

# Site-selective tagging of proteins by pnictogen-mediated self-assembly

Christoph Nitsche,\* Mithun C. Mahawaththa, Walter Becker,<sup>†</sup> Thomas Huber, and Gottfried Otting\*

Research School of Chemistry, Australian National University, Canberra, ACT 2601, Australia

*Supporting Information Placeholder*

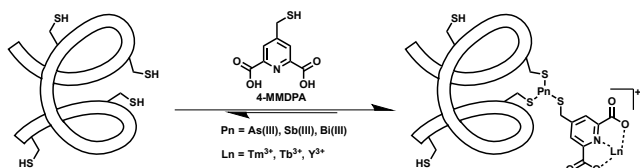
**ABSTRACT:** Only few methods exist for site-specific covalent modifications of proteins in aqueous solution at neutral pH. All require mutations of the protein that have a high potential to impede its function or are too costly to be practical. Here we describe a new strategy for site-selective covalent chemical modification of proteins, which uses trivalent pnictogens (As, Sb, or Bi) for the spontaneous assembly of stable complexes between a vicinal di-cysteine motif in a protein and a thiol-probe in water and at neutral pH. Owing to the reversibility of the reaction, the self-assembly is site-specific and inert towards other solvent-exposed single cysteine residues. A minimal mutation to generate two vicinal cysteine residues with appropriate side chain orientations is shown to be necessary and sufficient for the ligation reaction. The method is demonstrated by attaching a lanthanide tag to the Zika virus NS2B-NS3 protease, enabling paramagnetic nuclear magnetic resonance (NMR) spectroscopy and luminescence experiments. The equilibrium of tag assembly is relatively rapid with As(III) and Sb(III), whereas the complex with Bi(III) can be detected even after polyacrylamide gel electrophoresis. The approach enables site-specific protein modifications with minimal impact on protein structure.

Selective protein tagging strategies are critically important for probing their function, cellular localization and dynamics. It is difficult, however, to achieve this in a site-specific manner, as is often required to avoid impeding the natural function of the protein in an uncontrolled manner. N- or C-terminal fusions, e.g. with green fluorescent protein, tend to alter the biophysical properties of the target protein significantly. Cycloaddition reactions, often referred to as ‘click’ chemistry, require the selective incorporation of costly unnatural amino acids and thus depend on the availability of a suitable overexpression system for the protein of interest. Using a single solvent-exposed cysteine residue to a tag by formation of a disulfide or thioether linkage requires the prior mutation of all other naturally occurring cysteine residues that may compete in the reaction. For many proteins, so many mutations are detrimental to their structure and func-

tion. The present work describes a simple strategy whereby the chemical probe is attached to *two* cysteine residues by preferential and reversible binding, without interference by single cysteine residues.

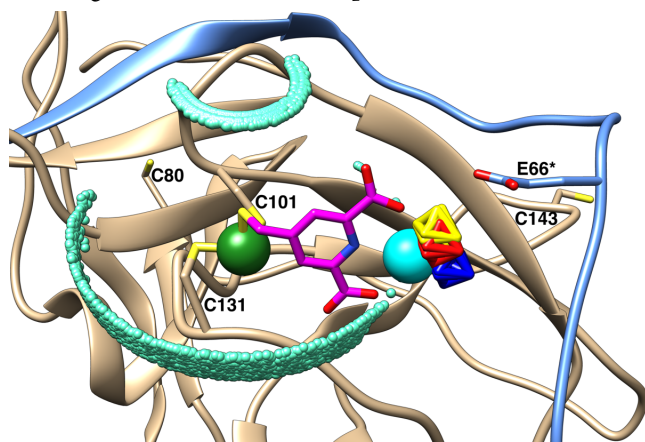
Our method is based on the unique thiophilic characteristics of trivalent pnictogens (As, Sb, or Bi). As(III) is well-known to prefer binding to multiple over single thiol groups.<sup>1,2</sup> For example, biarsenical fluorescent probes, such as FAsH-EDT<sub>2</sub> or ReAsH-EDT<sub>2</sub>, selectively bind to CCXXCC peptide motifs.<sup>3,4</sup> Recently, monoarsenical probes that bind to two vicinal cysteines have been used to link protein and peptide chains or detect vicinal dithiol-containing proteins (VDPs) in cells.<sup>5,6</sup> Unfortunately, the chemical synthesis of the requisite organoarsenic compounds is challenging and fraught with toxicity concerns, which severely limits broader availability and applicability. While the heavier pnictogens such as antimony or bismuth share the thiophilic character of arsenic and are less toxic, organic antimony and bismuth compounds are even more challenging to synthesize and are unlikely to be sufficiently stable in aqueous solution. Here we report a strategy for pnictogen-mediated tagging of VDPs without requiring the chemical synthesis of organic pnictogen tags, which greatly expands the scope of suitable probes toward readily available thiol compounds.

