Supporting Information

Malojčić *et al.* 10.1073/pnas.0806997105

SI Methods

Protein Concentrations. The concentration of ASST was determined according to ref. 1 via its specific absorbance at 280 nm (ε = 93350 $M^{-1}\cdot cm^{-1}$ for the ASST monomer). All concentrations in this paper are expressed per monomer.

Molecular Cloning, Protein Expression, and Purification. Cloning and expression of ASST was performed as described (2). ASST variants were constructed with the QuikChange site-directed mutagenesis kit (Stratagene) by using the plasmid for coexpression of ASST, DsbL, and DsbI (2) as a template and the mutagenesis primers listed in Table S3. After mutagenesis, the whole DNA sequence encoding the ASST-DsbL-DsbI operon was verified. ASST and its variants were purified as described (2). Fractions containing pure ASST were pooled, concentrated to 350 μM, and stored at 4 °C. Protein identity and purity was verified by mass spectrometry, Edman sequencing, and activity assays.

Selenomethionine-labeled ASST was expressed in the same bacterial strain by using metabolic inhibition in M9 minimal medium supplemented with ampicillin, glucose, and all amino acids, but with selenomethionine instead of methionine, as described in detail in ref. 3. Because of reduced rates of cell growth, a prolonged expression time (24 h) was used. All subsequent purification steps were identical to those for the nonlabeled protein.

Crystallography. ASST crystals were obtained using the sitting drop vapor-diffusion method by equilibrating 1.5 μL of protein solution (22 mg/mL in 20 mM 4-morpholinepropanesulfonic acid/NaOH pH 7.5, 100 mM NaCl) with 0.5 μL of reservoir solution consisting of 1.8 M Li_2SO_4 and 100 mM cacodylic acid/NaOH pH 6.5. Crystals were cryoprotected by dragging them through paraffin oil (Hampton Research) before flash-cooling in a nitrogen cryostat. Native X-ray diffraction data were collected at beamline X10SA of the Swiss Light Source (SLS), at a wavelength of 1.000 Å. Single wavelength anomalous data were collected from a selenomethionine-labeled protein crystal at the Se K-edge (0.978 Å) at beamline X06SA of the SLS. Diffraction data were integrated using MOSFLM (4) and scaled using SCALA (5). All of the expected 16 selenium sites (8 per monomer) were found by using Phaser (6), and experimental phases were then extended and improved by using solvent-flattening and noncrystallographic symmetry (NCS) averaging in the CCP4 program DM (7). The resulting electron density maps were readily interpretable, and ARP/wARP (8) was used to build an initial model, which was then further improved and refined by using Coot (9), Refmac (10), and Phenix (11) (R_{work}/R_{free} = 0.207/0.248). PROCHECK (12) indicated that 86.5% of residues were located in the most favored regions of the Ramachandran plot, and 12.9% and 0.2% were located in additionally and generously allowed regions, respectively. Notably, the only residues in the disallowed regions were His-252 and Ser-502, forming the active site.

Crystals of catalytic intermediates of ASST were obtained by soaking native crystals with a modified reservoir solution containing the substrate (PNS or MUS) at a concentration of 5 mM. These solutions were added stepwise and left to equilibrate for 15 min before cryo-cooling as described above. X-ray diffraction data were collected at beamline X06SA of the SLS. Initial maps and models of ASST intermediates were obtained by refinement of the substrate-free ASST against crystallographic data from the soaking experiments using Refmac (10). Further refinement, including the occupancy of the sulfuryl donor, was performed by using Phenix (11) and Coot (9) (R_{work}/R_{free} = 0.187/0.232 and 0.175/0.216 for PNS and MUS soaks, respectively). All X-ray data collection, phasing, and refinement statistics are presented in Table S1. Figures were prepared using Pymol (13).

Electrospray Ionization Mass Spectrometry (ESI-MS). ESI-MS of ASST was performed at the Functional Genomics Center Zurich by using standard measurement techniques. Protein samples were desalted before recording of mass spectra by adsorption to ZipTips (Millipore), and eluted and analyzed in 50% acetonitrile/0.2% formic acid (pH 2).

ASST-Catalyzed Reactions. Steady state kinetics experiments, as described in *Methods* in the main text, were performed at pH 6.0, 7.0, 8.0, 8.5, 9.0, and 10.0. The initial rates were measured and evaluated as described in *Methods*. A mixed buffer system composed of Mes, Pipes, Hepes, and glycine (20 mM each), adjusted with HCl or NaOH to the respective pH, was used for all measurements.

To probe the kinetic stability of sulfo-ASST, we followed the reaction of ASST (10 μM) with excess *p*-nitrophenylsulfate (PNS) (100 μM) as sulfuryl donor at pH 7.0. After a fast initial reaction in the dead time of manual mixing (corresponding to the stoichiometric sulfurylation of ASST), no significant further increase in the absorbance at 405 nm (formation of *p*-nitrophenylate) could be detected (rate of PNS hydrolysis $<10^{-8} M\cdot s^{-1}$ under these conditions), demonstrating that spontaneous hydrolysis of sulfo-ASST can be neglected on the time scale of the enzymatic assays.

