WILEY-VCH

COMMUNICATION

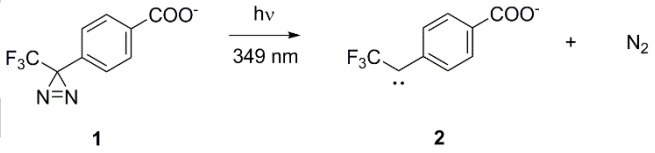
Carbene footprinting reveals binding interfaces of a multimeric membrane spanning protein

Lucio Manzi,[a] Andrew S. Barrow,‡[a] Jonathan T.S. Hopper,[b] Renata Kaminska,[c] Colin Kleanthous,[c] Carol V. Robinson,[d] John E. Moses,*‡ [a] and Neil J. Oldham*[a]

Abstract: Mapping the interaction sites between membrane spanning proteins is a key challenge in structural biology. In this study a carbene footprinting approach is developed and applied to identify the interfacial sites of a trimeric, integral membrane protein, OmpF, solubilised in micelles. The diazirine-based footprinting probe is effectively sequestered by, and incorporated into, the micelles leading to efficient labelling of the membrane-spanning regions of the protein upon irradiation at 349 nm. Areas associated with protein-protein interactions between the trimer subunits remained unlabelled, thus revealing their location.

Genes coding for membrane proteins comprise 20-30 % of the genome of most organisms,[1] and integral membrane proteins (IMPs), which are permanently attached to the lipid bilayer, perform many key roles in cellular signalling, reception, transport and metabolism. Despite their undoubted importance in basic cell biology, and as therapeutic targets for drugs,[2] our understanding of the structure of membrane proteins significantly lags behind that of their soluble counterparts. Structural characterization of IMPs by high-resolution techniques such as X-ray crystallography is complicated by their hydrophobic nature and resulting propensity to aggregate and precipitate. The use of detergents, or similar agents, is essential for solubilisation of most membrane proteins, but the presence of micelles can often lead to difficulties in crystallization. As a consequence, membrane proteins are underrepresented in structural databases.[3] Methods such as cryo-EM and solid state NMR are finding wider application in the study of membrane protein structure, where they show significant promise.[4] Nonetheless, there is a paucity of techniques for probing the structure of IMPs embedded in a bilayer or micelle.

The Robinson group has pioneered the use of mass spectrometry (MS) for the interrogation of membrane protein structure and interactions.[5] By electrospraying micelle-solubilised proteins, and removing the bound detergent through gas-phase collisional activation, the group has succeeded in detecting highly complex multi-subunit assemblies such as ATP synthase.[6] MS, being based on the measurement of mass-to-charge (*m/z*), is ideally suited for determining subunit stoichiometry and for detecting small molecule ligand binding (including endogenous lipids and drug molecules).[7] Complimentary methodology that allows the binding interfaces between subunits of multimeric IMPs to be determined with similar speed and sensitivity would provide a valuable new tool in the study of these analytically-demanding biomolecules.



1 2

Scheme 1. Photolysis of diazirine **1** forms the carbene footprinting probe **2**.

The use of reactive probes, such as radicals or carbenes, to covalently label proteins for mapping, or 'footprinting' is attracting increased attention in structural studies. This approach uses labelling to identify accessible protein surfaces in order to probe binding sites or to study protein folding. Hydroxyl radical labelling is a very promising method for protein footprinting, especially in the form of fast photochemical oxidation of proteins (FPOP) developed by Gross and co-workers.[8] A few studies have applied oxidative methodology to IMPs, where significant labelling of amino acid sidechains in solvent accessible regions of the proteins has been detected.[9] Whilst this mode of labelling provides information on the cytosolic and extracellular domains, as well as any polar pores, it is not best suited for mapping the hydrophobic membrane-spanning regions of IMPs. Recently, we have reported the use of photoactivatable aryl diazirine **1** for the efficient footprinting of soluble proteins.[10] Given the amphipathic nature of the probe, we postulated that it may interact with detergent molecules and footprint outer transmembrane regions of micelle-solubilised IMPs to provide data complementary to that obtained by hydroxyl radical labelling.