Our strategy involves the spontaneous attachment of trivalent inorganic arsenic, antimony or bismuth to two vicinal cysteine residues of the protein followed by addition of a thiol group containing probe to form a covalent thiolate complex (Scheme 1). The ternary complex between VDP, tag and pnictogen selectively self-assembles in aqueous solution. In the present work, we added, in addition, a lanthanide ion to confirm the location of the tag on the protein and explore the chemical stability of the probe by paramagnetic NMR spectroscopy.<sup>7</sup> The approach is demonstrated with the Zika virus NS2B-NS3 protease (ZiPro), which is an established drug target. The wild-type protein contains two solvent-exposed cysteine residues.<sup>8</sup>



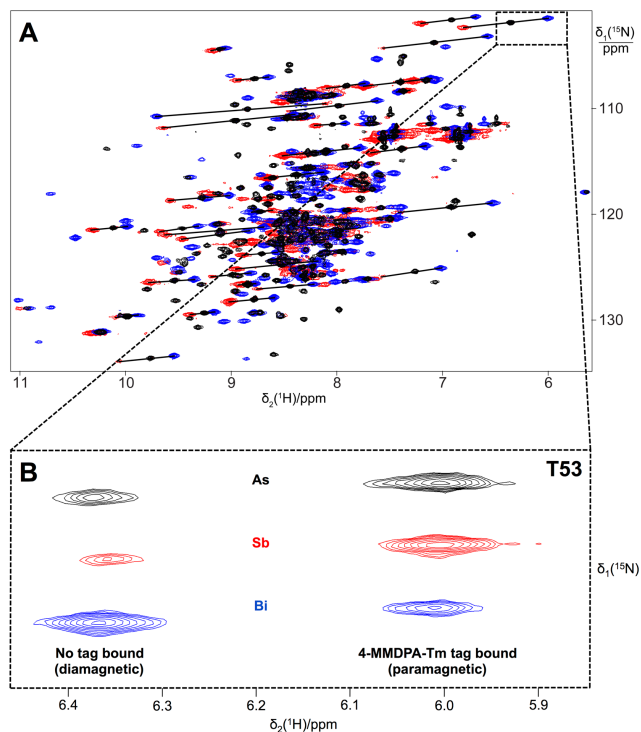
**Scheme 1.** Pnictogen-mediated assembly of the lanthanide binding tag 4-MMDPA and a protein that contains solvent-exposed cysteine residues in addition to an engineered binding site with two vicinal thiols.

Self-assembly of pnictogen thiolates has been described previously for all pnictogens except nitrogen in supramolecular chemistry protocols conducted in organic solvents.<sup>9,10</sup> The driving force is the formation of labile covalent bonds between sulfur and Pn(III), and additional pnictogen- $\pi$  interactions have been proposed to stabilize the complex in the case of mercaptomethylaryl-based ligands.<sup>11,12</sup> Consequently, we chose a simple paramagnetic lanthanide tag with a mercaptomethylaryl moiety, 4-mercaptopethyl-dipicolinic acid (4-MMDPA; Scheme 1) to favour binding of the probe to vicinal thiol groups over binding to competing single cysteine residues and to outcompete other thiol reagents such as glutathione, which are aliphatic.<sup>13</sup>