Structure-Based Sequence Alignments and Evolutionary Analyses. To identify conserved regions in the structure of ASST, a BLAST search was performed using the amino acid sequence of the mature ASST from *E. coli* CFT073 [algorithm blastp (14) using default parameters as implemented on the National Center for Biotechnology Information (NCBI) server]. The search against all nonredundant GenBank coding sequence translations, Protein Database, SwissProt, Protein Information Resource, and Protein Research Foundation entries provided a number of homologous protein sequences. The entries with an *E*-value $<10^{-100}$ were retrieved and further analyzed. For the multiple sequence alignment, redundant hits were excluded, as well as proteins that were too dissimilar (with a different function assigned or with <40% sequence identity). Multiple sequence alignment was generated by using ClustalW (15). Structure-based sequence alignment was prepared in ESPript 2.2 (http://espript.ibcp.fr/ESPript/cgi-bin/ESPript.cgi), by using the alignment file generated in ClustalW (aln) and the atomic coordinate file (pdb).

To obtain an independent evolutionary insight into the conserved regions of ASST based on its structure and genomic sequence, a phylogenetic analysis using the program Selecton (http://selecton.bioinfo.tau.ac.il/) was performed (16) with default parameters using the DNA sequences of the homologous proteins together with the coordinates of ASST. For a detailed explanation, see ref. 16. The sequences of ASST from *Citrobacter freundii*, *Salmonella enterica* subsp. enterica serovar Typhi str. CT18, *Enterobacter amnigenus*, *Shewanella pealeana* ATCC 700345, *Campylobacter fetus* subsp. fetus 82–40, *Campylobacter curvus* 525.92, *Azobacter vinelandii* AvOP, *Pseudomonas putida*, *Pseudomonas entomophila* L48, *Campylobacter lari* RM2100, *Geobacter metallireducens* GS-15, and *Yersinia frederiksenii* ATCC 33641 were included in this analysis.

Analytical Gel Filtration Chromatography. ASST (10 μM) was incubated with PNS (5 mM) or phenol (5 mM) in 100 mM phosphoric acid/NaOH pH 7, 0.1 mM EDTA. PNS or phenol were removed by iterative dilution and concentration by ultrafiltration (Amicon YM50, Millipore) with 100 mM phosphoric acid/NaOH pH 7, 0.1 mM EDTA. A 200 μL sample of ASST (10 μM) was loaded onto a Superdex 200 HR10/30 column (GE Healthcare) and equilibrated with the same buffer, but without PNS or phenol at 1 mL/min. Known retention volumes of thyroglobulin (669 kDa), ferritin (440 kDa), BSA (67 kDa), β-lactoglobulin (35 kDa), ribonuclease A (13.7 kDa), cytochrome C (13.6 kDa), aprotinin (6.512 kDa), and vitamin B_{12} (1.355 kDa) were used as references. Calibration of the column with the above mass standards yielded the equation

$$y = -0.204x + 7.72 \text{ mL},$$

where y is the logarithm of molecular mass (kDa) and x is the retention volume (mL). As ASST elutes at 12.7 mL, the experimentally determined molecular mass of ASST is 134 kDa, which is in good agreement with the calculated mass of the homodimer (127.6 kDa).

Activity Assays for ASST and Its Variants. Kinetics of sulfotransfer by ASST and the variants Tyr96Phe, Tyr208Phe, His252Leu, Tyr559Phe, Tyr-208,559Phe, and Cys322Ala was measured in the presence of 30 μM PNS, 2 mM phenol, and 5 nM ASST (monomer) in 100 mM phosphoric acid/NaOH pH 7.0, 0.1 mM EDTA at 25 °C by following the absorbance at 405 nm (ε_{PN} = 9026 $M^{-1} \cdot cm^{-1}$). Because of their significantly impaired catalytic activities, the His356Leu and Arg374Leu variants were assayed at ASST concentrations (monomer) between 30nM and 900 nM. The initial slope of the recorded absorbance increase was used to quantify enzyme activity. Proper folding of all variants was verified by far UV CD spectra (data not shown).

Tryptic Digestion and Liquid Chromatography/Mass Spectrometry (LC/MS) Analysis of ASST. Samples of sulfo-ASST were denatured, digested with trypsin, and the tryptic peptides were analyzed by mass spectrometry. LC/MS analyses were performed on an Agilent microLC system coupled to an LTQ-XL instrument with electron transfer dissociation (Thermo Fischer Scientific).

