Carbene footprinting accurately maps binding sites in protein–ligand and protein–protein interactions

Lucio Manzi1,*, Andrew S. Barrow1,*, Daniel Scott1,2, Robert Layfield2, Timothy G. Wright1, John E. Moses1 & Neil J. Oldham1

Specific interactions between proteins and their binding partners are fundamental to life processes. The ability to detect protein complexes, and map their sites of binding, is crucial to understanding basic biology at the molecular level. Methods that employ sensitive analytical techniques such as mass spectrometry have the potential to provide valuable insights with very little material and on short time scales. Here we present a differential protein footprinting technique employing an efficient photo-activated probe for use with mass spectrometry. Using this methodology the location of a carbohydrate substrate was accurately mapped to the binding cleft of lysozyme, and in a more complex example, the interactions between a 100 kDa, multi-domain deubiquitinating enzyme, USP5 and a diubiquitin substrate were located to different functional domains. The much improved properties of this probe make carbene footprinting a viable method for rapid and accurate identification of protein binding sites utilizing benign, near-UV photoactivation.
Interactions between proteins, and between proteins and small molecules are at the heart of practically all biochemical processes\(^1\). Methods that allow detection and mapping of these interactions are consequently essential both in understanding basic biology at the molecular level, and in developing drug discovery strategies based on rational design. Frequently used techniques, such as X-ray crystallography and nuclear magnetic resonance (NMR) spectroscopy deliver high-resolution structural data, but are often time-consuming and require relatively large amounts of material\(^2\). In contrast, methods that employ mass spectrometry (MS) to detect and map protein interactions are rapid and sensitive, but produce relatively low-resolution data\(^3\). MS-based strategies that provide reliable and high-resolution mapping of protein interaction sites are, therefore, highly sought after\(^4\). Towards this goal, a number of approaches have been employed, including hydrogen-deuterium exchange (HDX), chemical modification and cross-linking, and rapid footprinting strategies such as hydroxyl radical or carbene labelling. HDX, one of the first and most widely used methods, probes the mass shift associated with deuterium uptake\(^5\). This procedure has provided promising results, particularly with stable complexes, but suffers from two main drawbacks: unwanted (and variable) back-exchange of deuterium under quenching/digestion/analysis conditions\(^6\), and H/D scrambling under collisional MS/MS activation\(^7\). The latter problem arises when peptides are fragmented, using thermal activation methods, to gain sub-peptide-level information. Scrambling of the label can be reduced by non-thermal activation\(^8\), but the general labile nature of the deuterium label complicates data interpretation in HDX experiments. Chemical modification and cross-linking are, again, widely utilized and have been successfully used to map protein structure and interactions\(^9\). The relatively low resolution of the data (the labelling density can be low, and reliant on the chemical reactivity of specific amino acid sidechains), and the often slow rate of chemical reaction leading to labelling/crosslinking, are drawbacks of this approach. For stable complexes this may not be problematic, but for more dynamic systems the temporal resolution afforded by the reaction is limited. Hydroxyl radical labelling is a very promising method for protein footprinting\(^10\), especially in the form of fast photochemical oxidation of proteins (FPOP) developed by the Gross lab\(^11\), and a similar approach described by Sze and coworkers\(^12\). Recently, the methodology has been employed to explore the free-energy folding landscape of the bacterial immunity protein Im7 (ref. 18). Although powerful, drawbacks of the strategy include the need to expose the protein to hydrogen peroxide, which may cause oxidative damage, and the use of middle-ultraviolet laser irradiation (typically 266 nm) to photolyse the hydrogen peroxide to produce OH, which lies within the absorption region of the aromatic amino acids, and hence may be damaging to protein structure. More recently, a FPOP variant has been produced based on iodine photochemical labelling, but this still operates at 266 nm\(^19\).

Diazirine-based footprinting agents, in contrast, offer the potential to use near-ultraviolet wavelengths (\(\lambda \approx 350\) nm)–outside the absorbance window of amino acids–to generate highly reactive carbones, which rapidly label the surface of a protein\(^20\,21\). It should be noted that this generic strategy is quite distinct from that of photoaffinity labelling, where a specific photoactive ligand is employed, and which therefore requires a different probe for each system studied\(^22\). The use of methane gas itself has been reported, but its utility is limited by low solubility in water\(^20\). Recently, Schriemer and colleagues\(^23\) described the use of commercially available photoleucine 1 and 4,4-azipentanoic acid for diazirine-based protein carbene footprinting. The former has been used to map the calmodulin-M13 interaction site\(^24\), and the calcium-induced conformational change in calmodulin\(^25\). However, high concentrations (100 mM) of the probes were required to obtain reasonable levels of protein labelling. In investigating other protein systems, we found the labelling efficiency of photoleucine 1 to be highly protein dependent, and that labelled peptides exhibited, in some cases, sub-optimal behaviour during MS/MS fragmentation. Despite this, we believed that the carbene-labelling strategy offered considerable promise owing to its use of benign, near-ultraviolet photoactivation by compact low-cost laser systems, and clear mass shifts upon protein labelling. To provide a platform for high-resolution mapping of protein interactions by MS, we have designed and synthesized a bespoke diazirine-based footprinting agent for differential protein labelling experiments. This probe labels proteins with much higher efficiency than photoleucine at both lower concentrations, and lower irradiation times. We have successfully used the probe to map the lysozyme substrate-binding site accurately at high resolution, and identify changes in intramolecular interactions associated with binding. In a second example of the technology we have identified the substrate-binding sites of a large, multi-domain deubiquitinating protease and used the results to clarify uncertainties raised by the X-ray crystal structure.

