Organoarsenic probes to study proteins by NMR spectroscopy

Mithun C. Mahawaththa,‡[a] Henry W. Orton,‡[a] Ibidolapo Adekoya,[a] Thomas Huber,[a] Gottfried Otting*[a] and Christoph Nitsche*[a]

Arsenical probes enable structural studies of proteins. We report the first organoarsenic probes for nuclear magnetic resonance (NMR) spectroscopy to study proteins in solutions. These probes can be attached to irregular loop regions. A lanthanide-binding tag induces sizable pseudocontact shifts in protein NMR spectra of a magnitude never observed for small paramagnetic probes before.

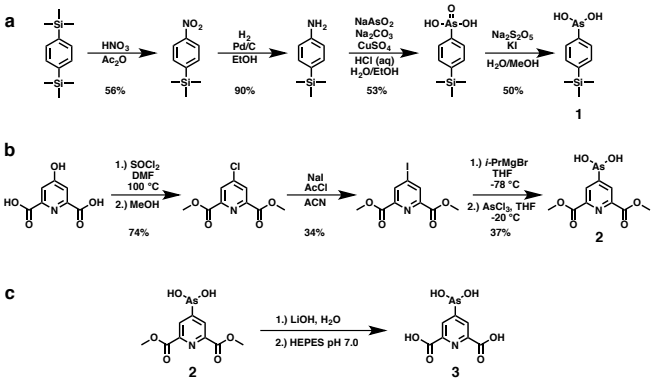
Trivalent arsenic compounds bind to spatially adjacent cysteine residues.[1] In the past, this entropically favourable interaction has been exploited to fluorescently label proteins with organoarsenic probes, such as FlAsH-EDT_2 or ReAsH-EDT_2 pioneered by Tsien and co-workers.[2] These established biarsenical probes bind to four cysteine residues and require engineered CCXXCC peptide motifs. Alternatively, fluorescent monoarsenical probes can be attached to vicinal dithiol motifs.[3]

Here we report the first monoarsenical probes specifically designed for applications in protein nuclear magnetic resonance (NMR) spectroscopy. NMR spectroscopy is an established tool in protein structural biology and drug discovery[4] and two kinds of probes recently gained increasing attention: (i) *tert*-butyl- and trimethylsilyl-based probes to monitor protein responses to binding from simple 1D ^{1}H NMR spectra[5] and (ii) lanthanide-binding tags (LBTs) to generate structural restraints from paramagnetic NMR spectra.[6] The present work presents organoarsenic probes for both applications with proteins. We introduce a trimethylsilyl (TMS) tag (**1**) for simple 1D ^{1}H-NMR experiments and a small LBT (**3**) for paramagnetic measurements (Scheme 1).

The ^{1}H chemical shift of the TMS group is near 0 ppm in a spectral region with few signals from the target protein, enabling NMR experiments at low concentrations without isotope-labelling. Previous TMS tags were either attached to single cysteine residues[5a] or incorporated as part of unnatural amino acids.[5d]

LBTs allow the attachment of paramagnetic ions to proteins in order to elicit paramagnetic relaxation enhancements (PREs), pseudocontact shifts (PCSs) and other paramagnetic effects in protein NMR spectra.[7] PCSs are particularly attractive, as they are easy to measure and the effect is long-ranging.[6a] To derive structural information from PCSs, it is important that the movement of the metal ion in the tag relative to the protein is minimized. Dipicolinic acid (DPA) derivatives are amongst the smallest LBTs and have previously been used successfully to generate PCSs following attachment to single cysteine residues.[8]

TMS probe **1** was synthesized by reacting the diazonium intermediate of 4-(trimethylsilyl)aniline with sodium arsenite[9] and subsequent reduction of As(V) to As(III) (Scheme 1a). The key step in the synthesis of DPA probe **3** involves a halogen-magnesium exchange[10] in dimethyl 4-iododipicolinate and subsequent reaction with arsenic(III) chloride to yield compound **2**, which can be isolated in moderate yield (Scheme 1b). The actual probe **3** is generated *in situ* from **2** via an ester cleavage step before usage and storage in buffer (Scheme 1c).



