A photocleavable surfactant for top-down proteomics

Kyle A. Brown¹, Bifan Chen¹, Tania M. Guardado-Alvarez¹, Ziqing Lin²,³, Leekyoung Hwang¹, Serife Ayaz-Guner², Song Jin¹ and Ying Ge¹,²,³*

We report the identification of a photocleavable anionic surfactant, 4-hexylphenylazosulfonate (Azo), which can be rapidly degraded by ultraviolet irradiation, for top-down proteomics. Azo can effectively solubilize proteins with performance comparable to that of sodium dodecyl sulfate (SDS) and is compatible with mass spectrometry. Azo-aided top-down proteomics enables the solubilization of membrane proteins for comprehensive characterization of post-translational modifications. Moreover, Azo is simple to synthesize and can be used as a general SDS replacement in SDS-polyacrylamide gel electrophoresis.

A comprehensive analysis of ‘proteoforms’ that arise from genetic variations and post-translational modifications (PTMs) is essential for deciphering biological systems at a functional level. The conventional ‘bottom-up’ proteomics analyzes peptides from protein digests, which does not directly identify proteoforms and is suboptimal for characterizing PTMs and sequence variants. By contrast, top-down mass spectrometry (MS)-based proteomics analyzes intact proteins and is the most powerful method to comprehensively characterize proteoforms deciphering the PTMs together with sequence variations. However, despite its promise, top-down proteomics still faces major challenges.

One challenge in top-down proteomics is protein solubility, especially for membrane proteins, which constitute a large proportion of the proteome, play a critical role in many cellular functions and are important drug targets. To effectively extract proteins from cells or tissues, surfactants (also known as detergents) are commonly included in the extraction buffer. Unfortunately, conventional ionic surfactants are not compatible with MS because they greatly suppress protein MS signal. Therefore, surfactants need to be removed before MS analysis, which may result in protein loss and degradation. Developing MS-compatible surfactants that can be quickly degraded into innocuous non-surfactant byproducts before MS analysis can help address the protein solubility challenge in top-down proteomics. Efforts have been made in developing various acid-labile surfactants, which have been effective for bottom-up proteomics; however, none have demonstrated direct compatibility with intact protein MS for top-down proteomics.

Here we evaluated photocleavable surfactants by inserting a photocleavable moiety in between the surfactant hydrophilic head and hydrophobic tail that can be rapidly cleaved and degraded upon ultraviolet (UV) irradiation before MS analysis. Degradation via a photochemical reaction has the advantages of being simple and fast, and one can easily control it by turning on a UV lamp on and off. Our goal was to identify a strong photocleavable surfactant that can effectively solubilize proteins during sample preparation with performance similar to that of sodium dodecyl sulfate (SDS), but which is also compatible with top-down proteomics.

We performed a systematic screening of many synthesized candidates (Supplementary Notes 1–3 and Supplementary Table 1) and identified 4-hexylphenylazosulfonate (Fig. 1b and Supplementary Figs. 1–2), hereinafter referred to as Azo, as the top-performing surfactant, as it not only is water soluble, but also greatly improves protein extraction (Supplementary Table 1). Notably, Azo was simple to synthesize, requiring only two steps (Fig. 1c), and could be effectively purified by recrystallization, making it an ideal candidate for general use as a surfactant in biochemical applications. For instance, we have used Azo instead of SDS to perform polyacrylamide gel electrophoresis (PAGE) (Supplementary Fig. 3), demonstrating that Azo could be used as an SDS replacement in SDS–PAGE.

We further investigated the photodegradation kinetics of the Azo dissociating into 4-hexylphenol, 4-hexylbenzene, nitrogen and hydrogen sulfate upon irradiation with a 100-W high-pressure mercury lamp for 0, 10, 30, 60, 90 and 120 s using UV-vis spectroscopy (Fig. 1d). By comparing several degradation conditions, we found that the presence of organic solvent and acid facilitates rapid degradation of Azo (Supplementary Fig. 4).

Next, we examined the efficacy of Azo for solubilizing proteins from cardiac tissues using a direct side-by-side comparison with SDS, and its acid-labile mimic, MS-compatible slowly degradable surfactant (MaSDeS), as well as n-dodecyl-β-D-maltoside (DDM), a commonly used surfactant for native MS. The SDS–PAGE gel (Fig. 1e) and protein assay (Fig. 1f) show that the addition of 0.5% Azo to the extraction buffer, labeled as E3(Azo), drastically improved the solubilization of proteins when compared with the control without surfactant, E3(NS), which barely solubilized proteins after the depletion of soluble proteins using HEPES buffer, E1 and E2. Overall, the anionic surfactants, Azo, SDS and MaSDeS, are highly effective in solubilizing proteins compared with the non-ionic surfactant, DDM (Fig. 1e,f). Furthermore, a western blot analysis confirmed the presence of common cardiac membrane proteins in E3(Azo), demonstrating the successful extraction of integral membrane proteins by Azo (Supplementary Fig. 5).