Results

Design and synthesis of a new footprinting probe. To produce an efficient carbene-based probe for protein footprinting, we elected to incorporate an aromatic ring to provide a non-polar surface for interaction with hydrophobic patches on the protein, and an ionic group both to associate with polar amino acid sidechains and to improve water solubility of the probe. The trifluoromethylaryl diazirine 2, based on a sodium benzoate core was considered an appropriate choice, where the presence of the –CF\(_3\) group has the benefit of reducing side reactions associated with the diazoisomer\(^26\). The free acid form of 2 has been synthesized some years previously, for incorporation into photoaffinity probes, but not for use as a footprinting tool\(^27\). The synthesis of aryldiazirine 2 was achieved following a modified literature method (see the ‘Methods’ section, and Supplementary Figs 1–6) in five steps from 4-bromotoluene. It has been shown that, once formed, the lifetime of aryl trifluoromethylcarbenes, such as 3, is of the order of a few ns (ref. 28), making 2 a promising reagent for fast protein labelling. We note that at the time of publication the free acid of 2 is commercially available.

Footprinting efficiency of aryldiazirine 2. A comparison of the protein labelling efficiency of the aryldiazirine salt 2 with that of photoleucine 1 showed that, for a range of proteins, 2 exhibited far superior properties to the existing probe (Fig. 1). Photo-activated labelling (see the ‘Methods’ section) with diazirine 1, required 100 mM concentrations of the probe and 16 s irradiation times to give modest levels of labelling of the proteins tested, whereas aryldiazirine 2 footprinted with much higher efficiency at only 10 mM and 1–4 s irradiation, following flash freezing (Fig. 1c). This property was found to be reproducible for a range of peptides and proteins tested, making 2 a much more efficient footprinting tool compared with 1 (Fig. 1c and Supplementary Figs 7–12). This may be a result of the increased hydrophobicity of 2, and its lack of zwitterion character. Indeed, the only protein efficiently labelled by 1 was calmodulin. The degree of labelling for the remaining proteins using 1 was found to be too low for footprinting purposes.

An important consideration when using probes such as 2 to study protein interactions is the potential effect of the diazirine on protein structure. Although we proposed to use a differential
approach, where a comparison of labelling in the presence and absence of a binding partner should take into account minor structural effects, we wished to establish that the probe caused no significant perturbation to protein structure. To provide evidence that the presence of $2$ in mM quantities was not detrimental to the native structure of the enzyme hen egg white lysozyme (HEWL), enzyme activity assays were performed in the presence and absence of the probe. No change in enzyme activity was detected upon addition of $10$ mM $2$ to HEWL, indicating that the enzyme’s structure was unaffected by the diazirine (Supplementary Fig. 13). These encouraging results led us to footprint differentially HEWL in the presence and absence of a penta-$N$-acetylchitopentaose ($\text{NAG}_5$) substrate to highlight regions of HEWL masked by NAG$_5$. Interactions between enzyme and substrate in this system have been studied in detail, and a large number of high-resolution structures are available$^{29}$, thus it is an excellent model to test the efficacy of $2$ as part of a carbene labelling strategy.

Mapping the $\text{NAG}_5$ binding site of HEWL. Differential footprinting of HEWL by photoactivation of $2$ was performed in the presence and absence of NAG$_5$ (see the ‘Methods’ section). Irradiation of the sample following flash freezing had the dual benefit of halting enzyme turn-over and reducing diffusion of the probe. Indeed, we found freezing a necessary step for efficient labelling. Following proteolysis of HEWL with either AspN or trypsin, the resulting peptides (representing $79\%$ sequence coverage) were analysed by liquid chromatography-tandem MS (LC-MS/MS) to quantify the degree of labelling at each position (see the ‘Methods’ section, Supplementary Figs 14–26, and Supplementary Tables 1–38). For some peptides, chromatographic separation of isomers derived from labelling at different positions was observed; in these cases, spectra were combined over all signals for each peptide to give an overall average. In general, the probe was found to label basic and hydrophobic residues preferentially, which may be expected from its structure. The results (Fig. 2) showed that the extent of carbene attachment at some residues was significantly ($\text{Student’s } t\text{-test, } P<0.05$) reduced in the presence of the NAG$_5$ ligand (Fig. 2a). Pleasingly, when mapped onto the surface of HEWL these positions were found to be located in and around the enzyme binding cleft (Fig. 2b), including the well-characterized NAG$_n$-interacting site W62. More distal regions of the protein showed no significant difference in labelling in the presence or absence of NAG$_5$. Depending upon the efficiency of MS/MS fragmentation, labelling was mapped to single amino acid residue resolution in some regions of the protein. Two AspN peptides, corresponding to K1-L17 and D49-T52 of HEWL were not detected with sufficient intensity to allow accurate determination of labelling. In addition to the masking of residues in the HEWL binding cleft, two additional structural details were revealed by carbene footprinting with $2$. First, a reproducible and significant increase in labelling at S60/R61 was observed on addition of the NAG$_5$.  