Scheme 1 (a) Synthetic procedure for organoarsenic probe **1**. (b) Synthetic procedure for the organoarsenic precursor probe **2**. (c) *In situ* generation of lanthanide-binding tag **3** from precursor **2** prior to protein tagging reaction.

[a] *ARC Centre of Excellence for Innovations in Peptide and Protein Science, Research School of Chemistry, Australian National University, Canberra, ACT 2601, Australia. Email: christoph.nitsche@anu.edu.au, gottfried.otting@anu.edu.au*

‡ These authors contributed equally.

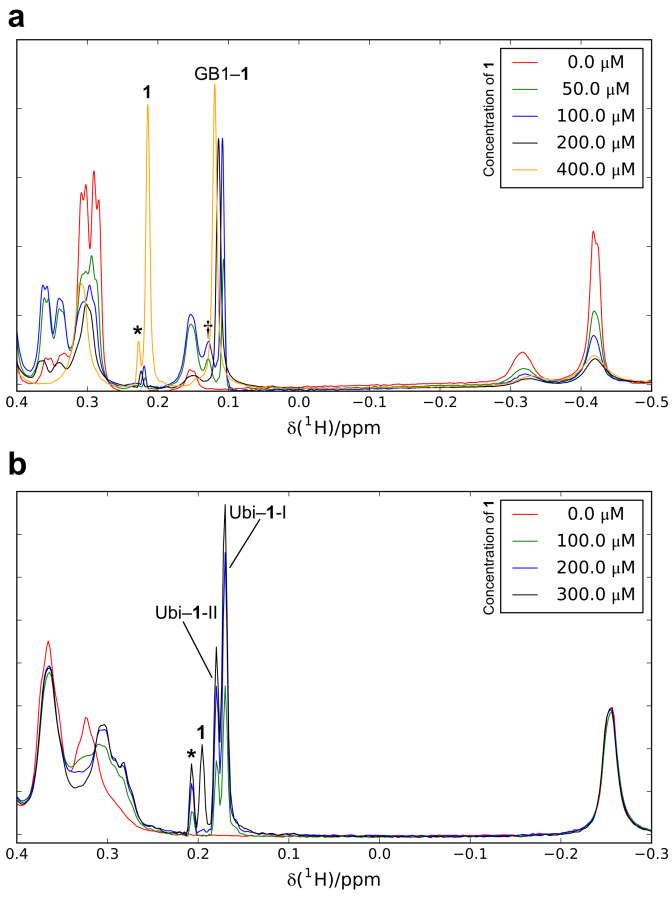


Fig. 1 ^{1}H-NMR titrations of TMS probe **1** with double-cysteine mutants of the proteins GB1 and ubiquitin in NMR buffer (50 mM HEPES, pH 7.0, 1 mM TCEP). Spectra were recorded at 25 °C on a Bruker 800 MHz NMR spectrometer using a jump-return sequence[11] for water suppression. (a) Titration of a 300 μM solution of GB1 K10C/T11C with **1** leads to the appearance of a singlet near 0.1 ppm (labelled GB-**1**). In excess of **1**, an additional singlet appears near 0.2 ppm, indicating that the exchange between dissociated **1** and **1** bound to GB1 is slow on the chemical shift time scale. During titration, also a minor peak (labelled with a dagger) appeared that vanished at higher **1**:GB1 ratios. The asterisk marks to an inert impurity of compound **1**. (b) Titration of a 250 μM solution of ubiquitin E18C/S20C/Q62E. During titration, two distinct ubiquitin–**1** complexes (labelled Ubi–**1**-I and Ubi–**1**-II) appear in an approximate ratio of 2:1. In excess of **1**, an additional peak appears that corresponds to dissociated **1**.

