Supporting Information

Crystal structures of the LsrR proteins complexed with phospho-AI-2 and two signal-interrupting analogs reveal distinct mechanisms for ligand recognition

Jung-Hye Ha[1,2], Yumi Eo[1,2], Alexander Grishaev[3], Min Guo[4], Jacqueline A. I. Smith[4], Herman O. Sintim[4], Eun-Hee Kim[1], Hae-Kap Cheong[1], William E. Bentley[5,6], Kyoung-Seok Ryu[1,2,7]

[1]Division of Magnetic Resonance Research, Korea Basic Science Institute, Yangcheong-Ri 804-1, Ochang-Eup, Cheongwon-Gun, Chungcheongbuk-Do 363-883, [2]Department of Bio-Analytical Science, University of Science and Technology, Daejon 305-333, South Korea. [3]Laboratory of Chemical Physics, National Institute of Diabetes and Digestive and Kidney Diseases, National Institutes of Health, Bethesda, Maryland 20892, [4]Department of Chemistry and Biochemistry, [5]Graduate Program in Molecular and Cell Biology, [6]Fischell Department of Bioengineering, University of Maryland, College Park, Maryland 20742, USA.

[7]Correspondence to: Kyoung-Seok Ryu, E-mail (ksryu@kbsi.re.kr)

1 Table & 6 Figures

Supporting materials and methods

References

Table S1. Statistics of X-ray data and structure refinements for the ligand-free and -bound C-LsrR and LsrR proteins.

	C-LsrR (SeMet)	C-LsrR	C-LsrR/R5P	C-LsrR/D5P'
Resolution (Å)	50.0~2.4 (2.49~2.40)	50.0~1.9 (1.97~1.90)	50.0~1.9 (1.97~1.90)	50.0~2.3 (2.38~2.30)
Space group	P65	P65	P65	P65
Unit-cell parameters	116.34, 116.34, 79.99,	115.81, 115.81, 80.89,	116.61, 116.61, 79.74,	117.69, 117.69, 79.37,
(a, b, c, α, β, γ)	90.0, 90.0, 120.0	90.0, 90.0, 120.0	90.0, 90.0, 120.0	90.0, 90.0, 120.0
Measured reflections		214,246	294,245	133,082
Unique reflections (#)	23,254	46,931	48,351	26,903
# of proteins in asymmetric unit	2	2	2	2
Redundancy	14.9 (7.4)	4.6 (2.6)	6.1 (4.6)	4.9 (3.0)
R_{merge} (%)	8.2 (31.7)	5.9 (33.3)	6.8 (34.7)	5.9 (26.1)
Mean I/Iσ	16.9 (4.9)	15.4 (2.4)	11.8 (3.7)	12.4 (2.1)
Completeness (%)	96.9 (88.4)	96.4 (90.1)	99.4 (98.4)	96.6 (91.3)
R_{cryst}/R_{free}		0.179/0.222	0.168/0.204	0.175/0.233
RMSD bond length (Å)		0.019	0.019	0.016
RMSD bond angle (°)		1.926	2.060	1.942
Ramachandran plot				
Favored (%)		98.29	98.77	92.93
Outliers (%)		0.64	0.62	2.42[b]
Poor rotamers		4.75	4.80	8.22
Average B factor (Å^2)				
Chain A		31.92	25.58	41.52
Chain B		31.53	24.18	40.37
Chain C				
Chain D				
Water		39.29	32.24	42.07
Ligand		-	28.06	41.11

	C-LsrR/D8P'	LsrR/R5P	LsrR
Resolution (Å)	50.0~2.1 (2.18~2.10)	50.0~2.6 (2.69~2.60)	50.0~3.2 (3.31~3.20)
Space group	P65	C2	C2
Unit-cell parameters	116.81, 116.81, 79.98,	189.33, 80.61, 118.70,	176.74, 80.54, 103.64,
(a, b, c, α, β, γ)	90.0, 90.0, 120.0	90.0, 118.0, 90.0	90.0, 112.2, 90.0
Measured reflections	177,355	145,729	82,311
Unique reflections (#)	35,349	45,073	19,907
# of proteins in asymmetric unit	2	4	4
Redundancy	5.0 (3.1)	3.2 (1.8)	4.1 (2.6)
R_{merge} (%)	5.2 (23.4)	6.2 (32.0)	9.0 (29.2)
Mean I/Iσ	19.9 (2.5)	19.41 (2.05)	11.1 (2.0)
Completeness (%)	97.3 (92.5)	92.9 (78.4)	90.3 (79.1)
R_{cryst}/R_{free}	0.169/0.218	0.217/0.277	0.224/0.289
RMSD bond length (Å)	0.019	0.007	0.009
RMSD bond angle (°)	1.983	1.250	1.310
Ramachandran plot			
Favored (%)	96.92	92.65	82.23
Outliers (%)	0.82	0.41	5.00
Poor rotamers	6.93	5.31	10.24
Average B factor (Å^2)			
Chain A	35.42	62.68	63.98
Chain B	36.26	67.74	63.13
Chain C		63.63	63.66
Chain D		67.71	63.83
Water	38.84	45.69	-
Ligand	33.01	62.29	-