**Figure 1 | Protein labelling efficiency of diazirine probe 2.** (a) Structures of the footprinting probes used in this study, (b) electrospray ionization-MS of HEWL ($9\%$ charge state) showing the high efficiency of labelling achieved with the aryldiazirine $2$ (lower spectrum) over photoleucine $1$ (upper spectrum), (c) extent of labelling of a range of proteins with $1$ (100 mM, 16 s irradiation), and $2$ (10 mM, 4 s irradiation (*1 s in the case of calmodulin)).
ligand (Fig. 2a,b). This apparently counterintuitive result can be explained by detailed examination of HEWL’s structure in the bound and unbound states. In the absence of substrate, R61 participates in an ionic bond with D48 (Fig. 2c). This intramolecular interaction is disrupted upon NAG5 binding, making R61 more accessible for labelling by 2. Thus, carbene footprinting with our probe is able to reveal established subtle structural changes associated with binding events. Second, a clear reduction in labelling was seen at M105, which—although not directly part of the binding cleft itself—is known to be affected by substrate binding. Termed the hydrophobic box, the region of HEWL around M105 is known to become more rigid when a NAGn molecule is bound30, which is highly likely to reduce the carbene labelling efficiency at this site.

Mapping the diubiquitin binding site of USP5. Following our successful proof-of-concept study mapping the model HEWL-NAG5 protein–ligand system, we sought to apply carbene footprinting to a less well characterized, and more complex protein interaction. The deubiquitinating enzyme ubiquitin specific protease 5 (USP5), of current interest in our laboratory, provides an excellent example of such a protein. The enzyme is responsible for the disassembly of unanchored poly-ubiquitin (poly-Ub) chains31, which are isopeptide-linked polymers of ubiquitin that arise in the cell during proteasomal degradation of poly-Ub modified proteins, or via de novo assembly. Recent work has identified important physiological roles for unanchored poly-Ub in regulating the 26S proteasome32, acting as second messengers in NF-κB signalling33, and controlling innate immune response signalling34. As a result, USP5 has attracted considerable interest as a potential drug target. This 100 kDa monomeric, multi-domain cysteine protease has three canonical ubiquitin-binding domains: two ubiquitin associated (UBA) domains and a high-affinity Zn-finger ubiquitin-binding protein (ZnF-UBP) domain (Fig. 3b)35. The latter possesses a deep
binding pocket to accommodate the free C-terminus of the proximal Ub moiety within unanchored poly-Ub chains (a key recognition feature), whilst the former are believed to bind more distal Ub subunits in longer chains. USP5’s active site, which is located between these domains and also has an affinity for Ub, cleaves the Ub–Ub isopeptide bond proximal to the C-terminus. A recent crystal structure of USP5 has raised a number of questions about this mechanism: in particular it shows the enzyme in a form where the ZnF-UBP domain is pointing away from the active site35. This orientation would prevent a poly-Ub substrate in a form where the ZnF-UBP domain is pointing away from the active site to deliver the substrate. To test this hypothesis, footprinting using diazirine 2 was performed on the catalytically inactive C335A USP5, both in the presence and absence of one equivalent of K48-linked di-Ub. Following irradiation and subsequent trypsin digestion, the differential labelling pattern of USP5 was quantified by LC-MS (see the ‘Methods’ section, Supplementary Figs 27–85 and Supplementary Tables 39–41 for details). Given the large sizes of the enzyme and substrate (ca. 100 kDa and 17 kDa, respectively), peptide-level analysis was found to be more than adequate to identify binding patches on the former. Figure 3a shows the amount of labelling on each detected USP5 peptide with and without the addition of di-Ub. Large (>10%) and statistically significant differences in labelling were observed for some peptides, and these results mapped on to the X-ray structure of USP5 (Fig. 3c). The data clearly show reduced labelling on the high-affinity ZnF-UBP domain and, importantly, on the active site domain, which is a biologically rational result. Since previous work by us and others has shown that USP5 binds di-Ub with a stoichiometry of 1:1 (refs 31,36), this result strongly suggests that the protein undergoes a conformational change allowing di-Ub to bind both at the ZnF-UBP and the catalytic site (in contrast to the conformation seen in the X-ray structure). This proposal is supported by mutagenesis and isothermal titration calorimetry findings that show cooperative binding of di-Ub to both the high-affinity ZnF-UBP site and the lower affinity catalytic domain31. In addition to labelling differences in these two regions of USP5, a very small (5%), but statistically significant, masking of peptide G606-K630 was seen. This corresponds to a flexible loop adjacent to the two UBA domains, and may reflect conformational change associated with substrate binding. Although the differential nature of the experimental design means that differences in the data should depend upon the presence of the binding partner alone, as with HEWL (see above), we tested against any detrimental effects of 10 mM 2 on the structure of USP5 using an enzyme assay (Supplementary Fig. 68). No deterioration of the protein activity was seen, which indicated the preservation of native protein structure in the presence of probe 2.

