

Metal-free and pH-controlled introduction of azides in proteins

S. Schoffelen, M. B. van Eldijk, B. Rooijackers, R. Raijmakers, A. J. R. Heck and J. C. M. van Hest

Electronic Supporting Information

Materials

Ampicillin and chloroamphenicol were purchased from MP Biomedicals. Imidazole (>99%), egg-white lysozyme (>90%), *para*-nitrophenyl butyrate and lyophilized *Micrococcus lysodeikticus* cells were purchased from Sigma Aldrich. Sulfuryl chloride, sodium azide, IPTG and diethanolamine were obtained from Acros. $\text{CuSO}_4 \cdot 5\text{H}_2\text{O}$ and Na_2CO_3 were purchased from Merck and Baker, respectively. Ultrapure Milli-Q water (MQ) was purified using a WaterPro PS polisher (Labconco, Kansas City, MO) set to 18.2 M Ω /cm.

Protein mass spectrometry

Protein samples were dialyzed against MQ (3 times 0.5 mL) using 10 kDa MWCO centrifugal filter units (Millipore) and analysed by electrospray ionization time-of-flight (ESI-TOF) on a JEOL AccuTOF. 5-10 μM protein in formic acid (0.1-1 %) was injected. Deconvoluted spectra were obtained using MagTran 1.03b2. Isotopically averaged molecular weights were calculated using the "Protein Calculator v3.3" at <http://www.scripps.edu/~cdputnam/protcalc.html>.

Production of CalB

The pET22b-calb plasmid has been described previously (*1*). The recombinant CalB that is produced has the following amino acid sequence:

MGLPSGSDPAFSQPKSVLDAGLTCQGASPSVSKPILLVPGTGTTGPQSFDSNWIPLSA
QLGYTPCWISPPPFMLNDTQVNTEYMVNAITTLYAGSGNKNLPVLTWSQGGLVAQW
GLTFFPSIRSKVDRMLMAFAPDYKGTVLGAPLDALAVSAPSVWQQTTGSALTTALRNA
GGLTQIVPTTNLYSATDEIVQPQVSN SPLDSSYLFNGKNVQAQAVCGPLFVIDHAGSL
TSQFSYVVGRSALRSTTGQARSADYGITDCNPLPANDLTPEQKVAAAALLAPAAAAI
VAGPKQNCEPDLMPYARPFVAVGKRTCSGIVTPLEHHHHHH

50 mL 2xYT medium supplemented with ampicillin (100 mg/L) and chloroamphenicol (50 mg/L) was inoculated with a single colony of *E. coli* strain B834(DE3)pLysS (Novagen) containing pET22b-calb, and incubated overnight at 37 °C. The culture was diluted in 1 L 2xYT medium to an OD₆₀₀ of 0.1 and grown at 37 °C. When an OD₆₀₀ of 0.6-0.7 was reached, isopropyl-β-D-thiogalactopyranoside (IPTG) was added to a final concentration of 1 mM. Protein production was allowed to take place for 20 hours at 25 °C. Purification was performed via Ni²⁺ NTA affinity chromatography followed by size-exclusion chromatography as described before (1). Protein concentration was determined by measuring the absorbance at 280 nm on a NanoDrop ND-1000 spectrometer (Thermo Scientific). The yield was 3.5 mg of purified protein from a 1 L bacteria culture. Purity was confirmed by SDS-PAGE. ESI-TOF: CalB calculated 34269.7, found 34269.0.

Production of ELP

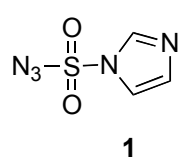
The construction of plasmids pEt15b-ELP-lys1[V₅L₂G₃-90] and pEt15b-ELP-lys3[V₅L₂G₃-90] has been described before (2). Amino acid sequences of the produced ELPs are as follows:

ELP-lys1: MGSSHHHHHHSSGLVPRGSHMLEKREAEAGP
(VPGGGVPGVGVPGVGVPGGGVPGLGVPGVGVPGVGVPGVGVPGGGVPGGLG)₉
VPGGGA

ELP-lys3: MGSSHHHHHHSSGLVPRGSHMLEKREAEAKKGP
(VPGGGVPGVGVPGVGVPGGGVPGLGVPGVGVPGVGVPGVGVPGGGVPGGLG)₉
VPGGGA

Both ELPs were expressed in the *E. coli* strain BLR(DE3)pLysS (Novagen) and purified via inverse transition cycling (2), yielding 5 mg of purified protein per litre culture. Protein concentrations were determined by BCA assay (Pierce). Purity was confirmed by SDS-PAGE. ESI-TOF: ELP-lys1 calculated 39670.9, found 39671.7; ELP-lys3 calculated 39927.2, found 39929.9.