[a] L. Manzi, Dr. A. S. Barrow, Prof. J. E. Moses, Prof. N. J. Oldham
School of Chemistry, University of Nottingham
University Park, Nottingham NG7 2RD (UK)
E-mail: neil.oldham@nottingham.ac.uk
[b] Dr. J. T. S. Hopper
OMass Technologies Ltd., Centre for Innovation and Enterprise, Begbroke Science Park, Woodstock Road, Oxford, OX5 1PF (UK)
[c] Prof. Colin Kleanthous
Department of Biochemistry, University of Oxford, South Parks Road, Oxford OX1 3QU (UK)
[d] Prof. Dame Carol V. Robinson
Chemistry Research Laboratory, University of Oxford, South Parks Road, Oxford OX1 3QZ (UK)
[‡] Current address: La Trobe Institute for Molecular Science, La Trobe University, Melbourne, Victoria 3086 (Australia)

Supporting information for this article is given via a link at the end of the document.((Please delete this text if not appropriate))

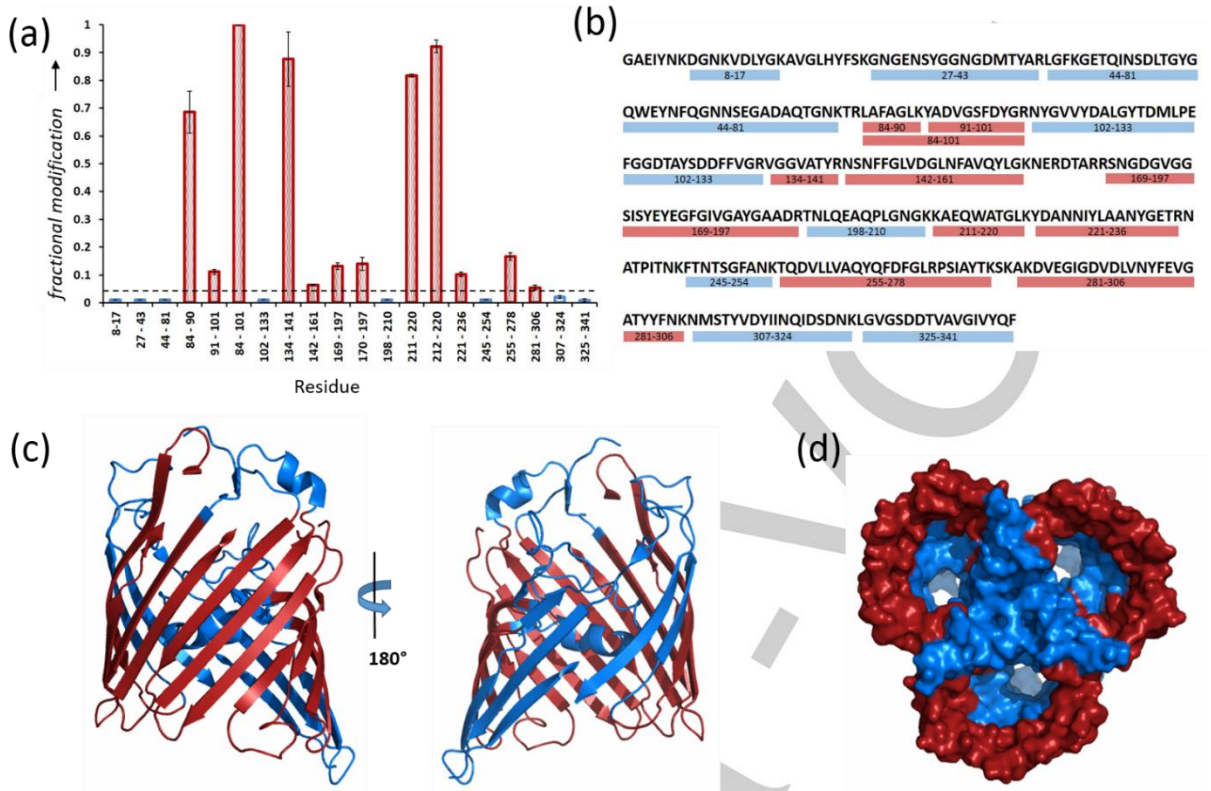


Figure 1. Carbene footprinting of OmpF using aryl diazirine probe **1**, with significantly (> 0.05) and insignificantly (< 0.05) labelled peptides represented in red and blue, respectively. (a) Fractional modification of each OmpF tryptic peptide by carbene **2** (peptide 84-101 arises due to a missed cleavage). (b) Amino acid sequence of OmpF, with associated tryptic peptides. (c) Monomer structure of OmpF (taken from PDB file 3pox) showing labelled and masked regions of the protein. (d) Structure of the OmpF trimer revealing the presence of carbene labelling on the outer surface of the structure, and the masking of the interfacial regions.

Here we report the successful application of diazirine **1** in footprinting the *Escherichia coli* outer membrane porin protein OmpF, thereby demonstrating the utility of carbene-based footprinting in characterising IMPs.

Responsible for the passive diffusion of a range of small molecules across the bacteria's outer membrane, the OmpF β-barrel IMP forms a well characterised trimer assembly in bilayers and micelles. We show that **1** inserts into micelles and efficiently labels accessible hydrophobic transmembrane regions of OmpF within the micelle. The absence of labelling from subunit interfacial binding regions in trimeric OmpF demonstrates that the probe is able to reveal these areas and thereby map protein-protein interaction sites in membrane proteins.