Discussion

Methods that permit high-resolution mapping of protein interaction sites with improved throughput and sensitivity are highly sought after. We have discovered and demonstrated that differential carbene footprinting with diazirine 2, in the presence and absence of a binding partner, together with MS-based analysis can provide reliable, high-quality data for both protein-small molecule ligand, and protein–protein systems. Using a bespoke, highly reactive carbene label, this approach has the benefit of fast, efficient and irreversible labelling. These
properties confer significant advantages over: HDX, where back-exchange can complicate interpretation; covalent labelling, where slow reaction rates and specific reactivity can give poor spatial and temporal resolution; and photoleucine-based carbene foot-printing, which requires high probe concentrations for acceptable levels of labelling. Hydroxy radical foot-printing, and FPOP in particular, is a very promising technique, but does require exposure of the protein to potentially damaging hydrogen peroxide and shorter wavelength ultraviolet radiation. Diazirine 2 appears benign until photophysical activation, which occurs rapidly at near-ultraviolet wavelengths (typically $\sim 350$ nm), and its structural features make it a highly efficient labelling reagent. It was not designed to be an accurate measure of solvent accessibility of a protein, but rather for use in a differential, $\pm$ binding partner assay. Once formed, the carbene has a lifetime of a few ns, reacting with either protein, buffer or water molecules. By conducting labelling experiments on flash frozen samples we were able to conduct diffusion effects 3,24, and hence it was thought likely that the sites of protein labelling were reflective of preferential diazirine $\pm$ amino acid sidechain interactions, rather than the inherent reactivity profile of the carbene. This notion was supported by the amino acid labelling-distribution of the hydrophobic and anionic probe 2 (on both HEWL and USP5), which generally favoured hydrophobic and basic (cationic) residues. The spin state of aryltrifluoromethyl carbenes, such as 3, is relatively complicated; $p$-toluyl(trifluoromethyl)carbene, for example, has a triplet ground state, but it principally reacts through its low-lying singlet state 28. This property allows aryltrifluoromethyl carbene to insert into the C-H bonds of non-polar amino acid sidechains, as well as the X-H bonds (where X = O, N, S) found on polar amino acid sidechains, and hence expands their utility in protein labelling.

The binding of carbohydrate substrates, such as NAGn oligomers, to HEWL has been studied in detail over the last 50 years, and the precise protein–ligand interactions are well described. The importance of W62, for example, in making van der Waals contacts with a sugar ring in the oligomeric substrate has been demonstrated by X-ray and NMR studies 37. This residue is highly masked in our differential carbene labelling measurements, as are nearby Q57, I58 and N59, which also lie within the binding cleft. In contrast, R60, positioned between these groups of residues, exhibits an increase in both labelling and solvent exposure upon substrate binding (see the ‘Results’ section). These findings are in good agreement with X-ray data, and highlight the accuracy and resolution of the footprinting technique applied to this system.

In contrast, the interactions between USP5 and its poly-Ub substrates are more complex and less well studied than those of lysozyme. USP5 is a large, dynamic enzyme with specific Ub-binding regions in addition to the active-site domain itself. The ZnF-UBP domain has a particularly high affinity for these regions, and this interaction is key in the recognition of unanchored poly-Ub chains over anchored forms. A recent X-ray crystal structure of USP5 shows the ZnF-UBP domain pointing away from the active site in a manner seemingly incompatible with substrate delivery. ZnF-UBP being flanked by flexible loop regions, can show significant domain motion, and potentially enable it to assist in substrate delivery to the active site. The carbene footprinting results described here show that peptides found on both the ZnF-UBP and catalytic domains of inactivated USP5 are masked from labelling on addition of K48-linked di-Ub. Since the stoichiometry of enzyme-substrate binding is known to be 1:1 (refs 31,36,38), and isothermal titration calorimetry data show cooperative binding of the proximal and distal units of di-Ub at the ZnF-UBP and catalytic site, respectively, this finding provides further evidence for an alternative conformation of USP5 where the ZnF-UBP binding pocket is in close spatial proximity to the active site such that di-Ub is bound at both sites. Thus, the conformation of USP5 seen in the crystal structure may represent a form different from the catalytically active species. This finding shows that carbene protein footprinting can be used to provide important insights into complex systems.

In summary, we have shown that differential carbene labelling of proteins with diazirine 2 in the presence and absence of small molecule or protein-binding partners allows high-quality mapping of interaction sites.

**Methods**

**Synthesis of diazirine probe 2.** Unless otherwise stated, reactions were carried out in flame-dried glassware, under an atmosphere of argon with magnetic stirring. Tetrahydrofuran (THF) was freshly distilled over sodium and benzophenone; dichloromethane (DCM) was freshly distilled over CaH2 under a nitrogen atmosphere. NEt3 was distilled from CaH2 and stored over 4 Å molecular sieves.