[a]R5P, ribose-5-phosphate; D5P, phospho-(S)-4,5-Dihydroxy-2,3-pentandione; D8P, phospho-isobutyl DPD; D5P' and D8P', the hydrated forms of D5P and D8P, respectively. The final structure refinements of all C-LsrR proteins were done with the Refmac5 program, but those of the LsR proteins were done with the CNSsolve program.

[b]The $2F_oF_c$ map in the region of α7 segment was not clear, and thus the dihedral angles of several residues in this region were not clearly defined.

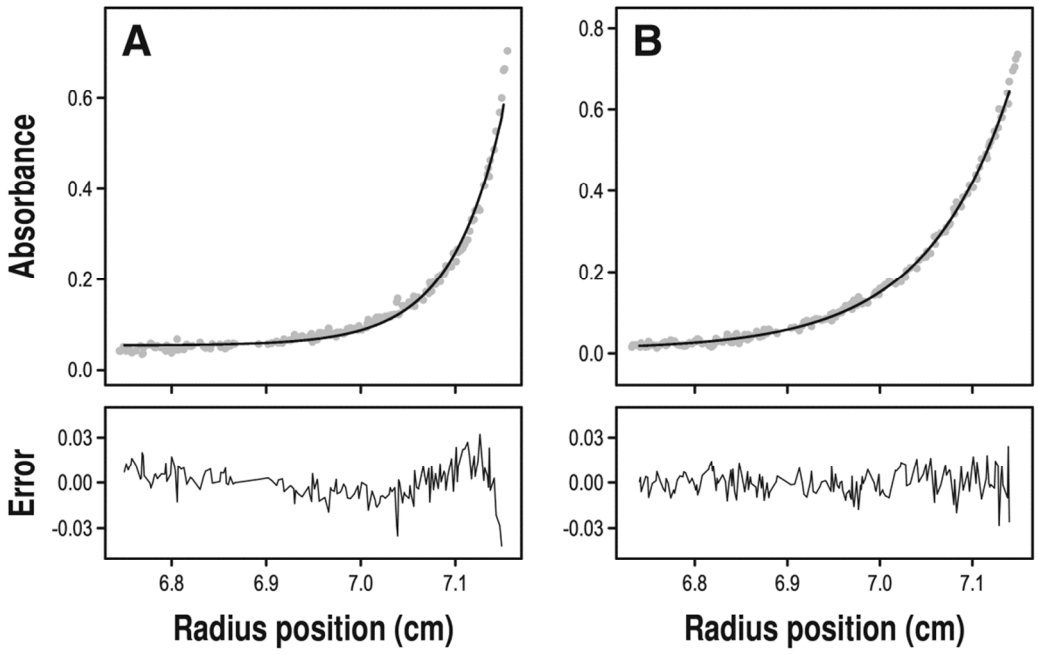

Figure S1. Absolute MW values of the LsrR (A) and C-LsrR (B) proteins were estimated by sedimentation equilibrium AUC experiments. (A) The AUC radius profile of LsrR does not agree with a single component but rather indicates the presence of an oligomeric equilibrium. Nevertheless, if applied, the single component analysis shows that 20 μM LsrR mainly exists as a dimer. The estimated MW of LsrR varied from 69.6 to 62.4 kDa when the partial specific volume was changed from 0.730 to 0.700, respectively. (B) The profile of C-LsrR (40 μM) is described well by a single component. Its molecular weight varied from 36.0 to 32.4 kDa when decreasing its partial specific volume from 0.730 to 0.700, respectively. Although the obtained MW values of C-LsrR are higher than the theoretical MW of C-LsrR monomer (28.0 kDa), these values are still lower than that of the dimeric C-LsrR (56.0 kDa). The presence of the long and unstructured N-terminal tail in the C-LsrR protein (residues 53-66) might be the reason for the decrease in the partial specific volume of the C-LsrR protein.