Laser irradiation (349 nm) of aryl diazirine probe **1** in the presence of OmpF (solubilised in 1 % n-octyl β-glucoside, OG) resulted in significant carbene labelling of the porin (Figure 1a). Tryptic peptides corresponding to 86 % coverage of the protein sequence were identified (Figure 1b, Figure S1), and their fractional modification (f_m) by the aryl carbene **2** determined. Of the twenty peptides detected, twelve showed significant labelling (f_m > 0.05, assuming identical detector response for labelled and unlabelled peptides) by the probe. Of these, peptides 84-90, 84-101, 134-141, 211-220, and 212-220 exhibited very high carbene modification, with f_m in excess of 0.6. In contrast, seven of the twenty peptides showed no detectable labelling, and one (307-324) possessed low labelling (f_m = 0.02). Mapping these unlabelled peptides onto the (monomer) structure of OmpF revealed their localization to essentially one face of the β-barrel structure (Figure 1c). Comparison of the labelling pattern seen on the protein with the trimeric arrangement of subunits in the OmpF complex showed a clear correlation between outer surface regions and the presence of significant carbene attachment (Figure 1d). The buried core of the trimer interface is dominated by hydrophobic residues that are present within unlabelled peptides (8-17, 44-81, and 325-341). Thus, in principle, using f_m as a reporter, the interfacial areas of the protein surface could be determined.

Analysis of OmpF footprinting by diazirine **1** at the peptide-level was found to be sufficient to map the buried and surface accessible regions of the porin in OG; however to gain insight into the location of modification within the labelled peptides, MS/MS was performed on their precursor ions. Example data is provided in Figure 2. This reveals the location of labelling on tryptic peptide 211-220. Plots of absolute modification at the sub-peptide for all labelled peptides are presented in Figure S2, and the MS/MS spectra in Figure S3. These data show that labelling is mainly located on hydrophobic residues: an observation consistent with the physicochemical properties of the probe, and our previous observations of footprinting soluble proteins. Although it is

important to point out that the probe is capable of labelling hydrophilic residues: particularly the basic Lys, Arg and His.

Highlighting, on the structure of OmpF, the positions of carbene labelling identified by MS/MS analysis shows a higher resolution view of protein modification by **2** (Figure 3). Using data from the OMP database, which models the membrane bilayer in a way consistent with experimental results for OmpF,[11] the position of the porin in a membrane can be observed. It is readily apparent that residues in the extensive extracellular portion of the protein remain largely unlabelled, whereas those in the accessible membrane spanning region and cytosolic loops are significantly modified.

