Mapping native disulfide bonds at a proteome scale

Shan Lu[1,2,7], Sheng-Bo Fan[3,4,7], Bing Yang[2,7], Yu-Xin Li[2], Jia-Ming Meng[3,4], Long Wu[3,4], Pin Li[5], Kun Zhang[3,4], Mei-Jun Zhang[2], Yan Fu[6], Jincai Luo[5], Rui-Xiang Sun[3], Si-Min He[3] & Meng-Qiu Dong[1,2]

We developed a high-throughput mass spectrometry method, pLink-SS (http://pfind.ict.ac.cn/software/pLink/2014/pLink-SS.html), for precise identification of disulfide-linked peptides. Using pLink-SS, we mapped all native disulfide bonds of a monoclonal antibody and ten standard proteins. We performed disulfide proteome analyses and identified 199 disulfide bonds in *Escherichia coli* and 568 in proteins secreted by human endothelial cells. We discovered many regulatory disulfide bonds involving catalytic or metal-binding cysteine residues.

Protein cross-linking, a special class of post-translational modifications, plays a fundamental role in all life forms. It builds up the extracellular matrix[1] and regulates protein function or half-life in and outside the cell[2,3]. The chemical nature of endogenous protein cross-links varies from the ubiquitous types such as the disulfide bond[2] or isopeptide bond[3] to the specialized linkages seen in collagens and elastin[4]. In this work we focus on disulfide bonds, which are indispensable for numerous secreted proteins including hormones, growth factors and cell surface receptors, but these bonds have long been a challenge to analyze[5]. We report two mass spectrometry (MS)-based workflows, equipped with automated data analysis and false discovery rate (FDR) control, to map protein disulfide linkages in simple or complex protein samples.

We built a standard data set using 72 cysteine-containing peptides (**Supplementary Table 1** and Online Methods). After comparing different fragmentation methods (**Supplementary Fig. 1**), we selected 2,287 high-quality HCD (higher-energy collisional dissociation) spectra of disulfide-linked peptide pairs and used 1,146 of them for training and 1,141 for testing pLink-SS, an adaptation of the pLink software for identifying chemically cross-linked peptides[6]. Statistical analyses of the training data revealed fragment-ion types that are specific for disulfide-linked peptides (**Supplementary Fig. 2a–c**) and those that are common for all cross-linked peptides. Both of these fragment-ion types helped pLink-SS correctly identify 1,126 (98.7%) of the 1,141 test spectra that were searched along with other 46,526 HCD spectra against a database containing >4,000 *E. coli* proteins and the 72 synthetic peptides (FDR = 5%, **Supplementary Fig. 2d**).

To avoid disulfide scrambling, a serious artifact that readily occurs at alkaline pH[7,8] or in the presence of free cysteine[7], we digested samples at pH 6.5 with 2 mM *N*-ethylmaleimide (NEM) (**Fig. 1a**). We also used different proteases, including nonspecific proteases, to increase the number of disulfide-bond identifications (**Supplementary Fig. 3**). A workflow for low-complexity samples is shown in **Supplementary Figure 4a**.

Mapping all the disulfide bonds of a monoclonal antibody is a pressing need in the pharmaceutical industry. We analyzed a monoclonal antibody and identified two unknown disulfide bonds in the variable regions, one in the heavy chain and one in the light chain (**Fig. 1b** and **Supplementary Fig. 5**), as well as all the known ones in the constant regions of the IgG2 antibody (**Supplementary Fig. 6**).

We further verified the method using a mixture of ten standard proteins and successfully mapped all 74 disulfide bonds using multiple protease digestions (**Supplementary Fig. 7** and **Supplementary Table 2**). These disulfide bonds were identified in interlinked (**Supplementary Fig. 6a–f**), loop-linked (**Supplementary Fig. 8a**) or complex forms that involve two or more disulfide bonds (**Supplementary Figs. 6g** and **8b**).

We next adapted the workflow for complex protein mixtures (**Supplementary Fig. 4b**). Nonspecific proteases are not suitable for samples containing thousands of proteins because the search space is prohibitively large; instead, we recommend combinations of Lys-C, trypsin and Glu-C (LTG) or of Lys-C and Asp-N (**Supplementary Fig. 9**). Disulfide bonds between different proteins are extremely rare (**Supplementary Fig. 10a**,**b**). To avoid the interference of numerous random interprotein matches, which are the result of a large database, we filtered inter- and intraprotein results separately (**Supplementary Fig. 4b**). The formula to calculate FDR for intraprotein disulfide bonds was adjusted (**Supplementary Fig. 10c**).

Next, we carried out proteome-level disulfide mapping. In Gram-negative bacteria, the thiol-disulfide oxidoreductase DsbA catalyzes the formation of protein disulfide bonds between two consecutive cysteine residues, and the protein disulfide isomerase DsbC catalyzes the formation of nonconsecutive protein disulfide bonds[9]. To identify the substrates of DsbA and DsbC, we compared the disulfide proteomes of wild type, *dsbA*$^-$ and *dsbC*$^-$*E. coli* BW25113 strains. The BL21 strain was also analyzed. Together,

[1]College of Life Science, Beijing Normal University, Beijing, China. [2]National Institute of Biological Sciences, Beijing, China. [3]Key Lab of Intelligent Information Processing, Institute of Computing Technology, Chinese Academy of Sciences (CAS), Beijing, China. [4]University of Chinese Academy of Sciences, Beijing, China. [5]Institute of Molecular Medicine, Peking University, Beijing, China. [6]National Center for Mathematics and Interdisciplinary Sciences, Key Laboratory of Random Complex Structures and Data Science, Academy of Mathematics and Systems Science, CAS, Beijing, China. [7]These authors contributed equally to this work. Correspondence should be addressed to M.-Q.D. (dongmengqiu@nibs.ac.cn) or S.-M.H. (smhe@ict.ac.cn).