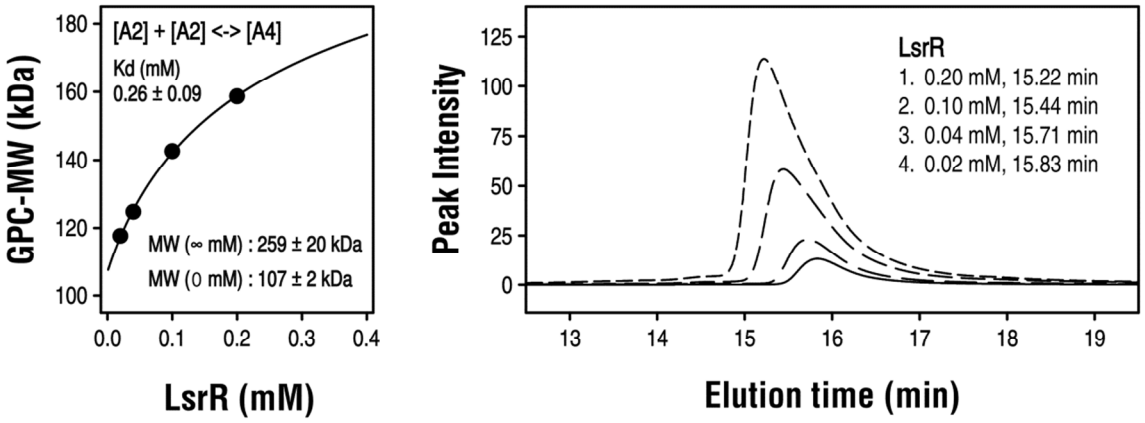

Figure S2. Concentration-dependent molecular sizes of the LsrR protein. GPC analysis was done using a Wyatt GPC column (WTC-03005) at varied LsrR concentration. **(Right panel)** The elution of LsrR became faster with the increase in the concentration, meaning that an apparent molecular size (MW^{GPC}) of the LsrR increased at higher concentration. **(Left panel)** The concentration-dependent MW^{GPC} values of LsrR were analyzed using a simple monomer-to-dimer equilibrium model (dimer-to-tetramer equilibrium in our case). The measured MW^{GPC} is similar to the intensity-averaged MW since one dimeric protein has the same UV absorption as two momomers. The $MW_M{}^{GPC}$ and $MW_D{}^{GPC}$ are apparent molecular sizes of the dimeric and tetrameric LsrR proteins, respectively.

(1) $$MW^{GPC} = \frac{[M]}{P_T} \cdot MW_M^{GPC} + \frac{2 \cdot [D]}{P_T} \cdot MW_D^{GPC} = \left(1 - \frac{2 \cdot [D]}{P_T}\right) \cdot MW_M^{GPC} + \frac{2 \cdot [D]}{P_T} \cdot MW_D^{GPC}$$

(2) $$P_T = [M] + 2 \cdot [D]$$

The monomer-to-dimer equilibrium process can be implemented into the MW^{GPC} equation that describes the concentration-dependent molecular sizes.

(3) $$K_d = \frac{[M] \cdot [M]}{[D]} = \frac{(P_T - 2 \cdot [D]) \cdot (P_T - 2 \cdot [D])}{[D]}, \quad \frac{2 \cdot [D]}{P_T} = \left(1 + \frac{K_d}{4 \cdot P_T}\right) - \sqrt{\left(1 + \frac{K_d}{4 \cdot P_T}\right)^2 - 1}$$

(4) $$MW^{GPC} = \left(\left(1 + \frac{K_d}{4 \cdot P_T}\right) - \sqrt{\left(1 + \frac{K_d}{4 \cdot P_T}\right)^2 - 1}\right) \cdot \left(MW_D^{GPC} - MW_M^{GPC}\right) + MW_M^{GPC}$$

The final equation can be used for fitting the SEC data of LsrR, in which P_T is the concentration of the LsrR dimer. Although we could not obtain the elution time of the LsrR protein at concentrations higher than 0.2 mM due to its limited solubility, our data suggest an equilibrium between the dimer and tetramer. The K_d values obtained by fitting to the above equation could be higher than the real K_d value since the protein bands dissipated during the elution through the SEC column.