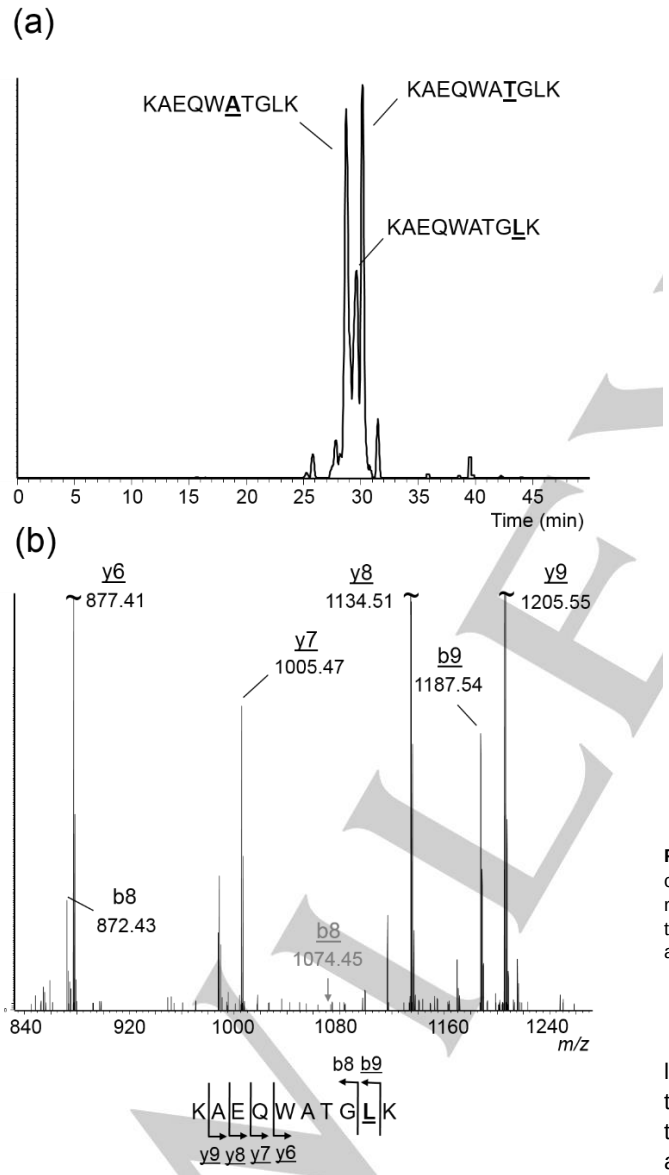


Figure 2. Example LC-ESI-MS/MS analysis for OmpF labelled with carbene **2** showing (a) an extracted in chromatogram for *m/z* 667.8, the $[M+2H]^{2+}$ ion for labelled tryptic peptide 211-220, and (b) a portion of the MS/MS spectrum of this ion taken from the central chromatographic peak. Ions exhibiting carbene attachment are underlined. The presence of labelling on b9, and its absence from b8 places the label on L219.

Examination of the 15 single amino acid-resolved positions that exhibited significant labelling (a threshold of 0.01 was applied to the absolute per residue labelling to indicate significance, as below this value quantitation became imprecise, and often indistinguishable from zero.) revealed that all but one of the side-chains were located on the outer surface of the β–barrel. This suggested that modification by the carbene probe was favoured at sites exposed to the detergent micelle. Ile-315 was the only residue with its side-chain inside the pore to which labelling could be definitively attributed, although there may well be others in cases where single amino acid resolution was not achieved. Moreover, we cannot entirely rule-out labelling on the main-chain, rather than side-chain of the amino acid.