RECEIVED 26 SEPTEMBER 2014; ACCEPTED 3 DECEMBER 2014; PUBLISHED ONLINE 9 FEBRUARY 2015; DOI:10.1038/NMETH.3283

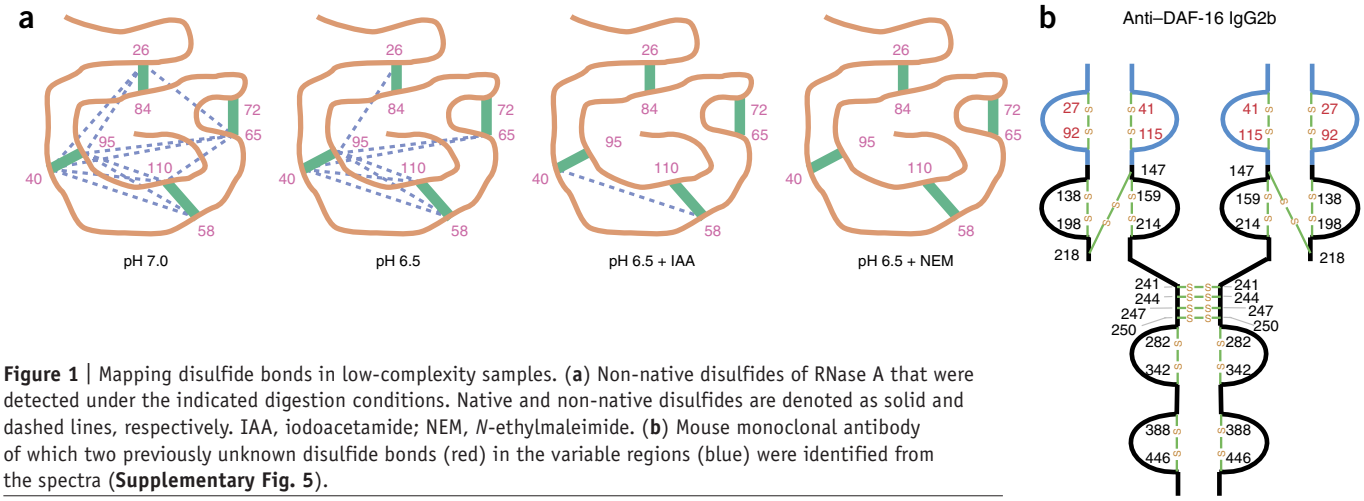


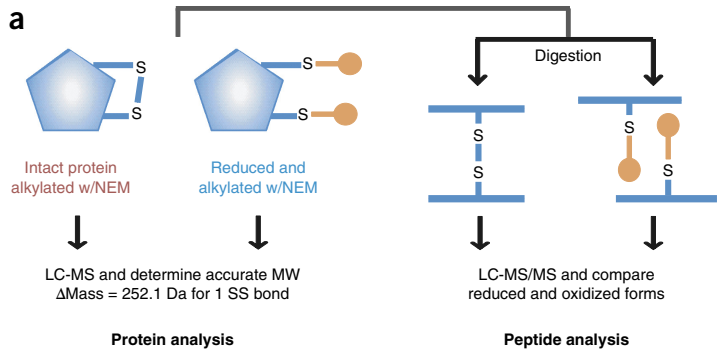
Figure 1 | Mapping disulfide bonds in low-complexity samples. (**a**) Non-native disulfides of RNase A that were detected under the indicated digestion conditions. Native and non-native disulfides are denoted as solid and dashed lines, respectively. IAA, iodoacetamide; NEM, *N*-ethylmaleimide. (**b**) Mouse monoclonal antibody of which two previously unknown disulfide bonds (red) in the variable regions (blue) were identified from the spectra (**Supplementary Fig. 5**).

199 disulfides in 150 proteins were identified using LTG digestion (**Supplementary Fig. 11** and **Supplementary Table 3**).

Out of 15 previously identified[10] and 23 suggested[11–14] protein substrates of DsbA, 23 have disulfide bonds detected in the BW25113 strains (**Supplementary Table 4**). Not surprisingly, 14 of them showed a >50% decrease in spectral counts of disulfide-bonded peptides in the *dsbA*⁻ mutant, and so did 30 new candidate substrates of DsbA. Moreover, for OmpA—a known DsbA substrate[10]—and four other proteins, deletion of *dsbA* did not significantly decrease the abundance of disulfide-linked peptides, nor the abundance of the protein, but greatly increased the spectral counts of peptides containing a reduced half-cystine (identified as NEM-modified forms). In addition, for three known DsbA substrates, OmpA, OppA and RcsF[10], although the canonical disulfide bonds OmpA Cys311-Cys323, OppA Cys297-Cys443 and RcsF Cys74-Cys118 showed no abundance decrease in the *dsbA*⁻ mutant, abnormal disulfide bonds appeared (OmpA Cys311-Cys311 and Cys323-Cys323, OppA Cys297-Cys297 and RcsF Cys109-Cys118) and suggested aberrant disulfide-mediated homodimers or homo-oligomers in the absence of DsbA. Thus, deletion of *dsbA* can cause multiple abnormalities on DsbA substrates, including lower levels of correct disulfide bonds, higher levels of reduced forms and the appearance of abnormal disulfide bonds (**Supplementary Table 4**).

We identified 17 nonconsecutive disulfide bonds in 14 proteins, some of which are known substrates of DsbC and others are new candidates (**Supplementary Table 5**). Many of them are also DsbA substrates.

To validate the novel disulfide bonds identified from *E. coli*, we purified eight proteins and analyzed them at both the protein



c

Verification of newly identified SS bonds in eight proteins

Protein	SS bond IDed	# SS bond from protein analysis	Verification by peptide analysis
GltI (2 Cys)	C111-C236	1	Yes
ArtI (2 Cys)	C55-C62	1	Yes
YjjA (2 Cys)	C82-C152	1	Yes
YicS (2 Cys)	C45-C94	1	Yes
EfeO (2 Cys)	C41-C103	1	Yes
LuxS (3 Cys)	C83-C128	0	Yes (low level)
BtuE (3 Cys)	C37-C83	0 (⅔) & 1 (⅓)	Yes
YhdW (4 Cys)	C39-C65 C150-C192	2	Yes

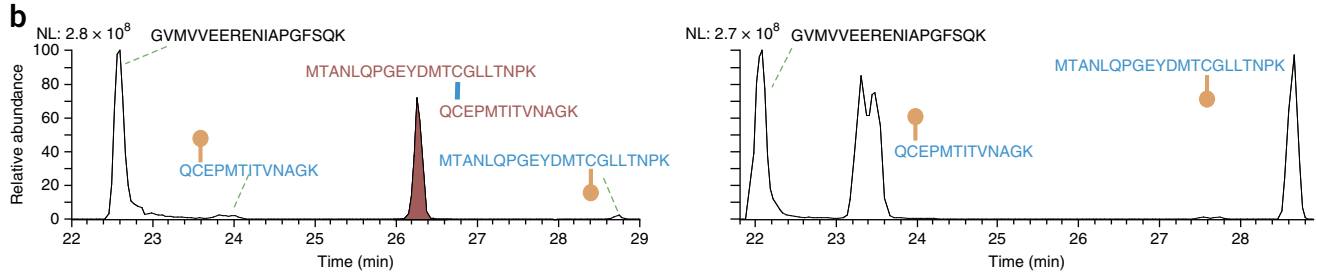


Figure 2 | Verification of newly identified disulfide (SS) bonds in *E. coli* proteins. (**a**) Two workflows for disulfide-bond verification at the protein and peptide levels. NEM, *N*-ethylmaleimide; LC, liquid chromatography; MW, molecular weight. (**b**) Extracted ion chromatograms of EfeO peptides. A cysteine-free peptide eluting at 22.4 ± 0.5 min is shown as a reference. NL, normalized level. (**c**) Summary of verification results.