Synthesis of imidazole-1-sulfonyl azide (1)



The hydrochloric salt of imidazole-1-sulfonyl azide hydrochloride was synthesized as described by Goddard-Borger *et al.* (3). Sulfuryl chloride (4.0 mL, 50 mmol) was added dropwise to an ice-cooled suspension of NaN₃ (3.3 g, 50 mmol) in MeCN (50 mL) and the mixture was stirred overnight, allowing it to warm up to room temperature. Then the mixture was recooled on ice and imidazole (6.5 g, 95 mmol) was added portion-wise. After the addition, the reaction mixture

was stirred for 3h while the temperature was allowed to warm up to room temperature. Subsequently, the mixture was diluted with EtOAc (100 mL), washed twice with MilliQ water (2x 100 mL) and twice with saturated NaHCO₃ (2x 100 mL) and then dried over MgSO₄. Then the HCl salt was formed via dropwise addition of 4M HCl in EtOAc (40 mL) to the ice-cooled mixture. The resulting precipitate was filtered and washed with EtOAc (3x50 mL) to obtain imidazole-1-sulfonyl azide hydrochloride in a yield of 59%. LRMS: (*m/z*) [M-Cl]⁺ calculated for C₃H₃N₅O₂S 174.0, found: 173.9. ¹H-NMR (400 MHz, D₂O): δ 7.57 (s, 1H) δ7.99 (s, 1H) δ9.22 (s, 1H).

It should be noted that diazotransfer reagent **1** is shelf stable, but should not be kept as a concentrated aqueous solution for prolonged time due to risk of decomposition.

Tryptic digest mass analysis

1 μg trypsin (Promega, mass spectrometry grade) was added to approximately 40 μg CalB in 30 μL 10 mM NH₄HCO₃. Upon incubation overnight at 37 °C, the tryptic digestion was quenched by adding 0.25 % formic acid and the digest was analyzed by ESI-TOF. The sequence of the N-terminal fragment with residue numbers 1-15 is MGLPSGSDPAFSQPK. In control experiments in which multiple azides were introduced, no cleavage was observed at diazotized lysine 15. When tryptic digestion resulted in the formation of the N terminal 15-mer and an azide was introduced in this fragment, the α-amine at the N-terminus of this fragment had to be converted.

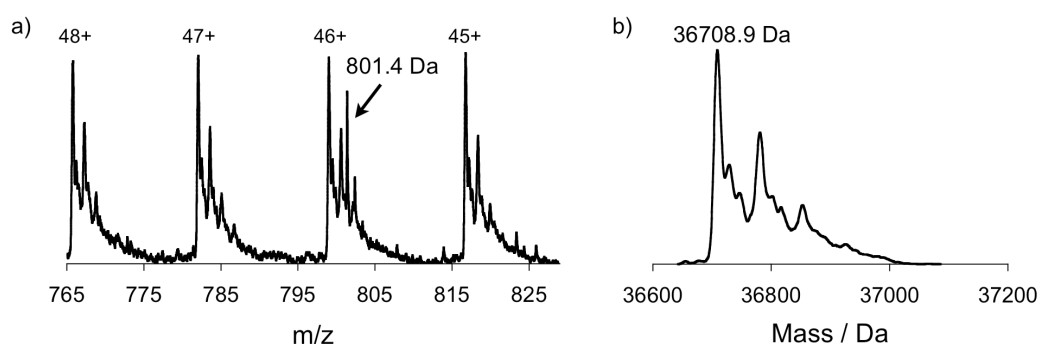


Fig. S1 Mass spectrum of mono-functionalized ELP-lys3 after tryptic digestion. a) Non-deconvoluted spectrum in which the peak at 801.4 Da is visible. b) Spectrum derived from deconvolution on the multiple-charged peaks shown in a).