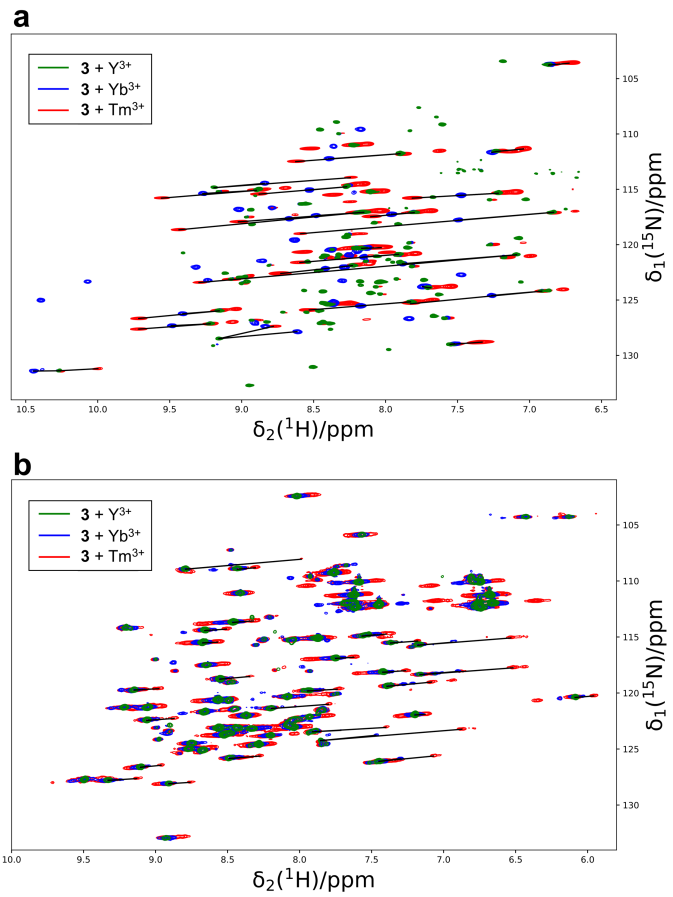


Fig. 2 Superimposition of [^{15}N,^{1}H]-HSQC spectra of 0.3 mM solutions of (a) GB1 K10C/T11C and (b) ubiquitin E18C/S20C/Q62E in the presence of 0.3 mM of **3** and paramagnetic Tm^{3+} (red), Yb^{3+} (blue) or diamagnetic Y^{3+} (green), as indicated. The spectra were recorded at 25 °C in NMR buffer.

To demonstrate the approach, we used double-cysteine mutants of human ubiquitin and the B1 immunoglobulin-binding domain of streptococcal protein G (GB1). Initial attempts to attach probes **1** and **3** to α-helices in ubiquitin and GB1 (using mutants with two cysteine residues in positions *i* and *i*+1, or *i* and *i*+4) yielded multiple TMS signals and no PCSs in NMR spectra, suggesting that the binding of the tag affected their structures (data not shown). As(III)-binding to two cysteine residues in an α-helix has previously been reported to destabilise the helix in all substitution patterns except *i*, *i*+4.[12] In order to develop a generally applicable method, we stepped beyond well-defined secondary structural elements and designed double-cysteine mutations in flexible loop regions of ubiquitin and GB1. The resulting flexibility provides an optimal geometry favourable for arsenic binding without risking major distortion of the protein structure.

We explored the K10C/T11C double mutant of GB1 (an *i*,*i*+1 mutation) and the E18C/S20C double mutant of ubiquitin (an *i*,*i*+2 mutation) that were suggested to be suitable for arsenic binding by modelling. An additional Q62E mutation distant from the probe-binding site was introduced in ubiquitin to maintain the overall protein charge.

High-resolution mass-spectrometry confirmed quantitative binding of probes **1** and **3** to GB1 K10C/T11C and ubiquitin E18C/S20C (Table S1). 1D ^{1}H-NMR titration experiments of both protein mutants with TMS probe **1** indicated tight binding and slow exchange on the NMR time scale between bound and dissociated probe **1** (Fig. 1). In the case of GB1 K10C/T11C, only a single TMS species was observed for the protein-tag complex, whereas for ubiquitin E18C/S20C two TMS peaks were observed in a 2:1 ratio, indicating two different protein-tag species. The latter observation is unsurprising, as trivalent arsenic compounds become chiral in the presence of three different substituents, owing to the free electron pair in As(III). Consequently, the achiral probe **1** can result in two diastereomeric protein-**1** complexes upon binding to two vicinal cysteine residues.