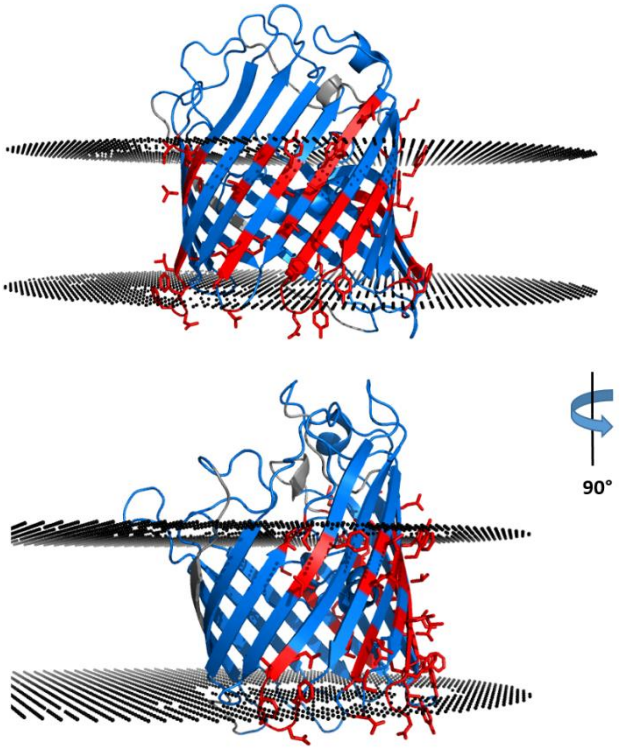


Figure 3. Sub-peptide level labelling of OmpF (shown as the monomer) with carbene **2** showing that modification is largely limited to the membrane spanning regions of the protein. The position of OmpF within the lipid bilayer is taken from the OMP database. Labelled residues are shown in red, unlabelled in blue and areas of no coverage in grey.

The amphiphilic nature of diazirine **1**, together with the high level of labelling observed for some of the OmpF peptides led us to speculate that the probe may be inserting into the OG micelle to produce a mixed micelle: a property that would be highly advantageous in an IMP-labelling probe. To investigate this possibility aqueous solutions of **1** were irradiated in the presence and absence of OG, and the products analysed by LC-MS. Photolysis of **1** alone produced a relatively simple mixture (Figure S4a). Significant peaks were attributed to (i) alcohol **3** (see Figure S4 for structures), which can result from the direct insertion of carbene **2** into the O-H bond of water, (ii) **4**, which can form from

reaction of carbene **2** with diazirine **1**, and (iii) a small amount of unreacted diazirine **1**. Additionally, a peak exhibiting a mass consistent with the peroxide **5** was seen (Figure S5). Interestingly, the presence of **4** suggests some level of interaction between molecules of **1** in aqueous solution.

When diazirine **1** was irradiated in a solution of 1 % OG the product profile changed significantly. LC peaks due to unreacted diazirine **1**, and – in particular – the dimer **4** were seen to decrease, and new peaks between 4 and 5 min were visible (Fig S4b). Their mass spectra (Figure S5) were consistent with reaction products, **6**, between carbene **2** and OG ($[M-H]^-$ = 493.2). Identification was confirmed by accurate mass and MS/MS analysis (measured for $[M+H]^+$ 495.2196±0.0001 (n = 10) required for $C_{23}H_{34}F_3O_8$ 495.2200, error 0.8 ppm; see Figure S6). The MS/MS spectra for all chromatographic peaks of **6** showed identical product ions resulting from the loss of a dehydrated glucose neutral. Carbene labelling was only observed on MS/MS product ions corresponding the octanol portion of the glucoside (Figure S6b). This finding strongly suggested that the carbene reacted with the alkyl chain portion of OG; presumably by C-H bond insertion. The presence of four partially resolved chromatographic peaks, all of which exhibited very similar MS/MS spectra, further suggested that labelling occurred at several sites along the octyl chain of OG (we were unable to find diagnostic ions in the MS/MS spectrum that unequivocally located the position of carbene attachment along the chain). Taken together, these observations are consistent with a mixed micelle model where the hydrophobic trifluoromethylaryldiazirine portion of **1** is aligned with the hydrophobic octyl chain of OG and the polar carboxylate group of **1** sits at the micelle-solution interface in proximity to the glucosyl groups.

Further evidence for the interaction between OG and diazirine probe **1** was obtained by labelling studies with the small soluble protein ubiquitin (Ub) in the presence and absence of OG. Without the addition of OG, Ub was found to be highly modified by the carbene (an average of 1.4 labels per Ub molecule at 10 mM probe concentration). With the addition of 1 % OG, however, the average number of labels fell to 0.16 per Ub molecule. Increasing the concentration of probe **1** to 25 mM and 50 mM in the presence of OG did increase Ub labelling to 0.29 and 0.43 labels per molecule, respectively, but it still did not reach the levels seen in the absence of the detergent.

