

## Supplementary Information:

### Site-specific incorporation of photocrosslinker and bioorthogonal amino acids into enteric bacterial pathogens

Shixian Lin<sup>†</sup>, Zhenrun Zhang<sup>†</sup>, Hao Xu<sup>§</sup>, Lin Li<sup>§</sup>, She Chen<sup>§</sup>, Jie Li<sup>†</sup>, Ziyang Hao<sup>†</sup> and Peng R. Chen<sup>\*,†,‡</sup>

<sup>†</sup>Beijing National Laboratory for Molecular Sciences, College of Chemistry and Molecular Engineering, Peking University, Beijing 100871, China and <sup>‡</sup>Peking-Tsinghua Center for Life Sciences, Beijing, China; <sup>§</sup>National Institute of Biological Sciences (NIBS), Beijing 102206, China.

## Supplementary Methods

### *General materials and methods*

Bacterial cells were grown in LB broth or on LB agar medium (Sigma). Antibiotics were used at final concentrations of 100  $\mu\text{g ml}^{-1}$  for ampicillin and 34  $\mu\text{g ml}^{-1}$  chloramphenicol (Sigma). Antibodies against shHdeA, SurA, PhoA and OspF were raised in rabbits as described.<sup>1</sup> Antibodies for phosphor-Erk1/2, Erk1/2 were purchased from Cell Signaling Technology. Epitope antibodies Myc and 6xHis were purchased from Santa Cruz Biotechnology. EPEC2348/69, *Shigella flexneri* 2a pcp301 and *Salmonella typhimurium* strains were used for the expression of model proteins and their mutants; *Shigella flexneri* 2a pcp301 ospF deletion strain was generated as described before.<sup>2</sup> Compounds DiZPK and ACPK were synthesized as described previously.<sup>1,3</sup> Congo Red and TCA (trichloroacetic acid) were obtained from Sigma. Biotin-(PEG)<sub>4</sub>-alkyne, Alexa Fluor 488, Streptavidin-HRP and Streptavidin-Alexa Fluor 488 were obtained from Invitrogen.

### *Plasmids construction*

The OspF gene (accession number GI: 56383075) was subcloned into pBAD24 with epitope tag Myc at C-terminus. The N-terminus 50 amino acids truncated version was constructed by subcloning the truncation gene into pBAD24. The pBAD24-OspF-EGFP was constructed by inserting the OspF gene into pEGFP-N1, followed by cloning the entire OspF-EGFP gene into pBAD24. The pSupAR-sal plasmid with *Salmonella* lpp promoter (sal-lpp) was generated from original pSupAR plasmid by replacing the proK promoter with sal-lpp. All mutations used in this study were generated by QuickChange Site-Directed Mutagenesis Kit.

### *Expression and purification of wild-type proteins.*

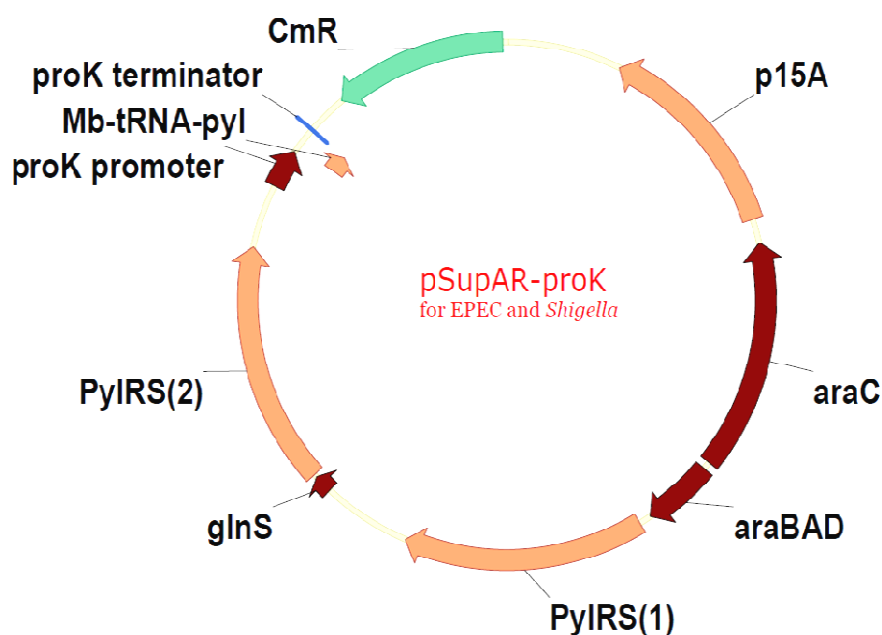
Alkaline phosphatase (AP) protein was expressed in the cytoplasm of *E. coli* BL21(DE3) cells and purified by anion exchange column Q-FF (GE Healthcare). GST-Erk2 and phosphor-GST-Erk2 were