and the peptide levels (**Fig. 2** and **Supplementary Fig. 12**). All nine disulfide bonds were fully confirmed except LuxS, which was validated at only the peptide level (**Fig. 2c**). LuxS is a key enzyme responsible for the synthesis of autoinducer 2 (AI-2), a cross-species chemical communication signal between bacterial cells[15]. The crystal structure of a LuxS homodimer of *Bacillus subtilis*[16] suggested that the Cys83-Cys128 disulfide bond of *E. coli* LuxS could be intermolecular (**Supplementary Fig. 13a**). We verified this by mutating Cys83 or Cys128 to serine (**Supplementary Fig. 13b**). Cys83 is critical for catalysis and Cys128 for coordinating a divalent metal ion next to the catalytic site[15]. Therefore, the Cys83-Cys128 disulfide bond should inactivate LuxS in a redox-sensitive manner. Consistent with this idea, there was a reversible increase of the LuxS covalent dimers and a decrease of AI-2 production upon treatment with H_2O_2 (**Supplementary Fig. 13c**,**d**). This potentially enables bacteria to adjust to environmental changes quickly and economically, without having to degrade and then synthesize LuxS.

Last, we analyzed the disulfide proteomes of human cells after LTG digestion. From untreated and diamide-treated human lung epithelial A549 cells, we identified 91 and 309 protein disulfide bonds, as well as 74 and 1,738 glutathionylated sites, respectively (**Supplementary Fig. 14** and **Supplementary Tables 6** and **7**). Additionally, 568 disulfide bonds were identified from proteins secreted by human umbilical vein endothelial cells (HUVECs) (**Supplementary Table 8**).

In addition to disulfide bonds functioning as structural elements, we also found many catalytic and regulatory disulfide bonds. Analysis of the distance between two disulfide-forming cysteine residues revealed that CXXC is the most prominent type among disulfide bonds identified from *E. coli* (**Supplementary Fig. 15**). Many disulfide bonds, often but not limited to the CXXC type, involve metal-binding cysteine residues. The formation of such disulfide bonds may expel metal ions and alter protein conformation[17]. Additionally, we found 30 new candidates of allosteric disulfide bonds[18] in HUVEC proteins (**Supplementary Table 9**). Further, some disulfide bonds contain catalytic cysteine residues and should cause inactivity, for example, Cys83-Cys128 of LuxS and Cys37-Cys83 of the glutathione peroxidase BtuE[19], in which Cys83 and Cys37 are catalytic residues, respectively. Certain members of the thioredoxin-like superfamily proteins catalyze disulfide exchange reactions with a characteristic CXXC motif that switches between the reduced and oxidized states. Such catalytic disulfides were identified in bacterial AhpF, DsbA, DsbC, DsbG, GrxB and TrxA.

For disulfide proteome analysis in prokaryotic or eukaryotic cells, we expect more disulfide bonds to be identified when a disulfide enrichment method becomes available and is incorporated into the current workflow. pLink-SS and some data files can be downloaded at http://pfind.ict.ac.cn/software/pLink/2014/pLink-SS.html. Additional data files are available at http://www.huanglab.org.cn/donglab/ or upon request.

METHODS

Methods and any associated references are available in the online version of the paper.

Note: Any Supplementary Information and Source Data files are available in the online version of the paper.

ACKNOWLEDGMENTS

We thank the National BioResource Project of Japan for providing bacterial strains; the antibody center of National Institute of Biological Sciences, Beijing, for the purified anti-DAF-16 IgG2 antibody; and X.-Z. Dong, A. Hühmer, D. Horn and Z. Hao for bringing the disulfide-bond problem to our attention and for encouragement. We also thank all members of the pFind group, X. Xiong, Y. Xia, S. Chen, K. Ye, Y. Zhou and members of the Dong lab for discussion and experimental support. This work was funded by the National Scientific Instrumentation Grant Program (2011YQ09000506 to M.-Q.D.), National Natural Science Foundation of China (grant no. 21475141 to S.-M.H.), Ministry of Science and Technology of China (973 grants 2013CB911203, 2012CB910602 to R.-X.S. and 2010CB912701 to S.-M.H.), CAS Knowledge Innovation Program (grant #KGCX1-YW-13 and ICT-20126033 to S.-M.H.), Strategic Priority Research Program of CAS (XDB13040600 to Y.F.), NCMIS CAS and municipal government of Beijing.

AUTHOR CONTRIBUTIONS

S.L. and B.Y. performed experiments, analyzed data and prepared the manuscript; S.-B.F. analyzed data and developed pLink-SS; J.-M.M., L.W., K.Z., Y.F. and R.-X.S. helped software development and data analysis; Y.-X.L. performed statistical analysis, pathway analysis and protein network analysis; M.-J.Z. helped with informatics analysis; P.L. and J.L. prepared the HUVEC samples; S.-M.H. directed software development and edited the manuscript; M.-Q.D. directed wet-lab experiments and wrote the manuscript.

COMPETING FINANCIAL INTERESTS

The authors declare no competing financial interests.