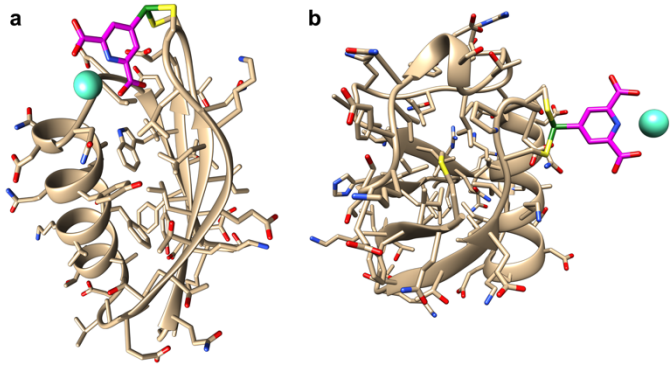


Fig. 3 Models of the complexes between **3**, lanthanide and (a) GB1 K10C/T11C and (b) ubiquitin E18C/S20C/Q62E. Compound **3** is shown in magenta and the lanthanide ion in cyan. Bonds with nitrogen are highlighted in blue, oxygen in red, sulfur in yellow and arsenic in green. The lanthanide positions were determined using pseudocontact shifts measured with Tm^{3+} and Yb^{3+} to fit magnetic susceptibility anisotropy ($\Delta\chi$) tensors (Table S2) to the crystal structures of GB1 (PDB ID: 2QMT)[13] and ubiquitin (PDB ID: 1UBQ).[14]

To explore the probe-protein interactions in more detail, we attached probe **3** to GB1 K10C/T11C and ubiquitin E18C/S20C and recorded [^{15}N,^{1}H]-HSQC NMR spectra in the presence of different lanthanide ions (Fig. 2). Sizeable PCSs were observed for Tm^{3+} and Yb^{3+} ions, using Y^{3+} as the diamagnetic reference (Fig. 2). In agreement with the observations made with the TMS probe **1**, a minor species was evidenced by a second set of weaker cross-peaks for the ubiquitin E18C/S20C complex. To determine the lanthanide positions, we used the PCSs of well-resolved cross-peaks (Tables S3, S4) and fitted $\Delta\chi$ tensors to the crystal structures of ubiquitin and GB1 (Table S2).[15] Excellent correlations between measured and back-calculated PCSs (Fig. S1) with quality factors ranging from 0.03 to 0.08 (Table S2) indicated minimal change in the protein structure upon probe binding. The magnitude of the axial $\Delta\chi$ tensor component for the GB1-**3**-Tm^{3+} complex ($-39.3 \cdot 10^{-32}$ m^3) exceeded that of any other previously reported DPA-based LBTs.[8] Axial $\Delta\chi$-tensor components of that magnitude were previously only observed for much larger cyclen-based tags.[7] This indicates superior immobilisation of the metal ion by the two-point anchoring and rigidity of probe **3**. Modelling revealed tag conformations for GB1 K10C/T11C and ubiquitin E18C/S20C compatible with the lanthanide position calculated from the PCSs measurements (Fig. 3). In contrast to previous studies with a DPA-based tag,[16] the accurate positioning of the lanthanide ions was achieved without additional assistance by carboxylate side-chains from aspartate or glutamate.