obtained as described.<sup>2</sup> shHdeA-Myc-His<sub>6</sub> proteins with DiZPK incorporated into positions 58 was obtained as described.<sup>1</sup> Protein concentrations were determined by the BCA (bicinchoninic acid) method (Pierce), with BSA as a standard.

### **Native PAGE analysis.**

The native PAGE gel was prepared using Bio-Rad Mini-PROTEAN Tetra Electrophoresis System (Figure. S8A). SDS was removed from the ingredients of both the stacking gel (pH 6.8, 5%) and the resolving gel (pH 8.8, 15%). All the protein samples were prepared in Tris buffer containing bromophenol blue, native running buffer (1 L) was consisted of 14.4 g glycine and 3.03 g Tris base plus ddH<sub>2</sub>O. Samples were run under ice-cold condition (150 V, 400 mA, 60 min). Soybean trypsin inhibitor (from Beijing Biodee Biotechnology Co.,Ltd, pI=4.6, MW~21 kDa) was used as the marker protein (for wild-type shHdeA protein, pI=5.2).

### **Supplementary Results**



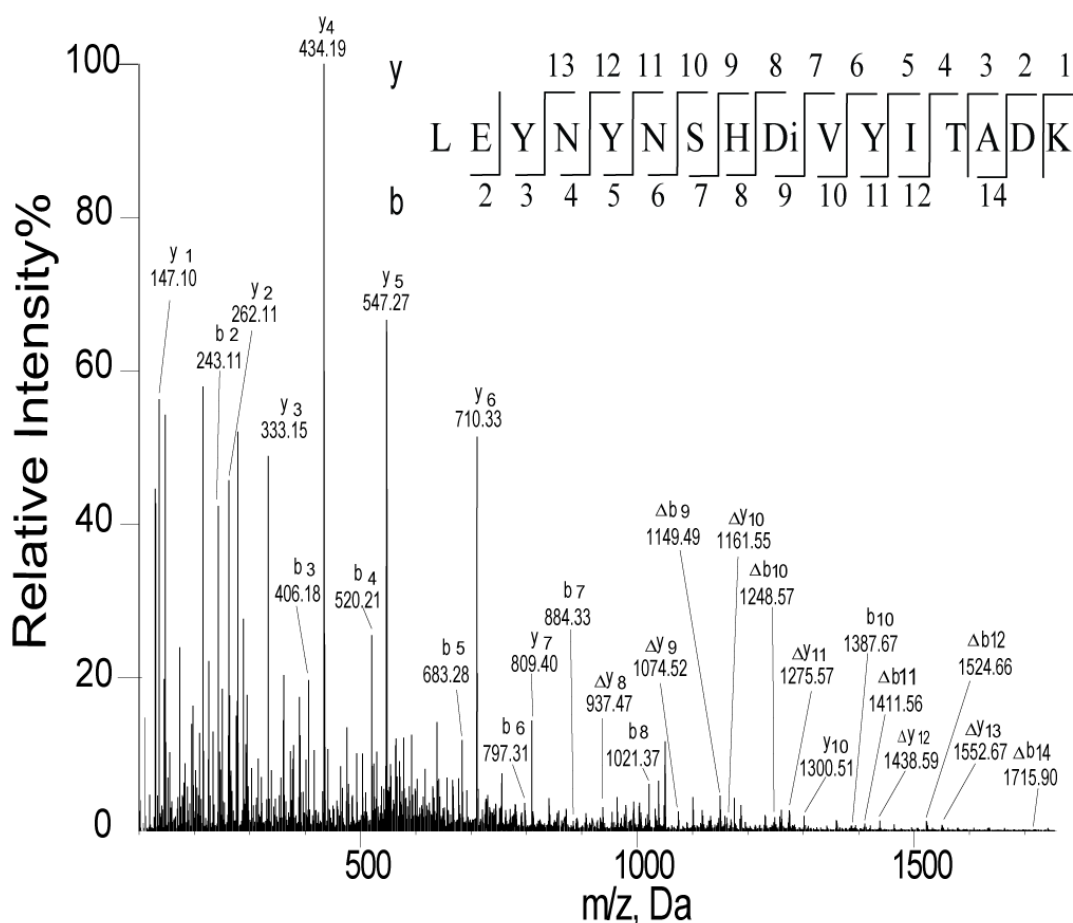
**Figure S1.** Plasmid map of pSupAR used in EPEC and *Shigella*.