More important, Azo surfactant is MS-compatible. We first performed direct infusion electrospray ionization (ESI)–MS analysis using ubiquitin (Ubi) in the presence of 0.1% of a chosen surfactant, without an additional desalting step (Fig. 1g). The results showed that the presence of 0.1% SDS completely suppressed the MS signal and 0.1% MaSDeS significantly suppressed the MS signal. In contrast,
0.1% DDM and 0.1% Azo yielded comparable MS signals, showing minimal signal suppression when compared with the control with no surfactant (Fig. 1g). We also examined the effect of the UV irradiation on MS analysis of proteins and found that Azo had minimal effect on the qualitative and quantitative analysis of intact proteins in the presence of reducing agent, Tris(2-carboxyethyl)phosphine hydrochloride (TCEP), or free methionine (Supplementary Note 4 and Supplementary Figs. 6–9). Furthermore, we have performed a systematic comparison of Azo with a broader range of commonly used surfactants to evaluate their ability to solubilize proteins from the insoluble cardiac tissue pellets and subsequently assessed their MS compatibility (Supplementary Note 5 and Supplementary Figs. 6–9). Additionally, we have performed experiments (E3) following the initial HEPES buffer extractions (E1 and E2) to deplete the cytosolic proteins from the cardiac tissue. Error bars represent s.e.m. for protein assay experiments (n = 3).

We observed various PTMs, including acetylation, methylation, phosphorylation and palmitoylation (Supplementary Tables 5–10). Azo greatly improved the depth of the detection and revealed many proteins that were undetectable in the control sample (Supplementary Fig. 14). For example, Azo enabled the detection and identification of multiple proteoforms of an important Z-disk protein, calspartin-1 (Supplementary Fig. 14d).

We further showed that Azo can effectively extract and enable the top-down proteomic analysis of membrane proteins from cardiac tissues (Fig. 2 and Supplementary Fig. 15), as well as human embryonic kidney (HEK293T) cells (Supplementary Fig. 16 and Supplementary Tables 9 and 10). Under the optimal UV-degradation conditions (which include organic solvent at low pH), many hydrophobic proteins were soluble after Azo degradation. Using cardiac tissue as an example, we identified several important integral membrane proteins such as phospholamban (PLN), receptor-expressing enhancer, CKE7, and succinate dehydrogenase cytochrome b560 with one, two and three transmembrane domains (TMDs), respectively.
Fig. 2 | Photocleavable Azo-enabled top-down membrane proteomics. a–c, MS and tandem MS analysis of representative membrane proteins from Azo-aided extraction of the cardiac tissue. a, PLN and palmitoylated PLN (palmPLN) with palmitoylation localized at cysteine 36 residue (a), and receptor expression-enhancing protein 5 (b). The sequences below the spectra represent the fragmentation maps with sequence coverage and PTM localization based on online RPLC–MS/MS analysis. The regions representing the transmembrane domains are highlighted by blue shading. The data represent +16 Da. c, Complete analysis of ATP synthase subunit proteins from cardiac tissue. Overall, all ATP synthase subunits (e, f, g, ATP6, ATP8, DAPIT, C and 6.8PL) that exist in the inner membrane space (IMS) and the subunits (a, b, c, δ, OCSP, F6, d and γ) located in the mitochondrial matrix were detected. The schematic of ATP synthase was modified on the basis of a previous publication by He et al.20. Data are representative of two independent experiments.

(See Fig. 2a,b and Supplementary Fig. 17). Notably, we detected not only intact unmodified PLN, but also its highly abundant palmitoylated proteoform (Fig. 2a). We confidently localized the palmitoylation modification to cysteine 36 within the transmembrane region on the basis of the unmodified b33 ion and the palmitoylated b38 ion (Fig. 2a). We confidently localized the palmitoylation modification to cysteine 36 with TMDs (Supplementary Table 7) directly from cardiac tissue. Overall, all ATP synthase subunits (e, f, g, ATP6, ATP8, DAPIT, C and 6.8PL) that exist in the inner membrane space (IMS) and the subunits (a, b, c, δ, OCSP, F6, d and γ) located in the mitochondrial matrix were detected. The schematic of ATP synthase was modified on the basis of a previous publication by He et al.20. Data are representative of two independent experiments.

Moreover, we confidently identified 46 subunits of the electron transport chain (Supplementary Table 6) and 51 proteins with TMDs (Supplementary Table 7) directly from cardiac tissue. Notably, all the subunits of the endogenous ATP synthase complex (e, f, g) that exist in the inner membrane space (IMS) and the subunits (a, b, c, δ, OCSP, F6, d and γ) located in the mitochondrial matrix were detected. The schematic of ATP synthase was modified on the basis of a previous publication by He et al.20. Data are representative of two independent experiments.

In summary, we report the application of a photocleavable MS-compatible surfactant to increase protein solubility for high-throughput top-down proteomics. Among all the surfactants we evaluated, we found Azo to be the only strong surfactant capable of effectively solubilizing proteins, including membrane proteins, without hindering downstream top-down MS analysis. Azo has the potential to further enhance top-down global proteomics when coupled to multidimensional separation, complementary fragmentation techniques and improved data acquisition strategies21 (Supplementary Note 6). We expect that Azo will facilitate a myriad of proteomic studies for understanding disease mechanisms and clinical diagnosis. Given the instrumental roles of surfactants in biochemical research, we envision that this photocleavable surfactant will have a broader impact beyond proteomics. Notably,
because Azo can be easily synthesized and purified, it can be used as a cleavable SDS replacement in general biochemical applications, for example, in SDS–PAGE.

Online content
Any methods, additional references, Nature Research reporting summaries, source data, extended data, supplementary information, acknowledgements, peer review information; details of author contributions and competing interests; and statements of data and code availability are available at https://doi.org/10.1038/s41592-019-0391-1.

