Selective N-terminal acylation of peptides and proteins with a Gly-His tag sequence

Manuel C. Martos-Maldonado, Christian T. Hjuler, Kasper K. Sørensen, Mikkel B. Thygesen, Jakob E. Rasmussen, Klaus Villadsen, Søren R. Midtgaard, Stefan Kol, Sanne Schoffelen & Knud J. Jensen

Methods for site-selective chemistry on proteins are in high demand for the synthesis of chemically modified biopharmaceuticals, as well as for applications in chemical biology, biosensors and more. Inadvertent N-terminal gluconoylation has been reported during expression of proteins with an N-terminal His tag. Here we report the development of this side-reaction into a general method for highly selective N-terminal acylation of proteins to introduce functional groups. We identify an optimized N-terminal sequence, GHHH\textsuperscript{n}— for the reaction with gluconolactone and 4-methoxyphenyl esters as acylating agents, facilitating the introduction of functionalities in a highly selective and efficient manner. Azides, biotin or a fluorophore are introduced at the N-termini of four unrelated proteins by effective and selective acylation with the 4-methoxyphenyl esters. This Gly-His\textsubscript{n} tag adds the unique capability for highly selective N-terminal chemical acylation of expressed proteins. We anticipate that it can find wide application in chemical biology and for biopharmaceuticals.
Methods for site-selective modification of peptides and proteins are required for different fields, such as the development of biopharmaceutical conjugates (e.g., PEGylation, lipiddation, and antibody-drug conjugates\(^2,3\)), bioimaging\(^4\), medical diagnostics\(^5\), and material sciences\(^6\). Most proteins display multiple copies of the same side-chain at different locations. Reactions that target a particular functional group, e.g., the primary amine in Lys or thiol in Cys, potentially modify all occurrences of this residue, leading to the formation of a heterogeneous mixture of modified protein. Reactions involving proteins must also proceed in aqueous solutions under mild pH and temperature. These factors make it challenging to modify proteins in a regioselective manner. Several strategies have been developed to address this challenge. Some methods rely on genetic code expansion\(^7\)-\(^9\), which allows for the site-specific, ribosomal incorporation of non-canonical amino acids that exhibit suitable functionalities for bioorthogonal chemistry\(^8\). However, selective chemical methods applicable to proteins consisting of only canonical amino acids are an attractive alternative, as these proteins can be produced using standard, higher-yielding expression methods.

N-termini of proteins tend to reside on the surface of proteins\(^10\) and are thus often well accessible to chemical modification\(^11\). Also, they can often be extended with additional amino acids without interference with protein function\(^11\). The N-terminal α-amine can be targeted for selective, pH-controlled acylation or alkylation due to its pK\(_a\) value that is lower than that of Lys side-chain ε-amines (pK\(_a\) 10.5 ± 1.1)\(^12\). Nevertheless, the selectivity is often challenged by the presence of a high number of competing ε-amines on Lys residues and the fact that their pK\(_a\) can be lower due to the local environment. Also, direct acylation with NHS esters and reductive alkylation require a low-to-neutral pH. Alternatively, particular amino acids can be specifically targeted when located at the N-terminus, such as Cys via native chemical ligation\(^13\) or 2-cyanobenzoazolozole (CBT) condensation chemistry\(^14\), Trp via Pictet–Spengler reactions\(^15,16\), Ser and Thr via periodate oxidation to yield aldehydes for subsequent conjugation with α-nucleophiles\(^17\)-\(^20\), or Pro via oxidative coupling with aminophenols\(^11,21\). A small number of more general approaches for site-selective N-terminal modification without specific residue requirements have been described. A site-selective diazotransfer reaction for azide introduction has been achieved with imidazole-1-sulfonyl azide at pH 8–8.5\(^22\), a phenyl ketene derivative has been used for alkylene introduction\(^23\), and a transamination reaction with 2-pyrindinocarboxylaldehydes has been reported\(^24\), as well as reductive alkylation\(^25\). In addition to these methods for N-terminal labeling, the use of a four amino acid sequence (Phe-Cys-Pro-Phe) that enhances the reactivity of its cysteine residue for side-selective reaction with perfluoroaryl compounds has been reported\(^26\). Other small peptide sequences with high affinity for molecules of interest, e.g., fluorophores, have been reported. For instance, a tetracysteine tag (Cys-Cys-Xxx-Xxx-Cys-Cys) for selective reaction with an arsenic-modified fluorescein derivative\(^26,27\) and a tetraserine motif (Ser-Ser-Pro-Gly-Ser-Ser) that binds a rhodamine-derived bisboronic acid\(^28\).

Polyhistidine tags (His tags) are widely used for protein purification by immobilized-metal ion affinity chromatography\(^29\). However, it has been reported that during their expression His-tagged proteins can undergo N-terminal acylation with β-gluconic acid δ-lactone (GDL, 1) as an inadvertent side-reaction\(^30\). Geoghegan et al. observed glucosylation of proteins when expressed in *Escherichia coli*, while using His tags with the N-terminal sequence GSHHHHHSGLVPFR. They also reported that synthetic peptides GSSHSSHSSHSSGLVPFR, GSAHHHAAR, GASHHHHAAAR, and GAAGHHHAAR could be modified with GDL in HEPES buffer at pH 7.5\(^30\).

Here, we convert an undesirable side-reaction into a highly selective chemical method for modification of peptides and proteins (Fig. 1). We identify an optimal N-terminal sequence (GHHH\(_{n}\)) using GDL as inexpensive, water-soluble acylating agent. Studying the reactivity with other acylating agents, 4-methoxyphenyl esters gave good selectivity for acylation of this N-terminal tag. 4-Methoxyphenyl esters facilitate the introduction of small reactive groups, such as azides, or the direct conjugation of functional molecules such as biotin. We demonstrate the potential on several peptides and proteins. The GHHH\(_{n}\) tag, which can be fused to the N-terminus of any protein of interest, offers a dual functionality His tag as it can still be used for affinity purification. We believe that this methodology will be a valuable contribution to the toolbox for N-terminal modification. We propose the name His tag acylation to describe this method.

Results

D-Gluconic acid δ-lactone for peptide modifications. We hypothesized that highly site-selective glucosylation of the N-amine could be achieved with an optimized N-terminal sequence. We synthesized 11 peptides with different N-terminal tags attached to the sequence LRFKFY-NH\(_2\) (Table 1). All peptides included one Lys residue to test the selectivity of the reaction for the N-terminal α-amine. The peptides were treated at a concentration of 1 mM in 200 mM HEPES buffer at pH 7.5 with 25 equiv. of GDL (1) at room temperature. The base peptide sequence LRFKFY-NH\(_2\) (2) by itself, which has an N-terminal Leu and a Lys with an ε-amine side-chain, was practically resistant to N-terminal acylation (Table 1). Addition of a Gly residue with a non-hindered α-amine (3, 4) increased the degree of N-acylation somewhat. Addition of N-terminal His residues to give peptides 5 and 6 significantly improved N-acylation. Also, the presence of three His was beneficial (5 vs 6). N-terminal addition of Gly to His improved the N-acylation (5 vs 7) further, while a Ser\(^30\) in between Gly and His was detrimental (7 vs 8). Increasing the number of His from Gly–His to Gly–His–His (4, 9 vs 10) provided a clear improvement in propensity to be acylated by GDL. Additional control experiments confirmed the importance of a sterically non-hindered Gly α-amine N-terminally of the His–His–His, as corresponding peptides with N-terminal Ala and Val proved increasingly resistant to N-acylation (6, 11, 12 vs 10).