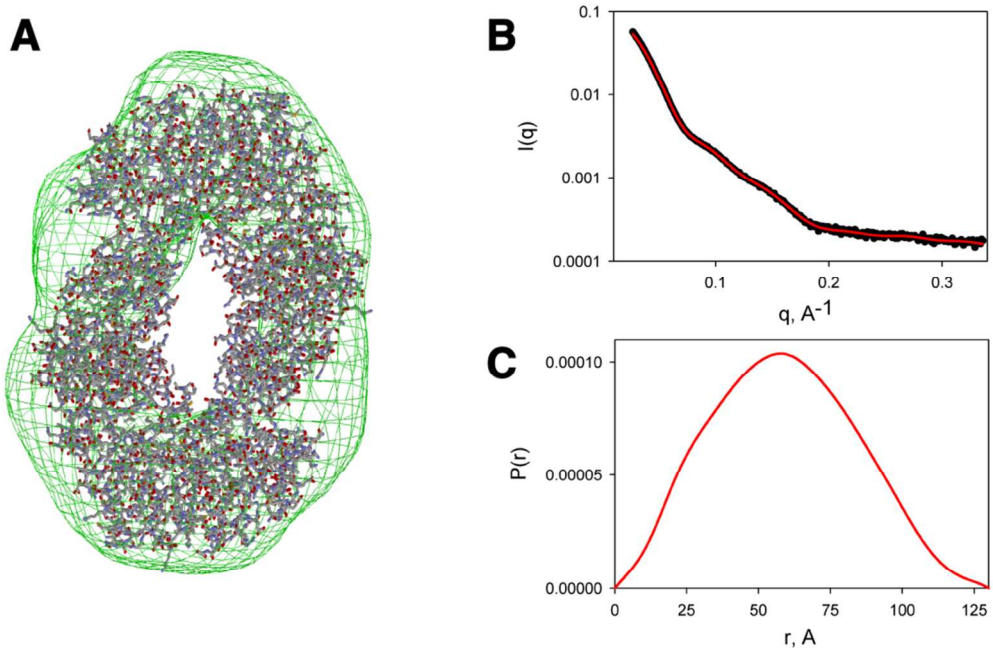

Figure S3. Analysis of SAXS data for 100 μM LsrR protein. The SAXS data were Fourier-transformed using the GNOM program[1] up to q_{max}=0.34Å^{-1}, yielding a radius of gyration of 45 Å and a maximum dimension of 130 Å. The fitted P(r) distribution hints at a hollow particle shape as evident from the nearly symmetric appearance of the curve. (A) The low-resolution model is superimposed with the crystal structure of the LsrR tetramer. (B) The scattering data of 100 μM LsrR (black dots) was fitted by using the GNOM program (red line) showing the corresponding distribution of the inter-atomic distances within the particle. (C) The envelope of the low-resolution reconstructions was determined via the DAMMIN program[2] and was aligned via SUPCOMB[3].

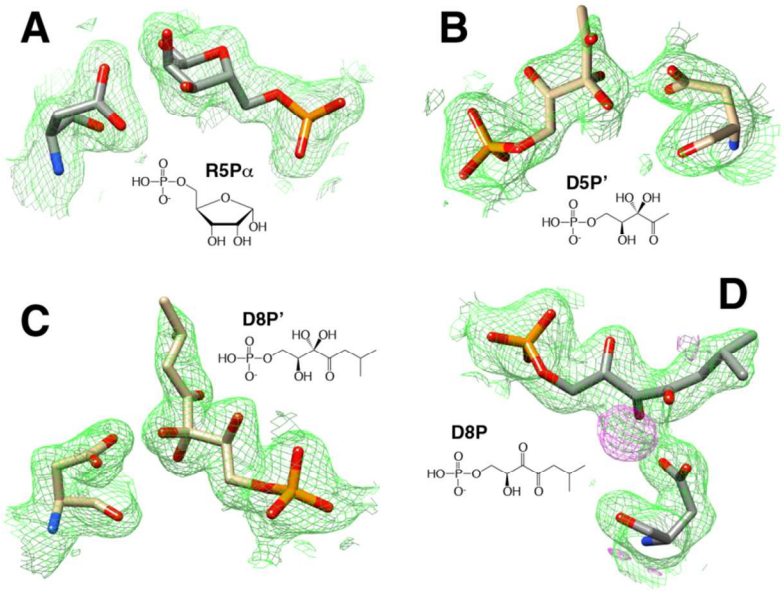

Figure S4. Electron-density maps of the bound ligands. (A) The 1.9 Å resolution $2F_o$-F_c map clearly shows that the bound R5P is a cyclic α-anomer (R5Pα). The $2F_oF_c$ maps of 2.3 and 2.1 Å resolution show that both bound D5P' (B) and D8P' (C) are not diketone forms but rather the hydrated forms (D5P' and D8P', respectively). The $2F_o$-F_c map of D5P' is less clear than that of D8P', which likely resulted from the lower occupancy of D5P' compared to D8P'. (D) The fitting of D8P molecule into the electron-density map shows the presence of an extra-map in the region of C3 atom (the pink colored F_o-F_c map).