1. Alberts, B. *et al. Molecular Biology of the Cell* 3rd edn. **1**, 978–993 (Garland Science, 1994).
2. Hogg, P.J. *Trends Biochem. Sci.* **28**, 210–214 (2003).
3. van der Veen, A.G. & Ploegh, H.L. *Annu. Rev. Biochem.* **81**, 323–357 (2012).
4. Monnier, V.M. *et al. Ann. NY Acad. Sci.* **1043**, 533–544 (2005).
5. Tsai, P.L., Chen, S.-F. & Huang, S.Y. *Rev. Anal. Chem.* **32**, 257–268 (2013).
6. Yang, B. *et al. Nat. Methods* **9**, 904–906 (2012).
7. Zhang, W., Marzilli, L.A., Rouse, J.C. & Czupryn, M.J. *Anal. Biochem.* **311**, 1–9 (2002).
8. Xu, H., Zhang, L. & Freitas, M.A. *J. Proteome Res.* **7**, 138–144 (2008).
9. Depuydt, M., Messens, J. & Collet, J.F. *Antioxid. Redox Signal.* **15**, 49–66 (2011).
10. Kadokura, H., Tian, H., Zander, T., Bardwell, J.C. & Beckwith, J. *Science* **303**, 534–537 (2004).
11. Leichert, L.I. & Jakob, U. *PLoS Biol.* **2**, e333 (2004).
12. Hiniker, A. & Bardwell, J.C. *J. Biol. Chem.* **279**, 12967–12973 (2004).
13. Agudo, D., Mendoza, M.T., Castanares, C., Nombela, C. & Rotger, R. *Proteomics* **4**, 355–363 (2004).
14. Vertommen, D. *et al. Mol. Microbiol.* **67**, 336–349 (2008).
15. Zhu, J., Knottenbelt, S., Kirk, M.L. & Pei, D. *Biochemistry* **45**, 12195–12203 (2006).
16. Hilgers, M.T. & Ludwig, M.L. *Proc. Natl. Acad. Sci. USA* **98**, 11169–11174 (2001).
17. Fan, S.W. *et al. Protein Sci.* **18**, 1745–1765 (2009).
18. Cook, K.M. & Hogg, P.J. *Antioxid. Redox Signal.* **18**, 1987–2015 (2013).
19. Mishra, S. & Imlay, J. *Arch. Biochem. Biophys.* **525**, 145–160 (2012).

ONLINE METHODS

Chemicals. Acetonitrile, formic acid, acetic acid and ammonium bicarbonate were purchased from JT Baker. Dimethyl sulfoxide (DMSO), 1,1′-azobis(*N*,*N*-dimethylformamide) (diamide), ammonium acetate and other general chemicals were purchased from Sigma-Aldrich. *N*-Ethylmaleimide (NEM) was purchased from Pierce.

Bacterial strains and plasmids. Bacterial strains and plasmids are listed in **Supplementary Table 10**.

Peptides, purified proteins and proteases. 95% pure peptides were synthesized by GL Biochem. RNase A, haptoglobin, ovalbumin, transferrin, fetuin and BSA were purchased from Sigma-Aldrich. Lysozyme was purchased from Amresco. Insulin was purchased from Gibco. PNGase F was purchased from New England BioLabs. Elastase, subtilisin and proteinase K were purchased from Sigma-Aldrich. Trypsin was purchased from Promega. Glu-C and Lys-C were purchased from Roche. A mouse anti–DAF-16 monoclonal antibody (IgG2 subtype) was generated at the antibody center of National Institute of Biological Sciences (NIBS), Beijing, purified from the ascites fluid by protein G affinity chromatography. The variable-region cDNAs were sequenced using the mouse Ig-Primer Sets (Novagen).

Library of synthetic peptides linked through disulfide bonds. 47 peptides containing one cysteine and 25 peptides containing two cysteines were synthesized. 1,312 combinations of two 1-C peptides (0.8 mM each, final concentration (f.c.)) and one 2-C peptide (0.4 mM, f.c.) were allowed to oxidize in 20% DMSO at 37 °C for 17 h to form disulfide bonds[20]. After an 80-fold dilution in 1% HAc, 1 µL of sample was analyzed on an LTQ-orbitrap-ETD instrument (Thermo Fisher Scientific) connected to an Easy nano-LC II (Thermo Fisher Scientific). The products of each reaction were separated over a 30-min gradient from 100% buffer A (0.1% formic acid (FA) in water) to 35% buffer B (80% acetonitrile (ACN), 0.1% FA) on a 10-cm analytical column (75-µm inner diameter (i.d.)) with a 4-cm trap column (100-µm i.d.), both packed with 3-µm, 100-Å Luna C18 resin (Phenomenex). MS/MS spectra were acquired for the top three most intense ions in HCD-CID-ETD triple-play mode; +1 and +2 precursors were excluded; R = 60,000 in full scan, R = 7,500 in HCD scan, CID and ETD were analyzed in the linear ion trap; AGC targets were 5e5 for FTMS full scan, 2e4 for LTQ MS2 and 3e5 for reagent ion; minimal signal threshold for MS2 = 5,000; normalized collision energy was set to 35 for CID, 45 for HCD, and the reaction time was 100 ms for ETD.

Analysis of disulfide bonds of purified proteins or simple protein mixtures. RNase A, the mouse anti–DAF-16 monoclonal antibody and an equal-mass mixture of nine proteins (**Supplementary Fig. 7a**) were denatured at a protein concentration of 2 µg/µL in 8 M guanidine hydrochloride (GndCl, for proteinase K digestion) or 8 M urea (for other protease digestions) buffered by 100 mM Tris, pH 6.5, in the presence of 2 mM NEM at 37 °C for 2 h. The samples in 8 M urea were digested with Lys-C at a 1:100 enzyme/substrate ratio at 37 °C for 4 h. Then all the samples were fourfold diluted with 100 mM Tris, pH 6.5, digested with trypsin (1:20, 12 h), trypsin (1:20, 12 h) followed by Glu-C (1:40, 10 h), subtilisin (1:20, 4 h), elastase (1:20, 8 h) or proteinase K (1:20, 4 h) at 37 °C or, for Glu-C, 25 °C. To remove glycans from glycoproteins, we added PNGase F (112 NEB units per 6 µg proteins) to the nine-protein mixture 2 h before the digestion was quenched with 5% FA (f.c.). For the nine-protein mixture digested with proteinase K, an equal volume of 100 mM Tris, pH 6.5, was added to further dilute GndCl to 1 M before PNGase F was added. For RNase A and the antibody, 0.5 µg of protein digests were analyzed by LC-MS/MS. For the protein mixture, 2 µg of protein digests were analyzed.

Preparation of *E. coli* and human cell samples. The *E. coli* periplasmic fraction was prepared using a slightly modified osmotic-shock method[21]. A cold hypotonic solution of 5 mM $MgSO_4$ supplemented with 2 mM NEM was used to release proteins from the *E. coli* periplasm. Released proteins (200 µL) were precipitated on ice with 6 volumes of acetone or with 25% trichloroacetic acid (TCA) followed by cold acetone wash twice. Precipitated proteins were air dried, r-suspended in 100 mM Tris, pH 6.5, 8 M urea, 2 mM NEM to a concentration of 1–2 µg/µL and digested sequentially with Lys-C, trypsin and Glu-C as described above for RNase A digestion.