Formation of di-gluconoylated products of these peptides was not observed in any case, which pointed towards a site-selective modification of the N-terminal α-amine. To further substantiate this, the selectivity of the reaction was analyzed in the case of peptide 9. NMR spectroscopy confirmed that the reaction took place selectively at the N-terminal α-amine and not on the ε-amine of the Lys residue (Supplementary Fig. 1).

Encouraged by these observations, we evaluated the optimized GHHH sequence in the 18-mer peptide 13 (DWLKAHYDKE]=[VEKLEAF, Beltide-1), which has four Lys residues. Beltide-1 is an amphipathic helical peptide that self-assembles to give lipid filled nanodiscs\(^31\). The obtained results with model peptides 2–12 suggested that the optimal N-terminal sequence for site-selective glucosylation was GHHH\(_n\) (where \(n\) ≥ 1). Commonly used His tags for protein purification by immobilized-metal ion affinity chromatography can be placed N- or C-terminally and typically contain six His residues\(^29\). We hypothesized that adding extra His residues to GHHH could improve the ability to be acylated. We compared GHHH– with GHHHHH–tags on a Beltide-1 peptide sequence (Table 2). First, control experiments with non-tagged Beltide-1 (13) and its acetylated analog (13Ac) did not show any peptide acylation. Next, GHHH–Beltide-1 (14) was N-acylated in
73%, while GHHHHHH-Beltide-1 (15) showed higher reactivity and reached full conversion to the corresponding mono-gluconoylated derivative after 1 h, when treated at a concentration of 1 mM with 100 equiv. of GDL in 200 mM HEPES buffer at pH 7.5 at room temperature (Supplementary Fig. 2). Remarkably, formation of di-gluconoylated peptides was not detected. Importantly, in a control experiment we observed that N-terminally acetylated peptide Ac-GHHHHHH-Beltide-1 (15) was not acylated by GDL, which again indicated that the reaction occurred selectively at the N-terminal α-amine, when available.

**Phenyl esters as acylating agents.** While other acylating agents, including lactones, thioesters, and a N-hydroxysuccinimide ester were also tested (Supplementary Fig. 3), they only gave limited N-terminal acylation or low selectivity. We then focused on tuning the properties of representative phenyl esters 16–18 (Fig. 2). The 4-nitrophenyl and unsubstituted phenyl esters proved too reactive and were not selective in the acylation, whereas 4-methoxyphenyl ester 18 provided good selectivity while retaining reactivity. The half-life of 18 in 200 mM HEPES buffer, pH 7.5, at 4 °C was 3.8 h (Supplementary Fig. 4).

When GHHHHHHH-Beltide-1 (15) at 1 mM was reacted with 2.5 equiv. of 18 at 4 °C for 24 h, full conversion was achieved (92% mono-acylated (15Az) and 8% di-acylated peptide) (Table 3, Supplementary Fig. 5 and 6). Shorter reaction time or lower equiv. led to exclusive formation of mono-acylated peptide, although with lower conversion. The same conditions were tested on other Beltide-1 derivatives (Table 3). GHHHHHHH-Beltide-1 (15) and GHHHHHHHG-Beltide-1 (15G) showed higher reactivity and selectivity than GHHH-Beltide-1 (14). The beltide sequence has four Lys (position 4, 9, 13, and 15). Trypsin digestion of the diacylated product (15Az-Az) showed that the second acylation had occurred chiefly at Lys-4, located most closely to the His acylation tag (Supplementary Fig. 7 and 8). The small difference between 15 and 15G could point to a minimal effect of the additional spacing of Lys-4 and the His acylation tag provided by the Gly, in this particular case.

**Fig. 1** Concept for N-terminal His tag acylation of Gly-Hisₙ sequences. **a** Optimization of the His tag sequence for acylation with GDL; **b** identification of 4-methoxyphenyl esters as functional acylating agents; and **c** application of the His tag acylation to proteins.
In 15K and 15K-Ac an additional Lys was placed C-terminal of the GHHHHHH (GH6) and the conditions for acylation were then fine-tuned. The optimized N-terminal sequence (amino acids 1–14) (Fig. 3b). Only the unmodified version of tryptic fragment #29–276 was detected by ESI-TOF MS, indicating that none of the 20 Lys residues located in this fragment were modified. Furthermore, digestion with chymotrypsin showed that fragment 20–29 (m/z = 1158.623) was detected but not its glucosynlated version, which revealed that Lys21, the only Lys residue in the first 28 amino acids of the protein, had remained unmodified. Taken together, these observations indicated that the modification with GDL occurred exclusively at the N-terminus of EGFP.

Interestingly, for both 15 and GH6-EGFP the glucosynlation proved to be reversible (Supplementary Fig. 9 and 10). We hypothesize that the reversibility can be due to the polyol structure in the glucosynlated protein, however, this would require additional studies. In contrast, the acylation with 4-methoxyphenyl esters of acetic acid derivatives gave stable products (Supplementary Fig. 11 and 12).

For comparison, additional experiments were carried out with GSSH6-EGFP, which has a conventional His tag, as well as non-tagged GH6-EGFP. In both cases, some formation of mono-glucosynlated product was observed, although with significantly lower conversion in comparison to GH6-EGFP (Fig. 3c). These results highlighted the importance of our His acylation tag for the reaction to occur. GH6-EGFP and GSSH6-EGFP are identical except for the two inserted Ser residues; thus, these data confirmed the importance of an uninterrupted N-terminal Gly–His sequence.

Next, the 4-methoxyphenyl ester 18 was used to modify GH6-EGFP. Gratifyingly, 18 was effective in labeling the protein at 4 °C. When a 35 µM solution of GH6-EGFP in 200 mM HEPES buffer at pH 7.5 was treated with 40 equiv. of 18 for 4 days, followed by the addition of two aliquots of 10 equiv. of 18 in the next two days, 88% conversion was observed by ESI-MS (71% conversion to mono-functionalized products) (Fig. 4b). Here it should be noted that the 4 days incubation time can be shortened to one day. In follow-up experiments and in agreement with the observed half-life of 18 at 4 °C (see above), it was observed that no additional acylation occurred after one day, unless a fresh portion of phenyl ester 18 was added. In a second scenario for optimization, fewer equivalents and shorter reaction times resulted in a somewhat lower overall conversion, but also in a higher selectivity towards the mono-functionalized species. For example, when the protein was treated with 20 equiv. of 18 for 1 day, 51% was converted and 45% was mono-functionalized (Supplementary Fig. 13).