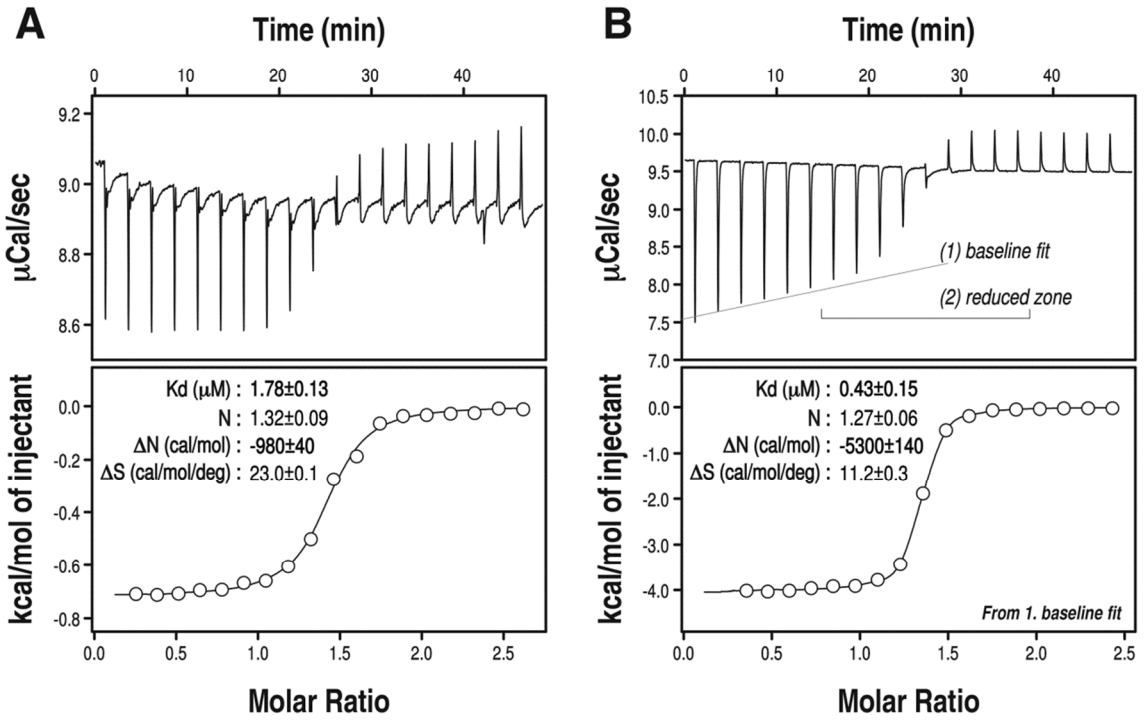

Figure S5. The binding affinities of LsrR were estimated for D5P (A) and D8P (B), respectively, by the ITC experiments. The raw data of heat changes following the time-coursed injections of the ligand are shown in the upper panel, and the processed data appear in the lower panel. The amount of the released heat by the binding of D5P to LsrR was too small to analyze via ITC. Therefore, the concentration of LsrR was increased to 176 μM for the experiment with D5P, and to 158 μM with D8P. The adjustments of baseline were done manually to for more reliable integrations of the heat changes. The blank injection of D5P or I5P solutions in the sample cell without the LsrR protein resulted in a constant and positive heat changes. This effect was subtracted before the analysis. (B) The initial heat release by the addition of D8P was not constant and linearly decreased before the main transition. The same ITC pattern was consistently identified for the separate experiment. Therefore, we obtained thermodynamic parameters for the binding of LsrR and D8P by using the linear correction of the baseline (1) or by reducing the analyzed zone (2).