1. Gill SC, von Hippel PH (1989) Calculation of protein extinction coefficients from amino acid sequence data. *Anal Biochem* 182:319–326.
2. Grimshaw J, *et al*. (2008) DsbL and DsbI form a specific dithiol oxidase system for periplasmic arylsulfate sulfotransferase in uropathogenic E. coli. *J Mol Biol*, 10.1016/j.jmb. 2008.05.031.
3. Van Duyne GD, Standaert RF, Karplus PA, Schreiber SL, Clardy J (1991) Atomic structure of FKBP-FK506, an immunophilin–immunosuppressant complex. *Science* 252:839–842.
4. Leslie AG (1999) Integration of macromolecular diffraction data. *Acta Crystallogr D* 55:1696–1702.
5. Evans P (1993) in *CCP4 Daresbury Study Weekend* (Daresbury Laboratory, Warrington, UK), pp. 114–122.
6. McCoy AJ, Grosse-Kunstleve RW, Storoni LC, Read RJ (2005) Likelihood-enhanced fast translation functions. *Acta Crystallogr D* 61:458–464.
7. Cowtan KD, Zhang KY (1999) Density modification for macromolecular phase improvement. *Prog Biophys Mol Biol* 72:245–270.
8. Perrakis A, Morris R, Lamzin VS (1999) Automated protein model building combined with iterative structure refinement. *Nat Struct Biol* 6:458–463.
9. Emsley P, Cowtan K (2004) Coot: Model-building tools for molecular graphics. *Acta Crystallogr D* 60:2126–2132.
10. Murshudov GN, Vagin AA, Lebedev A, Wilson KS, Dodson EJ (1999) Efficient anisotropic refinement of macromolecular structures using FFT. *Acta Crystallogr D* 55:247–255.
11. Adams PD, *et al*. (2002) PHENIX: Building new software for automated crystallographic structure determination. *Acta Crystallogr D* 58:1948–1954.
12. Laskowski RA, MacArthur MW, Moss DS, Thornton JM (1993) PROCHECK: A program to check the stereochemical quality of protein structures. *J Appl Cryst* 26:283–291.
13. DeLano WL (2002) *The PyMOL Users Manual* (DeLano Sci, CA).
14. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ (1990) Basic local alignment search tool. *J Mol Biol* 215:403–410.
15. Chenna R, *et al*. (2003) Multiple sequence alignment with the Clustal series of programs. *Nucleic Acids Res* 31:3497–3500.
16. Doron-Faigenboim A, Stern A, Mayrose I, Bacharach E, Pupko T (2005) Selecton: A server for detecting evolutionary forces at a single amino-acid site. *Bioinformatics* 21:2101–2103.

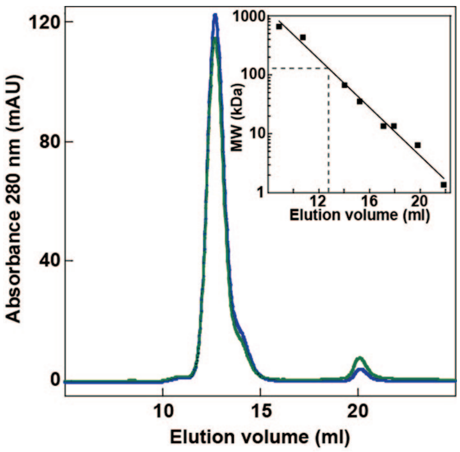

Fig. S1. Analytical gel filtration of sulfurylated (blue) and desulfurylated (green) ASST at pH 7.0 and 25 °C on a Superdex 200 HR 10/30 column. ASST elutes with a retention volume of 12.7 mL, corresponding to the mass of the homodimer. The sulfurylation state of ASST does not influence its oligomerization state. (*Inset*) Calibration of the gel filtration column. The dotted line indicates the elution volume and the logarithm of the apparent molecular mass of ASST.

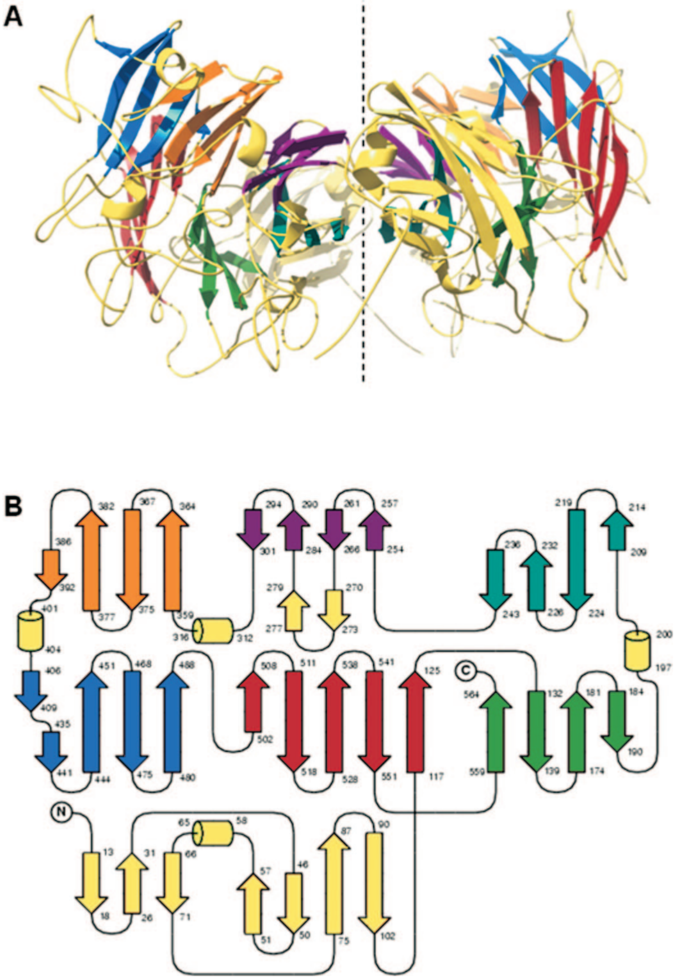

Fig. S2. (*A*) Side view of dimeric ASST, obtained by rotating the view in Fig. 2*A* by 90° around a horizontal axis lying in the plane of projection. (*B*) Topology diagram of ASST using the same coloring and numbering convention as in Fig. 2. The disulfide bond Cys-418-Cys-424 is located in the loop formed by the residues 410–434 of blade 6 (shown in detail in Fig. S3). The first and last residue of each secondary structure element is indicated.