**Figure 1.** Model of the self-assembled complex between ZiPro P101C/P131C, a trivalent pnictogen (green), 4-MMDPA (magenta) and lanthanide (cyan), highlighting bonds with nitrogen in blue, oxygen in red, and sulfur in yellow. Small cyan balls identify all sterically allowed metal positions determined by modeling the protein-pnictogen-4-MMDPA-lanthanide complex with variable dihedral angles of the 4-MMDPA moiety. The lanthanide coordinates found by NMR are shown as octahedrons for the complexes assembled with arsenic (yellow), antimony (red), and bismuth (blue). These metal positions were determined using pseudocontact shifts measured with  $\text{Tm}^{3+}$  and  $\text{Tb}^{3+}$  to fit magnetic susceptibility anisotropy ( $\Delta\chi$ ) tensors to the crystal structure (PDB ID: 5LC0)<sup>8</sup> of the protein (Table S2). NS3 is shown in beige and the NS2B cofactor in light blue. The carboxyl group of Glu 66\* (where the star identifies residues of NS2B) is near the metal site and may assist in positioning the lanthanide ion. Cys 80 and Cys 143 are part of the wild-type amino acid sequence of ZiPro.

Suitable sites for pnictogen installation were predicted computationally, starting from the crystal structure of the Zika virus protease (PDB ID: 5LC0)<sup>8</sup> and mutating residues to cysteines, then checking whether standard rotamers could form bonds with an arsenic atom with close to 90° S-As-S bond angles. This identified the double-mutation P101C/P131C as a possible site near the active site of the protease (Figure 1).

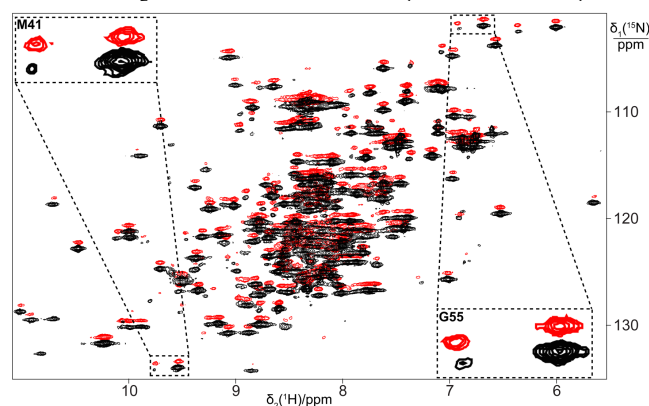


**Figure 2.** Superimposition of  $[\text{}^{15}\text{N}, \text{H}]$ -HSQC spectra of 0.3 mM solutions of ZiPro P101C/P131C in the presence of 0.3 mM  $\text{SbCl}_3$ , 4-MMDPA and paramagnetic  $\text{Tm}^{3+}$  (blue),  $\text{Tb}^{3+}$  (red) or diamagnetic  $\text{Y}^{3+}$  (black). The spectra were recorded at 25 °C in 20 mM MES pH 6.5, 150 mM NaCl, 1 mM TCEP on a Bruker 800 MHz NMR spectrometer. A) Overview of the spectra. Black lines connect corresponding cross-peaks observed with diamagnetic and paramagnetic lanthanides. B) Selected spectral region showing the cross-peak of the backbone amide of Thr 53 in the complexes assembled with  $\text{NaAsO}_2$  (black),  $\text{SbCl}_3$  (red) or  $\text{BiBr}_3$  (blue), in the presence of 0.3 mM 4-MMDPA and  $\text{Tm}^{3+}$ . The persistence of diamagnetic cross-peaks illustrates the equilibrium between protein with and without probe molecules, which is most strongly shifted towards the complex for the sample assembled with antimony. The spectra were shifted in the  $\delta_1$  dimension for improved visibility.