Having explored well-structured proteins by paramagnetic NMR, we were interested to know if the probe **3** could also be applied to more flexible protein regions like protein termini, as this would allow selective spin labelling of polypeptide segments without prior knowledge of their 3D structure. As PCSs are highly sensitive to tag flexibility, we used EPR spectroscopy to measure the distance between two Gd(III) ions by double electron–electron resonance (DEER) in the homodimer of the main protease of SARS-CoV-2. A tag binding site was created by fusing the peptide sequence GSGCCHHHHHH to the C-terminus of the protease. As expected, DEER data analysis revealed a broad distance distribution, centred at about 4.5 nm and approximately 2 nm FWHM (Fig. S3), reflecting the flexibility of the fusion peptide. In contrast, no DEER effect was observed in a construct without the GSG spacer between the C-terminus of the wild-type protein and the double-cysteine motif, confirming the specificity of the probe for flexible, solvent-exposed double-cysteine motives.

Probes **1** and **3** are the first organoarsenic probes developed for protein NMR spectroscopy. Pairs of cysteine residues can be readily introduced in proteins by site-directed mutagenesis. Probe attachment to flexible protein regions, such as loops, allows backbone and sidechain thiol groups to conform to the geometric requirements for As(III) coordination, while at the same time avoiding distortion of the rest of the protein structure. The double anchoring approach results in well-defined probe positions. For the first time, this has enabled a small LBT, such as DPA, to elicit large $\Delta\chi$ tensors in the target protein, opening a useful alternative to large cyclen-based tags. Given the convenience of spontaneous, tight yet reversible, covalent attachment, organoarsenic tags carry unique promise in protein NMR spectroscopy.

Acknowledgments

We thank Professor Daniella Goldfarb and Dr Akiva Feintuch (Weizmann Institute, Israel) and Professor Nicholas Cox and Ms Martyna Judd (Research School of Chemistry, Australian National University) for EPR measurements. We thank Gizem Bilgin for assisting with data analysis. C.N. thanks the Alexander von Humboldt Foundation for a Feodor Lynen fellowship and the Australian Research Council (ARC) for a Discovery Early Career Research Award (DE190100015). G.O. gratefully acknowledges an ARC Laureate Fellowship (project FL170100019) and research support (project CE200100012). Financial support by the Australian Research Council for project funding (DP200100348) is gratefully acknowledged.

Conflicts of interest

There are no conflicts to declare.

Notes and references

1 a) B. Chen, Q. Liu, A. Popowich, S. Shen, X. Yan, Q. Zhang, X.-F. Li, M. Weinfeld, W. R. Cullen and X. C. Le, *Metallomics*, 2015, **7**, 39-55; b) S. Shen, X. F. Li, W. R. Cullen, M. Weinfeld and X. C. Le, *Chem. Rev.*, 2013, **113**, 7769-7792.

2 a) S. R. Adams, R. E. Campbell, L. A. Gross, B. R. Martin, G. K. Walkup, Y. Yao, J. Llopis and R. Y. Tsien, *J. Am. Chem. Soc.*, 2002, **124**, 6063-6076; b) B. A. Griffin, S. R. Adams and R. Y. Tsien, *Science*, 1998, **281**, 269-272.

3 a) C. Huang, T. Jia, M. Tang, Q. Yin, W. Zhu, C. Zhang, Y. Yang, N. Jia, Y. Xu and X. Qian, *J. Am. Chem. Soc.*, 2014, **136**, 14237-14244; b) Y. Wang, X.-F. Yang, Y. Zhong, X. Gong, Z. Li and H. Li, *Chem. Sci.*, 2016, **7**, 518-524; c) P. Wilson, A. Anastasaki, M. R. Owen, K. Kempe, D. M. Haddleton, S. K. Mann, A. P. Johnston, J. F. Quinn, M. R. Whittaker, P. J. Hogg and T. P. Davis, *J. Am. Chem. Soc.*, 2015, **137**, 4215-4222.