Received: 5 August 2018; Accepted: 13 March 2019; Published online: 15 April 2019

References
1. Aebersold, R. et al. Nat. Chem. Biol. 14, 206–214 (2018).
2. Siuti, N. & Kelleher, N. L. Nat. Methods 4, 817–821 (2007).
3. Cai, W. X., Tucholski, T. M., Gregorich, Z. R. & Ge, Y. Expert Rev. Proteomics 13, 717–730 (2016).
4. Chen, B., Brown, K. A., Lin, Z. & Ge, Y. Nat. Chem. Biol. 14, 206–214 (2018).
5. Barrera, N. P. & Robinson, C. V. Anal. Chem. 90, 110–127 (2016).
6. Speers, A. E. & Wu, C. C. Rev. Biochem. 107, 3687–3714 (2007).
7. Loo, R. R., Dales, N. & Andrews, P. C. Protein Sci. 3, 1975–1983 (1994).
8. Wisniewski, J. R., Zougman, A., Nagaraj, N. & Mann, M. Nat. Methods 6, 359–362 (2009).
9. Laganowsky, A., Reading, E., Hopper, J. T. S. & Robinson, C. V. Nat. Protoc. 16, 359–362 (2013).
10. MacLennan, D. H. & Kranias, E. G. Nat. Rev. Mol. Cell Biol. 4, 566–577 (2003).
11. He, J. et al. Proc. Natl Acad. Sci. USA 115, 2988–2993 (2018).
12. Loo, R. R., Dales, N. & Andrews, P. C. Protein Sci. 3, 1975–1983 (1994).
13. MacLennan, D. H. & Kranias, E. G. Nat. Rev. Mol. Cell Biol. 4, 566–577 (2003).
14. He, J. et al. Proc. Natl Acad. Sci. USA 115, 2988–2993 (2018).

Acknowledgements
This research is supported by National Institutes of Health R01 GM117058 (to S.J. and Y.G.). Y.G. acknowledges R01 HL109810, R01 HL096971, R01 GM125085 and S10 OD018475. MaSDeS was a gift from S. Saveliev (Promega Corporation). We thank A. Chen, E. Chang and W. Tang for their assistance in the early stage of the project, S. Mitchell and T. Tucholski for the help with graphics, and T. Hacker for providing the swine hearts. We thank M. Willetts at Bruker for his assistance with DataAnalysis software. We also acknowledge A. Carr, E. Bayne and J. Melby for their help testing the Supplementary Protocol to ensure reproducible results.

Author contributions
K.A.B. designed and performed experiments, analyzed the data and wrote the manuscript. B.C. designed and performed experiments, analyzed the data and wrote the manuscript. T.M.G.-A. designed and performed experiments, analyzed the data and wrote the manuscript. Z.L. performed experiments and analyzed the data. L.H. performed experiments and analyzed the data. S.A.-G. performed experiments and analyzed the data. S.J. designed the experiments, supervised the project and wrote the manuscript. Y.G. conceived the idea, designed the experiments, supervised the project and wrote the manuscript.

Competing interests
The University of Wisconsin–Madison has filed a provisional patent application P180335US01, US serial number 62/682027 (7 June 2018) on the basis of this work. Y.G., S.J., K.B. and T.M.G.-A. are named as inventors on the provisional patent application.

Additional information
Supplementary information is available for this paper at https://doi.org/10.1038/s41592-019-0391-1.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher’s note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

© The Author(s), under exclusive licence to Springer Nature America, Inc. 2019
Aldrich. ProteaseMax (PM) was obtained from Promega. RapiGest (RG, also from equine heart and cytochrome c (TCEP), dithiothreitol (DTT), 2-mercaptoethanol (2-ME), -d-glucopyranoside (OG), SDS, digitonin saline (TBS) blocking buffer, Tween 20 and molecular weight cutoff (MWCO) from Bioss. Fetal bovine serum was purchased from Life Technologies. Mini-gels anion-selective channel antibody was purchased from Biovision. Mitochondrial Assay Dye Reagent Concentrate was purchased from Bio-Rad. Voltage-dependent organic layer was washed with H2O (6 ml) and dried over MgSO4. The dry organic layer was filtered and concentrated in vacuo to yield 74% of 4-(hydroxyethyl)-3-nitrobenzoic acid. Synthesis of intermediate product I. Using a traditional EDC coupling, 270 mg 4-(bromomethyl)-3-nitrobenzoic acid, octylamine, decylamine, dodecylamine, N-ethylidiospropionamide, piperidine, 1.4, butaneasolute, anhydrous N,N-dimethylformamide (DMF), NNN-betamethy-1-ol (1H-benzotriazole-1-yl) uronium hexafluophosphate (HBTU) and dinitrophenylhydrazine were obtained from TCI America. Fmoc-phatable linker was purchased from Advanced Chembect. Tetrahydrofuran (THF), ammonium hydroxide (NH4OH), hexafluorosilicophenol (HFIP), dichloromethane, heptane, acetone, trimethylamine, magnesium sulfate (MgSO4), sodium carbonate and silica were purchased from Sigma-Aldrich. Extraction solutions were made in nanopure deionized water (H2O) from Milli-Q water (Millipore). HEPES, ammonium bicarbonate (NH4HCO3), sucrose, sodium fluoride (NaF), phenylmethanesulfonyl fluoride (PMSF), EDTA, DDM, octyl-β-D-glucopyranoside (OG), SDS, digitonin (DGT), protease inhibitor cocktail, tri(2-carboxyethyl) phosphine hydrochloride (TCEP), dithiothreitol (DTT) 2-mercaptoethanol (2-ME), β-casein from bovine milk, Ubi from bovine erythrocytes, bovine serum albumin (BSA), myoglobin from equine heart and cytochrome c (Cyt c) from equine heart, ribonuclease A (RNase A) and guanidine hydrochloride (GuHCl) were purchased from Sigma-Aldrich. ProteaseMax (PM) was purchased from Promega. Rapigest (RG, also known as ALS) was purchased from Waters. Sodium orthovanadate, HPLC-grade H2O, acetonitrile (ACN), methanol, ethanol, optima LC–MS-grade formic acid (FA) and acetic acid were purchased from Sigma-Aldrich. Octylamine, octylamine, decylamine, dodecylamine, N,N,N',N''-tetramethyl-O-(1H-benzotriazol-1-yl) uronium hexafluophosphate (HBTU) and dinitrophenylhydrazine were obtained using a Waters (Microsorb) LCT mass spectrometer. O-nitrobenzyl (ONB) C12 (C8H11NO3SNa), [M – Na + H+]$^+$: calculated m/z, 547.6; observed m/z, 547.3. ONB C8 (C6H7NO3SNa), [M – Na + H+]$^+$: calculated m/z, 491.5; observed m/z, 491.3. ONB C6 (C4H3NO3SNa), [M – Na + H+]$^+$: calculated m/z, 463.5; observed m/z, 463.3. See also Supplementary Note 1.