Twenty 10-cm plates of primary cultured human umbilical vein endothelial cells (HUVECs) were washed twice with PBS, starved in 6 mL starvation medium (M199 medium supplemented with 10 mM HEPES, pH 7.4, 10 µM hydrocortisone, 10 µg/mL insulin) for 4 h and washed with 5 mL M199 twice. Then cells were incubated with starvation medium (5 mL/plate), half (ten plates) with and half without 50 ng/mL VEGF. After 2 h, the medium was harvested very carefully to avoid touching the cells. After centrifugation at 3,320*g* for 5 min at 4 °C, the supernatants were collected and dialyzed in 8-kDa-pore-size tubing against a buffer reservoir of 1 mM Tris, pH 6.8, 1.3 mM NaCl, 0.5 mM NEM for 24 h. After dialysis, samples were frozen at −80 °C and lyophilized until the volume was reduced to about 100 µL. About 50 µg of proteins were obtained from ten plates of cells, which were precipitated overnight with 900 µL of 10% TCA in acetone at −20 °C. After centrifugation and a cold acetone wash, protein precipitates were dissolved in 8 M urea, 100 mM Tris, pH 6.5 and 2 mM NEM and digested sequentially with Lys-C, trypsin and Glu-C as above.

A549 cells (adenocarcinomic human alveolar basal epithelial cells) were cultured in 15-cm dishes to 90–95% confluency in RPMI-1640 medium supplemented with 10% FBS at 37 °C under 5% CO_2. After being washed twice with 15 mL PBS, pH 7.4, one plate of cells was treated with 15 mL PBS containing 1 mM diamide for 10 min at 37 °C, and another plate was incubated with PBS only. Diamide is a drug that quickly drains the cellular supply of reduced glutathione[22]. For the purpose of breaking the cells and quenching thiol reactions, both plates were washed with PBS and then incubated with 1.5 mL of 20% TCA at 4 °C. After 20 min, cells were scraped off plates and centrifuged at 20,000*g* for 30 min at 4 °C. The pellets were washed once with 10% TCA, once with 5% TCA, twice with cold acetone, air dried and then resuspended in 8 M urea, 4 mM NEM and 100 mM Tris, pH 6.5. The protein concentration was measured using the BCA assay and adjusted to 2 µg/µL in 8 M urea, 2.5 mM NEM and 100 mM Tris, pH 6.5. The Lys-C, trypsin and Glu-C digestion was done sequentially as above. The cell lines were not tested for mycoplasma contamination or authenticated for these experiments.

doi:10.1038/nmeth.3283

Mass spectrometry analysis. For RNase A, IgG2 and the ten-protein mixture, digestion products were directly loaded onto a 100-μm i.d., 10-cm-long analytical column packed with 3-μm, 100-Å Luna C18 resin and analyzed by LTQ-Orbitrap-ETD combined with an Agilent 1200 quaternary pump. The HPLC gradients were adjusted according to sample complexity: for RNase A and IgG, a 100 μL/min flow rate, 0–50 min from 100% buffer A (0.1% FA in water) to 35% buffer B (80% ACN, 0.1% FA), increasing to 100% buffer B in the next 5 min, then returning to 100% buffer A in 3 min and ending with a 13-min buffer A wash with a gradient flow rate from 0.1 mL/min to 0.2 mL/min; for the standard protein mixture, the 0–35% buffer B gradient was lengthened to 95 min and the 35–100% buffer B gradient was lengthened to 10 min. The MS parameters were similar to those used for synthetic peptides as above, except that only HCD spectra were collected for the top six (RNase A and IgG2) or ten (ten-protein mix) most intense precursors.

For the *E. coli*, human A549 and HUVEC samples, the digested peptides were off-line fractionated using a strong-cation exchange (SCX) column (250 μm × 2 cm) that was preceded by a section (250 μm × 2 cm) of reverse-phase C18 resin as previously described[6]. For the *E. coli* periplasmic fraction, seven fractions were collected by eluting with 0.025, 0.05, 0.1, 0.2, 0.35, 0.5 and 1.0 M ammonium acetate, 5 μL each. For the HUVEC samples, nine 5-μL SCX fractions were collected at 0.025, 0.05, 0.075, 0.1, 0.15, 0.2, 0.35, 0.5 and 1.0 M ammonium acetate. For the A549 samples, eight 5-μL fractions were collected at 0.025, 0.05, 0.1, 0.15, 0.2, 0.35, 0.5 and 1.0 M ammonium acetate. These were analyzed on a Q-Exactive MS instrument (Thermo Fisher Scientific) interfaced with an Easy nano-LC 1000 liquid chromatography system (Thermo Fisher Scientific). Each fraction was loaded onto a 75 μm × 6 cm trap column that was packed with 10-μm, 120-Å ODS-AQ C18 resin (YMC Co.) and connected through a micro-Tee to a 75 μm × 10 cm analytical column packed with 1.8-μm, 120-Å UHPLC-XB-C18 resin (Welch Materials). After desalting, peptides were separated over a 100-min linear gradient from 100% buffer A (0.1% FA) to 30% buffer B (100% ACN, 0.1% FA), then a 10-min gradient from 30% to 80% buffer B, reaching 100% buffer B in the next 1 min and maintaining at 100% buffer B for 3 min before returning to 100% buffer A in 2 min and ending with a 4-min 100% buffer A wash. The flow rate was 250 or 300 nL/min.

Following each full scan (R = 140,000), HCD spectra (R = 17,500) were collected for the ten most intense precursors carrying +3, +4, … or +7 positive charges; to increase the identification of loop-linked disulfide bonds, we carried out a technical repeat for each fraction in which +2 precursors were included; AGC target was set to 1e6 for FTMS full scan and 5e4 for MS2; minimal signal threshold for MS2, 4e4; normalized collision energy, 27%; peptide match preferred.

Training pLink for disulfide-bond analysis. *Comparing HCD and ETD.* Disulfide-linked peptides made of chemically synthesized peptide were identified using pLink by searching the spectra collected for a given reaction against a database consisting of only the peptide sequences used in the reaction and their reversed sequences. The *in silico* digestion was not limited to any specific proteases. Only a^{1+}, a^{2+}, b^{1+}, b^{2+}, y^{1+} and y^{2+} ions were considered for HCD spectra and c^{1+}, c^{2+}, $[c-H]^{1+}$, $[c-H]^{2+}$, z^{1+}, z^{2+}, $[z+H]^{1+}$, and $[z+H]^{2+}$ ions for ETD spectra. For each peptide-spectrum match, pLink first calculates a *p*-value, which estimates how likely it is a random match; then it calculates the corresponding *E*-value, which is the product of the *p*-value and the number of candidate cross-links falling within the mass-tolerance window in the searched database. The number of disulfide-linked peptides identified from HCD spectra (1,371 unique pairs, requiring *E*-value <0.01) far exceeded that from ETD spectra (412 unique pairs, *E*-value <1) (**Supplementary Fig. 1a**).