Azido-functionalized N2-Ch2-C(O)-GH6-EGFP was reacted with DBCO-OEG4-biotin (19, Fig. 4) to demonstrate the potential of the strategy for subsequent strain-promoted alkyn–azide conjugation. ESI-MS analysis displayed predominant formation of the mono-labeled protein (Fig. 4c). The sample was digested with trypsin and subsequently analyzed by MALDI-TOF MS. The labeled N-terminal fragment (amino acids #1–14) was observed (Supplementary Fig. 14). Furthermore, as observed for glucosynlated GH6-EGFP (Fig. 3b), a peak corresponding to unmodified fragment #29–276 was present, while the labeled version was not detected. This indicated that no modification took place at this region of the protein. Azido-functionalized N2-Ch2-C(O)-GH6-EGFP was also conjugated to a 5 kDa DBCO-PEG. SDS-PAGE analysis revealed formation of mono-PEGylated protein with high conversion, while di-PEGylated species were only observed as very faint bands (Fig. 4d).

For comparison, the potential reaction between compound 18 and GSH-EGFP and GSSH6-EGFP were studied. The former displayed negligible formation of the mono-acetylated species when treated at 35 µM with 40 equiv. of 18 in 200 mM HEPES buffer at pH 7.5 with 10% acetonitrile for 2 days. Under the same

Table 1 Acylation of a series of peptides with GDL

| Peptide | Peptide sequence | Acylation (%)a |
|---------|-----------------|---------------|
| 2       | LRFKFY-NH2      | 3.0 (± 0.7)   |
| 3       | GLRFKFY-NH2     | 17.3 (± 0.7)  |
| 4       | GLRFKFY-NH2     | 15.7 (± 1.6)  |
| 5       | HLRFKFY-NH2     | 29.8 (± 0.7)  |
| 6       | HHLRFKFY-NH2    | 64.4 (± 13.9) |
| 7       | GHHLRFKFY-NH2   | 45.2 (± 1.7)  |
| 8       | GSHHLRFKFY-NH2  | 33.4 (± 1.9)  |
| 9       | GHHHLRFKFY-NH2  | 63.7 (± 8.6)  |
| 10      | GHHHHHLRFKFY-NH2| 81.9 (± 3.5)  |
| 11      | AHHHHLRFKFY-NH2 | 27.4 (± 4.3)  |
| 12      | VHHHHHLRFKFY-NH2| 4.3 (± 7.5)   |

*aConversion of peptides 2–12 (1mM) after treatment with 25 equiv. of GDL (1) in 200 mM HEPES buffer at pH 7.5 at room temperature for 2 h, as determined by LC–MS (UV, 215 nm). Standard deviations are based on triplicate measurements. The His acylation tags and compound numbers are indicated in bold.

Table 2 Acylation of a series of Beltide-1 derivatives with GDL

| Beltide-1 (13) | Acylation (%)a |
|---------------|---------------|
| No reaction   |               |
| No reaction   |               |
| ~ 73          |               |
| 100           |               |
| No reaction   |               |

*aConversion of Beltide-1 derivatives (1mM) after treatment with 100 equiv. of GDL (1) in 200 mM HEPES buffer at pH 7.5 at room temperature for 1 h, as determined by LC–MS (UV, 215 nm). The His acylation tags and compound numbers are indicated in bold.
To exemplify the optimization of the selective N-terminal acylation, we reacted GH₆-tagged SUMO with 4-methoxyphenyl ester 18. Under all conditions tested mono-functionalized products were the major species (35 μM protein, 36 equiv. 18, 4 °C). Conversion to mono-labeled protein was determined by ESI-MS as 59% for GH₆-MBP (Fig. 4e).

To further evaluate the generality of the selective acylation of the N-terminal segment, we reacted GH₆-tagged MBP with 4-methoxyphenyl ester 18. Remarkably, under the conditions tested mono-functionalized products were the major species (35 μM protein, 36 equiv. 18, 4 °C). Conversion to mono-labeled protein was determined by ESI-MS as 59% for GH₆-MBP (Fig. 4e).

Table 3 Acylation of a series of Beltide-1 derivatives with 4-methoxy phenyl ester 18a

| Peptide | N-Terminal mono-acylated product (%) | Di-acylated product (%) |
|---------|--------------------------------------|-------------------------|
| Beltide-1 (13) | No reaction | No reaction |
| Ac-Beltide-1 (13-Ac) | – | – |
| GHHH-Beltide-1 (14) | 63 | 13 |
| GHHHHH-Beltide-1 (15) | 92 | 8 (±1) |
| Ac-GHHHHHH-Beltide-1 (15Ac) | – | 9 (±0.5) |
| GHHHHHHG-Beltide-1 (15G) | 97 | 3 (±0.5) |
| GHHHHHHK-Beltide-1 (15K) | 87 | 13 (±1) |
| Ac-GHHHHHHK-Beltide-1 (15K-Ac) | – | 31 (±4) |

*Conversion of a series of Beltide-1 derivatives after being treated at a concentration of 1 mM with 2.5 equiv. of 4-methoxy-phenyl ester 18 at 4 °C for 24 h as determined by LC-MS (UV, 215 nm). Standard deviations are based on duplicate measurements. Beltide-1 sequence DWKAYDKVABKEAF, Lys underscored. The His acylation tags and compound numbers are indicated in bold.

In order to demonstrate the application of the method on a biologically more relevant protein, we selected X-linked inhibitor of apoptosis proteins (IAPs) constitute a class of proteins that play an essential role in the anti-apoptotic and pro-survival signaling pathways. They are characterized by having one or more baculovirus IAP repeats called BIR domains. These domains of ~70 amino acids interact with different kinds of proteins. IAPs are upregulated in various cancers and associated with tumor growth and resistance to treatment. Therefore, IAPs are attractive targets for antitumor drug discovery. The site-selective, N-terminal introduction of a fluorophore or biotin moiety in XIAP will facilitate studies aiming at the identification and characterization of interaction partners or inhibitors of this class of proteins. Here, the XIAP fragment [124–240] was selected which comprises the linker region between BIR1 and BIR2 and the BIR2 domain, and which is known to potently inhibit caspase-3 and -7.

A GH₆-tagged version of the protein (for simplicity referred to as GH₆-BIR2) was expressed, purified by Ni²⁺ affinity chromatography, and treated with compound 18. It was found that treatment with 20 equiv. of compound 18 was optimal, leading to 83% conversion and 65% mono-functionalization (Supplementary Fig. 17). In order to test whether acylated GH₆-BIR2 had retained its function, a caspase-7 inhibition assay was performed using the tetrapeptide DEVD linked to 7-amino-4-trifluoromethylcoumarin (AFC) as substrate. Acylated GH₆-BIR2 was as potent in inhibiting caspase-7 as its non-modified counterpart (Fig. 5b). Next, the possibility to fluorescently label azido-functionalized N₃-CH₂-C(O)-GH₆-BIR2 through the Cu(I)-catalyzed alkyne–azide cycloaddition was demonstrated, as depicted in Fig. 5c. Furthermore, in order to evaluate whether 4-methoxyphenyl esters could be used to directly attach larger moieties, biotin derivative 20 (Fig. 5d) was reacted with GH₆-BIR2. Treatment with 20 equiv. of compound 20 at room temperature for 24 h led to a ~50% conversion to mono-labeled protein as determined by ESI-MS. Addition of an extra aliquot of 20 equiv. for 1 more day gave 80% conversion with 60% mono-functionalization. The ability of biotinylated GH₆-BIR2 to bind streptavidin was demonstrated by western blot (Fig. 5d).