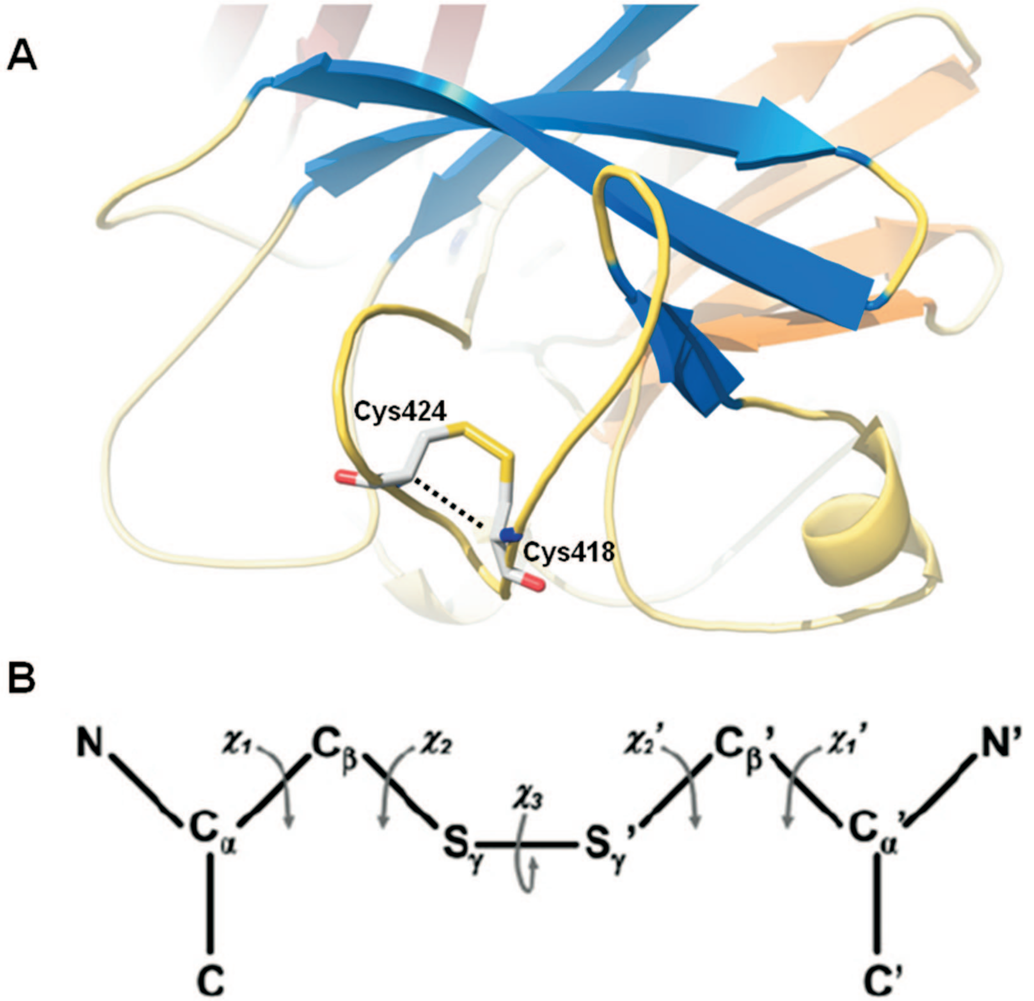

Fig. S3. The disulfide bond Cys-418-Cys-424 of ASST shows unusual geometry. (*A*) The disulfide bond Cys-418-Cys-424 in ASST is situated in a loop in the outer part of blade 6 of the propeller. The C_α (Cys-418)–C_α (Cys-424) distance (dashed line) is only 3.8 Å. The same color scheme as in Fig. 2 is used. (*B*) Dihedral angles of a disulfide bond. The values for the Cys-418-Cys-424 disulfide in ASST are: χ_1, −37.1 °; χ_2, −110.9 °; χ_3, 86.5 °; χ_2', −77.7 °; and χ_1', −53.9 °. The angle χ_1 is defined by the planes of (N, C_α, C_β) and (C_α, C_β and S_γ).