To analyze why HCD outperformed ETD, we focused on 2,121 pairs of HCD/ETD spectra of disulfide-linked synthetic peptides. Those consisting of two identical peptides were not included. Intensities of precursor ions, p ions (peptide ions, resulting from the breaking of the disulfide bond but no cleavage on the peptide backbone) and the other fragments relative to the total ion intensity were calculated for each spectrum. The ETD spectra, especially those of +3 or +4 peptide pairs, were dominated by precursor ions (**Supplementary Fig. 1b**). In HCD spectra, the proportion of fragment ions (those that are not precursor or p ions as defined above) were close to 100% of all ions, whereas in ETD spectra, it was much lower (**Supplementary Fig. 1b**). We thus conclude that ETD suffers from inefficient fragmentation of precursors and that HCD is better than ETD for disulfide-bonded peptide pairs.

ETD MS2 followed by CID MS3 has been used to map disulfide bonds of purified proteins[23]. A statistical analysis of our standard data set finds that this strategy has serious limitations. In 49% (1,044/2,121) of the ETD spectra, especially those of +3 or +4 precursors, the p-ion intensity was less than 10% of the total ion intensity and sometimes even 0 (133 spectra, 6.2% of the total) (**Supplementary Fig. 1b**). This means that for about half of the peptide pairs, the CID MS3 spectra of separated peptides would be of very low intensity or sometimes nonexistent.

Optimizing pLink for identification of disulfide-linked peptides from HCD spectra. The HCD search results above were further filtered by requiring an *E*-value lower than 1×10^{-5} and no more than two best-scoring spectra for each unique identification (different peptide sequences, positions of disulfide bonds, or charge states). We obtained 2,287 high-quality HCD spectra of disulfide-linked peptide pairs, and they were randomly divided into a training group (1,146) and a testing group (1,141).

The 1,146 spectra were preprocessed as described[6] and examined for the presence of various ion types of different charge states, including: (i) the common a, b and y ions; (ii) the yb and ya ions generated by two backbone-cleavage events[6]; (iii) p and $p_{(-2)}$ ions from the breaking of the S-S bond (**Supplementary Fig. 2a**); (iv) $p_{(-34)}$ and $p_{(+32)}$ ions from the breaking of the C-S bond (**Supplementary Fig. 2b**); (v) p_b, $p_{(-2)b}$, $p_{(-34)b}$, $p_{(+32)b}$, p_y, $p_{(-2)y}$, $p_{(-34)y}$ and $p_{(+32)y}$ ions from the breaking of both the disulfide linkage and the peptide backbone (**Supplementary Fig. 2a**,**b**); and (vi) precursor and neutral-loss ions.

For every ion type, the count ratio, gain ratio and average intensity were first calculated for individual spectra[6]. Then they were summed and divided by the number of spectra containing the ion type in question to obtain averaged values. The match significance of an ion type is the product of the three averaged values. Different types of ions were ranked by match significance.

The 25 most prominent fragment-ion types were y^{1+}, b^{1+}, yb^{1+}, y^{2+}, a^{1+}, b^{2+}, yb^{2+}, p^{1+}, p^{2+}, $p_{(+32)}{}^{2+}$, $p_{(+32)}{}^{1+}$, ya^{2+}, p^{3+}, y^{3+}, a^{2+}, $p_{(-2)}{}^{2+}$, $p_{(-34)}{}^{1+}$, b^{3+}, $p_{(+32)b}{}^{1+}$, $p_{(+32)}{}^{3+}$, $p_y{}^{1+}$, yb^{3+}, $p_{(-2)}{}^{1+}$, $p_{(-2)}{}^{3+}$ and ya^{1+}. Precursor and neutral-loss ions were omitted because they do not contribute to peptide-spectrum matching. In a normal pLink search, they are removed from the peak list in the preprocessing step.

To optimize the use of ion types in pLink search for disulfide bonds (pLink-SS), we tested combinations of different ion types on the remaining 1,141 spectra of known identities. These spectra were mixed with 46,526 HCD spectra of regular peptides identified from an *E. coli* lysate and searched under different ion-type settings against a database consisting of the 72 synthetic peptides (**Supplementary Table 1**) appended to an *E. coli* protein database. To the basic group of a^{1+}, a^{2+}, b^{1+}, b^{2+}, b^{3+}, y^{1+}, y^{2+} and y^{3+} ions, we systematically added other ion types from the ranking list above, compared the *E*-value distributions of the forward- and reverse-database matches, recall and precision. The addition of yb and ya ions substantially improved the identification of disulfide-bonded peptides, and further adding the peptide ions brought only a modest improvement. In the end, multiple ion-type combinations led to similar results as the improvement plateaued (not shown). In fact, we could achieve the best performance for disulfide bond analysis by using the ion types for BS3–cross-linking (a^{1+}, a^{2+}, b^{1+}, b^{2+}, b^{3+}, y^{1+}, y^{2+}, y^{3+}, ya^{1+}, yb^{1+}) together with ten ion types that are specific for disulfide-bond peptides (p^{1+}, p^{2+}, $p_{(+32)}{}^{1+}$, $p_{(+32)}{}^{2+}$, $p_{(-34)}{}^{1+}$, $p_{(-34)}{}^{2+}$, $p_b{}^{1+}$, $p_y{}^{1+}$, $p_{(-2)b}{}^{1+}$, $p_{(-2)y}{}^{1+}$) (**Supplementary Fig. 2d**). This is the finalized ion-type setting for pLink-SS.

Data analysis by pLink-SS. The hardware requirement of pLink-SS is a Microsoft Windows PC or workstation, preferably with 64-bit architecture and a minimum of 4 GB RAM.

Operating system requirements are pLink, Microsoft Windows Vista or above, and Microsoft .Net Framework 3.5 or above. Thermo Scientific Xcalibur 2.0 or above (or the free version Thermo MSFileReader) should be installed to extract MS2 data from RAW files.

For low-complexity samples (purified RNase A, anti-DAF-16 IgG, or the ten-protein mix), the database consisted of the sequences of the analyte protein(s) and all the proteases used. For human cell samples, the database for pLink search contained only the proteins identified from the samples using Prolucid[24] and DTASelect 2 (ref. 25) (≤1% FDR at the peptide level, requiring only one peptide for a protein identification). For *E. coli* samples, a full *E. coli* protein database was used. The raw data of the ten-protein sample were preprocessed using pParse[26], which was set to exclude coeluting precursor ions and precursors of +1, +2 charge.