*E.coli K12* -AGGCATTTTGCTATTAAGGGATTGACGAGGGCGTTC-TGCGCA---GTAAGATGCGCCCGCATT  
*EPEC 2348/69* -AGGCATTTTGCTATTAAGGGATTGACGAGGGCGTTC-TGCGCA---GTAAGATGCGCCCGCATT  
*Shigella flexneri 2a* -AGGCATTTTGCTATTAAGGGATTGACGAGGGCGTTC-TGCGCA---GTAAGATGCGCCCGCATT  
*Salmonella typhimurium* ATTGTATTTTGCGACTAAGGGATTGACGGGCGTTCGCGCTGACCA---GTAAGATGCGCCC-GCATT  
*Yersinia petis* --TGC-TCATCCCAACCAGTTTCACACCGCAGCCACACTCCACAAACGACAAACCAACCCGAAT-

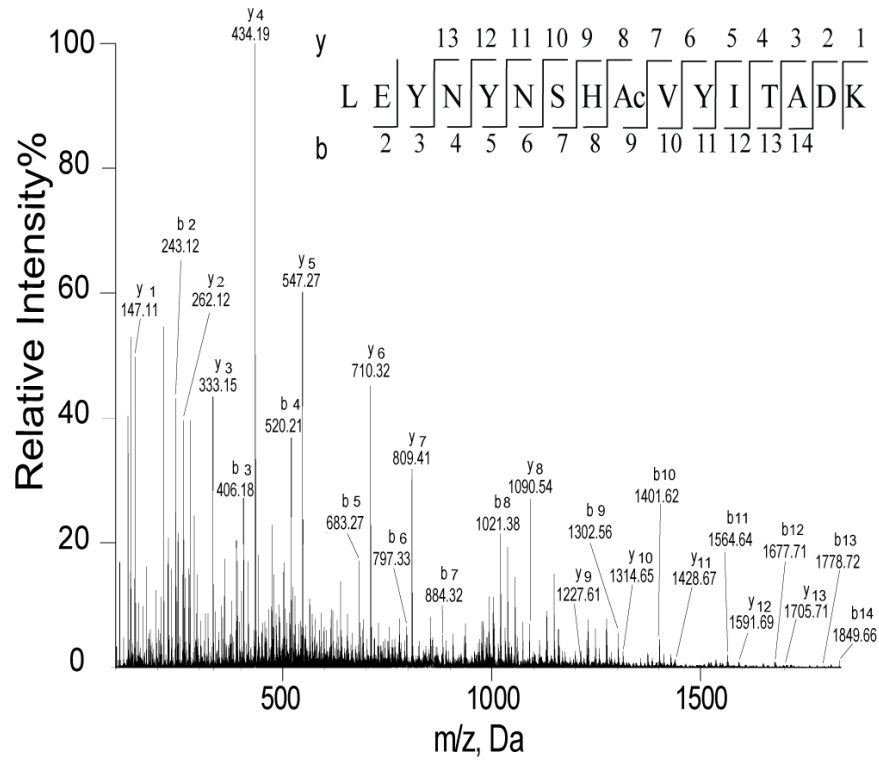
**Figure S2.** Alignment of proK tRNA promoters from *E.coli K12*, EPEC 2348/69, *Shigella flexneri 2a*, *Salmonella typhimurium* and *Yersinia petis*. The alignment was performed with Vector NTI (Invitrogen).