4 a) A. D. Gossert and W. Jahnke, *Prog. NMR Spectrosc.*, 2016, **97**, 82-125; b) C. Nitsche and G. Otting, *Curr. Opin. Struct. Biol.*, 2018,

48, 16-22; c) K. Takeuchi and G. Wagner, *Curr. Opin. Struct. Biol.*, 2006, **16**, 109-117.
5 a) W. Becker, L. A. Adams, B. Graham, G. E. Wagner, K. Zangger, G. Otting and C. Nitsche, *J. Biomol. NMR*, 2018, **70**, 211-218; b) W. N. Chen, K. V. Kuppan, M. D. Lee, K. Jaudzems, T. Huber and G. Otting, *J. Am. Chem. Soc.*, 2015, **137**, 4581-4586; c) S. Jabar, L. A. Adams, Y. Wang, L. Aurelio, B. Graham and G. Otting, *Chem. Eur. J.*, 2017, **23**, 13033-13036; d) C. T. Loh, L. A. Adams, B. Graham and G. Otting, *J. Biomol. NMR*, 2018, **71**, 287-293; e) Q. Liu, Q.-t. He, X. Lyu, F. Yang, Z.-l. Zhu, P. Xiao, Z. Yang, F. Zhang, Z.-y. Yang, X.-y. Wang, P. Sun, Q.-w. Wang, C.-x. Qu, Z. Gong, J.-y. Lin, Z. Xu, S.-l. Song, S.-m. Huang, S.-c. Guo, M.-j. Han, K.-k. Zhu, X. Chen, A. W. Kahsai, K.-H. Xiao, W. Kong, F.-h. Li, K. Ruan, Z.-j. Li, X. Yu, X.-g. Niu, C.-w. Jin, J. Wang and J.-p. Sun, *Nat. Commun.*, 2020, **11**, 4857.
6 a) C. Nitsche and G. Otting, *Prog. NMR Spectrosc.*, 2017, **98-99**, 20-49; b) G. Pintacuda, M. John, X. C. Su and G. Otting, *Acc. Chem. Res.*, 2007, **40**, 206-212; c) W.-M. Liu, M. Overhand and M. Ubbink, *Coord. Chem. Rev.*, 2014, **273-274**, 2-12.
7 C. Nitsche and G. Otting, in *Paramagnetism in Experimental Biomolecular NMR*, The Royal Society of Chemistry, 2018, pp. 42-84.
8 a) X.-C. Su, B. Man, S. Beeren, H. Liang, S. Simonsen, C. Schmitz, T. Huber, B. A. Messerle and G. Otting, *J. Am. Chem. Soc.*, 2008, **130**, 10486-10487; b) B. Man, X.-C. Su, H. Liang, S. Simonsen, T. Huber, B. A. Messerle and G. Otting, *Chem. Eur. J.*, 2010, **16**, 3827-3832; c) X. Jia, A. Maleckis, T. Huber and G. Otting, *Chem. Eur. J.*, 2011, **17**, 6830-6836; d) Q.-F. Li, Y. Yang, A. Maleckis, G. Otting and X.-C. Su, *Chem. Commun.*, 2012, **48**, 2704-2706.
9 C. S. Palmer and R. Adams, *J. Am. Chem. Soc.*, 1922, **44**, 1356-1382.
10 D. S. Ziegler, B. Wei and P. Knochel, *Chem. Eur. J.*, 2019, **25**, 2695-2703.
11 P. Plateau and M. Gueron, *J. Am. Chem. Soc.*, 1982, **104**, 7310-7311.
12 D. J. Cline, C. Thorpe and J. P. Schneider, *J. Am. Chem. Soc.*, 2003, **125**, 2923-2929.
13 H. L. Frericks Schmidt, L. J. Sperling, Y. G. Gao, B. J. Wylie, J. M. Boettcher, S. R. Wilson and C. M. Rienstra, *J. Phys. Chem. B*, 2007, **111**, 14362-14369.
14 S. Vijay-Kumar, C. E. Bugg and W. J. Cook, *J. Mol. Biol.*, 1987, **194**, 531-544.
15 C. Schmitz, M. J. Stanton-Cook, X. C. Su, G. Otting and T. Huber, *J. Biomol. NMR*, 2008, **41**, 179-189.
16 C. Nitsche, M. C. Mahawaththa, W. Becker, T. Huber and G. Otting, *Chem. Commun.*, 2017, **53**, 10894-10897.