As an alternative approach, we tested whether GH₆-EGFP could be biotinylated with a 4-methoxyphenyl ester derivative of biotin formed in situ by coupling biotin reagent 19 to compound 18 via the strain-promoted alkyne–azide cycloaddition. Approximately 45% mono-labeled protein was obtained when GH₆-EGFP was incubated with 60 equiv. of the in situ formed ester for 48 h at 4 °C (Supplementary Fig. 18).

GH₆-SUMO was found to react with compound 20 to a similar degree and with comparable selectivity as GH₆-BIR2. Moreover, biotinylated SUMO could still be processed by SUMO protease as confirmed by ESI-MS (Supplementary Fig. 19). This result suggests that also other enzymes involved in the SUMOylation process will still recognize the modified SUMO.
Mechanistic studies. The Gly-His$_n$ sequence catalyzes the acylation of the Gly N-$\alpha$-amine. Fundamental studies of acyl transfer catalysis by imidazoles have shown that the catalytic mechanism is modulated by variations in the leaving group of the acylating agent$^{34-40}$. Strong electrophiles, such as acetic anhydride and 4-nitrophenyl acetates, favor a nucleophilic mechanism of imidazole catalysis involving intermediate formation of neutral N-acyl imidazoles. In contrast, weak electrophiles, such as alkyl esters, react by a specific base catalytic mechanism. Interestingly, 4-methoxyphenyl acetate is known to be a borderline case between nucleophilic and specific base catalysis$^{40}$. Both of the above mentioned types of mechanisms may be envisioned in...
the present case of His tag acylation. Initial interaction with the imidazole ring or base catalysis seems to be a requirement, since no reaction was observed in control experiments with non-tagged Beltide-1 (13) or its acetylated derivative 13Ac. In the case of a hypothesized nucleophilic catalysis mechanism, the reaction would start by formation of an N-acyl imidazole intermediate by reaction of the ester with the imidazole side-chains of the His residues. Then, the acyl imidazole could react with amines or undergo hydrolysis. When the N-terminal Gly α-amine is available, the reaction would occur preferentially on this amine, which can be explained by its closer proximity and lower pK_{α} in comparison with amines of Lys side-chains. Alternatively, for a specific base catalytic mechanism, the imidazoles of the His side-chains would assist in deprotonation of the charged, tetrahedral addition complex between the acylating agent and the N-terminal Gly α-amine. Once again, the close proximity of the His side-chain would further promote catalysis. In order to elucidate the mechanism, we conducted a set of control experiments where the reactivity of 4-methoxyphenyl ester 18 and acetic anhydride (as positive control) with different imidazoles was studied at pH 7.5. The hypothetical N-acyl derivatives of imidazole or Ac-GHHHHHHH-NH_{2} were not observed in any experiments involving 18, neither by UV (245 nm)\textsuperscript{90} or \textsuperscript{1}H-NMR spectroscopy. Interestingly, nor was the N-acyl imidazole derivative detectable in the presence of the more reactive 4-nitrophenyl or phenyl esters, 16 or 17, under the applied reaction conditions. In contrast, full conversion to N-acetylimidazole was evident when acetic anhydride was used as the acylating agent (Supplementary Fig. 20 and 21).

Furthermore, we compared hydrolysis rates of 18 in the presence of imidazole or 2-isopropylimidazole in order to reveal the involvement of any transiently formed N-acyl imidazole. The steric hindrance of the 2-isopropyl group is known to affect nucleophilic catalytic processes negatively, whereas specific base catalytic processes of 2-isopropylimidazole are virtually unaffected, as compared to imidazole\textsuperscript{34,35}. The observed hydrolysis rate of 18 with 6 mM 2-isopropylimidazole was essentially identical to that of 6 mM imidazole or 1 mM Ac-GHHHHHHH-NH_{2} in HEPES buffer, pH 7.5 (Supplementary Fig. 22).

**Discussion**

We present a chemical method for highly selective and efficient N-terminal acylation of proteins. This method, for which we propose the name His tag acylation, is based on a short N-terminal peptide sequence, GHHHHH, and is complemented by the use of 4-methoxy phenyl esters as finely tuned acylating agents. The method proceeds in aqueous medium, at mild temperature and neutral pH. In comparison with the oxidation of an N-terminal Ser or Thr residue or sortase-mediated N-terminal modification, the reaction avoids the use of harsh oxidants or expensive enzymes. Moreover, the method, which was successfully applied to four unrelated proteins, is anticipated to be highly versatile. Structural variation among proteins is known to affect the yield of chemical N-terminal modification strategies, such as the transamination reaction using pyridoxal-5'-phosphate or 2-pyridinecarboxaldehyde and the diazotransfer reaction. In contrast to these strategies, the hydrophilic nature of our GHHHHHH sequence will contribute to an increased accessibility of the N-terminus and we anticipate that it can provide a high yield over a wide range of proteins. Furthermore, oxidative methods require two steps for introducing an azide. The Gly-His\textsubscript{α} sequence will remain in the protein after acylation, however, the sequence can be rather short. Other methods for N-terminal modification, such as sortase mediated reactions, also rely on short peptide sequences, which remain in the final protein conjugate.

The optimized sequence constitutes a unique version of one of the most widely used affinity tags for protein purification, the His tag. In commonly used bacterial expression vectors, N-terminal His tags are typically preceded by 3–4 amino acids. Here, we show that by direct attachment of the HHHHHH segment to an N-terminal Gly residue instead, a tag with dual functionality is generated, facilitating both purification and efficient, site-specific modification of recombinant proteins. The non-hindered N-terminal α-amine of Gly guarantees high conversion, while the proximity of the His segment directs the selectivity of the reaction. As demonstrated, proteins containing the GH6 tag can easily be obtained by bacterial expression. One can rely on the esicin of the N-terminal methionine, which is known to occur during expression when the second amino acid is small. Alternatively, in case an N-terminal signal peptide is required, for example, for secretion to the periplasm, this signal peptide can be removed using the commonly employed TEV protease, thereby revealing the needed N-terminal Gly.