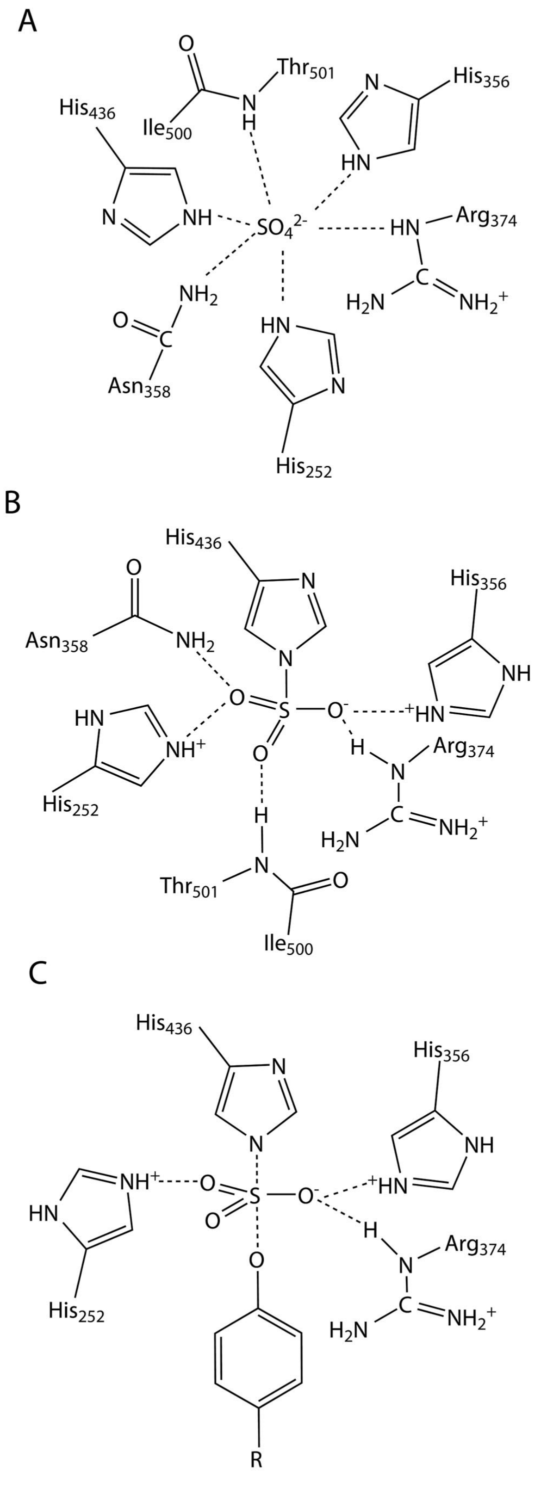

Fig. S4. A schematic representation of the hydrogen bond network stabilizing the sulfate bound to the active site of substrate-free ASST, the sulfohistidine intermediate, and the proposed transition state (See *Results*). (*A*) An ordered sulfate dianion is coordinated by nitrogen atoms from the side chains of His-252 (3.0 Å), His-356 (3.1 Å), Asn-358 (2.9 Å), Arg-374 (3.8 Å), His-436 (2.8 Å), and the backbone nitrogen of Thr-501 (3.0 Å) (numbers in brackets refer to mean distances between side chain nitrogens and sulfate oxygens). The Tyr-208 and Tyr-559 side chains are also within 10 Å of the sulfate. The previously proposed active site residue Tyr-96 (see text) is in the β-sandwich domain, ≈40 Å distant from the sulfate. (*B*) A schematic representation of the hydrogen bond network stabilizing the sulfuryl moiety of the sulfohistidine in the active site of PNS-soaked ASST. (*C*) The proposed structure of the transition states in the ASST reaction cycle deduced from the presented biochemical and crystallographic data. The presumed trigonal-bipyramidal transition state is stabilized by the 3 essential histidines and the essential arginine. In our activity assays with PNS and phenol, R corresponds to NO_2 or H, respectively.

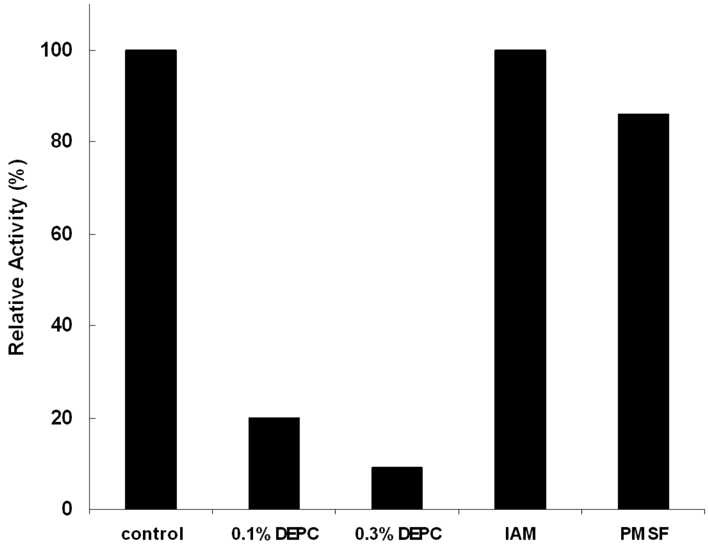

Fig. S5. Relative activity of ASST following treatment with different reagents that block potential active site residues. ASST (5 μM, monomer) was incubated with an excess of the respective reagent (DEPC, diethylpyrocarbonate at 0.1% and 0.3%; IAM, iodoacetamide, 1 mM; PMSF, phenylmethylsulfonylfluoride, 1 mM) for 120 min (at pH 6 for DEPC and pH 7 for all other samples) at 25 °C. Excess reagent was subsequently removed by desalting (NAP5 column, GE Healthcare). ASST activity assays were performed with 20 nM ASST, 1 mM phenol, and 60 μM PNS, and normalized by comparison with the activity of the unmodified enzyme.