*E.coli K12* CCCATC AAAAAAATATTCTCAACATAAAAAAC TTTGTGTAATACTTGTAAACGCT  
*EPEC 2348/69* CCCATC AAAAAAATATTCTCAACATAAAAAAC TTTGTGTAATACTTGTAAACGCT  
*Shigella flexneri 2a* CCCATC AAAAAAATATTCTCAACATAAAAAAC TTTGTGTAATACTTGTAAACGCT  
*Salmonella typhimurium* CCCATC AAAAAAATATTCTCAACATAAAAAAG TTTGTGTAATACTTGTAAACGCT  
*Yersinia petis* ----- AAAAAATTATTCTCTTTCG AAAAAAC TTTGTGTAATACTTGTAAACGCT

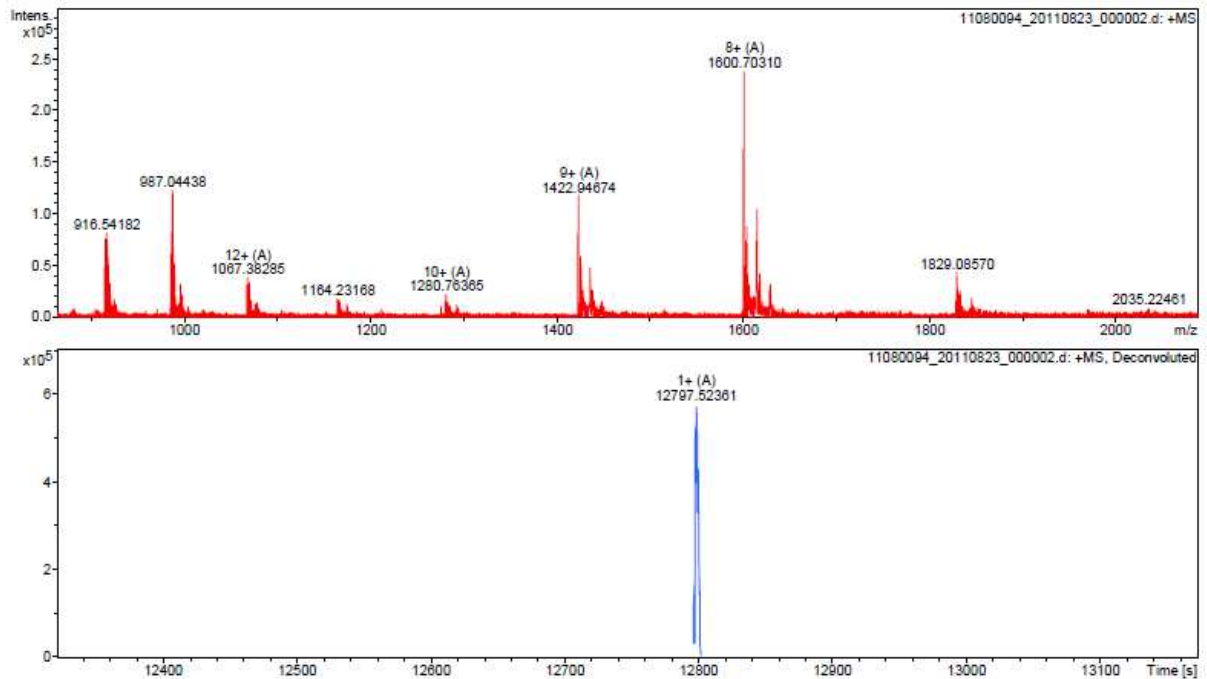
**Figure S3.** Alignment of lpp tRNA promoters from *E.coli K12*, EPEC 2348/69, *Shigella flexneri 2a*, *Salmonella typhimurium* and *Yersinia petis*. The alignment was performed with Vector NTI (Invitrogen).