**2,2,2-trifluoro-1-(1-p-toly)ethane 4.** To a solution of crude 2,2,2-trifluoro-1-(1-p-toly)ethane (from the previous step) in pyridine (10 ml) and ethanol (5 ml) was added hydroxylamine hydrochloride (1.22 g, 17.5 mmol). The reaction mixture was heated at 50°C for 16 h, cooled and diluted with EtOAc (100 ml), and the organic phase washed with 2 M HCl(aq) (50 ml), dried over MgSO$_4$, and concentrated under reduced pressure. Volatiles were removed under reduced pressure, the product was extracted into the organic layer of a work up between EtOAc (50 ml) and 2 M HCl(aq) (50 ml). The organic layer was dried over MgSO$_4$, concentrated under reduced pressure to yield a colourless oil, which was used without further purification. A separate reaction was purified for analytical data.

**2,2,2-trifluoro-1-(p-toly)ethane oxime 5.** To a solution of crude 2,2,2-trifluoro-1-(p-toly)ethane (from the previous step) in pyridine (10 ml) and ethanol (5 ml) was added hydroxylamine hydrochloride (1.22 g, 17.5 mmol). The reaction mixture was heated at 50°C for 16 h, cooled and diluted with EtOAc (100 ml), and the organic phase washed with 2 M HCl(aq) (50 ml), dried over MgSO$_4$, and concentrated under reduced pressure. Purification by column chromatography (5% EtOAc/Petrol) yielded the title compound as a white solid (2.77 g, 78% over two steps). m.p. 77–80°C. 1H-NMR (400 MHz) of the title compound in CDCl$_3$, DMSO-D$_6$ or CD$_3$OD. Infrared spectra were obtained using a Bruker Tensor 27 FT-IR spectrophotometer as a solution in CHCl$_3$, with the peaks recorded as $\nu_{max}$ (cm$^{-1}$) or using an attenuated total reflectance (ATR) adaptor for the Nicolet IsF Spectrometer. High-resolution mass spectra (HRMS) were obtained using a Bruker MicroTOF time of flight mass spectrometer operating in electrospay ionization mode, or a Fisons AutoSpec magnetic sector mass spectrometer in electron impact ionization (EI) mode. Meltting point data was collected using a Stuart SMP3. Spectral data showed agreement with the literature 27.
To a solution of 2,2,2-trifluoro-1-(p-toly)ethanone oxime (5.25 g, 12.3 mmol) and DMAP (75 mg, 0.62 mmol) in anhydrous dichloromethane (30 ml) was added triethylamine (4.29 ml, 30.8 mmol) followed by tosyl chloride (2.82 g, 14.80 mmol) and the reaction stirred for 16 h at 0 °C. Purification by column chromatography (20% EtOAc/Petrol) yielded the title compound as a white solid (1.92 g, 11.9 mmol) and the reaction stirred for 5 min. The suspension was filtered, and the filtrate concentrated under reduced pressure to yield the title compound as a white solid (98 mg, 98%).

3-(p-toly)-3-(trifluoromethyl)diaziridine 7.

To a solution of 2,2,2-trifluoro-1-(p-toly)ethanone oxime (1.30 g, 6.40 mmol) and 4-(dimethylamino)pyridine (DMAP) (39 mg, 0.32 mmol) in pyridine (10 ml) and H2O (10 ml) was added KMnO4 (101 mg, 101.5 mmol) for 16 h. The solution was cooled, diluted with H2O (50 ml), before the organic layer was separated, basified with 2 M NaOH/aq (50 ml), and the reaction stirred for 16 h at 0 °C. The reaction mixture was stirred at room temperature for 1 h before the organic layer was washed with H2O, dried over MgSO4 and concentrated under reduced pressure. Purification by column chromatography (5% EtOAc/Petrol) yielded the title compound as a white solid (2.21 g, 97%). m.p. 101–103 °C/C176 C (Et2O/hexane).

Sodium 4-((trifluoromethyl)-3H-diazirin-3-yl)benzoate 2.

To a solution of sodium hydroxide (16 mg, 0.39 mmol) in H2O (1 ml) was added 4-((trifluoromethyl)-3H-diazirin-3-yl)benzoic acid (8, 100 mg, 0.43 mmol), and the suspension stirred for 5 min. The suspension was filtered, and the filtrate concentrated under reduced pressure to yield the title compound as a white solid (98 mg, 89%).

Protein footprinting. HEWL, penta-N-acetylchitotetrasaccharide (NAG5), diithiotherol (DTT), iodocadaverine, calmodulin, equine cytochrome c, melittin and ubiquitin were purchased from Sigma-Aldrich (Poole, UK). Trypsin and AspN were obtained from Promega (Southampton, UK). Lys48-poly-ubiquitin ladder, wild-type USPs and Lys48 diubiquitin were purchased from Biochrom (Cambridge, MA, USA). C335A USP5 was expressed and purified as previously described.\(^4\) Photoreaction 1 was purchased from Thermo Fisher Scientific, Loughborough, UK, and aryldiazirine 2 was synthesized as described above.