These results demonstrate that the labelling of a soluble protein is inhibited by the presence of OG, and indicate that the diazirine probe is being sequestered by the detergent micelles, and made less available in solution.

Finally, we compared the labelling of OmpF by **2** with that obtained using the diazirine photoleucine, which has previously been used to label soluble proteins as a footprinting agent,[12] and also for incorporation into the proteome as a photo-cross-linking tool for detecting membrane protein interactions.[13] As a footprinting agent, the latter was found not to be nearly so efficient at labelling OmpF as **2** (see Figure S7).

In summary, we have shown that carbene **2**, derived from diazirine **1** upon irradiation at 349 nm, is an efficient labelling tool for footprinting the integral membrane protein OmpF, solubilised in OG detergent. Footprinting reveals high levels of labelling on hydrophobic membrane-spanning regions of the protein, but little or no labelling at the trimer interface sites, or on the extracellular portion of the protein. This makes the technique highly complementary to hydroxyl radical footprinting, which tends to label solvent exposed areas. Straightforward analysis of MS data at the peptide level is shown to provide an adequate degree of resolution to map successfully the trimer interface site of OmpF. Furthermore, MS/MS interrogation is demonstrated to deliver a level resolution that approaches individual amino acids. Evidence is presented to show incorporation of the probe into micelles, which represents a highly favourable property for a probe used in membrane protein footprinting.

Experimental

Experimental details and supporting information are available via the following link. Supporting Information.

The mass spectrometry data have been deposited to the ProteomeXchange Consortium via the PRIDE[14] partner repository with the dataset identifier PXD007207.

Acknowledgements

LM, ASB, JEM and NJO are grateful to the University of Nottingham for funding, and to T. Wright for facilitating preliminary laser irradiation experiments. CK acknowledges BBSRC for the award BB/L021234/1.

Keywords: carbene footprinting • mass spectrometry • integral membrane protein • protein structure • protein labelling

[1] a) A. Krogh, B.R. Larsson, G. Von Heijne, E.L.L. Sonnhammer, *J. Mol. Biol.* **2001**, *305*, 567-580; b) K. Liszewski, *Genet. Eng. Biotech. N*. **2015**, *35*, 16-17.

[2] a) J.P. Overington, B. Al-Lazikani, A.L. Hopkins, *Nat. Rev. Drug Discov.* **2006**, *5*, 993-996; b) R. Lappano, M. Maggiolini, *Nat. Rev. Drug Discov*. **2011**, *10*, 47-60.

[3] W.A. Hendrickson, *Nat. Struct. Mol. Biol*. **2016**, *23*, 464-467.

[4] a) K.N. Goldie, P. Abeyrathne, F. Kebbel, M. Chami, P. Ringler, H. Stahlberg, *Methods Mol. Biol*. **2014**, *1117*, 325-341; b) K.R. Vinothkumar, *Curr. Opin Mol. Biol*. **2015**, *33*, 103-114; c) C. Kang, Q. Li, *Curr. Opin. Chem. Biol*. **2011**, *15*, 560-569; d) S.A. Shahid, B. Bardiaux, W.T. Franks, L. Krabben, M. Habeck, B-J. van Rossum, D. Linke, *Nat. Methods*, **2012**, *9*, 1212-1217.

[5] a) N.P. Barrera, N. Di Bartolo, P.J. Booth, C.V. Robinson, *Science*, **2008**, *321*, 234-246; b) N.P. Barrera, C.V. Robinson, *Annu. Rev. Biochem*. **2011**, *80*, 247-271; c) A. Laganowsky, E. Reading, J.T.S. Hopper, C.V. Robinson, *Nat. Protoc*. **2013**, *8*, 639-651.