Addition of  $\text{NaAsO}_2$  in water, or  $\text{SbCl}_3$  or  $\text{BiBr}_3$  in DMSO to uniformly  $^{15}\text{N}$ -labeled ZiPro P101C/P131C indeed yielded new cross-peaks in the  $[\text{}^{15}\text{N}, \text{H}]$ -HSQC spectrum, indicating slow chemical exchange between bound and free protein, as expected for the formation of covalent-reversible bonds (data not shown). Isothermal titration calorimetry (ITC) with As(III) indicated a dissociation constant  $K_d$  of about 260  $\mu\text{M}$  (Figure S4). Subsequent NMR-titrations of

the pnictogen-protease complexes with 4-MMDPA also indicated slow chemical exchange for the association and dissociation of the ternary complex (data not shown). Mass spectrometry confirmed the formation of the ZiPro-As(III)-4-MMDPA assembly in equilibrium with free ZiPro (Figure S5). In the case of bismuth, the samples precipitated during longer NMR measurements at 25 °C, presumably by the formation of insoluble Bi<sub>2</sub>O<sub>3</sub>. A dimeric coordination complex of Bi(III) with dipicolinic acid has been described<sup>14</sup> and a corresponding complex with 4-MMDPA may assist the disassembly of the complex in the absence of lanthanide.

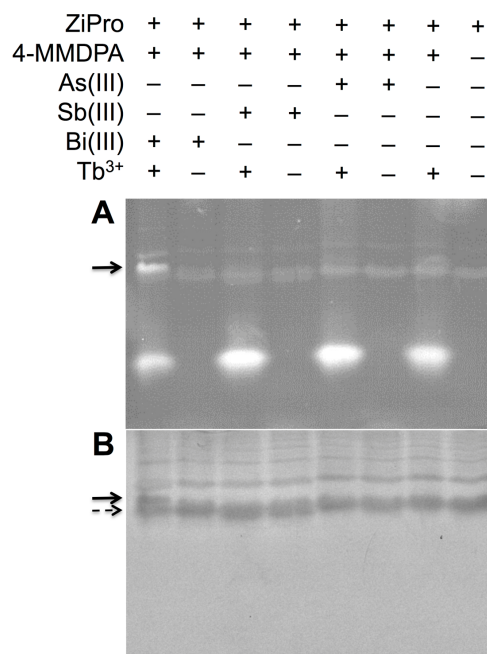
Addition of lanthanide ions produced the quaternary complex of ZiPro P101C/P131C, pnictogen, 4-MMDPA and lanthanide, with the paramagnetic lanthanide ions Tm<sup>3+</sup> and Tb<sup>3+</sup> generating pseudocontact shifts (PCS) of opposite sign in the [<sup>15</sup>N,<sup>1</sup>H]-HSQC spectrum (Figures 2, S2 and S3). The chemical exchange between the paramagnetic quaternary complex and the dissociated diamagnetic species was slow on the NMR time scale for all three pnictogens, with the equilibrium favouring the fully assembled complex for arsenic and, even more so, antimony (Figure 2B). The latter assembly showed the highest degree of association in our experiments. In contrast, the complex with bismuth appeared to be more dissociated, but it is unclear how far this was a consequence of limited solubility or lower affinity.



**Figure 3.** Superimposition of [<sup>15</sup>N,<sup>1</sup>H]-HSQC spectra of 0.3 mM solutions of ZiPro P101C/P131C, SbCl<sub>3</sub>, 4-MMDPA and paramagnetic Tm<sup>3+</sup>, in the absence (black) and presence (red) of 1 mM glutathione (GSH). The cross-peaks of Met 41 and Gly 55 illustrate the persistence of the paramagnetic peaks in the presence of excess GSH. For improved visual representation, the black spectrum was shifted in the  $\delta_1$  dimension by 0.6 ppm.