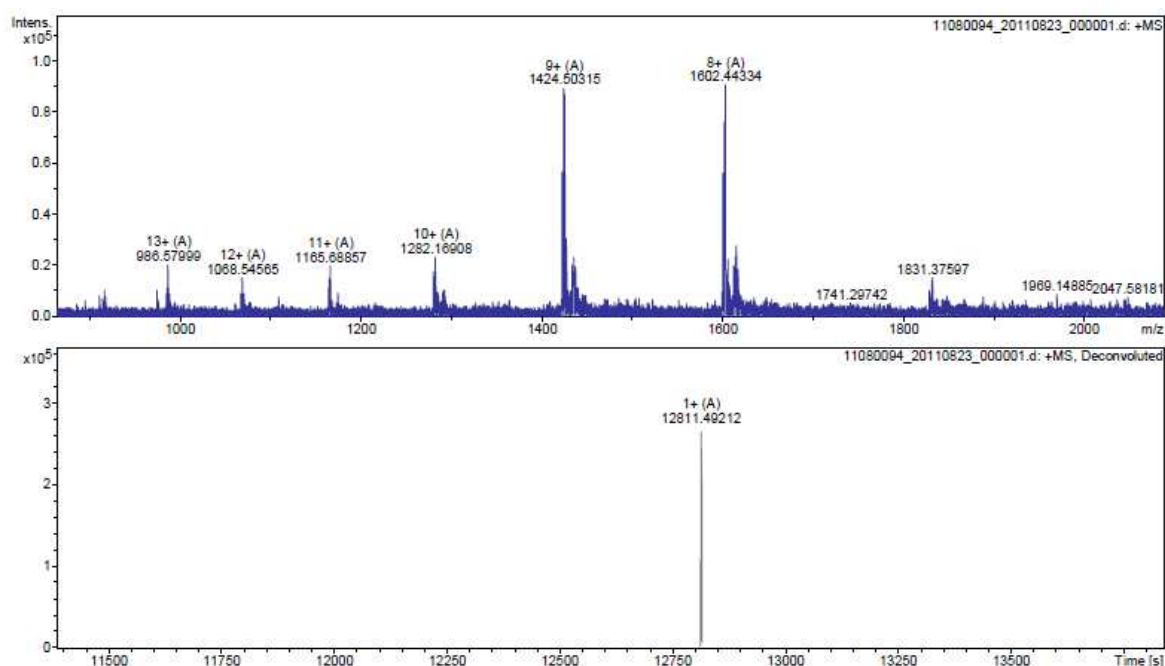
**Figure S4.** Mass spectrometry analysis of GFP protein carrying DiZPK expressed in *Salmonella*. MS/MS of GFP149-DiZPK. Amino acid **Di** represents the DiZPK. The b and y ion are marked in the spectrum. The amino acid sequence of the peptide is shown at the top of the figure. Δ: neutral loss of a 139 Da ( $C_6H_9N_3O$ ) .