Our results indicate that the mechanism behind the very high selectivity of the His tag acylation was specific base catalysis, in which a His side-chain assists deprotonation during the direct acylation of the Gly α-amine. Facilitation of this type of proton transfer is a key driving force for reaction rate enhancements in aqueous, neutral medium\textsuperscript{41}, including for enzymatic reactions. The ester reacts preferentially with assistance from His side-chain imidazoles since they are not protonated (pK_{α} ∼ 6.0) at the pH of the reaction, in contrast with the N-terminal α-amine (pK_{α} ∼ 7.6–8.0) and Lys side-chains (pK_{α} ∼ 10.5). The presence of the additional five His residues in the His tag may serve to modulate the basicity of the imidazole nitrogen of the catalytic residue. A recent study has shown that the pK_{α} values of individual His side-chains in a His\textsubscript{8} tag span a range from 4.8 to 7.5\textsuperscript{42}. This implies that imidazoles with different acid-base properties are to some extent organized within the His tag. A His side-chain in an α-helix can in some cases catalyze the acylation of a flanking Lys Ne-amino in position i–3 and i + 4\textsuperscript{43}. However, this requires an engineered α-helix, whereas the present method provides a general method for N-terminal Ne-amine acylation.
4-Methoxyphenyl esters, such as azido reagent 18 and biotin reagent 20, were found to provide the best performance in terms of reactivity vs selectivity. These reagents are easy to prepare and allow for the one-step introduction of an N-terminal functionality. As demonstrated, azido acetyl moieties can be smoothly introduced and used for subsequent azide–alkyne couplings. Functional moieties, such as a biotin, can be incorporated in a single step as well. Importantly, the reactivity of the acylating agent and the catalytic properties of the His tag are optimized for a selective reaction at the N-terminus. The selectivity for N-terminal acylation of proteins with our His acylation tag should be seen in view of the number of competing Lys residues.

For example, a 9:1 ratio for acylation of the ε-amine in EGFP over any of the ε-amines would correspond to an effective 189:1 selectivity for the ε-amine over each of the ε-amines.

The focus of our study was to develop a general method for N-terminal chemical modification of proteins for use in chemical biology and in development of biopharmaceuticals. The chemical modification of the His tag with acetic acid derivatives is stable, in contrast to some other methods. The fact that ester could introduce with high selectivity an azido moiety at 4 °C is a particularly attractive feature. We anticipate that the method also has the potential to be used for selective labeling of proteins in a cell lysate or even in cells for bioimaging.

**Methods**

**Materials.** Synthesis and purification: N,N′-dicyclohexylcarbodiimide (DCC), dichloromethane (DCM), 4-(dimethylamino)pyridine (DMAP), heptane, ethyl acetate (EtOAc), N,N-dimethylformamide (DMF), benzyl bromoacetate, magnesium sulfate (MgSO₄), palladium hydroxide on carbon (Pd(OH)₂/C), Celite, triethylsilane (TES), diethyl ether (Et₂O), acetonitrile (MeCN), Sodium carbonate (Na₂CO₃), and gluconolactone (GDL) were purchased from Sigma-Aldrich (Brondby, Denmark). N-[(1H-azabenzotriazol-1-yl)(dimethylamino)methylene]-N-methylmethanaminium hexafluorophosphate N-oxide (HATU), N-[(1H-benzotriazol-1-yl)[(dimethylamino)methylene]-N-methylmethanaminium hexafluorophosphate N-oxide (HBTU), and 1-Hydroxy-7-azabenzotriazol (HOAt) were purchased from GL Biochem Ltd. (Shanghai, China). N,N′...
Synthesis of phenyl 2-azidoacetates 16-18. The phenol (200 mg) was dissolved in CH₂Cl₂ (5 mL), and 4-dimethylaminopyridine (0.1 equiv.) was added. 2-Azidoacetic acid (1.2 equiv.) was dissolved in CH₂Cl₂ (5 mL) with N,N'-disopropylcarbodiimide (1.2 equiv.) and stirred for 15 min, before it was added carefully to the phenol solution. The reaction mixture was stirred under N₂. After 2 h, additional 2-azidoacetic acid (1.2 equiv.) and N,N'-disopropylcarbodiimide (0.6 equiv.) were added to the reaction mixture, and stirring was continued for 2 h. Then, the reaction mixture was concentrated by rotary evaporation, and the crude product was purified using an Isolera One instrument from Biotage equipped with a 25 g KP-Si column. The product was eluted with CH₂Cl₂ and fractions containing the pure product were combined and concentrated by rotary evaporation. The product was dried in vacuo.

Analysis of 4-nitrophenyl 2-azidoacetate (16). Colorless oil, 294 mg (78%). ¹H NMR (300 MHz, CDCl₃) δ 8.36–8.19 (m, 2H, 2 × H-Ar), 7.45–7.27 (m, 2H, 2 × H-Ar), 4.19 (s, 2H, CH₂). ¹³C NMR (75 MHz, CDCl₃) δ 166.18 (s, C = O), 154.67 (C-Ar), 145.86 (C-Ar), 125.48 (2 × C-Ar), 122.27 (2 × C-Ar), 50.52 (CH₂). HR-MS (Q-TOF): m/z calcld. for chemical formula C₉H₉O₃N₃: 245.0243 (Supplementary Fig. 23).

Fig. 5 Functionalization of GH6-BIR2 with an azide, a fluorophore and biotin. a 3D structure of XIAP(124–240) with the linker region known to interact with caspase-3 and -7 depicted on the left and the BIR2 domain on the right. The reaction scheme of the inhibition assay is shown as well. The tetrapeptide DEVD is hydrolyzed by caspase-7 between the second aspartic acid and the AFC reporter group. AFC = 7-amino-4-trifluoromethylcoumarin. b Kinetic curves of the hydrolysis of Ac-DEVD-AFC by caspase-7, without inhibition (filled triangles) or when inhibited by unmodified (horizontal lines) or acylated (open circles) GH6-BIR2. The substrate only (crosses) was included as reference as well. c Two-step fluorescent labeling of GH6-BIR2 through acylation with 4-methoxyphenyl ester 18 followed by Cu(I)-catalyzed conjugation of alkyne cyanine dye 718, visualized by SDS-PAGE analysis. A fluorescence image and an image of the Coomassie-stained version of the same gel are shown. Lane 1: GH6-BIR2 treated with alkyne cyanine dye 718 and Cu(I) (negative control), lane 2: azido-functionalized GH6-BIR2 treated with alkyne cyanine dye 718 and Cu(I). d Direct biotinylation of GH6-BIR2 with 4-methoxyphenyl ester 20, visualized by Western blot analysis. Binding of streptavidin to biotinylated GH6-BIR2 (lane 2) was detected by incubating the blot with HRP-SAv, a conjugate of streptavidin and horseradish peroxidase (HRP), followed by the addition of a chemiluminescent substrate of HRP. Unmodified GH6-BIR2 was loaded in lane 1 as negative control.