Synthesis of O-nitroveratryl surfactant family. The O-nitroveratryl (ONV) surfactants were synthesized according to previously reported procedures. In brief, to a solution of Fmoc–ONV–COOH (0.57 mmol) and HBTU (0.69 mmol) in 3.5 ml of anhydrous DMF, N-ethylidiospropionamide (1.2 mmol) was added dropwise. The solution was cooled on ice and added to a solution of dodecylamine in 0.5 ml of ice-cold ethanol. After stirring for 30 min at 0°C, the mixture was stirred overnight at room temperature. The resulting precipitate was filtered and washed with DMF and then subjected to vacuum drying. Intermediate product I (n = 10, C12) was obtained as an amorphous white powder. Similar procedures were followed for n = 8, C10 and n = 6, C8. See also Supplementary Note 2.

Synthesis of NH2-ONV-CH$_2$(CH$_2$)$_n$(CH$_3$)$_i$ (intermediate product II). Piperidine was added dropwise to a solution of intermediate product I (0.6 mmol) in anhydrous DMF (3 ml) to reach a final concentration of 2 M. The solution was stirred at room temperature for 2 h, and then DMF was removed by evaporation. The residue was dissolved in methanol, and the resulting precipitate was removed by filtration. A pale yellow solid was obtained after evaporation of the filtered solution. See also Supplementary Note 2.

Synthesis of sodium–ONV-CH$_2$(CH$_2$)$_n$(CH$_3$)$_i$ (final product). 1.4 Butanesulfonyl (2.1 eq, 0.74 mmol) was added to a solution of intermediate product II (1.0 eq, 0.35 mmol) with trimethylamine (2.0 eq) in ACN (2 ml) and then the flask was sealed. The mixture was stirred and heated to approximately 90°C for 48 h. After the solvent was removed by evaporation, a light yellow viscous oil was obtained quantitatively. The oil was suspended in water, and an NH$_4$OH(aq) solution was added dropwise until approximately pH 8 was reached. The surfactant solutions were centrifuged. The final product was confirmed by ESI-MS. ONV C10 (C8H$_7$NO$_3$SNa), [M–Na]: calculated m/z, 600.3; observed m/z, 600.3. ONV C12 (C10H$_9$NO$_3$SNa), [M–Na]: calculated m/z, 572.3; observed m/z, 572.2. ONV C8 (C6H$_7$NO$_3$SNa), [M–Na]: calculated m/z, 544.3; observed m/z, 544.2. See also Supplementary Note 2.

Synthesis of the azobenzene surfactant family. The azobenzene surfactant family was synthesized according to similar procedures as previously described. Specifically, 4 mmol of 4-nitrohexane (n = 4, C6) was stirred in a mixture of 4.8 ml of 10% hydrochloric acid and 8 ml of deionized H$_2$O. Then 4 mmol of NaNO$_2$ dissolved in 4 ml of cold water was added dropwise to this solution. During the addition of the NaNO$_2$, the solution was cooled to 10°C. After the addition was completed (15 min), the solution was stirred for another 15 min at 5°C. A similar procedure was carried out for 4-n-octylamine (n = 6, C8). For 4-n-decylamine (n = 8, C10) and 4-n-dodecylamine (n = 10, C10), the solution of 4-n-alkylamine was heated to 70°C and then cooled in an ice bath to 10°C under vigorous stirring. NaNO$_2$ was added dropwise starting at 20°C and concluded at 25°C, after which the solution was stirred for 15 min at 5°C. For the coupling reaction, the freshly prepared diazonium salt was filtered into a stirred and cooled solution (temperature = 5–10°C) of 8 mmol of Na$_2$SO$_4$ and 12 mmol of Na$_2$CO$_3$, in 20 ml of deionized H$_2$O. To complete the precipitation of the surfactant, the solution was refrigerated at 4°C overnight. The yellow compounds were purified by precipitation with a yield of about 50%, and no impurities were detected by NMR. We made surfactant solutions by gently heating the surfactant at 37°C and then bringing it to room temperature after no further solid remained. The working concentration was 0.5%–1% in 25 mM NH$_4$HCO$_3$ Kraft temperature (a clear 1% surfactant solution) was previously reported at 24.5°C (ref. ). A high-resolution mass spectrum of azobenzene (C6), referred to as AzO (Supplementary Fig. 1), was taken as follows: a solution of 1% AzO in 25 mM NH$_4$HCO$_3$ was diluted from 1:100 to 500 m/z. ESI–MS for AzO (C$_8$H$_7$NO$_3$SNa), [M–Na]: calculated m/z, 269.096; observed m/z, 269.098. A Hemes-Varian Mercury Plus 300 operating at 300 MHz was utilized for 1H-NMR spectroscopy with chemical shifts reported as ppm (parts per million). 1H NMR: 6.74 (2H, ddd, –Ar–H), 7.37 (2H, d, –Ar–H), 2.67–2.48 7.53 (2H, d, –Ar–H).
Tissue handling. Swine hearts were excised from healthy Yorkshire domestic pigs, snap-frozen in liquid N$_2$, and stored under −80°C before use. All homogenization and centrifugation steps were performed at 4°C.