Photochemical labelling of lysozyme. An aqueous solution of lysozyme (160 μM) in 25 mM ammonium acetate or 20 mM Tris/150 mM NaCl was mixed with an equal volume of a 20 mM solution of the diazirine labelling agent (1 or 2) in 20 mM Tris/150 mM NaCl to give a final label concentration of 10 mM. The mixture was left equilibrating for 5 min at room temperature. Aliquots (6 μl) of this solution were placed in 11 mm PP vial crimp snap 250 nl sample vials (Thermo Fisher Scientific). To avoid enzymatic cleavage of the pentasaccharide substrate, NAG5 was added to the mixture, to a final concentration of 80 μM, just before snap-freezing the samples in liquid nitrogen (77 K). Freezing was found to be important for efficient and reproducible labelling. For lysozyme footprinting in the absence of substrate, the NAG5 solution was substituted by an equal volume of buffer. The labelling reaction was initiated by irradiation (16 s) of the mixture using a Spectra Physics Explorer 349 laser (an actively Q-switched Nd:YLF laser operating at 349 nm, with a repetition frequency of 1,000 Hz, and a pulse energy of 125 μJ, Newport, Didcot, UK). The laser beam was directed into the open top of the sample vial using a small 45° mirror. Photochemical labelling of C335A USP5. A solution of C335A USP5 (5 μM), Lys48 di-ubiquitin (30 μM each, 10 mM ammonium acetate, or 20 mM Tris/150 mM NaCl) and diazirine 2 (in 20 mM Tris/150 mM NaCl, to a final concentration of 10 mM) was snap-frozen and irradiated at 349 nm as described for lysozyme.

Labeling efficiency. Solutions of calmodulin, myoglobin, cytochrome c, ubiquitin or melittin (70 μM in 25 mM Tris/150 mM NaCl) and photoleucine 1 (100 mM in 20 mM Tris/150 mM NaCl) or diazirine 2 (10 mM in 20 mM Tris/150 mM NaCl) were snap-frozen and irradiated at 349 nm for 4 s in the case of diazirine 2 (except calmodulin, which required only 1 s), and 15 s in the case of photoleucine 1. Samples were diluted to 5 μM with mobile phase A (H2O/Acetone 95/5 + 0.1% formic acid) and analysed directly by LC-MS.

Protein digestion. Following separation by SDS–polyacrylamide gel electrophoresis (SDS–PAGE; 15% acrylamide), protein bands were excised, reduced (DTT, 10 mM in 100 mM ammonium bicarbonate), alkylated (iodoacetamide, 55 mM in 10 mM ammonium bicarbonate) and digested at 37 °C with either trypsin (1:15 protease:protein ratio in 10 mM ammonium bicarbonate) or AspN (1:70 protease/protein ratio in 5 mM Tris-HCl) for 16 h (ref. 39). Formic acid (0.5 μl) was added to the supernatant (50 μl) to inactivate the proteases. The resulting solutions were analysed by LC/MS without further dilution.

LC-MS analysis. Measurement of intact mono-ubiquitin, lysine or protein digests was carried out on a Dionex Ultimate 3000 Nano LC (Dionex, Camberley, UK) system using a commercially available C4 trapping column (Dionex) and a custom packed WP C4 column (5μm, 100 Å, Phenomenex, Macclesfield, UK) with a Picofrit tip (75 μM i.d. × 150 mm, new objective, supplied by Aquilant, Basingstoke, UK). The mobile phases A and B consisted of 95/5 water/acetonitrile (v/v) and 5/95 water/acetonitrile (v/v), respectively, and both contained 0.1% formic acid. The samples (1 μl) were injected in load-trapping mode. Peptides were eluted using a 30 min linear gradient of mobile phase B from 0–70% at a flow rate of 0.3 μl min\(^{-1}\) followed by column equilibration. The high-performance liquid chromatography system was coupled to a Thermo Scientific (Hemel Hempstead, UK) LTQ Ultra mass spectrometer equipped with a Cap-LC/Finnigan nanoelectrospray ionization source. A 1.7 kV voltage was applied to a coated PicorTip emitter (New Objective). The capillary temperature was set at 275 °C, with inner capillary voltage set at 37 V and tube lens voltage of 145 V. Spectra were acquired in positive ion mode over a 400–2,000 m/z range at a nominal resolution of 100,000 (m/z = 400). The instrument was controlled by Xcalibur software (Thermo Fisher). Identification of USP5 peptides was conducted using an automated data-dependent acquisition mode, followed by manual examination of
the raw data. For data-dependent acquisition the four most intense 
ions for each labelled and unlabelled peptide ion. Partial, or even 
completely exchanged positions were rejected. The data were 
searched against a custom database including the 

cannot be related to the indicated 

Data analysis: the fractional modification per peptide, and the average number 
of labels per residue were determined using the approach described by Jumper 
et al.24, to uncover poly-UB linkage selectivity of different deubiquitinating 

