23.7	38.2	23.7
Space group	$P2_12_12_1$	$P12_12_1$	$P2_12_12_1$
Cell dimensions, Å	$a = 32.181$; $\alpha = 90.00$ $b = 60.425$; $\beta = 90.00$ $c = 73.124$; $\gamma = 90.00$	$a = 30.931$; $\alpha = 90.00$ $b = 60.603$; $\beta = 93.92$ $c = 38.421$; $\gamma = 90.00$	$a = 34.090$;$\alpha = 90.00$ $b = 59.843$; $\beta = 90.00$ $c = 73.558$; $\gamma = 90.00$
Refinement			
Resolution, Å	46.6–2.0 (2.1–2.0)	30.3–2.15 (2.25–2.15)	31.1–1.9 (2.0–1.9)
No. of nonhydrogen atoms	1,046	1,039	1,075
No. of unique reflections	10,218 (1,019)	7,757 (750)	12,337 (1,215)
No. of reflections in test set	1,019 (62)	751 (76)	1,215 (91)
R_{work}	0.200 (0.187)	0.208 (0.220)	0.191 (0.170)
R_{free}	0.258 (0.218)	0.276 (0.300)	0.252 (0.264)
rms distance from ideal geometry			
Bonds, Å	0.008	0.012	0.015
rms angles, °	1.23	1.30	1.43
Average B factors, Å^2			
Protein	12.5	32.4	15.5
Solvent	18.9	34.6	23.3
Ion	11.3	21.4	42.8
Ramachandran plot			
Most favored	102 (88.7)	96 (84.2)	98 (83.8)
Additionally allowed, %	12 (10.4)	17 (14.9)	18 (14.4)
Generously allowed, %	0 (0.0)	0 (0.0)	0 (0.0)

	Δ+PHS/I72R	Δ+PHS/A90R	Δ+PHS/A109R
Disallowed, %	1 (0.9)	1 (0.9)	1 (0.9)
No. of nonglycine, nonproline, and nonend residues	115	114	117
No. of glycine, proline, and end residues	15	15	15
Total no. of residues *	130	129	132
PDB ID Code	3D8G	3DHQ	3D4W

*Values in parentheses correspond to the highest resolution shell.

Table S3. Microenvironments of guanidinium moieties of Arg residues

Residue	Atom	Solvent accessible surface area, Å^2	Fractional accessibility*	Interactions	Distance, Å
Arg-72	Nε	0.1	0.3%	Val-66 O	2.9
	Nη1	32.7	54.8%	water-187	2.8
	Nη2	26.3	51.9%	Glu-67 O	2.9
Arg-90	Nε	0.1	0.7%	water-216	2.9
	Nη1	0.0	0.0%	Arg-81 O	2.7
				Leu-89 O	3.0
	Nη2	8.4	16.5%	Tyr-27 OH	3.3
				water-179	2.9
Arg-109	Nε	0.0	0.0%	Asp-21 O δ_1	2.8
	Nη1	0.6	1.0%	Asp-21 O δ_1	3.1
				water-168	3.1
	Nη2	12.3	24.3%	Leu-108 O	3.0
				water-222	2.8
				water-223	3.3

*Accessibility calculated relative to the atom in an extended, blocked Ala-Arg-Ala peptide.