**Protein extraction and LC–MS analysis of cardiac tissue.** The frozen tissue samples (approximately 500 mg) were cut into small pieces and washed with PBS buffer containing protease inhibitors and reducing agent (5 mM DTT, 1 mM PMSE, 1× protease inhibitor cocktail). The tissue was then homogenized in HEPES buffer with both protease and phosphatase inhibitors (25 mM HEPES, 250 mM sucrose, 50 mM NaF, 1 mM PMSE, 2.5 mM EDTA, 1 mM Na$_3$VO$_4$, 1 mM PMSE, 5 mM DTT, 1× protease inhibitor cocktail) and a Polytron electric homogenizer (model PT-3000 or PT-1800) set to the lowest speed (tissue) to deplete soluble proteins as described previously25. The homogenate was centrifuged at 21,750 g using Beckman Ultracentrifuge and a Ti-80 rotor for 1 h. The supernatant was separated by centrifugation and labeled as ‘E1’. The cell pellets were centrifuged and labeled as ‘E2’. After the second HEPES extraction, the tissue pellets were suspended in 25 mM NH$_4$HCO$_3$, and centrifuged to separate smaller aliquots. In one aliquot, 25 mM NH$_4$HCO$_3$ buffer with no surfactant serving as controls was used in a 1:1 ratio (homogenate:buffer) and labeled as ‘E3/NS’. The remaining aliquots were individually added to the other aliquots in a 1:1 ratio (homogenate:surfactant) and labeled as ‘E3(NS)’ following incubation with both protease and phosphatase inhibitors (25 mM HEPES, 250 mM sucrose, 500 mM NaF, 1 mM PMSE, 2 mM EDTA, 1 mM Na$_3$VO$_4$, 5 mM DTT, 25 μg/mL DGT, 1× protease inhibitor cocktail) with a Polytron electric homogenizer set to the lowest speed (tissue). The supernatant samples were centrifuged at 21,750 g using Beckman Ultracentrifuge and a Ti-80 rotor for 1 h. The supernatant was set from 600 to 2,000 μg/mL. The methods described here correspond to data presented in Fig. 1g.

**Protein extraction and LC–MS analysis of sarcosplasmic reticulum and mitochondria enrichment from cardiac tissue.** Around 170 mg of tissue had been cut into small pieces, the tissue was homogenized in HEPES buffer containing both protease and phosphatase inhibitors (50 mM HEPES, 0.6 M KCl, 250 mM sucrose, 500 mM NaF, 1 mM PMSE, 2 mM EDTA, 1 mM Na$_3$VO$_4$, 5 mM DTT, 25 μg/mL DGT, 1× protease inhibitor cocktail) with a Polytron electric homogenizer set to the lowest speed (tissue) to deplete soluble proteins as described previously25. The homogenate was centrifuged at 21,750 g using Beckman Ultracentrifuge and a Ti-80 rotor for 1 h. The supernatant was removed and labeled as ‘E1’. The pellet was suspended in the buffer (25 mM NH$_4$HCO$_3$, 500 mM NaF, 1 mM PMSE, 2 mM EDTA, 1 mM Na$_3$VO$_4$, 5 mM DTT, 25 μg/mL DGT, 1× protease inhibitor cocktail) to remove residual proteins and labeled as ‘E2’. The resulting tissue pellet was suspended in 25 mM NH$_4$HCO$_3$, and centrifuged into smaller aliquots at 20,000 g, and the supernatant was removed. NH$_4$HCO$_3$ buffer (25 mM) or 1% Azo in 25 mM NH$_4$HCO$_3$ was added to the aliquots. After homogenization and incubation, the samples were centrifuged and the supernatant was collected.

Enriched sarcosplasmic reticulum (50 μl) and mitochondria lysate from cardiac tissues were diluted with 440 μl of 0.5% IPA and 10 μl of TCEP (1 M). The sample was irradiated for 3 min and concentrated to a final volume of 150 μl MWCO (10 kDa in run 1 or 30 kDa in run 2). Proteins were separated using the following gradient: 0–1 min 5% B, 1–5 min 5–30% B, 5–55 min 30–60%, 55–57 min 60–95% B, 57–65 min 95%, 65–67 min 95% B, 67–80 min 5% B. Column temperature was 35°C. For ATP synthase subunit α, a single charge state was isolated and fragmented with 5, 10, 16, 18 and 20 eV, respectively, using an isolation window of 3 m/z during targeted CID MS/MS experiments. The methods described here correspond to data presented in Supplementary Figs. 13–15 and 17–19, and Supplementary Tables 2–8.