Data Availability 

References 

1. Barabási, A. L. & Oltvai, Z. N. Network Biology: understanding the cell's 
fundamental organization. Nat. Rev. Genet. 5, 101–113 (2004).
2. Ilari, A. & Savino, C. Protein structure determination by x-ray crystallography. 
Methods Mol. Biol. 452, 63–87 (2008).
3. Benesch, J. L. P. & Ruotolo, B. T. Mass spectrometry: an approach come-of-age 
for structural and dynamical biology. Curr. Opin. Struct. Biol. 21, 641–649 
(2011).
4. Heck, A. J. Native Mass Spectrometry: a bridge between interactomics and 
structural biology. Nat. Methods 5, 927–933 (2008).
5. Konermann, L., Vahidi, S. & Sowole, M. A. Mass spectrometry methods for 
structural biology. Methods 54, 32–38 (2011).
6. Konermann, L., Pan, J. & Liu, Y.-H. Hydrogen exchange mass spectrometry for 
studying protein structure and dynamics. Chem. Soc. Rev. 40, 1224–1234 
(2014).
7. Sheff, J. G., Rey, M. & Schriemer, D. C. Peptide-column interactions and their 
influence on back exchange rates in hydrogen/deuterium exchange-MS. J. Am. 
Chem. Soc. 129, 1096–1105 (2007).
8. Ferguson, P. L. et al. Hydrogen/deuterium scrambling during quadrupole 
time-of-flight MS/MS analysis of a zinc-binding protein domain. Anal. Chem. 
79, 153–160 (2007).
9. Pan, J., Han, J., Borchers, C. H. & Konermann, L. Electron capture dissociation 
of electrosprayed protein ions for spatially resolved hydrogen exchange 
measurements. J. Am. Chem. Soc. 130, 11574–11575 (2008).
10. Mendoza, V. L. & Vachet, R. W. Probing protein structure by amino acid-
specific covalent labeling and mass spectrometry. Mass Spectrom. Rev. 28, 
785–815 (2009).
11. Sinz, A. Chemical cross-linking and mass spectrometry to map three-
dimensional protein structure and protein-protein interactions. Mass Spectrom. 
Rev. 25, 663–682 (2006).
12. Livak, A., Espina, V. & Liotta, L. A. Protein painting reveals solvent-excluded 
drug targets hidden within native protein-native interfaces. Nat. Commun. 5, 
4143 (2014).
13. Tullius, T. D. & Dombroski, B. A. Hydroxyl radical `footprinting': high-
resolution information about DNA-protein contacts and application to lambda 
repressor and Cro protein. Proc. Natl Acad. Sci. USA 83, 5469–5473 (1986).
14. Sclavi, B., Woodson, S. A., Sullivan, M., Chance, M. & Brenowitz, M. in 
Energetics of Biological Macromolecules, Part B. 1st edn Vol. 295 (eds Ackers, G. K. & 
Johnson, M. L.) (Academic Press, 1998).
15. Ralston, C. Y. et al. Time-resolved synchrotron X-ray footprinting and its 
application to RNA folding. Methods Enzymol. 317, 353–368 (2000).
16. Hambly, D. M. & Gross, M. L. Laser flash photoysis of hydrogen peroxide to 
oxidize protein solvent-accessible residues on the microsecond timescale. J. Am. 
Soc. Mass Spectrom. 16, 2057–2063 (2005).
17. Aye, T. T., Low, T. Y. & Sze, S. K. Nanosecond laser-induced chemical 
oxidation method for protein surface mapping with mass spectrometry. Anal. 
Chem. 77, 5814–5822 (2005).
18. Calabrese, A. N., Ault, J. R., Radford, S. E. & Ashcroft, A. E. Using hydroxyl 
radical footprinting to explore the free energy landscape of protein folding. 
Methods 89, 39–44 (2015).
19. Chen, J. W., Rempel, D. L., Gau, B. C. & Gross, M. L. New protein footprinting: 
fast photochemical iodination combined with top-down and bottom-up mass 
spectrometry. J. Am. Chem. Soc. 134, 18724–18731 (2012).
20. Richards, F. M., Lamed, R., Wynn, R., Patel, D. & Olack, G. Methylene as a 
possible universal footprinting reagent that will include hydrophobic surface 
areas: overview and feasibility: properties of diazirine as a probe. Protein 
Sci. 9, 2506–2517 (2000).
21. Gomez, G. E., Monti, J. L. E., Mundo, M. R. & Delfino, J. M. Solvent 
mimicry with methylene carbene to probe protein topography. Anal. Chem. 
87, 16008–16007 (2015).
22. Kotzyba-Hilbert, F., Kapfer, I. & Goedl, M. Recent trends in Photoaffinity 
labeling. Angew. Chem. Int. Ed. 34, 1296–1312 (1995).
23. Jumper, C. C. & Schriemer, D. C. Mass spectrometry of laser-initiated 
carbene reactions for protein topographic analysis. Anal. Chem. 83, 2913–2920 
(2011).
24. Cintrón, C. C., Bomgaard, R., Rogers, J., Etienne, C. & Schriemer, D. C. High-resolution mapping of carbene-based protein footprints. Anal. Chem. 
84, 4411–4418 (2012).
25. Zhang, B., Rempel, D. L. & Gross, M. L. Protein footprinting by carbones on a 
fast photochemical oxidation of proteins (FPPO) platform. J. Am. Soc. Mass 
Spectrom. 27, 552–555 (2016).
26. Schriemer, D. C., Senn, H. & Richards, F. M. 3-Trifluoromethyl-3-phenyl 
diazirine, a new carbene generating reagent for photo labeling reagents. J. Biol. 
Chem. 255, 3313–3318 (1980).
27. Nossal, M. A. 1-(Aziridinyl)benzoic acid, a highly photolabile 
carbene generating label readily fixable to biochemical agents. Eur. J. Org. 
Chem. 1983, 1510–1523 (1983).
28. Admassu, A. et al. A laser flash photoysis study of p-tolyl(trifluoromethyl) 
carbene. J. Chem. Soc. Perkin Trans 2 1998, 1093–1099 (1998).
29. Strynadka, N. J. C. & Games, N. M. G. in Lysosomes: model enzymes in 
biochemistry and biology. (ed. Jolles, P.) (Birkhäuser Verlag, 1996).
30. Cross, A. J. & Fleming, G. R. Influence of inhibitor binding on the internal 
movements of lysosome. Biophys. J. 50, 507–512 (1986).
31. Leyvraz-Turca, F. E., Shanks, J. R., Radford, S. E., Wilson, K. & Wilkinson, K. D. Recognition of 
polyubiquitin isoforms by the multiple ubiquitin binding modules of 
isopeptidase T. J. Biol. Chem. 283, 15981–15992 (2008).
32. Dayal, S. et al. Direct activation of protein kinases by unanchored 
polyubiquitin chains. Nature 461, 114–119 (2009).
33. Zeng, W. et al. Reconstitution of the RIG-I pathway reveals a signaling role 
of unanchored polyubiquitin chains in innate immunity. Cell 141, 315–330 
(2010).
34. Avvakumov, G. V. et al. Two ZnF-UBP domains in isopeptidase T (USP5). 
Struct. 1006–1015 (2013).
35. Mendoza, L. & Vachet, R. W. Probing protein structure by amino acid-
specific covalent labeling and mass spectrometry. Mass Spectrom. Rev. 28, 
785–815 (2009).
36. Scott, D., Layfield, R. & Oldham, N. J. Ion mobility-mass spectrometry reveals 
conformational flexibility in the deubiquitinating enzyme USP5. Proteomics 15, 
2835–2841 (2015).
37. Maenaka, K., Kawai, G., Wantanabe, K., Sunada, F. & Kumagai, I. Functional and structural role of a tryptophan generally observed in protein-carbohydrate interaction. TRP-62 of hen egg white lysozyme. J. Biol. Chem. 269, 7070–7075 (1994).