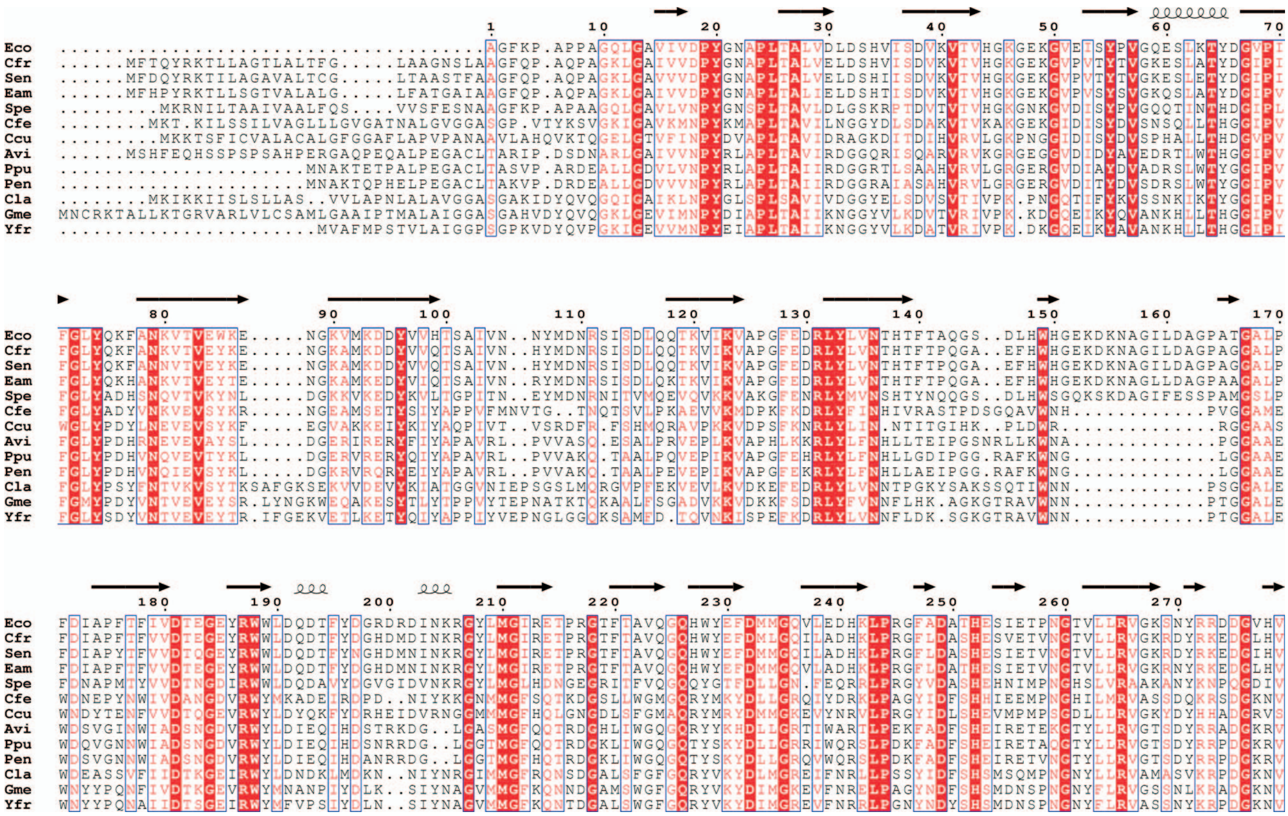

Fig. S6. Structure-based sequence alignment of ASSTs. Residues are numbered according to the mature *E. coli* CFT073 ASST. Conserved residues are boxed in red. The sequences from *E. coli* CFT073 (Eco), *Citrobacter freundii* (Cfr), *Salmonella enterica* (Sen), *Enterobacter amnigenus* (Eam), *Shewanella pealeana* (Spe), *Campylobacter fetus* (Cfe), *Campylobacter curvus* (Ccu), *Azotobacter vinelandii* (Avi), *Pseudomonas putida* (Ppu), *Pseudomonas entomophila* (Pen), *Campylobacter lari* (Cla), *Geobacter metallireducens* (Gme), and *Yersinia freundii* (Yfr) were aligned.