**Protein extraction and LC–MS analysis of endoplasmic reticulum and mitochondria enriched lysate from HEK293T cells.** Cells were grown on 10 cm plates in DMEM with 10% fetal bovine serum and 1× penicillin/streptomycin solution at 37°C and 5% CO$_2$. Cells from two 10 cm plates were washed twice with PBS and lysed in 500 μl of buffer (10 mM Tris, 2 mM DTT, 1 mM PMSE, 50 μg/mL DGT, 1× protease inhibitor cocktail) using 50 strokes with dounce homogenizer followed by five passages through a 27-gauge needle. Cells were then incubated for 10 min on ice, evenly divided into two aliquots and centrifuged at 1,000 g (4°C) to remove unbroken cells and the nuclei. The supernatant was mixed with 0.5 μl of 200 μg/mL of sucrose (50%) and centrifuged at 21,000 g (4°C). The pellet was washed with 1 ml of NH$_4$HCO$_3$ (E2). Finally, the pellets were dissolved in 100 μl of Azo (0.5% in 25 mM NH$_4$HCO$_3$) or 100 μl of 25 mM NH$_4$HCO$_3$, without surfactant serving as controls. Enriched endoplasmic reticulum (50 μl) and mitochondria lysate from HEK293T cells was diluted with 400 μl of 50% IPA:4% H$_2$O:1% formic acid and 50 μl of TCEP (1 M). The sample was irradiated for 3 min, then concentrated and exchanged into 10:10:80 ACN:IPA:1% formic acid in H$_2$O with a 10 kDa MWCO centrifugal filter. Proteins were separated using the following gradient: 0–5 min, 20%; 5–65 min 20–40% B; 65–75 min 70–75% B; 75–80 min, 75–95% B; 85–86 min, 95–20% B; 86–95 min, 20% B. The methods described here correspond to data presented in Supplementary Fig. 2, Supplementary Figs. 13–15 and 17–19, and Supplementary Tables 2–8.

**SDS–PAGE comparing Azo with SDS, DDM and MasDeS.** An equal volume (7 μl) of each extraction was subsequently resolved using 12.5% SDS–PAGE with a voltage of 50 V for 30 min and 120 V for approximately 75 min. Proteins were visualized using Coomassie Brilliant Blue R-250. The methods described here correspond to data presented to Fig. 1e.

**Western blot comparing Azo with SDS, DDM and MasDeS.** Equal volumes of tissue lysate (10 μl) were loaded and resolved on 12.5% SDS–PAGE gels. Proteins were transferred to a polyvinylidene difluoride membrane, fast semi-dry blotter (Fishier Scientific), using 20 V for 12 h at 4°C. The membrane was placed in a protein-free blocking buffer (Fishier Scientific) for 1 h at room temperature and incubated with primary antibodies for 1.5 h at room temperature. The membranes were then washed with TBS with 0.1% Tween, then incubated with the secondary antibodies for 1.5 h (room temperature). After five washes with TBS with 0.1% Tween, the membranes were developed using enhanced chemiluminescence detection (Fishier Scientific). The methods described here correspond to data presented in Supplementary Fig. 3.

**Comparison of the top-down MS compatibility of Azo with SDS, DDM and MasDeS.** Ubi was dissolved in a buffer containing 80:5:5:10 IPA:H$_2$O:formic acid:1% surfactant (Azo, SDS, DDM or MasDeS) with 10 mM DTT. The Azo sample was irradiated for 3 min. The MasDeS sample was degraded for 24 h at room temperature. The samples were then directly injected into a 7 T linear ion trap/Fourier transform ion cyclotron resonance mass spectrometer (LTQ/FT Ultra, Thermo Scientific) with a nano-ESI sprayer (TriVersa NanoMate; Advion Bioscience). A voltage of 1.4 kV versus the inlet was applied with 0.3 psi drying gas. Full MS scans were collected with the microscans in over scan mode. The maximum ion trap fragmentation was set from 600 to 2,000 m/z. The methods described here correspond to data presented in Fig. 1g.

**UV-vis degradation.** Azo (50 μl, 0.1%) in (1) H$_2$O, (2) 1% formic acid, (3) IPA, (4) 1% formic acid in IPA, (5) 2 M ethanol in H$_2$O and (6) 1% formic acid in IPA:H$_2$O were irradiated with a 100 W high-pressure mercury lamp (Nikon house) with Nikon HB-10101 AF power supply; Nikon for 0, 10, 30, 60, 90 and 120 s, respectively, in a quartz cuvette. The samples were diluted to a final volume of 1 ml in H$_2$O. A UV-vis spectrometer was taken from each sample with a Varian Cary 50 UV-Visible spectrophotometer (background correction, medium scan rate, 600–200 nm). The methods described here correspond to data presented in Supplementary Fig. 4.
Evaluation of the effect of reducing agents during Azo degradation.

Standard proteins, Ubi, RNase A, CytC, and BSA were dissolved in 49.5:49.5:1 H$_2$O:IPA:formic acid and kept on ice until analysis. Samples were irradiated with a 100 W lamp for 3 min. Sample (5 µl) was injected onto a trap column and eluted with 40:40:20 ACN:IPA:1% formic acid in H$_2$O after a 5 min wash with 2.5:2.5:95 ACN:IPA:1% formic acid in H$_2$O. DTT, TCEP, and 2-ME (all 50 mM) were added to each sample before irradiation. Additionally, a sample of CytC was kept at room temperature with no reducing agent and irradiated for 3 min with no reducing agent as a control (corresponding to Supplementary Figs. 6 and 7). This method was repeated to test 10 mM TCEP and 35 mM methionine (corresponding to Supplementary Fig. 8).

Protein extraction and LC–MS analysis for evaluating the effect of Azo on relative quantitation. Ten volumes of buffer (10 mM Tris, 500 mM NaF, 2 mM EDTA, 1 mM PMSE, 1 mM Na$_2$VO$_4$, 5 mM DTT) was added to swine heart tissue. The sample was homogenized with Teflon homogenizer, centrifuged at 16,000g, and the supernatant was collected. Protein extract was diluted to 100 µl with or without 0.2% Azo. The sample was irradiated for 3 min and exchanged into a 10% ACN, 10% IPA, with 0.2% formic acid using a 10kDa MWCO centrifugal filter. Proteins were separated using the following gradient: 0–5 min 20% B, 5–30 min 20–65% B, 30–35 min 65% B, 35–36 min 20% B, 36–40 min 20% B. Column temperature was 60°C. The methods described here correspond to the data presented in Supplementary Fig. 9.