The parameters for pLink-SS search were as follows: four missed cleavage sites for trypsin; five missed cleavage sites for trypsin/Glu-C; for elastase, subtilisin and proteinase K, no specificity; peptide length 4–25 aa; fixed modification of −1.007285 Da on cysteine and the disulfide mass was set to 0 to allow the identification of more than one disulfide bond in a peptide pair; candidate pairs satisfying Mα+ Mβ+ Mlinker < M ± 5 Da were scored for spectrum-peptide matching.

pLink-SS search results were filtered by requiring ≤10-p.p.m. deviation in the observed precursor mass from the monoisotopic or the first, second, third, or fourth isotopic mass of the matched candidate. For spectra preprocessed by pParse, the requirement was no more than 10-p.p.m. deviation from the monoisotopic mass. Additional filtering criteria are as follows.

For low-complexity samples, candidate disulfide-linked peptides were filtered with an *E*-value cutoff of 0.01 and a FDR cutoff of 0.05, in which the FDR was calculated as described before for chemically cross-linked peptides[6].

For complex samples, inter- and intraprotein disulfide bonds were filtered separately, so the small number of intraprotein disulfide-bond matches would be free from the noise of numerous random matches of interprotein disulfide bonds (**Supplementary Fig. 10b**). The FDR of interprotein disulfide-bond identifications was calculated as described for chemically cross-linked peptides[6]. For intraprotein disulfide-bond identifications, FDR was calculated by dividing the number of reverse-reverse database matches (F) with the number of forward-forward database matches (T), that is, FDR = F/T. A random match has an equal probability of being a F-F or a R-R pair (**Supplementary Fig. 10c**). Intraprotein disulfide bonds were filtered by requiring *E*-value ≤ 0.01 and spectral count ≥ 2, and the FDR was less than 0.01. Interprotein disulfide bonds were filtered by requiring *E*-value ≤ 0.01, FDR ≤ 0.05, and spectra count ≥ 2. The disulfide-bonded peptide pairs were further classified as interlinked or complex forms on the basis of the number of peptides (1 or 2), the number of cysteine residues, and the number of disulfide bonds. pLink-SS also standardizes the report format for complex forms in protein_name(site_1, site_2, site_3)-protein_name(site_x) or protein_name(site_1, site_2)-protein_name(site_x, site_y). The MS2 or MS1 spectra of disulfide-bonded peptides in complex forms or between two copies of the same peptide were further evaluated manually. For instance, in BSA(250,251,258)-BSA(205), it cannot be determined whether the interlinked disulfide bond is Cys250-Cys205, Cys251-Cys205 or Cys258-Cys205 (**Supplementary Fig. 8b**). In the original pLink-SS output, BSA(250,251,258)-BSA(205) may be reported as BSA(250)-BSA(205) or BSA(258)-BSA(205). Intrapeptide disulfide bonds (loop-linked forms) were filtered from linear peptides, with FDR <0.05, *E*-value <0.01, and spectral count 2. The spectra displayed in this paper are annotated using pLabel[6].

Comparing pLink-SS to MassMatrix and StavroX. As MassMatrix[8] does not support nonspecific protease digestion, the data of Lys-C/trypsin/Glu-C–digested RNase A and ten standard proteins were used to compare the performance of pLink-SS with MassMatrix. The protein sequence database contained, in addition to RNase A or the ten standard proteins, Lys-C, trypsin, Glu-C, elastase, subtilisin and proteinase K. The MassMatrix parameters were precursor mass tolerance = 10 p.p.m., 0.05-Da mass tolerance for fragments in the exploratory mode, peptide length 4–25 aa, sequence specificity = trypsin or Glu-C, up to four missed cleavages, minimal pp score = 5.0, minimal pp2 score = 5.0, and minimal pp_tag score = 1.3, and only the highest scoring match was used in statistical analysis. On the RNase A data, pLink-SS and MassMatrix gave similar results. Out of the 74 known SS bonds in the ten-protein mix, pLink-SS and MassMatrix identified 36 and 31, respectively, with the overlap of 21. However, MassMatrix made a great number of false identifications. In the pLink result, the correctly and wrongly identified spectra of SS

doi:10.1038/nmeth.3283

bonds number at 611 and 4, respectively. In the MassMatrix result, those numbers are 264 and 283, respectively, and 262 out of the 283 wrongly identified spectra are in the form of tripeptides. This shows that the tripeptide results of MassMatrix have a very high false positive rate.

We were able to compare pLink-SS with StavroX[27] using the RNase A data as above but not the ten-protein data because the latter caused StavroX to quit in the middle of the search. The protein sequence database was the same as above. The StavroX parameters were precursor mass tolerance = 10 p.p.m.; 20-p.p.m. mass tolerance for fragments; minimal peptide length is 4 aa; cleavage sites are lysine, arginine and glutamate, allowing 2, 1 and 2 miss-cleavages, respectively; no amino acid modifications. In the StavroX result, all except nine identified spectra had a score value greater than 0 (from 16 to 60), and only five spectra remained at 5% FDR.

Verification of newly identified disulfide bonds in *E. coli* proteins. A subset of *E. coli* proteins with novel disulfide bonds were verified with purified recombinant proteins. Their coding sequences were cloned into pET22b in-frame with a C-terminal 6× His tag. LuxS C83S and C128S mutants were constructed using a standard QuikChange site-directed mutagenesis procedure (Stratagene). Proteins were expressed in BL21. The periplasmic fractions were extracted as above and incubated with Ni-NTA beads (GE Healthcare) that had been equilibrated with wash buffer (40 mM Tris, pH 6.5, 30 mM imidazole, 100 mM NaCl) for 2 h at 4 °C. Then, the beads were washed five times with 1 mL wash buffer. Proteins were eluted with 50 mM Tris, pH 6.5, 500 mM imidazole on ice for 30 min. After TCA precipitation and cold acetone wash, proteins were brought into 8 M urea, 100 mM Tris, pH 7.0, to a concentration of 1 μg/μL. For detection of disulfide bonds in each protein, a 10-μL aliquot was reduced with 5 mM TCEP and then alkylated with 10 mM NEM, and another 10-μL aliquot was alkylated with 10 mM NEM without reduction. One quarter of each sample was acidified with 2.5 μL 5% FA for protein LC-MS analysis on LTQ-orbitrap, and another 2.5 μL was digested with Lys-C, Lys-C/trypsin, or Lys-C/Glu-C for peptide analysis.