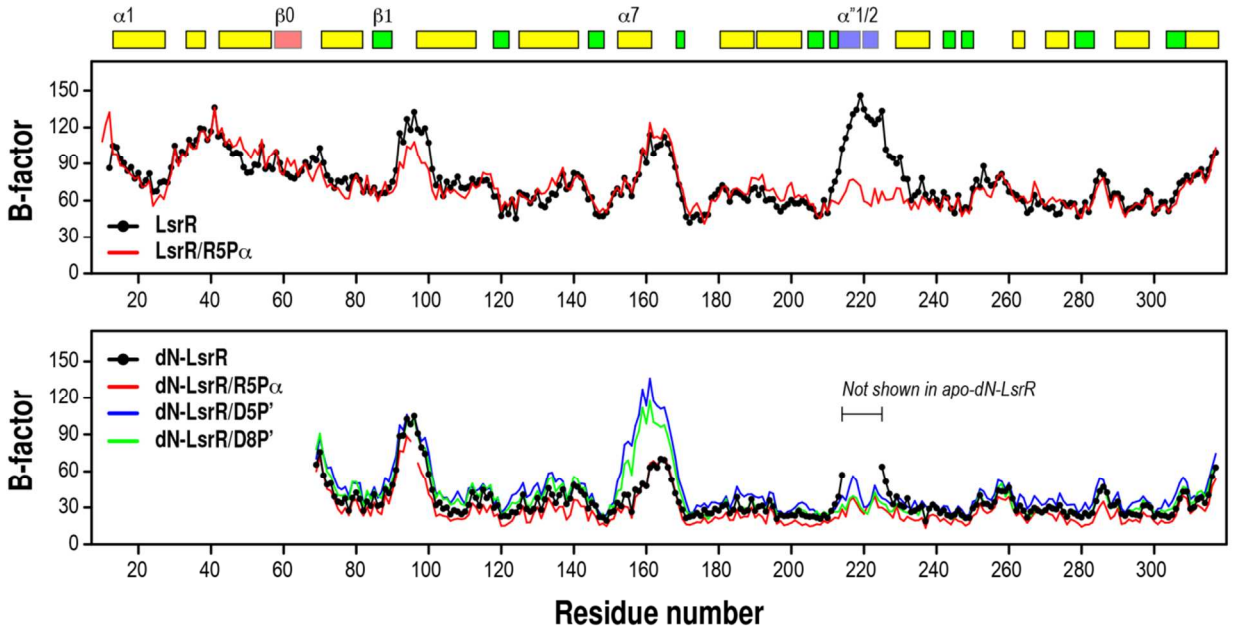

Figure S6. The averaged B-factor values were calculated for the subunit-A molecules of all determined structures. The boundaries of the secondary structures are presented for the subunit-A molecule of the LsrR/R5P structure.

MATERIALS AND METHODS

Protein expression and purification. The full length (1-317) and N-terminal deleted (53-317 & 61-317) *lsrR* genes were cloned into the pGEX-4T-1 vector (GE Healthcare) using restriction enzymes (*BamH* I and *Xho* I). All point mutations were generated using the Quick-Change method (Stratagene). The LsrR-expression plasmid was transformed into *E. coli* BL21 (DE3). Overnight seed culture was diluted 1:100 into fresh LB media containing 100 μg/mL ampicillin, 0.4% glucose, 0.25 M NaCl, and then was incubated at 37°C until the OD600 reached to 0.5. To increase the cellular protein before the induction[4], the culture was incubated at 42°C for 1-hr and then was quickly cooled on ice or 4°C. The protein expression was induced with 0.5 mM isopropyl β-D-thiogalactopyranoside (IPTG) for 20-hrs at 20°C. L-selenomethionine (SeMet)-substituted C-LsrR (residues 53-317) was obtained by growing *E coli* cells in M9 medium supplemented with 50 μg/ml SeMet (TCI), and each 100 μg/ml of Thr, Phe, Leu, Ile, Lys, and Val were additionally used to suppress the methionine bio-synthesis[5].

The harvested cells were re-suspended in buffer (pH 7.0) containing 50 mM Tris-HCl, 500 mM NaCl, 2 mM DTT, and 0.1% tween-20. 1.0 mM benzamidine and 1.0 mM ABESF were also added to suppress potential protease activity. The GST-fusion proteins were purified by GST-affinity column chromatography. The protein was eluted with buffer (pH 7.5) containing 25 mM Tris-HCl, 350 mM NaCl, and 10 mM L-glutathione (GSH). 5% glycerol was additionally added to the protein elution for decreasing the aggregation during the thrombin-digestion. After the thrombin-digestion for overnight at 4°C, the process was stopped by adding 0.5 mM PMSF or ABESF (Sigma-Aldrich). LsrR and C-LsrR were finally purified by gel permeation chromatography in buffer (pH 8.0, 10 mM Tris-HCl, 200 mM NaCl and 1 mM DTT) using HiLoad 16/60 Superdex 200 and HiLoad 16/60 Superdex 75 column (GE Healthcare), respectively. The remaining GST-tag was removed by passing the protein eluents through a GST-affinity column.