Analysis of 4-methoxyphenyl 2-azidoacetate (18). Colorless oil, 316 mg (95%). ¹H NMR (300 MHz, CDCl₃) δ 7.49–7.38 (m, 2H, 2 × H-Ar), 7.34–7.25 (m, 1H, H-Ar), 7.21–7.13 (m, 2H, 2 × H-Ar), 4.15 (s, 2H, CH₂). ¹³C NMR (75 MHz, CDCl₃) δ 166.95 (C = O), 150.23 (C-Ar), 129.71 (2 × C-Ar), 126.50 (C-Ar), 121.28 (2 × C-Ar), 50.54 (CH₂). HR-MS (Q-TOF): m/z calcld. for chemical formula C₁₀H₁₀N₂O₂: 177.0538; found: [M + Na⁺]⁺ 200.0426 (Supplementary Fig. 24).
evaporation. The resulting brown Celite adsorbate was loaded on top of a pre- conditioned (ethyl acetate/heptane 1:39) column for vacuum liquid chromatography (VLC) (phenyl 30 × 1 cm, silica gel 60, 0.040 mm). The product was eluted using a gradient of ethyl acetate/heptane (1:39)–(1:1). Fractions containing the pure product were combined and concentrated by rotary evaporation. The product was dried in vacuo to yield 1.63 g (95%) of 18 as a colorless oil.

Synthesis of biotin OEG 4-methoxyphenyl ester 20. A solution of 8 (0.09 mmol) in dry DCM (4 mL) was added to 2-chlorotrityl chloride polystyrene resin (1 g, 1 mmol of 20% piperidine in DMF (4 mL) for 15 min. The resin was then washed with DCM (5 × 4 mL) and DMF (5 × 4 mL), and the Fmoc group was removed by treatment with 20% piperidine in DMF (4 mL) for 5 min., followed by 20% piperidine in DMF (5 × 4 mL) for 15 min. The resin was then washed with DCM (5 × 4 mL), and DCM (5 × 4 mL). The product was then purified using HPLC (on a Dionex Ultimate 3000 system) using a preparative C18 column (Phenomenex Gemini, 110 Å 5 µm C18 particles, 21 × 100 mm). Solvent A, water containing 0.1% TFA, and solvent B, acetonitrile containing 0.1% TFA, were used with gradient elution (0–5 min: 5–100% 50 min at 1 ml/min, and then added to the above resin (0.5 mmol). The resin was washed with NMP (4×).

NATURE COMMUNICATIONS | DOI: 10.1038/s41467-018-05695-3 | www.nature.com/naturecommunications

ARTICLE NATURE COMMUNICATIONS | DOI: 10.1038/s41467-018-05695-3

Analysis of peptides and proteins. Analyses of peptides were performed by HUPLC-MS on a RSLC Dionex Ultimate 3000 (Thermo) coupled to a QTOF Impact HD (Bruker) on a kinetex 2.6 µm EVO 100 Å C18 column (50 × 2.1 mm, Phenomenex) which were used for the peptides and Aers 3.6 µm wide pore C4 column (50 × 2.1 mm, Phenomenex) for the proteins. The peptide system was used at a flow rate of 0.5 mL min⁻¹; solvent A, water containing 0.1% formic acid; solvent B, acetonitrile containing 0.1% formic acid. The column was eluted using a linear gradient from 5 to 100% of solvent B. Some short peptides (Table 1) were analyzed on analytical HPLC (Dionex Ultimate 3000) instrument with an analytical C18 column (Phenomenex Gemini NX 110 Å 5 µm, 4.60 × 50 mm) coupled to a ESI-MS (MSQ Plus Mass Spectrometer, Thermo) using a linear gradient flow rate of water-acetonitrile containing 0.1% formic acid (Supplementary Tables 2 and 3). MALDI-TOF MS spectra were recorded on a Bruker autoflex™ speed MALDI-TOF instrument.

For the peptide modifications, conversions were based on relative abundance values calculated from deconvoluted MS data. Given the large number of positive charges that the proteins pick up when analyzed by ESI-MS, the effect of the loss of a single positive charge upon acylation of the N-terminus is expected to be negligible. EFGP picks up 20–40 charges, MBP 20 to >50, SUMO 8–20, and Bir 2 10–24.

Expression of proteins. The expression plasmid for GhEHFR-EGFP was generated from the pET11b-EGFP plasmid encoding GSSH4-EGFP™. The two N-termini serine residues were deleted by PCR using the following primers: 5′-TTGGTTGATCGGAAAGGATATACCATGCGGATC-ATAC-3′ and 5′-CTAGTTATGGCTCAGCG-3′. The insert, treated with Ncol and BamHI (New England Biolabs), was ligated in the pET15 vector, digested with the same restriction enzymes. D305S cells (Invitrogen) were used for cloning. The plasmid was sequenced using the T7 forward primer (5′-TAATACGACTCA TATAGGGG-3′, Eurofins). E. coli BL21[DE3] cells were transformed with plasmid encoding GH6-tagged EGFP or GSSH4-tagged EGFP. Cultures were grown in 230 mL LB medium with 100 µg mL⁻¹ ampicillin and protein expression was induced with 1 mM IPTG for 4 h at 37 °C. Cells were harvested and resuspended in 15 mL of 50 mM Na₂HPO₄, 300 mM NaCl at pH 8. Lysozyme was added to a final concentration of 1 mg mL⁻¹ and lysis was allowed to take place by incubation for 30 min on ice. The lysate was sonicated and centrifuged at 18,000 × g at 4 °C for 30 min. The supernatant was incubated with 1 mL of Ni-NTA resin for 20 min under rotary incubation. The resin was washed with 20 × 10 mL of 50 mM NaCl and 300 mM NiCl₂, 20 mM imidazole at pH 8, and eluted in 3 mL elution buffer (50 mM NaH₂PO₄, 300 mM NaCl, 20 mM imidazole at pH 8). An additional purification step was performed on an AKTA™ pure system equipped with a HiLoad 16/600 Superdex 75 pg column (GE Healthcare). Fractions containing the protein (as confirmed by SDS-PAGE analysis) were pooled and concentrated using a centrifugal filter unit (Amicon, Ultra-15, MWCO 10 kDa). The protein concentration was determined by measuring absorbance at 490 nm on a Nanodrop 2000 (Thermo Scientific) using an extinction coefficient of 55,200 M⁻¹ cm⁻¹. The protein (4.1 mg mL⁻¹) was frozen in aliquots in liquid nitrogen and stored at −20 °C.

GSH-tagged EGFP was produced by incubation of GSSH₄-tagged EGFP (100 µg with thrombin (0.05 U, Thrombin cleavage capture kit, Novagen) for 16 h at room temperature. Efficient removal of the N-terminal peptide GSH₄HHHEHSSGLPVR was achieved by MALDI-TOF analysis. Biotinylated thrombin was removed by incubation with streptavidinagarose, according to the manufacturer’s protocol.