To determine the metal position, we measured the PCSs of well-resolved [<sup>15</sup>N,<sup>1</sup>H]-HSQC cross-peaks of 29 residues in the protease generated by Tm<sup>3+</sup> and Tb<sup>3+</sup> (Table S1) and fitted  $\Delta\chi$  tensors for all pnictogen complexes to the crystal structure of the protein (Table S2).<sup>15</sup> Excellent correlations between experimental and back-calculated PCSs were obtained with quality factors between 0.05 and 0.06 (Figure S1 and Table S2). This indicates that the binding of pnictogen,

tag or lanthanide caused no global distortion of the protein structure. The axial components of the  $\Delta\chi$  tensors (up to  $20 \cdot 10^{-32} \text{ m}^3$ ) were significantly larger than those observed previously for 4-MMDPA attached to cysteine or other DPA derivatives that form shorter linkers with cysteine residues, such as 4-mercapto-DPA (4-MDPA) or 3-mercapto-DPA (3-MDPA),<sup>13,16,17</sup> indicating good immobilisation of the lanthanide ions. Figure 1 shows the positions of the metal ions obtained by the  $\Delta\chi$  tensor determinations. Comprehensive modelling of the range of tag conformations compatible with the pnictogen coordination by Cys 101 and Cys 131 indicated two arcs of localization spaces, where the lanthanide ion could, in principle, be located (Figure 1). Proximity of the metal to the side chain carboxyl group of Glu 66\* suggests that coordination to this side chain helps immobilise the lanthanide ion.



**Figure 4.** Native PAGE for ZiPro C80S/C143S/P101C/P131C in the presence and absence of 4-MMDPA, trivalent pnictogens, and Tb<sup>3+</sup>. A) Terbium luminescence at 254 nm excitation wave length. B) Coomassie blue stain. The dashed arrow identifies the band of native ZiPro without tag. The bold arrow identifies the band of the quaternary complex of ZiPro with trivalent bismuth, 4-MMDPA and terbium.

Initially, we performed these studies with a ZiPro mutant, where the solvent-exposed residues Cys 80 and Cys 143 of the wild-type protein were substituted by serine to avoid any potential competition for pnictogen binding. This proved to be unnecessary. To probe the inertness of the quaternary complex with Sb(III) towards additional thiol compounds, we recorded [<sup>15</sup>N,<sup>1</sup>H]-HSQC spectra also in the presence of excess glutathione (GSH). The results showed that GSH competes with 4-MMDPA for binding to the pnictogen, but the 4-MMDPA complex clearly is more stable, which may in

part be explained by a pnictogen- $\pi$  interaction with the mercaptomethylpyridine group (Figure 3). Remarkably, when we probed the stability of the quaternary complex between ZiPro, pnictogen, 4-MMDPA and lanthanide by native polyacrylamide electrophoresis (PAGE) and terbium luminescence, the bismuth-mediated complex was found to survive intact, whereas the complexes with As(III) and Sb(III) decomposed, indicating a slower dissociation rate of the bismuth complex (Figure 4).

In summary, a VDP containing a pair of cysteine residues with appropriate geometry presents opportunities for pnictogen-mediated self-assembled tagging with mercaptomethylaryl containing molecules. The assembly from simple inorganic pnictogen compounds presents a facile way of tagging, which is hardly perturbed by the presence of single cysteine residues and other thiol-group containing biomolecules. Interestingly, the assembly appears to be more favoured with antimony than with arsenic (Figure 2B). The present work establishes a strategy to dispense with the difficult synthesis of toxic organoarsenical compounds and to expand the spectrum of pnictogen-mediated protein modifications from arsenic to less toxic antimony and non-toxic bismuth. The assembly is orthogonal to conventional tagging reactions of cysteine residues, can be performed in aqueous solution at neutral pH and produces surprisingly stable complexes. The approach opens a path to site-specific reversible tagging of sensitive proteins under very mild condition for use with a broad range of commercially available probes.

## ASSOCIATED CONTENT

**Supporting Information.** Detailed procedures and conditions for protein expression and purification, sample preparation, ITC, MS and PAGE experiments; details of computational design and molecular modeling; tables of PCSs and fitted  $\Delta\chi$  tensor parameters; correlations between back-calculated and measured PCSs;  $^{15}\text{N}, ^1\text{H}$ -HSQC spectra of ZiPro P101C/P131C in presence of As(III) or Bi(III), 4-MMPDA and  $\text{Y}^{3+}$ ,  $\text{Tm}^{3+}$  or  $\text{Tb}^{3+}$ ; ITC data; mass spectrometry data.

The Supporting Information is available free of charge on the ACS Publications website.