Tryptic digest analysis of the ELPs, was performed by adding 0.5 μg trypsin to 7 μg ELP in 12 μL 20 mM NH₄HCO₃. After 5 hours of incubation at 37 °C the samples were purified using ZipTip-C18 (Millipore) and eluted in 20 μL 75 % acetonitrile in 0.1 % trifluoroacetic acid / MQ. The lysine with residue number 23 is present in both ELP-lys1 and ELP-lys3. In

the tryptic digest of mono-functionalized ELP-lys3, fragment 17-23 (calculated $[M+H]^+$ 801.4 Da) was detected indicating that no azide had been introduced on this lysine. The two additional lysines in ELP-lys3 (residue numbers 29 and 30) were not diazotized either as proven by the detection of tryptic fragment 30-488 (calculated $[M+H]^+$ 36708.8 Da). Figure S1 shows part of the mass spectrum in which a peak at 801.4 Da is present together with several multiple-charged peaks, which upon deconvolution turned out to correspond to a mass of 36708.9 Da.

Protein digestion for LC-MS/MS

Prior to mass spectrometry analysis, 5 μ g of CalB and of Lysozyme was reduced with 1,4-dithiothreitol (6mM), alkylated with iodoacetamide (12mM) and digested at with trypsin, chymotrypsin or Lys-N (4) (each 1:50 protein/enzyme ratio) in ammonium bicarbonate 50mM at 37 °C for 16 hours. Samples were acidified to a concentration of 10% formic acid and analyzed by LC-MS/MS.

Nano LC-MS/MS

All LC-MS/MS analyses were performed on a nanoLC LTQ-Orbitrap (Thermo, San Jose, CA) mass spectrometer at a resolution of 60,000 at 400 m/z. For nanoLC, an Agilent 1200 series LC system was equipped with a 25 mm Aqua C18 (Phenomenex, Torrance, CA) trapping column (packed in-house, i.d., 100 μ m; resin, 5 μ m) and 400 mm ReproSil-Pur C18-AQ (Dr. Maisch GmbH, Ammerbuch, Germany) analytical column (packed in-house, i.d., 50 μ m; resin, 3 μ m). Solvents used were 0.6% HAc (buffer A) and 0.6% HAc/80% acetonitrile (ACN) (buffer B). Trapping was performed at 5 μ L/min for 10 min, and elution was achieved with a gradient of 13–32% B in 30 minutes, 32-40% B in 5 minutes, 40-100% B in 2 minutes and 100%B for 2 minutes. The flow rate was passively split from 0.6 mL/min to 100 nL/min. Nanospray was achieved using in-house made distally coated fused silica emitters (o.d., 375 μ m; i.d., 20 μ m) biased to 1.8 kV. The mass spectrometer was operated in the data dependent mode to automatically switch between MS and MS/MS. Survey full scan MS spectra were acquired from m/z 350 to m/z 1500 and the five most intense ions were fragmented in the linear ion trap using collisionally induced dissociation. The target ion setting was 5e5 for the Orbitrap, with a maximum fill time of 250 ms. Fragment ion spectra were acquired in the LTQ with a target ion setting of 1e4 and a maximum fill time of 100 ms. Dynamic exclusion for selected precursor ions was set at 30 seconds.

LC-MS/MS Data Processing and Analysis

Raw MS data were converted to peak lists using Proteome Discoverer (v1.2). For protein identification, MS/MS data were searched against a custom database containing the CalB and

Lysosyme sequences, as well as known contaminant and background proteins, using Mascot v2.2 (Matrix Science) with the appropriate enzyme, allowing up to 2 miss cleavages, a precursor tolerance of 15 ppm, and a product ion tolerance of 0.5 Da. Carbamidomethylation of cysteine was set as a fixed modification and oxidation of methionine, deamidation of asparagines and glutamine as well as azide moieties on lysines and protein n-termini were included as variable modifications. Peptide identifications were accepted with a minimum Mascot score of 25 and an expect value of <0.05, corresponding to an estimated false discovery rate (FDR) of around 1%, as determined by decoy database searching. All identification results are available for download in a Scaffold 3 file that may be downloaded from ProteomeCommons.org Tranche, <https://proteomecommons.org/tranche/>, using the following hash:

```
8x0MIGHIq9zIOOC3T9PTbkerc227XqNgyYDQFVHkVnaVRA/2+a2rpeHdzPrTC+  
6vRIqUfEMNRpfeTjehz/8qpTJwdKkAAAAAAAAAADEA==
```

The viewer is downloadable from <http://www.proteomesoftware.com>.