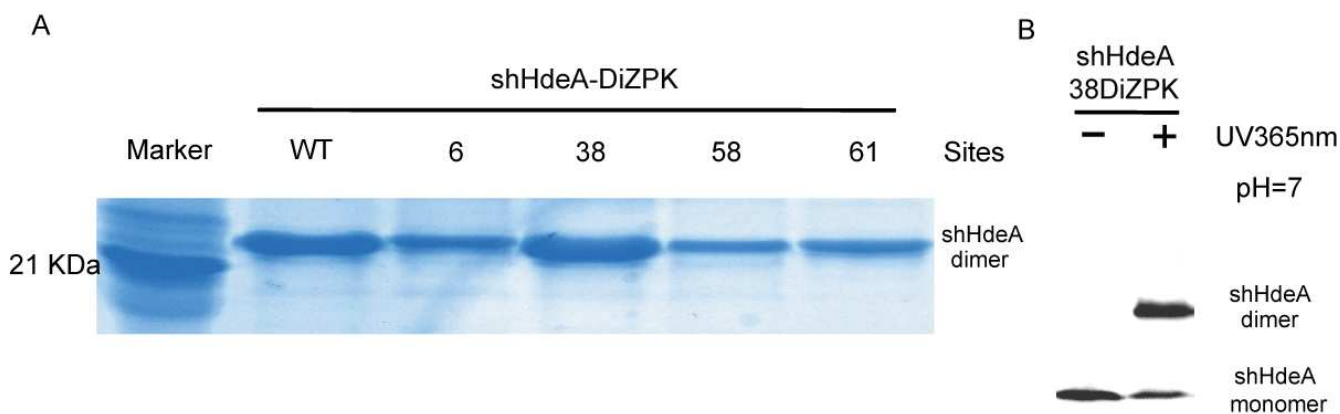
**Figure S5.** Mass spectrometry analysis of GFP protein carrying ACPK expressed in *Salmonella*. MS/MS of GFP149-ACP. Amino acid **Ac** represents ACPK. The b and y ion are marked in the spectrum. The amino acid sequence of the peptide is shown at the top of the figure.



**Figure S6.** Mass spectrometry of shHdeA protein carrying DiZPK expressed in *Shigella*. ESI-MS analysis of shHdeA58-DiZPK. To verify that DiZPK was encoded in *Shigella* without being modified, the mass of purified full length shHdeA carrying DiZPK at residue V58 site was measured using electrospray ionization mass spectrometry (ESI-MS), which gave a measured mass of 12798Da (expected: 12799 Da).



**Figure S7.** Mass spectrometry of shHdeA protein carrying ACPK expressed in *Shigella*. ESI-MS analysis of shHdeA58-ACP. To verify that ACPK was incorporated in the *Shigella* without being modified, the mass of purified full length shHdeA carrying ACPK at residue V58 site was measured using electrospray ionization mass spectrometry (ESI-MS), which gave a measured mass of 12811Da (expected: 12813 Da).



**Figure S8.** Incorporation of DiZPK into indicated positions of shHdeA. (A) Native PAGE analysis of shHdeA proteins carrying DiZPK at the indicated positions using shHdeA-WT as a positive control. All the variants formed dimmers with no detectable monomers. (B) Photocrosslinking of live *Shigella* cells harboring shHdeA-38DiZPK at pH =7 with or without UV irradiation. Majority of the shHdeA variant proteins became covalently linked dimers upon UV irradiation at pH =7.



**Figure S9.** The OspF protein concentrations in cell lysate were adjusted to the same level by judging from immunoblotting analysis using the anti-Myc antibody.