Comparison of the top-down MS compatibility of Azo to a broader range of commonly used surfactants. Ubi was dissolved in buffer containing 75:10:5:1 methanol:H$_2$O:formic acid:1% surfactant (MaSDeS, PM, RG, NS, SDS, Azo, OG, DDM, DGT) with 10 mM TCEP. The Azo sample was irradiated for 3 min. The acid-labile surfactants were incubated for 75 min (24 h for MaSDeS) at 37°C. The samples were then directly injected into a 12 T Fourier transform ion cyclotron resonance (solariX) mass spectrometer (Bruker Daltonics) with a nano-ESI sprayer (TriVersa NanoMate; Advion Bioscience). A voltage of 1.4 kV versus the inlet was applied with 0.3 psi drying gas. Two hundred scans were averaged for each sample. The mass range was set from 600 to 2,000 m/z with a 512,000 word transient. The file contained the following parameters: quality factor (0.4); signal-to-noise ratio (S/N) (3); intensity threshold (500); retention window (1.5 min). The file contained the following reference information: precursor mass, precursor charge, precursor mass followed by the fragment masses, intensities and charges. TopPIC$^2$ was utilized for intact protein identification on the basis of protein spectrum matches searching against the UniProt Sus scrofa (released on 22 November 2017; containing 26,817 protein sequences) or Homo sapiens (released on 20 December 2017; containing 20,244 reviewed protein sequences) databases$. Fragment mass tolerance was set to 15 ppm. All identifications were validated with statistically significant $P$ and $E$ values (<10$^{-7}$) and satisfactory numbers of assigned fragment ions (>6). Sequence mass determination and validation was performed using Mash Suit Pro$^{29}$ or ProSight Lite$. All masses reported are monoisotopic. The corresponding MS and MS/MS data were summarized in Supplementary Tables 4–10. UniProt gene ontology$^{32}$ was used to determine the subcellular locations of the identified proteins, which were then graphed in Microsoft Excel. String analysis software$^{33}$ was used to create an interactome map of identified proteins belonging to the electron transport chain.

The proteoform maps were generated as follows: (1) LC–MS scans were averaged every minute; (2) deconvoluted using maximum entropy algorithm (resolution: 80,000; mass range: 5,000–60,000 Da); (3) mass list outputs were generated using SNAP peak picking (quality factor: 0.8, S/N: 3, absolute intensity 1,000) as described previously$. A graphic map was then generated in Microsoft Excel based on the first retention time and the monoisotopic mass generated from SNAP. The methods described here correspond to the data presented in Supplementary Fig. 13.

Statistical analysis. For the protein solubility experiment (Fig. 1) comparing Azo, SDS, DDM, and MaSDeS, three independent protein assays (n = 3) were performed to evaluate surfactant performance. Error bars represent s.e.m. For the broader protein solubility comparison (Supplementary Fig. 10), data presented were based on three independent experiments (n = 3). Error bars represent s.e.m. For LC–MS analysis (Supplementary Fig. 9), three separate samples (n = 3) were prepared for each condition. Error bars represent standard error of the mean.

Reporting Summary. Further information on research design is available in the Nature Research Reporting Summary linked to this article.

Data availability
All data generated or analyzed during this study are presented in this article or in the provided supplementary materials. Raw gel, blot and mass spectra data are available as Supplementary Data, and source data for Fig. 1 and Supplementary Figs. 4, 9, 10 and 13 are available online. Proteomics data have been uploaded to the PRIDE repository via ProteomeXchange with identifier PXD010825.

References
21. Savelev, S. V. et al. Anal. Chem. 85, 907–914 (2013).
22. Lee, H. B. et al. J. Org. Chem. 69, 701–713 (2004).
23. Savelev, S., Simpson, D. & Wood, K. V. Cleavable surfactants. US patent 095628 A1 (2009).
24. Mezger, T., Nuyken, O., Meindl, K. & Wokaun, A. Prog. Org. Coat. 29, 147–157 (1996).
25. Wentzell, M. & Katz, S. J. Mol. Cell. Cardiol. 23, 1149–1163 (1991).
26. Peng, Y. et al. Mol. Cell. Proteomics 13, 2752–2764 (2014).
27. Kott, J., Xu, L. & Liu, X. Bioinformatics 32, 3495–3497 (2016).
28. Apweiler, R. et al. Nucleic Acids Res. 32, D115–D119 (2004).
29. Cai, W. et al. Mol. Cell. Proteomics 15, 703–714 (2016).
30. Fellers, R. T. et al. Proteomics 15, 1235–1238 (2015).
31. Ashburner, M. et al. Nature Genet. 25, 25–29 (2000).
32. Szklarczyk, D. et al. Nucleic Acids Res. 45, D362–D368 (2017).
33. Cai, W. et al. Anal. Chem. 89, 5467–5475 (2017).
Reporting Summary

Nature Research wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Research policies, see Authors & Referees and the Editorial Policy Checklist.

Statistical parameters

When statistical analyses are reported, confirm that the following items are present in the relevant location (e.g. figure legend, table legend, main text, or Methods section).