LC-MS/MS results

By LC-MS/MS analysis of CalB and lysozyme, both digested with trypsin, Lys-N (5) and chymotrypsin, 96% and 100% protein coverage was achieved for these proteins, respectively. For both proteins, all lysine residues were identified as part of one or more unique peptides. In the case of CalB (diazotransfer at pH 8.5, in absence of Cu(II)), extensive modification of peptides stemming from the N-terminus of the protein was observed, whereas only one of the nine lysine residues (Lys210) was found to be modified at a very low level (2 spectra corresponding to modified peptides versus 36 spectra matching peptides containing Lys210 without azide group). Lysozyme was analyzed by LC-MS/MS after diazotransfer in the absence as well as in the presence of Cu(II). In the absence of Cu(II), only limited modification of Lys13 and Lys116 was observed, whereas in the presence of Cu(II), there was significant conversion of Lys13, Lys96 and/or Lys97 (indistinguishable by LC-MS/MS), Lys116 as well as of the N-terminus of the protein.

Synthesis of lysyl-phenylalanyl-phenylalanine (KFF)

The tripeptide lysyl-phenylalanyl-phenylalanine (KFF) was synthesized using standard Fmoc solid-phase peptide synthesis. Cleavage from the resin and removal of the protecting group on the lysine side-chain was achieved by incubation in trifluoroacetic acid. A white powder was obtained upon lyophilization. Purity was confirmed by HPLC. LRMS: (m/z) $[M+H]^+$ calcd for $C_{24}H_{32}N_4O_4$ 441.2, found: 441.6. ^{13}C NMR (300 MHz, D_2O) δ 171.3, 168.5, 136.4, 135.6, 135.7, 128.8, 128.6, 128.3, 128.1, 126.8, 126.5, 54.5, 52.2, 38.5, 36.6, 29.9, 25.8, 20.6.

Diazotransfer to KFF

25 mg of KFF (57 μmol) was dissolved in 5 mL MQ or 30 mM diethanol amine (pH 8.5). 28.5 mg of diazotransfer reagent **1** (136 μmol) was dissolved in 136 μL 1 M NaOH. This solution was then added to the tripeptide. The reaction in MQ was supplemented with 1.6 mg Na_2CO_3 and 1.25 mg $\text{CuSO}_4 \cdot 5\text{H}_2\text{O}$. The reactions were stirred overnight at room temperature and lyophilized afterwards. The reactions were purified upon solvation in 2 mL MQ by high-performance liquid chromatography (HPLC) over a C18 column using a linear gradient from 5 % acetonitrile in MQ/0.1 % trifluoroacetic acid to 0.1% trifluoroacetic acid in acetonitrile. Peak fractions were analyzed by mass spectrometry. LRMS (m/z) $[\text{M}+\text{H}]^+$ calcd for $\text{C}_{24}\text{H}_{32}\text{N}_4\text{O}_4$ 441.2, found: 441.5; (m/z) $[\text{M}+\text{H}]^+$ calcd for $\text{C}_{24}\text{H}_{30}\text{N}_6\text{O}_4$ 467.2, found: 467.5; (m/z) $[\text{M}+\text{H}]^+$ calcd for $\text{C}_{24}\text{H}_{28}\text{N}_8\text{O}_4$ 493.2, found: 493.4. Fractions of interest were pooled. Acetonitrile was evaporated under vacuum and water was removed by lyophilization. $\text{C}_{24}\text{H}_{30}\text{N}_6\text{O}_4$ (product in the reaction performed at pH 8.5): ^{13}C NMR (300 MHz, D_2O) δ 173.8, 171.7, 136.0, 135.7, 128.8, 128.7, 128.2, 128.2, 126.6, 126.6, 62.2, 54.0, 53.5, 38.6, 36.5, 36.3, 30.1, 25.8, 20.8; $\text{C}_{24}\text{H}_{28}\text{N}_8\text{O}_4$ (product in the reaction performed at pH 11): ^{13}C NMR (300 MHz, $\text{CD}_3\text{CN}/\text{D}_2\text{O}$ 3:1) δ 172.6, 170.8, 169.7, 136.4, 128.9, 128.8, 128.0, 127.9, 126.4, 126.3, 62.2, 53.5, 53.2, 50.3, 36.6, 36.3, 30.3, 27.4, 21.5.

Click reaction and SDS-page analysis

20 equivalents of aza-dibenzocyclooctyne functionalized PEG2000 (1 mM in MQ) (**5**) were added to 6 μg of diazotized protein. Samples were incubated overnight at room temperature while gently shaken. The next day sample buffer was added, samples were incubated at 95 $^\circ\text{C}$ for 5 minutes and loaded on a 12% (w/v) polyacrylamide gel followed by Coomassie Brilliant Blue staining. The Precision Plus Dual Color Protein standard (Biorad) was used as marker.