The DNA coding for OneStep-TEV-Gly-His-Sumo (smtlQ12306) was ordered codon-optimized for E. coli from Geneart. Plasmid pIC828 Bsu4a was created by digestion with NdeI and BsiBI and ligation into the corresponding sites of plasmid pIC828 Bsu4a. After transformation of the plasmid into E. coli BL21[DE3], cells were cultured in 500 mL 2xYT medium containing 30 µg mL⁻¹ kanamycin at 37 °C until mid-exponential phase, at which time expression was induced with 1 mM IPTG for 4 h at 18 °C. After harvesting, cells were lysed by three passes through an Emulsiflex-C3 homogenizer (Aviaton) at 10–15 psi and any debris and unbroken cells were removed by centrifuging at 18,000 × g at 4 °C for 15 min. For protein purification, the supernatant was loaded onto a StrrepTrap HP column (GE Healthcare, Piscataway, NJ, USA) on an AKTA™ pure system. Purification was performed as per the specifications of the manufacturer. Cleavage of the OneStrep tag was performed with 1:100 (w/w) recombinant His-tagged TEV protease (produced in-house) overnight at 4 °C. TEV cleavage was assessed by SDS-PAGE and was found to be complete. TEV protease and SUMO protein with an N-terminal glycine residue followed by a His tag were subsequently separated by size exclusion chromatography using a HiLoad 16/600 Superdex 75 pg column (GE Healthcare). Competent E. coli BL21[DE3] cells were transfected with the pET28a plasmid containing the MBP coding sequence (null/P9AE9X). They were grown in LB
grown in 750 mL LB medium with 100 μg/mL ampicillin. E. coli BL21[DE3] cells were transformed with the respective plasmid. A culture was grown to an OD600 of 0.5 and induced with 1 mM isopropyl β-D-1-thiogalactopyranoside (IPTG) for 4 h at 30°C. The cells were lysed by two cycles of a French press at 35k psi. Cell debris was removed by centrifugation for 10 min at 18,000 × g. The resulting solution was incubated at 4°C for 15 min and the progress of the reaction was monitored by SDS-PAGE. The precipitated protein at 37°C for 1 day. An aliquot of 200 μL of a freshly prepared 25 mM solution of 4-methoxyphenyl 2-azidoacetate (18) in acetonitrile was added to 20 μL of a 35 μM solution of the GSH-EGFP and GSSH-EGFP variants in 200 mM HEPES buffer at pH 7.5 at 4°C. The resulting solution was incubated at that temperature and the progress of the reaction was monitored by LC-MS.

Two-step biotinylation and PEGylation of GH6-BIR2. An aliquot of 23 μL of a 14 mM solution of 4-methoxyphenyl 2-azidoacetate (18) in acetonitrile was added to 80 μL of a 5 mM solution of GH6-BIR2 variant in 200 mM HEPES buffer at pH 7.5 at 4°C. The resulting solution was incubated at that temperature for a period of 1 h. After the reaction was monitored by LC-MS, an aliquot of 200 μL of a 50 mM solution of 4-hydroxypropriazolomethylamine (THPTA) and 37.5 mM sodium ascorbate was added to catalyze the cyclodissociation. As a negative control, non-modified GH6-BIR2 was treated with alkyl cyanine 718 (1.5 μL of a 10 mM stock solution in DMSO, Sigma, Aldrich) and an equimolar amount to the activated enzyme, followed by an additional incubation step for 15 min at 37°C. The enzyme/inhibitor solution (50 μL) was mixed with a pre-heated solution of Ac-DEVD-AFC (50 μL, 200 μM in assay buffer, 2% DMSO), performed in duplicate. The formation of fluorescent product was followed at 37°C for 1 h using a SpectraMax i3x plate reader (λex = 405 nm, λem = 500 nm). Measurements were performed in duplicate.

Biotinylation of GH6-BIR2. Twenty equivalent of compound 20 (4 μL of a 10 mM stock solution in 1:1 acetonitrile:water) were added to 50 μL of a 40 μM GH6-BIR2 in 50 mM NaH2PO4, 150 mM NaCl (pH 7.5). The reaction solution was incubated at 4°C for 24 h and analyzed by LC-MS and western blot.

Trypsin and chymotrypsin digestion. Excess of GDL, phenyl ester 18 or DBCO-PEG4-biotin were removed by repeated (five times) centrifugal filtration using a 0.5-mL Amicon Ultra centrifugal filter (10 MCKW). Trypsin (sequencing-grade, modified, Sigma-Aldrich) digestions were performed overnight at 37°C in 0.1 M
NH₄HCO₃ (pH 8). Chymotrypsin (sequencing grade, Sigma-Aldrich) digests were performed overnight at room temperature in 0.1 M NH₄HCO₃ (pH 8). Digests were analyzed by MALDI-TOF and LC-MS.

**SDS-PAGE analysis.** Proteins (2–5 µg) were mixed with 4× Laemmlı sample buffer supplemented with 2-mercaptoethanol (BioRad), boiled at 95 °C for 2–5 min and loaded on an any kD™ MINI-PROTEAN® TGE™ precast protein gel (BioRad). The Dual Color Precision Plus Protein® Standard (BioRad) was used as reference. The gel was run at 200 V, stained in 0.1% (v/v) Coomassie Brilliant Blue R250 solution (10% (v/v) AcOH, 50% (v/v) MeOH in MilliQ water) and destained in destaining solution (10% (v/v) AcOH, 50% (v/v) MeOH in MilliQ water) (see Supplementary Fig. 28 for uncropped versions of the gels).

**Data availability.** All relevant data are available from the authors.

Received: 24 November 2017 Accepted: 17 July 2018 Published online: 17 August 2018

**References**

1. van Witteloostuijn, S. B., Pedersen, S. L. & Jensen, K. J. Half-life extension of recombinant protein molecules inside live cells. *Nature Mater.* **11**, 2474–2495 (2016).

2. Cui, L. & Rao, J. in *The Chemistry of Molecular Imaging* Ch. 2 (John Wiley & Sons Inc., Hoboken, NJ, 2014).

3. Krall, N., da Cruz, F. P., Boutureira, O. & Bernardes, G. J. Site-selective protein-modification chemistry for basic biology and drug development. *Nat. Chem.* **8**, 103–111 (2016).

4. Witus, L. S. & Francis, M. B. Using synthetically modified proteins to make new materials. *Acc. Chem. Res.* **44**, 774–783 (2011).

5. Davis, L. & Chin, J. W. Designer proteins: applications of genetic code expansion in cell biology. *Nat. Rev. Mol. Cell Biol.* **13**, 168–182 (2012).

6. Lang, K. & Chin, J. W. Cellular incorporation of unnatural amino acids and bioorthogonal labeling of proteins. *Chem. Rev.* **114**, 4764–4806 (2014).

7. Xie, J. & Schultz, P. G. A clinical toolkit for proteins—an expanded genetic code. *Nat. Rev. Mol. Cell Biol.* **17**, 775–782 (2006).

8. Gong, Y. & Pan, L. Recent advances in bioorthogonal reactions for site-specific protein labeling and engineering. *Tetrahedron Lett.* **56**, 2123–2132 (2015).

9. Sletten, E. M. & Bertozzi, C. R. Bioorthogonal chemistry: fishing for selectivity in a sea of functionality. *Angew. Chem. Int. Ed. Engl.* **48**, 6974–6998 (2009).