Preparation of phospho-(S)-4,5-Dihydroxy-2,3-pentanedione (D5P) and phospho-isobutyl-DPD (D8P). D/L-mixture form of DPD and isobutyl-DPD were synthesized following the previous reported method[6]. D5P and D8P were synthesized by an enzyme reaction using LsrK protein in buffer (pH 7.5, 5 μM LsrK, 10 mM DPD, 20 mM Tris-HCl, 100 mM NaCl, 1 mM $MgCl_2$, 6.5 mM ATP, and 10% D_2O). 0.2 mM 4,4-dimethyl-4-silapentane-1-sulfonic

acid (DSS) was also added to the reaction mixture as an internal concentration reference. Detailed method for the preparation of LsrK protein will be published elsewhere. Progress of the phosphorylation reaction was monitored by measuring the time-course of ^{1}H-NMR spectra using 800 MHz NMR spectroscopy (KBSI); we confirmed that half of DPD was completely converted into D5P. The reaction mixtures were frozen in liquid nitrogen and then were stored at -70°C for later usages.

X-ray diffraction data collections and model buildings. Diffraction data were collected at 100K at beamlines BL-1A and BL-17A of the Photon Factory (Tsukuba, Japan). The single wavelength anomalous dispersion (SAD) data of SeMet-substituted C-LsrR was collected at a wavelength of 0.9789 Å and its diffraction resolution was 2.4 Å. All data sets were processed using the HKL-2000 program package[7].

The analysis of SAD data and the automatic building of the initial structure were performed using the Solve/Resolve programs[8]. Detailed model building and structure refinement were done using the Coot program[9] and Refmac5 in the CCP4 program package[10], respectively. The structure of LsrR/R5P was first solved via molecular replacement (MR) by the Phaser program[11], since its diffraction resolution was much higher than that of the native LsrR crystal. The topology files of R5P, D5P, and D8P were prepared using the Dundee PRODRG web server (http://davapcl.bioch.dundee.ac.uk/prodrg)[12] and the final structure refinements of LsrR proteins were finished using the CNSsolve1.3 program[13] and those of C-LsrR proteins were done using the Refmac5 program to handle the alternative conformations[14]. The populations of the alternative conformations were estimated from the structure refinement using the Phenix program[15]. The evaluations of the calculated structures were performed *via* the Molprobity program[16]. Analysis of all structures and the generation of all figures were done using the Chimera program.

Determination of molecular size of various LsrR proteins. Apparent molecular sizes of proteins were first determined by size exclusion chromatography (SEC) using WTC-03005 SEC Protein column (7.8×300 mm) (Wyatt Technology) with 0.5 ml/min flow rate. The samples were eluted with buffer (pH 8.0, 25 mM Tris-HCl and 200 mM NaCl). The elution times of five different proteins (19.72 min, 17 kDa, horse myoglobin; 17.82 min, 44 kDa, chicken ovalbumin; 15.48 min, 158 kDa, bovine γ-globulin; 12.12 min, 670 kDa, bovine thyroglobulin) were used

for the reference curve of molecular size.

Absolute molecular weight of the native LsrR (20 μM) and C-LsrR (40 μM) were determined by analytical ultra-centrifugation (AUC). Sedimentation equilibrium experiments (18,000 rpm at 4°C) were performed using the ProteomeLab XL-1 analytical ultracentrifuge (Beckman Coulter) equipped with an An-60Ti analytical rotor in buffer (pH 7.5, 50 mM Tris-HCl and 100 mM NaCl). The curves of the protein concentration gradient were measured by using absorbance at 280 nm. The analysis of the AUC data were done using the SEDPHAT program[17].

Dynamic light scattering (DLS) experiments were used to monitor the molecular size of LsrR in the presence of D5P. DLS experiments were performed using a Viscotek 802 DLS (Houston, TX) at 20°C. LsrR samples (0.5 and 1.0 mg/ml) were prepared in buffer (pH 8.0, 50 mM Tris-HCl, 100 mM NaCl and 5 mM 2-mercaptoethanol) and filtered through a 0.22 μm centrifugal filter (Millipore) before the measurement. The concentrations of D5P, D8P and R5P that were 0.2, 0.2 and 20 mM, respectively.