38. Scott, D., Layfield, R. & Oldham, N. J. Structural insights into interactions between ubiquitin specific protease 5 and its polyubiquitin substrates by mass spectrometry and ion mobility spectrometry. Protein Sci. 24, 1257–1263 (2015).

39. Shevchenko, A., Tomas, H., Havlis, J., Olsen, J. V. & Mann, M. In-gel digestion for mass spectrometric characterization of proteins and proteomes. Nat. Protoc. 1, 2856–2860 (2006).

40. Komander, D. et al. Molecular discrimination of structurally equivalent Lys 63-linked and linear polyubiquitin chains. EMBO Rep. 10, 466–473 (2009).

41. Schneider, C. A., Rasband, W. S. & Eliceiri, K. W. NIH Image to ImageJ: 25 years of image analysis. Nat. Methods 9, 671–675 (2012).

Acknowledgements

We thank the BBSRC (grant number BB/I006052/1) and the University of Nottingham for funding. We are also grateful to Joe Harris and Anna Andrejeva for performing initial test photolysis experiments.

Author contributions

N.J.O. and J.E.M. designed experiments, A.S.B. synthesized and characterized the diazirine probe, L.M. conducted the labelling experiments and MS analysis. D.S. and R.L. expressed and purified recombinant USP5, and T.G.W. provided expertise for the laser experiments. N.J.O. wrote the paper with the assistance of the other coauthors. Furthermore, R.L. and T.G.W. made a substantial contribution to the organization of the study, and to the analysis of the data, as well as providing important input into the intellectual content of the paper.

Additional information

Supplementary Information accompanies this paper at http://www.nature.com/naturecommunications

Competing financial interests: The authors declare no competing financial interests.

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

How to cite this article: Manzi, L. et al. Carbene footprinting accurately maps binding sites in protein–ligand and protein–protein interactions. Nat. Commun. 7, 13288 doi: 10.1038/ncomms13288 (2016).

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

This work is licensed under a Creative Commons Attribution 4.0 International License. The images or other third party material in this article are included in the article’s Creative Commons license, unless indicated otherwise in the credit line; if the material is not included under the Creative Commons license, users will need to obtain permission from the license holder to reproduce the material. To view a copy of this license, visit http://creativecommons.org/licenses/by/4.0/