**Supplementary Table 1.** Summary of the mass spectrometry result on the identified shHdeA client proteins (protein photocrosslinking was performed on shHdeA58-DiZPK at pH 2.3 in living *Shigella* cells)

Accession number	Protein name	Number of matched peptides	MASCOT score	Brief Description
gil110805876	AraF	24	37795	a component of the high affinity L-arabinose transport system
gil56383414	OppA	39	10357	a component of the binding protein-dependent transport system
gil24112179	GlnH	24	6308	a component of glutamine transport system
gil24113539	MglB	23	6042	a periplasmic galactose binding protein
gil24113682	ArgT	19	5176	a periplasmic lysine arginine ornithine binding protein
gil56480072	GlpQ	23	3627	an enzyme that catalyzes the hydrolysis of glycerophosphodiester
gil24115375	MalE	27	3398	a periplasmic maltose binding protein
gil24111499	SurA	24	3320	a periplasmic peptidyl-prolyl cis-trans isomerase
gil24114813	DppA	31	2261	a periplasmic dipeptide transport system
gil24114218	AnsB (L-asparaginase)	18	2147	an enzyme that catalyzes the hydrolysis of asparagine
gil24115464	YtfQ	20	2028	ABC transporter periplasmic-binding protein
gil24113681	HisJ	15	1693	a periplasmic histine binding protein
gil24114521	DegP	18	1506	a periplasmic serine endoprotease

gil56383440	PRC	20	1432	A Periplasmic tail-specific protease
gil56479804	MdoG	27	1356	a periplasmic Glucans biosynthesis protein
gil56479781	OmpA	12	1010	an outer membrane protein which have a periplasmic domain
gil24112266	LolA	11	865	an outer membrane lipoprotein carrier protein which have lipoproteins anchored to the periplasmic side
gil83287997	YggE	9	862	a uncharacterized protein
gil24113600	OmpC	14	727	an outer membrane protein which have a periplasmic domain
gil24114713	UgpB	17	691	a periplasmic glycerol-3-phosphate binding protein
gil56480566	CpdB	21	689	a periplasmic 2',3'-cyclic phosphodiesterase (3'-nucleotidase)
gil24112526	PotD	17	670	a periplasmic spermidine/putrescine binding protein
gil24114795	YhjJ	16	610	a uncharacterized protein
gil24111612	YaeT	8	496	a membrane protein which helps the assemble of outer membrane proteins
gil24114842	XylF	10	446	a periplasmic D-xylose binding protein
gil24114147	DsbC	6	435	a periplasmic thiol and disulfide inter change protein
gil56383845	MDH	11	348	a periplasmic malate dehydrogenase
gil24114628	PpiA	6	346	a Periplasmic peptidyl prolyl cis-trans isomerases
gil161486476	ZnuA	9	342	a periplasmic Zinc uptake system component
gil56480517	NrfA	9	289	a component of periplasmic nitrite reductase complex
gil24115149	PDI	11	283	a periplasmic protein disulfide isomerase I
gil24114611	SlyD	10	263	a FKBP-type peptidyl prolyl isomerases
gil313648524	DegQ	12	261	a periplasmic serine endoprotease
gil24114741	NikR	4	258	a periplasmic nickel responsive regulator

gil24111633	MetQ	9	257	a D-methionine-binding lipoprotein
gil281601895	YfgC	7	243	a Tetratricopeptide repeat-containing protein
gil56480524	AphA	6	237	a periplasmic acid phosphatase class B
gil56480161	ProX	4	212	a periplasmic glycine betaine binding protein
gil24113046	SodB	5	196	a superoxide dismutase
gil24111893	PpiB	6	192	a periplasmic peptidyl-prolyl cis-trans isomerase B
gil24115625	CreA	2	177	a putative periplasmic protein
gil56479975	YebF	2	169	a protein contain a signal sequence for its secretion to periplasm
gil56480115	NlpB	3	166	a lipoprotein 34
gil24113809	BCP	4	157	a putative peroxiredoxin bcp
gil24112031	FUR	6	141	a ferric uptake regulation protein
gil56479691	TolB	9	135	a component of Tol-dependent translocation system
gil32699529	PtrA	7	125	a protease
gil24111953	AhpC	3	114	an alkyl hydroperoxide reductase subunit C
gil24051610	SapA	8	110	a peptide transport periplasmic protein
gil56480408	TnaA	4	100	a tryptophanase
gil56404025	UshA	1	92	a bifunctional UDP-sugar hydrolase/5'-nucleotidase periplasmic precursor
gil313646493	CueO	1	89	a blue copper oxidase cueO
gil56479896	GapA	2	82	a glyceraldehyde-3-phosphate dehydrogenase A
gil24114369	YqjI	1	81	a uncharacterized protein
gil24114342	RibB	2	76	a 3,4-dihydroxy-2-butanone 4-phosphate synthase