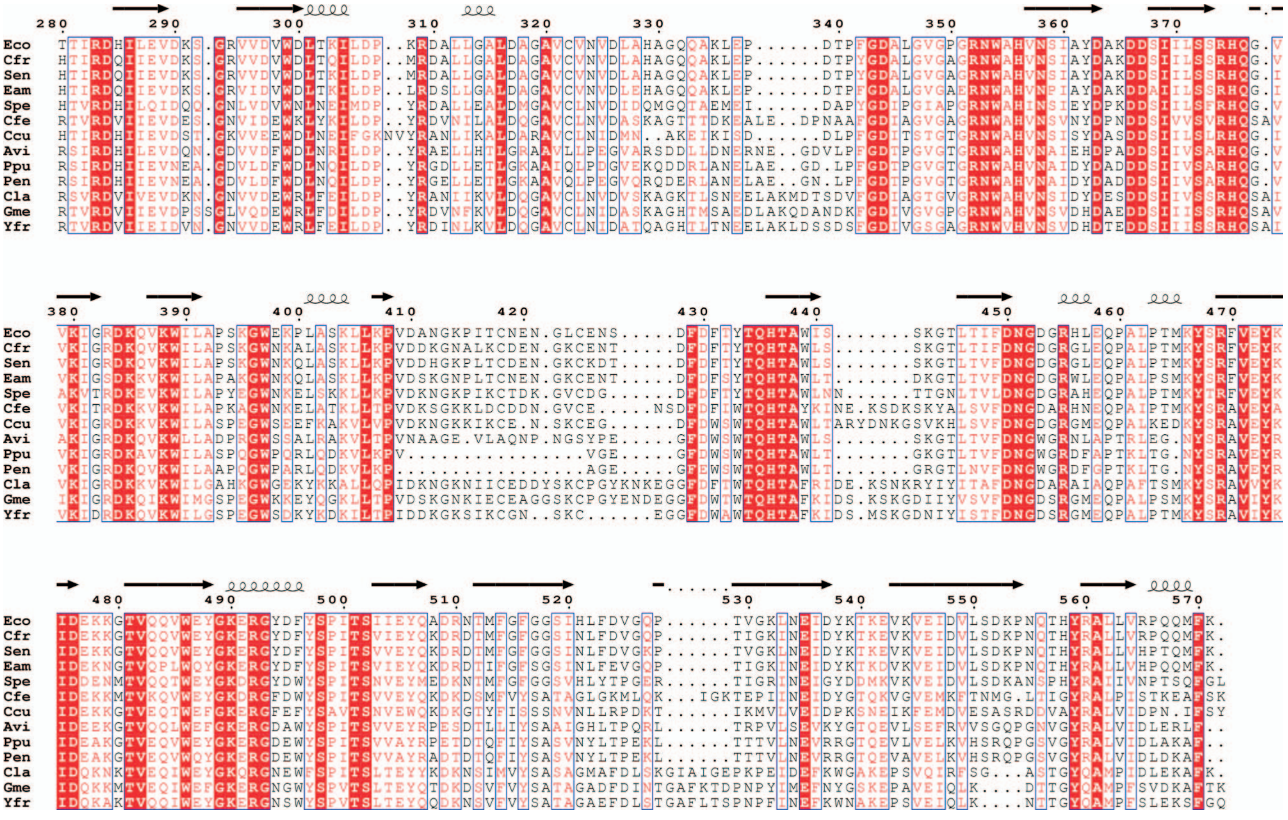

Fig. S6 continued.

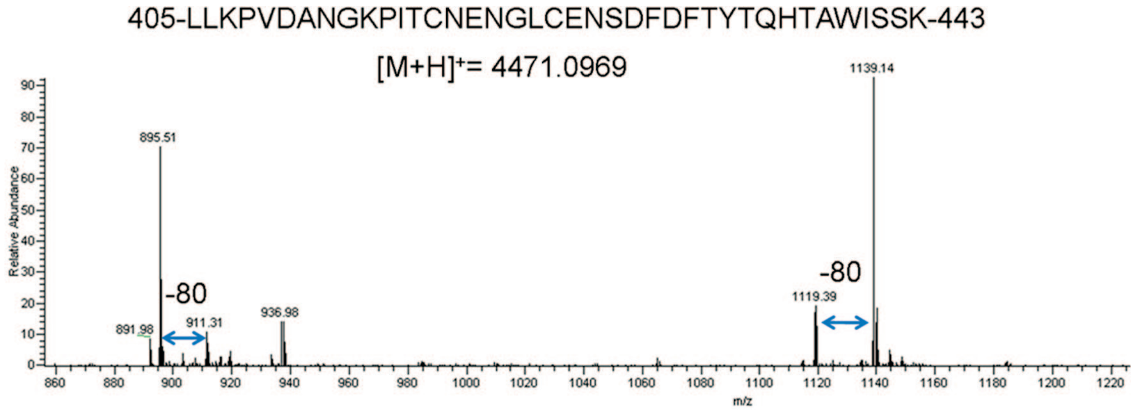

Fig. S7. Mass spectrum and amino acid sequence of the tryptic peptide Leu-405–Lys-443 of ASST containing the catalytic residue His-436. A difference of 80 Da is observed for both the 4+ and 5+ charge states of the peptide.