- The exact sample size (n) for each experimental group/condition, given as a discrete number and unit of measurement
- An indication of whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
- The statistical test(s) used AND whether they are one- or two-sided
  - Only common tests should be described solely by name; describe more complex techniques in the Methods section.
- A description of all covariates tested
- A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
- A full description of the statistics including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
- For null hypothesis testing, the test statistic (e.g. F, t, r) with confidence intervals, effect sizes, degrees of freedom and P value noted
  - Give P values as exact values whenever suitable.
- For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
- For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
- Estimates of effect sizes (e.g. Cohen’s d, Pearson’s r), indicating how they were calculated
- Clearly defined error bars
  - State explicitly what error bars represent (e.g. SD, SE, CI)

Our web collection on statistics for biologists may be useful.

Software and code

Policy information about availability of computer code

Data collection

Mass spectrometry data was collected using otofControl Version 4.3 (Build 60.11), ftmsControl 2.1.0, or LTQ FT Ultra 2.5. Protein assay data was generated using Gen5 reader control 2.05. UV-Vis collected on Varian Cary 50 UV-Visible spectrophotometer.

Data analysis

Mass spectrometry data was analyzed using DataAnalysis 4.3 (Build 110.102.1532) or Xcalibur. TopPIC v1.1.0 was used to determine protein and proteoform identifications. Mash Suite Pro v1.0.00.26946 and ProSight lite v1.4 were used for validation. Interactome analysis was performed using String Software V10.5. Plots were generated in Microsoft Excel 2016 and figures were made in Microsoft Powerpoint 2016.

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors reviewers upon request. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Research guidelines for submitting code & software for further information.

Corresponding author(s): Ying Ge
Data

Policy information about availability of data
All manuscripts must include a data availability statement. This statement should provide the following information, where applicable:
- Accession codes, unique identifiers, or web links for publicly available datasets
- A list of figures that have associated raw data
- A description of any restrictions on data availability

All data generated or analyzed during this study is presented in this manuscript or supplementary information with corresponding source data. Moreover, proteomics data has been uploaded to PRIDE repository via ProteomeXchange with identifier PXD010825.

Field-specific reporting

Please select the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

- Life sciences
- Behavioural & social sciences
- Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see nature.com/authors/policies/ReportingSummary-flat.pdf

Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

| Sample size | The overall goal of the study is to demonstrate Azo’s good protein solubility and MS-compatibility, which is demonstrated in multiple n=3 experiments. For the protein solubility experiment (Figure 1f) comparing Azo, SDS, DDM, and MaSDeS, three independent protein assays (n=3) were performed to evaluate surfactant performance. Error bars represent standard error of the mean. For the broader protein solubility comparison (Supplementary Fig. 10), data presented were based on three independent experiments (n=3). Error bars represent standard error of the mean. For LC-MS analysis (Supplementary Fig 9), three separate samples (n=3) were prepared for each condition. Error bars represent standard error of the mean. |
| Data exclusions | For the proteoform comparison (Supplementary Figure 13), the following parameters were set in DataAnalysis: quality factor: 0.8, S/N: 3, absolute intensity: 1,000. Data that did not meet these requirements were excluded. This mitigated artifacts from the deconvolution algorithm or adducts (i.e. the addition of sodium from sample vials or lines) which would inflate the number of detected proteoforms. Moreover a quality factor of 0.8 helped ensure that only quality proteoform spectra were included in the count. |
| Replication | All attempts at replication were successful. |
| Randomization | Randomization was not relevant to this study as no biological comparisons were made. Qualitative assessment of surfactant extraction and mass spectrometry performance are not time or sample order dependent. All comparisons were performed side-by-side using the same extraction conditions. |
| Blinding | Blinding was not relevant to this study as no biological comparisons or correlations were made. Additionally, different surfactants required different workups (e.g. acid or UV); therefore a blind test was not applicable. |

Reporting for specific materials, systems and methods

Materials & experimental systems

- n/a Involved in the study
- ☒ Unique biological materials
- ☒ Antibodies
- ☒ Eukaryotic cell lines
- ☒ Palaeontology
- ☒ Animals and other organisms
- ☒ Human research participants

Methods

- n/a Involved in the study
- ☒ ChIP-seq
- ☒ Flow cytometry
- ☒ MRI-based neuroimaging

Antibodies

Primary antibodies used: VDAC antibody- Biovision; Cat. 3594-100; Lot 40394 (1:1000 dilution). TOM20 Antibody (FL-145) is a
rabbit polyclonal IgG- Santa Cruz Biotechnology; Cat. sc-11415 (1:500 dilution), Anti-alpha 1 Sodium Potassium ATPase antibody- Abcam; Cat. ab76171; Lot.GR454517-3 (1:1000 dilution), pan Cadherin- Abcam; Cat. ab6528; Lot.GR103554-1 (1:1000 dilution), Phospholamban antibody- Bioss, Cat. bs-4197r (1:500 dilution). Secondary Antibodies: horseradish peroxidase-conjugated secondary antibodies were used Goat Anti- Antigen: Rabbit IgG (H+L) Horseradish Peroxidase (Thermo, 32460) was used for Tom20, VDAC1, and PLN while Goat anti-mouse IgG, (H+L), Peroxidase Conjugated (Thermo, 31430) was used for Na-K ATPase and cadherin. (1:10,000 dilution for all 2nd antibodies).

Validation
No validation was performed

Eukaryotic cell lines

Policy information about cell lines

Cell line source(s) HEK 293T ATCC® CRL-3216™
Authentication No authentication was performed
Mycoplasma contamination Cell lines were not tested for mycoplasma contamination.
Commonly misidentified lines (See ICLAC register) No commonly misidentified lines were used.