Small angle light scattering (SAXS) data collection. Solution X-ray scattering data for LsrR were acquired at protein concentrations of 15 and 100 μM at the Beam Line 12-IDB and 12-IDC, Advanced Photon Source (Argonne National Laboratory, Argonne, IL). At station 12-IDC, data collection was done using Pilatus 2M detector positioned 3.04 m from the sample capillary in a highly offset geometry with 12 keV incident radiation resulting in the observable q-range of 0.01~0.70 $Å^{-1}$. Scattered radiation was detected subject to an 11 keV low-energy cutoff. At station 12-IDC, data collection was carried out using a mosaic Gold CCD detector positioned in an on-center geometry, 3.08m and 0.48m from the sample capillary using 18 keV incident radiation, resulting in the observable q-range of 0.01~0.21 $Å^{-1}$ for the small-angle and 0.10~0.23 $Å^{-1}$ for the wide-angle data. Q-axis mapping was done using silver behenate standard samples. Totals of 20 sequential data frames with exposure times of 20 seconds (12-IDB station) and 2 seconds (12-IDC station) were recorded with the samples kept at 25°C throughout the measurement. In order to prevent radiation damage, volumes of 100 μl of samples and buffers were oscillating during data collection. Individual data frames were masked, corrected for the detector sensitivity, radially integrated and normalized by the corresponding incident beam intensities and sample transmissions. The final 1D scattering profiles and their uncertainties were calculated as means and mean uncertainties over the 20 individual frames. The buffers data were then subtracted from the samples.

REFERENCES

(1) Svergun, D. I. *J. Appl. Cryst.* **1992**, *25*, 495-503.

(2) Svergun, D. I. *Biophys. J.* **1999**, *76*, 2879-86.

(3) Kozin, M. B.; Svergun, D. I. *J. Appl. Cryst.* **2001**, *34*, 33-41.

(4) Oganesyan, N.; Ankoudinova, I.; Kim, S. H.; Kim, R. *Protein Expr. Purif.* **2007**, *52*, 280-5.

(5) Harrison, C. J.; Bohm, A. A.; Nelson, H. C. *Science* **1994**, *263*, 224-7.

(6) Roy, V.; Smith, J. A.; Wang, J.; Stewart, J. E.; Bentley, W. E.; Sintim, H. O. *J.Am. Chem. Soc.* **2010**, *132*, 11141-50.

(7) Otwinowski, Z.; Minor, W. *Methods Enzymol*. **1997**, 276, 307–326.

(8) Terwilliger, T. C. *Methods Enzymol.* **2003**, *374*, 22-37.

(9) Emsley, P.; Cowtan, K. *Acta Crystallogr. D Biol. Crystallogr.* **2004**, *60*, 2126-32.

(10) Collaborative Computational Project, Number 4 *Acta Crystallogr. D Biol. Crystallogr.* **1994**, *50*, 760-3.

(11) McCoy, A. J.; Grosse-Kunstleve, R. W.; Adams, P. D.; Winn, M. D.; Storoni, L. C.; Read, R. J. *J. Appl. Crystallogr.* **2007**, *40*, 658-674.

(12) Schüttelkopf , A. W.; van Aalten, D. M. *Acta crystallogr.D Biol. Crystallogr.* **2004**, *60*, 1355-63.

(13) Brunger, A. T.; Adams, P. D.; Clore, G. M.; DeLano, W. L.; Gros, P.; Grosse-Kunstleve, R. W.; Jiang, J. S.; Kuszewski, J.; Nilges, M.; Pannu, N. S.; Read, R. J.; Rice, L. M.; Simonson, T.; Warren, G. L. *Acta Crystallogr. D Biol. Crystallogr.* **1998**, *54*, 905-21.

(14) Murshudov, G. N.; Skubak, P.; Lebedev, A. A.; Pannu, N. S.; Steiner, R. A.; Nicholls, R. A.; Winn, M. D.; Long, F.; Vagin, A. A. *Acta Crystallogr. D Biol Crystallogr.* **2011**, *67*, 355-67.

(15) Adams, P. D.; Afonine, P. V.; Bunkoczi, G.; Chen, V. B.; Davis, I. W.; Echols, N.; Headd, J. J.; Hung, L. W.; Kapral, G. J.; Grosse-Kunstleve, R. W.; McCoy, A. J.; Moriarty, N. W.; Oeffner, R.; Read, R. J.; Richardson, D. C.; Richardson, J. S.; Terwilliger, T. C.; Zwart, P. H. *Acta Crystallogr. D Biol. Crystallogr.* **2010**, *66*, 213-21.

(16) Chen, V. B.; Arendall, W. B., 3rd; Headd, J. J.; Keedy, D. A.; Immormino, R. M.; Kapral, G. J.; Murray, L.

W.; Richardson, J. S.; Richardson, D. C. *Acta Crystallogr. D Biol. Crystallogr.* **2010**, *66*, 12-21.

(17) Schuck, P. *Anal. Biochem.* **2003**, *320*, 104-24.