## AUTHOR INFORMATION

### Corresponding Authors

\*christoph.nitsche@anu.edu.au

\*gottfried.otting@anu.edu.au

## ORCID

Christoph Nitsche: 0000-0002-3704-2699

Gottfried Otting: 0000-0002-0563-0146

## Present Address

<sup>†</sup>Institute of Chemistry, Karl Franzens Universität Graz, 8010 Graz, Austria

## Notes

The authors declare no competing financial interest.

## ACKNOWLEDGMENT

C.N. thanks the Alexander von Humboldt Foundation for a Feodor Lynen fellowship. Financial support by the Australian Research Council is gratefully acknowledged.

## REFERENCES

- (1) Shen, S.; Li, X. F.; Cullen, W. R.; Weinfeld, M.; Le, X. C. *Chem. Rev.* **2013**, *113*, 7769.
- (2) Chen, B.; Liu, Q.; Popowich, A.; Shen, S.; Yan, X.; Zhang, Q.; Li, X.-F.; Weinfeld, M.; Cullen, W. R.; Le, X. C. *Metallomics* **2015**, *7*, 39.
- (3) Griffin, B. A.; Adams, S. R.; Tsien, R. Y. *Science* **1998**, *281*, 269.
- (4) Adams, S. R.; Campbell, R. E.; Gross, L. A.; Martin, B. R.; Walkup, G. K.; Yao, Y.; Llopis, J.; Tsien, R. Y. *J. Am. Chem. Soc.* **2002**, *124*, 6063.
- (5) Wilson, P.; Anastasaki, A.; Owen, M. R.; Kempe, K.; Haddleton, D. M.; Mann, S. K.; Johnston, A. P.; Quinn, J. F.; Whittaker, M. R.; Hogg, P. J.; Davis, T. P. *J. Am. Chem. Soc.* **2015**, *137*, 4215.
- (6) Wang, Y.; Yang, X.-F.; Zhong, Y.; Gong, X.; Li, Z.; Li, H. *Chem. Sci.* **2016**, *7*, 518.
- (7) Nitsche, C.; Otting, G. *Prog. NMR Spectrosc.* **2017**, *98-99*, 20.
- (8) Lei, J.; Hansen, G.; Nitsche, C.; Klein, C. D.; Zhang, L.; Hilgenfeld, R. *Science* **2016**, *353*, 503.
- (9) Cangelosi, V. M.; Carter, T. G.; Crossland, J. L.; Zakharov, L. N.; Johnson, D. W. *Inorg. Chem.* **2010**, *49*, 9985.
- (10) Collins, M. S.; Choi, R. Y.; Zakharov, L. N.; Watson, L. A.; Hay, B. P.; Johnson, D. W. *Chem. Sci.* **2015**, *6*, 2444.
- (11) Cangelosi, V. M.; Pitt, M. A.; Vickaryous, W. J.; Allen, C. A.; Zakharov, L. N.; Johnson, D. W. *Cryst. Growth Des.* **2010**, *10*, 3531.
- (12) Watt, M. M.; Collins, M. S.; Johnson, D. W. *Acc. Chem. Res.* **2013**, *46*, 955.
- (13) Su, X. C.; Man, B.; Beeren, S.; Liang, H.; Simonsen, S.; Schmitz, C.; Huber, T.; Messerle, B. A.; Otting, G. *J. Am. Chem. Soc.* **2008**, *130*, 10486.
- (14) Stavila, V.; Davidovich, R. L.; Gulea, A.; Whitmire, K. H. *Coord. Chem. Rev.* **2006**, *250*, 2782.
- (15) Schmitz, C.; Stanton-Cook, M. J.; Su, X.-C.; Otting, G.; Huber, T. *J. Biomol. NMR* **2008**, *41*, 179.
- (16) Jia, X.; Maleckis, A.; Huber, T.; Otting, G. *Chem. Eur. J.* **2011**, *17*, 6830.
- (17) Man, B.; Su, X. C.; Liang, H.; Simonsen, S.; Huber, T.; Messerle, B. A.; Otting, G. *Chem. Eur. J.* **2010**, *16*, 3827.

TOC