[6] M. Zhou, N. Morgner, N.P. Barrera, A. Politis, S.C Isaacson, D. Matak-Vinković, T. Murata, R.A. Bernal, D. Stock, C.V Robinson, *Science*, **2011**, *334*, 380-3855.

[7] C. Bechara, C.V. Robinson, *J. Am. Chem. Soc*. **2015**, *137*, 5240-5247; b) J. Gault, J.A.C. Donlan, I. Liko, J.T.S. Hopper, K. Gupta, N.G. Housden, W.B. Struwe, M.T. Marty, T. Mize, C. Bechara, Y. Zhu, B. Wu, C. Kleanthous, M. Belov, E. Damoc, A. Makarov, C.V. Robinson, *Nat. Methods*, **2016**, *13*, 333-336.

[8] D. M. Hambly, M.L. Gross, *J. Am. Soc. Mass Spectrom*. **2005**, *16*, 2057-2063.

[9] a) Y. Pan, B.B. Stocks, L. Brown, L. Konermann, *Anal. Chem*. **2009**, *81*, 28-35; b) T.E. Angel, S. Gupta, B. Jastrzebska, K. Palczewski, M.R.

Chance, *Proc. Natl. Acad. Sci. USA.* **2009**, *106*, 14367-14372; c) Y. Zhu, T. Guo, J.E. Park, X. Li, W. Meng, A. Datta, M. Bern, S.K. Lim, S.K. Sze, *Mol. Cell Proteom.* **2009**, *8*, 1999-2010; d) T.G. Watkinson, A.N. Calabrese, J.R. Ault, S.E. Radford, A.E. Ashcroft, *J. Am. Soc. Mass Spectrom.* **2016**, *28*, 35-40; e) Y. Lu, H. Zhang, D.M. Niedzwiedzki, J. Jiang, R.E. Blankenship, M.L. Gross, *Anal. Chem.* **2016**, *88*, 8827-8834.

[10] L. Manzi, A.S. Barrow, D. Scott, R. Layfield, T.G. Wright, J.E. Moses, N.J. Oldham, *Nat. Commun*. **2016**, *7*, 13288.

[11] a) M.A Lomize, A.L. Lomize, I.D. Pogozheva, H.I. Mosberg, *Bioinformatics* **2006**, *22*, 623-625; b) E. Pebay-Peyroula, R.M. Garavito, J.P. Rosenbusch, M. Zulauf, P.A Timmins, *Structure* **1995**, *3*, 1051-1059; c) A.H O′Keeffe, J.M. East, A.G. Lee, *Biophys. J.* **2000**, *79*, 2066-2074.

[12] a) C.C. Jumper, D.C. Schriemer, *Anal. Chem*. 2011, *83*, 2913–2920; b) C.C. Jumper, R. Bomgaarden, J. Rogers, C. Etienne, D.C. Schriemer, *Anal. Chem.* **2012**, 84, 4411–4418.

[13] M. Suchanek, A. Radzikowska, C. Thiele, *Nat. Methods* **2005**, 2, 261.

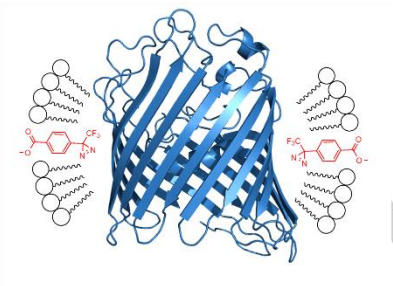
[14] J.A Vizcaíno, A. Csordas, N. del-Toro, J.A. Dianes, J. Griss, I. Lavidas, G. Mayer, Y. Perez-Riverol, F. Reisinger, T. Ternent, Q.W. Xu, R. Wang, H. Hermjakob, *Nucleic Acids Res.* **2016**, *44*, D447-D456.

Entry for the Table of Contents (Please choose one layout)

Layout 1:

COMMUNICATION

Diazirine probe 1 inserts into micelles to label integral membrane proteins efficiently and reveal binding interfaces.



Lucio Manzi, Andrew S. Barrow, Jonathan T.S. Hopper, Colin Kleanthous, Carol V. Robinson, John E. Moses, and Neil J. Oldham**

Page No. – Page No.

Title