10. Jacob, E. & Unger, R. A tale of two tails: why are terminal residues of proteins crucial? *Chem. Soc. Rev.* **44**, 782 (2015).

11. Rosen, C. B. & Francis, M. B. Targeting the N terminus for site-selective protein modification. *Tetrahedron Lett.* **56**, 2123–2132 (2015).

12. Grimsley, G. R., Scholtz, J. M. & Pace, C. N. A summary of the measured pKₐ values of the ionizable groups to peptides and proteins via periodate oxidation. *Angew. Chem. Int. Ed.* **113** (2016).

13. Ren, H. et al. A biocompatible condensation reaction for the labeling of recombinant protein molecules inside live cells. *Nat. Chem.* **10**, 288–292 (2018).

14. Chai, J. et al. Structural basis of caspase-7 inhibition by XIAP. *J. Biol. Chem.* **284**, 438–439 (2009).

15. Stoll, B. et al. Metal-free and pH-controlled introduction of azides in proteins. *Bioconjug. Chem.* **17**, 168–172 (2006).

16. Skolnik, S. et al. Metal-free and pH-controlled introduction of azides in proteins. *Chem. Sci.* **7**, 705–718 (2016).

17. Chen, D., Dsotouar, M. M., Xiong, X., Wang, Y. & Chou, D. H. Selective N-terminal functionalization of native peptides and proteins. *Chem. Sci.* **8**, 7663–7722 (2017).

18. Zhang, C. et al. Pr-Clamp-mediated cysteine conjugation. *Nat. Chem.* **8**, 120–128 (2016).

19. Griffin, B. A., Adams, S. R. & Tsien, R. Y. Specific covalent labeling of recombinant protein molecules inside live cells. *Science* **281**, 269–272 (1998).

20. Halo, T. L., Appelbaum, J., Hober, M. E., Balkin, D. M. & Schepartz, A. Selective recognition of protein tetraserine motifs with a cell-permeable, pro-fluorescent bis-boronic acid. *J. Am. Chem. Soc.* **131**, 438–439 (2009).

21. Block, H. et al. Immobilized-metal affinity chromatography (IMAC): a review. *Methods Enzymol.* **463**, 439–473 (2009).

22. Geoghegan, K. F. et al. Synthesis of a new acyclic lipopeptide or a fluorophore that stabilizes membrane proteins. *Soft Matter* **10**, 738–752 (2014).

23. Fuks, S. & Vucic, D. Targeting IAP proteins for therapeutic intervention in cancer. *Nat. Rev. Drug Discov.* **11**, 109–124 (2012).

24. Chai, J. et al. Structural basis of caspase-7 inhibition by XIAP. *Cell. Mol. Life Sci.* **104**, 769–780 (2001).

25. Akiyama, M., Hara, Y. & Tanabe, M. Effects of substituents on the hydrolysis of 4-nitrophenyl acetate catalyzed by 2-substituted imidazoles. *J. Chem. Soc., Perkin Trans. 2*, 288–292 (1978).

26. Akiyama, M., Ihjima, M. & Hara, Y. General base catalytic activity of 2-substituted imidazoles for hydrolysis of ethyl dichloroacetate. *J. Chem. Soc., Perkin Trans. 2*, 1512–1516 (1979).

27. Brecher, A. S. & Balls, A. K. The catalysis of the non-chymotryptic hydrolysis of p-nitrophenyl acetate. *J. Biol. Chem.* **272**, 845–851 (1957).

28. Bruce, T. C. & Schmir, G. L. Imidazole catalysis. I. The catalysis of the hydrolysis of phenyl acetates by imidazole. *J. Am. Chem. Soc.* **79**, 1663–1667 (1957).

29. Jencks, W. P. & Carrillo, J. Imidazole catalysis. III. General base catalysis and the reactions of acetyl imidazoles with thiols and amines. *J. Biol. Chem.* **234**, 1280–1285 (1959).

30. Jencks, W. P. & Gilchrist, M. Nonlinear structure-reactivity correlations. The reactivity of nucleophilic reagents toward esters. *J. Am. Chem. Soc.* **90**, 2622–2637 (1968).

31. Kirsch, J. F. & Jencks, W. P. Nonlinear structure-reactivity correlations. The imidazole-catalyzed hydrolysis of esters. *J. Am. Chem. Soc.* **86**, 837–846 (1964).

32. Jencks, W. P. *Catalysis in Chemistry and Enzymology* 163–170 (Dover Publications Inc., New York, 1987).

33. Wally, J. et al. Insight into the coordination and the binding sites of Cu2+ by the histidyl-6-Tag using experimental and computational tools. *Inorg. Chem.* **53**, 6675–6683 (2014).

34. Broo, K. et al. Site selectivity in self-catalysed functionalization of helical polypeptide structures. *J. Chem. Soc., Perkin Trans. 2*, 397–398 (1997).

35. Uyttendaele, S. et al. Click chemistry mediated functionalization of vertical nanowires for biological applications. *Chem. Eur. J.* **22**, 496–500 (2016).

**Acknowledgements**

The VILLUM FONDEN is acknowledged for funding the BioNano-scale Nanoscale Engineering Center (BioNEC), a VILLUM center of excellence, grant number VKR022710, and an HPLC-MS instrument (VKR9912 to K.J.). A DFF-Mobiles grant from DFF and the European Commission (4093-00083 to S.S.) is acknowledged. The Novo Nordisk Fonden is acknowledged for financial support (S.K). Helle Munck Petersen is acknowledged for technical assistance. Amine Arslan is acknowledged for expression and purification of caspase-7.

**Author contributions**

The present work was designed and supervised by K.J. K.K.S. and C.T.H. synthesized and purified all peptides, while C.T.H. performed experiments on the glycosylation of short peptides (Table 1). J.E.R. did additional experiments on glycosylation and on oxidation. S.K. expressed and purified the SUMO variant, S.R.M. expressed and purified the MBP variant, while S.S. expressed and purified the EGFP and BIR-2 variants. K.V. and M.B.T. synthesized phenyl esters. M.C.M.-M. performed experiments on bivalent peptides (Tables 2 and 3), SUMO, EGFP variants, and MBP. S.S. supervised the protein chemistry and performed experiments on acylation of EGFP, MBP, and BIR-2. K.K.S. and S.S. performed the mass spectrometry. M.B.T. performed the mechanistic studies on model compounds. K.K.S. and M.B.T. performed further optimization experiments on
SUMO. K.J.J., M.C.M.-M., and S.S. cowrote the draft manuscript. All authors edited the manuscript.

Additional information
Supplementary Information accompanies this paper at https://doi.org/10.1038/s41467-018-05695-3.

Competing interests: The authors declare no competing interests.

Reprints and permission information is available online at http://npg.nature.com/reprintsandpermissions/

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

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article’s Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article’s Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit http://creativecommons.org/licenses/by/4.0/.

© The Author(s) 2018