Activity of CalB and lysozyme

The activities of both CalB and lysozyme were determined after treatment with **1**. Upon dialysis against MQ, the CalB solution was diluted to 1 μM by adding an aqueous solution containing 50 mM NaH_2PO_4 , 150 mM NaCl (pH 7.0). Hydrolysis of para-nitrophenyl butyrate was followed by measuring absorbance at 405 nm as reported elsewhere (*1*). The initial slope was taken as a measure for activity. The small decrease in activity that was observed might be caused by the fact that the sample was incubated overnight at room temperature and dialysed by centrifugation, whereas the control was kept at 4 $^\circ\text{C}$ all the time (see Fig. S2).

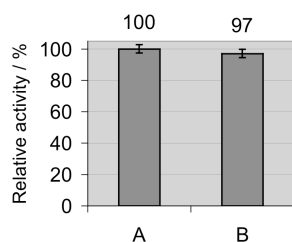


Fig. S2 Activity of untreated CalB (A), and mono-functionalized CalB reacted with 17.5 equivalents of **1** at pH 8.5 without addition of Cu(II) (B). Relative activities are expressed as an average for n=3 with the error bars indicating the standard deviation.

Lysozyme activity was determined by following lysis of lyophilized *Micrococcus lysodeikticus* cells. A bacterial suspension was prepared containing 0.074 mg/mL cells in 66 mM phosphate buffer (pH 7.0), resulting in an A_{405} between 0.5-0.75. 5 μ g lysozyme was added to 3 mL bacterial suspension and the decrease in absorption at 450 nm was followed for 3 minutes at 25 °C using a 10 mm cuvette. The slope was taken as a measure of activity. Lysozyme was significantly less active upon incubation at pH 11, especially in presence of Cu(II) (see Fig. S3). At pH 8.5 the presence of Cu(II) was less harmful, but some decrease in activity was still detected due to incubation with the metal catalyst. These findings underline the importance of being able to introduce azide moieties in proteins under milder conditions. Treatment with **1** led to an additional reduction in activity. The enzyme was almost entirely inactivated at pH 11, which suggested that under these conditions lysines were converted that were crucial for enzyme activity. It shows that selective conversion of lysines is essential to retain activity.

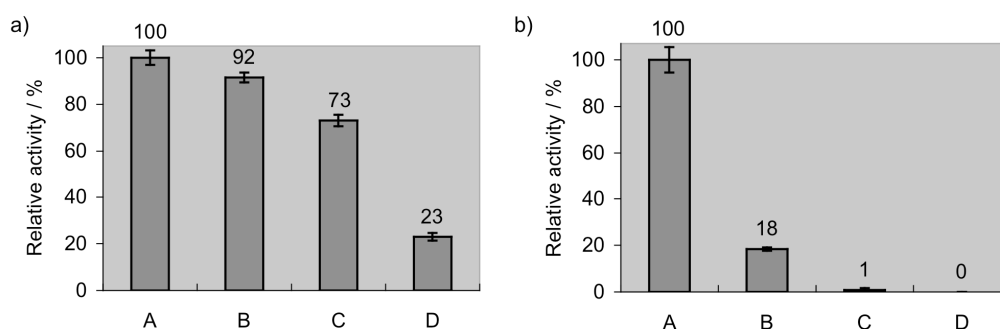


Figure S3 Relative activities of lysozyme after a) incubation at pH 8.5 and 11, with or without Cu(II), and b) in presence of **1** as well. (A) pH 8.5, no Cu(II), (B) pH 8.5, with Cu(II), (C) pH 11, no Cu(II), (D) pH 11, with Cu(II). Relative activities are expressed as an average for n=3 with the error bars indicating the standard deviation.

References

- (1) Schoffelen, S.; Lambermon, M. H. L.; van Eldijk, M. B.; van Hest, J. C. M. *Bioconjugate Chem.* **2008**, *19*, 1127.
- (2) Teeuwen, R. L. M.; de Wolf, F. A.; Zuilhof, H.; van Hest, J. C. M. *Soft Matter* **2009**, *5*, 4305.
- (3) Goddard-Borger, E. D.; Stick R. V. *Org. Lett.* **2007**, *9*, 3797.
- (4) Taouatas, N.; Heck, A.J.R.; Mohammed, S. *J Proteome Res.* **2010** *9*, 4282.
- (5) Debets, M. F.; van Berkel, S. S.; Schoffelen, S.; Rutjes, F. P. J. T.; van Hest, J. C. M.; van Delft, F. L. *Chem. Commun.* **2010**, *46*, 97.