gil189030834	NanM	3	76	a N-acetylneuraminate epimerase
gil24115188	FdoG	3	75	a formate dehydrogenase-O major subunit
gil24112719	TPX	1	69	a thiol peroxidase
gil24115187	FdoH	1	66	a formate dehydrogenase-O iron-sulfur subunit
gil56480284	LptA	1	61	a lipopolysaccharide export system protein
gil56480149	KgtP	1	59	an alpha-ketoglutarate permease
gil56479605	LpdA	4	56	a dihydrolipoyl dehydrogenase
gil24111599	Serine endoprotease	2	53	serine endoprotease
gil24114481	YrbC	4	52	a ABC transporter periplasmic binding protein MlaC
gil56479649	TSX	2	48	a nucleoside-specific channel-forming protein
gil24112270	DmsB	1	48	an anaerobic dimethyl sulfoxide reductase chain B
gil56480520	YjcO	2	47	a uncharacterized protein
gil24113186	Putative transport periplasmic protein	1	46	putative transport periplasmic protein
gil24111675	PhoA	2	46	A periplasmic alkaline phosphatase

\*Blue lanes represent the proteins are well known substrates of shHdeA for *in vitro* study before. Red lanes represent the proteins have been identified by our recent work in *E. coli*.<sup>1</sup> Black lanes represent additional client proteins identified from this study. These proteins were ranked by the MASCOT score and only client proteins with a score above 45 were selected.

## References

- (1) Zhang, M.; Lin, S.; Song, X.; Liu, J.; Fu, Y.; Ge, X.; Fu, X.; Chang, Z.; Chen, P. R. *Nat. Chem. Biol.* **2011**, 7, 671-677.
- (2) Li, H.; Xu, H.; Zhou, Y.; Zhang, J.; Long, C.; Li, S.; Chen, S.; Zhou, J.-M.; Shao, F. *Science*. **2007**, 315, 1000-1003.
- (3) Hao, Z.; Song, Y.; Lin, S.; Yang, M.; Liang, Y.; Wang, J.; Chen, P. R. *Chem. Comm.* **2011**, 47, 4502-4504.

Complete ref 35:

Reference Type: Journal Article

Record Number: 35

Author: M. T. G. Holden, E. J. Feil, J. A. Lindsay, S. J. Peacock, N. P. J. Day, M. C. Enright, T. J. Foster, C. E. Moore, L. Hurst, R. Atkin, A. Barron, N. Bason, S.

Year: 2004

Title: Complete genomes of two clinical Staphylococcus aureus strains: Evidence for the rapid evolution of virulence and drug resistance

Journal: Proc. Natl. Acad. Sci. USA

Volume: 101

Issue: 26

Pages: 9786-9791

Date: Jun 29

Short Title: Complete genomes of two clinical Staphylococcus aureus strains: Evidence for the rapid evolution of virulence and drug resistance

ISSN: 0027-8424

DOI: 10.1073/pnas.0402521101

Accession Number: WOS:000222405600049

Notes: Times Cited: 303

URL: <Go to ISI>://WOS:000222405600049