The dimer bands of LuxS proteins were cut for in-gel digestion as described before[28] except that reduction and alkylation were omitted, and dehydrated gel slices were rehydrated with 100 mM Tris (pH 6.5) containing trypsin and Glu-C at 10 ng/μL each for overnight digestion.

LC-MS analysis of proteins was conducted on an LTQ-orbitrap mass spectrometer coupled to an Agilent 1200 binary pump. Proteins were loaded onto a 10-cm analytical column packed with 10-μm, 90-Å C4 resin (YMC Co.) and subjected to the following gradient: constant flow rate of 0.1 mL/min, from 100% buffer A (0.1% FA in water) to 5% buffer B (80% ACN, 0.1% FA) over 1 min, increasing to 50% buffer B in the next 29 min, then to 100% buffer B over 5 min, returning to 100% buffer A in the next 2 min and ending with a 14-min buffer A wash. MS1 spectra were collected at R = 100,000 with three microscans. Averaging MS1 spectra were deconvoluted using the Xtract tool in QualBrowser with the following parameters: mass range = appropriate for the protein molecular ion of interest, S/N ratio cutoff = 10%, fit factor = 30, remainder value = 30%.

H_2O_2 treatment and autoinducer-2 bioassay. *E. coli* strains were grown at 37 °C in Luria-Bertani (LB) broth to OD_{600} of 1.5–2.0. Cells were collected by centrifugation at 1,900*g* for 20 min and then suspended in fresh medium. Cells were treated or not with 40 mM H_2O_2 for 15 min before the cells were centrifuged at 1,900*g* for 10 min. About half of the pelleted cells were precipitated with 10% TCA, washed with cold acetone twice and then lysed in 1× SDS loading buffer without reducing agents. The other half of the pelleted cells were resuspended in fresh LB medium without H_2O_2 and cultured for another 30 min before the supernatant and the cells were collected once again; the cells were TCA precipitated as above, and the supernatant was further centrifuged at 13,000*g* for 5 min and filtered with a 0.22-μm membrane (Millipore, Millex) for the AI-2 bioassay. The LuxS monomer and dimers were detected by western blot with anti-His mAb-HRP (Medical & Biological Laboratories, D291-7).

The autoindicer-2 bioassay was conducted according to an established method with slight modifications[29]. In brief, *Vibrio harveyi* BB170 was grown in Autoinducer Bioassay (AB) medium[30] for 16 h and then diluted 5,000-fold in fresh AB medium. 20 μL of cell-free culture fluid of the treated or untreated *E. coli* cells were added to 180 μL of diluted cells in Costar 96-well microtiter wells (Corning) and cultured at 30 °C in a rotary shaker at 250 r.p.m. Bioluminescence measurements were taken every 30 min with EnSpire Multimode Plate Reader. Because the early time-point readings were too low to be reliable, measurements taken after 5 h of incubation were used to estimate AI-2 production as previously suggested[31].

Bioinformatics analysis. Functional pathways enrichment analysis on proteins that had increased levels of disulfide bonds or glutathionylation after 1 mM diamide treatment was conducted using the ClueGO[32] plug-in of Cytoscape (version 3.0.2)[33]. The KEGG and Reactome databases were used. All the identified proteins from the A549 cells are served as the background. A right-sided hypergeometrical test with a Benjamini-Hochberg correction for multiple testing was used to estimate the significance of enrichment and to control the false discovery rate. To reduce redundancy, ClueGO groups the pathways that have similar associated genes using the kappa score, and the most significant GO term within a group is selected as the group leading term. Those with a corrected *P* value <0.01 were considered significantly enriched.

Existing knowledge about disulfide-forming cysteine residues was extracted from the UniProt database. Functional protein interaction network analysis was performed on the basis of the interaction data from the STRING database version 9.1 (ref. 34), and only the interactions with a STRING score higher than 0.7 was used. Cytoscape was used to visualize the data. Densely connected subnetworks that can be considered as protein complexes or functional modules were detected using the MCODE[35] (molecular complex detection) plug-in of Cytoscape.

Motif analysis of glutathionylation sites was performed using pLogo[36] (http://plogo.uconn.edu). A 13-aa sequence containing the glutathionylated cysteine and 6-aa flanking sequences on either side was extracted. Glutathionylated sites fewer than six residues away from either end of a protein were excluded. The unmodified cysteine residues from the same proteins and their flanking sequences were used as background.

20. Tam, J.P., Wu, C.R., Liu, W. & Zhang, J.W. *J. Am. Chem. Soc.* **113**, 6657–6662 (1991).
21. Nossal, N.G. & Heppel, L.A. *J. Biol. Chem.* **241**, 3055–3062 (1966).
22. Kosower, N.S. & Kosower, E.M. *Methods Enzymol.* **251**, 123–133 (1995).
23. Wu, S.L. *et al. Anal. Chem.* **81**, 112–122 (2009).
24. Xu, T. *et al. Mol. Cell. Proteomics* **5**, S174 (2006).
25. Tabb, D.L., McDonald, W.H. & Yates, J.R. *J. Proteome Res.* **1**, 21–26 (2002).
26. Yuan, Z.F. *et al. Proteomics* **12**, 226–235 (2012).
27. Götze, M. *et al. J. Am. Soc. Mass Spectrom.* **23**, 76–87 (2012).
28. Pandey, A., Andersen, J.S. & Mann, M. *Sci. Signal.* **2000**, pl1 (2000).
29. Bassler, B.L., Wright, M. & Silverman, M.R. *Mol. Microbiol.* **13**, 273–286 (1994).
30. Turovskiy, Y. & Chikindas, M.L. *J. Microbiol. Methods* **66**, 497–503 (2006).
31. Surette, M.G. & Bassler, B.L. *Proc. Natl. Acad. Sci. USA* **95**, 7046–7050 (1998).
32. Bindea, G. *et al. Bioinformatics* **25**, 1091–1093 (2009).
33. Smoot, M.E., Ono, K., Ruscheinski, J., Wang, P.-L. & Ideker, T. *Bioinformatics* **27**, 431–432 (2011).
34. Franceschini, A. *et al. Nucleic Acids Res.* **41**, D808–D815 (2013).
35. Bader, G.D. & Hogue, C.W. *BMC Bioinformatics* **4**, 2 (2003).
36. O'Shea, J.P. *et al. Nat. Methods* **10**, 1211–1212 (2013).

doi:10.1038/nmeth.3283