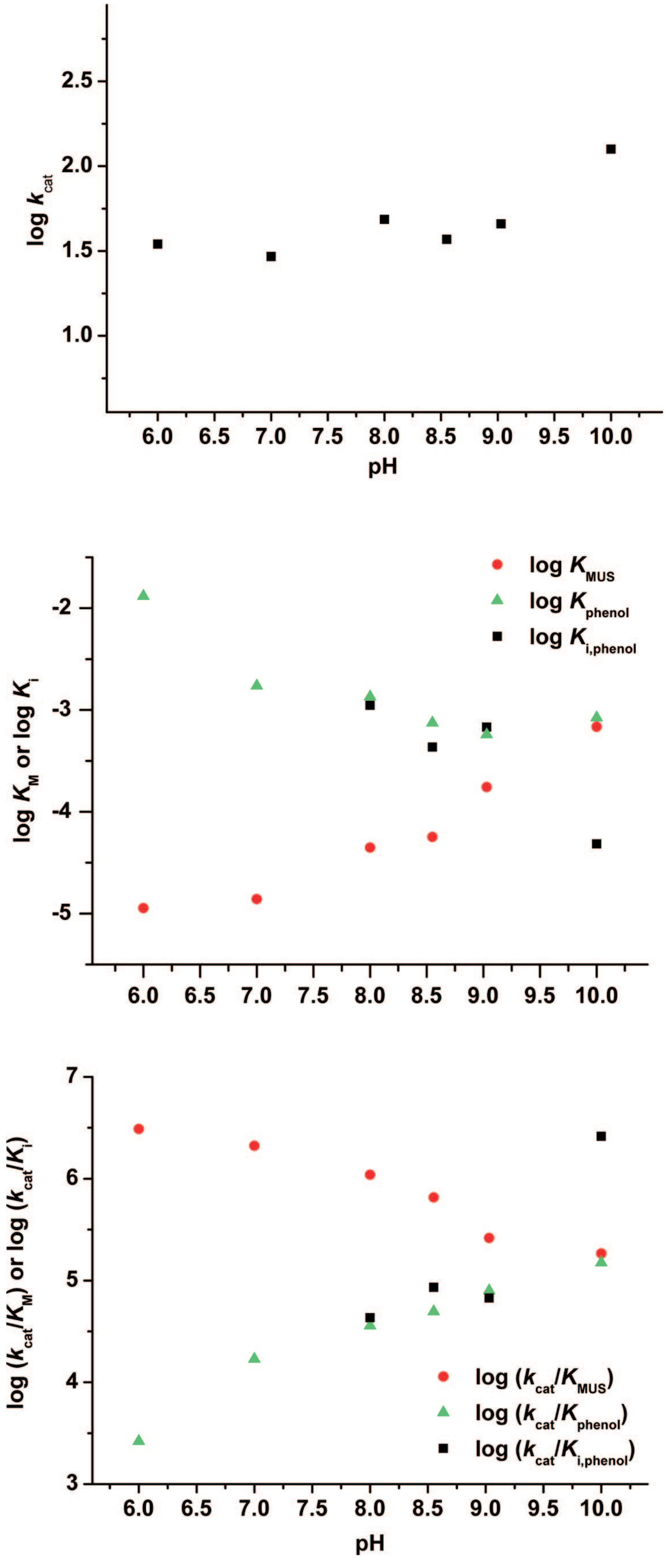

Fig. S8. Kinetic parameters for ASST-catalyzed sulfuryl transfer from MUS to phenol within the pH range 6.0–10.0. Although the apparent Michaelis constant for MUS increases with increasing pH, the one for phenol decreases and k_{cat} remains essentially invariant. The data sets for each individual pH value were fitted globally according to ping-pong kinetics as described in *Methods*.

Table S1. Data collection, phasing, and refinement statistics

Data collection	Native	SeMet (peak)	PNS	MUS
Space group	$P3_212$	$P3_212$	$P3_212$	$P3_212$
Cell dimensions				
a (=b), c (Å)	181.73,100.43	181.68, 100.36	181.39, 99.91	181.83, 100.66
γ, β, α (°)	90, 90, 120	90, 90, 120	90, 90, 120	90, 90, 120
Wavelength	1.000	0.978	1.000	1.000
Observed reflections	1501755	548262	403425	560960
Unique reflections	124546	38033	104994	73355
Resolution (Å)	60.0–2.0 (2.1–2.0)	50–3.0 (3.15–3.0)	40–2.1 (2.2–2.1)	50–2.4 (2.5–2.4)
R_{sym}	0.088 (0.40)	0.165 (0.65)	0.08 (0.40)	0.13 (0.51)
$I / \sigma(I)$	20.3 (3.5)	17.3 (4.6)	8.1 (1.6)	3.7 (1.4)
Completeness (%)	99 (94.9)	99.6 (99.6)	99.6 (99.6)	99.7 (100)
Redundancy	12.1 (6.0)	14.4 (14.1)	3.8 (3.6)	7.6 (7.6)
Refinement				
Resolution (Å)	33.9–2.0		35.0–2.1	40.0–2.4
No. reflections	180194		106194	72370
R_{work} / R_{free}	0.207/0.248		0.187/0.232	0.175/0.216
No. of atoms	10008		9723	9152
Mean B-factor (Å^2)	32.9		44.5	45.2
rms deviations				
Bond lengths (Å)	0.009		0.025	0.024
Bond angles (°)	1.188		1.647	1.376

Values in parentheses represent highest resolution shell.

Table S2. Relative catalytic activities of ASST variants

Mutation	Relative activity, %
Wild type	100
Previously proposed catalytic residue	
Tyr96Phe	103
Sulfate binding pocket variants	
His252Leu	4
His356Leu	0.06
Arg374Leu	0.10
His436Leu	ND*
Catalytic residue variant	
Tyr208Phe	197
Cys322Ala	100
Tyr559Phe	63
Tyr208,559Phe	40[†]

See *Methods* for details. The error in all experiments was ± 5%.

*ND, Not determined because this variant could not be expressed.

[†]This variant was particularly prone to proteolysis.

Table S3. Oligonucleotide primers used for site-directed mutagenesis of ASST

Mutation	Primers (5′-3′)
Tyr96Phe	FW:GGTCATGAAAGATGATTTTGTGGTGCACACTTCGG
	BW:CCGAAGTGTGCACCACAAAATCATCTTTCATGACC
His252Leu	FW:CGCGGATTTGCTGACGCCACTCTTGAGTCCATTGAGACGCC
	BW:GGCGTCTCAATGGACTCAAGAGTGGCGTCAGCAAATCCGCG
His356Leu	FW:CCGCAACTGGGCGCTCGTGAACTCTATCGC
	BW:GCGATAGAGTTCACGAGCGCCCAGTTGCGG
Arg374Leu	FW:CTATCATCCTCTCCTCTCTCCACCAGGGTGTTGTG
	BW:CACAACACCCTGGTGGAGAGAGGAGAGGATGATAG
His436Leu	FW:CGATTTTACCTACACCCAGCTTACCGCCTGGATTTCC
	BW:GGAAATCCAGGCGGTAAGCTGGGTGTAGGTAAAATCG
Tyr208Phe	FW:CAACAAGCGTGGTTTTCTGATGGGTATCCG
	BW:CGGATACCCATCAGAAAACCACGCTTGTTG
Cys322Ala	FW:GGCGCGCTGGATGCAGGTGCAGTTGCCGTTAACGTTGACCTTGCCCATGCAGGACAACAGGC
	BW:GCCTGTTGTCCTGCATGGGCAAGGTCAACGTTAACGGCAACTGCACCTGCATCCAGCGCGCC
Tyr559Phe	FW:CCCAATCAGACTCACTTCCGTGCGCTGTTAGTCCG
	BW:CGGACTAACAGCGCACGGAAGTGAGTCTGATTGGG