High-Resolution Mapping of Carbene-Based Protein Footprints

ARTICLE *in* ANALYTICAL CHEMISTRY · APRIL 2012

Impact Factor: 5.64 · DOI: 10.1021/ac300120z · Source: PubMed

analytical chemistry

Article

pubs.acs.org/ac

High-Resolution Mapping of Carbene-Based Protein Footprints

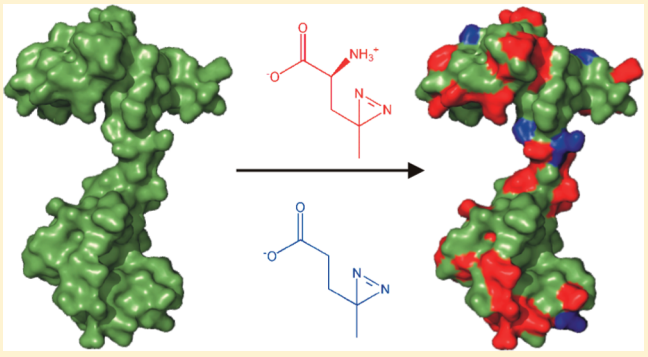
Chanelle C. Jumper,[†] Ryan Bomgarden,[§] John Rogers,[§] Chris Etienne,[§] and David C. Schriemer*[,†,‡]

Departments of [†]Chemistry and [‡]Biochemistry and Molecular Biology, University of Calgary, Calgary, Alberta, T2N 4N1, Canada

[§]Thermo Fisher Scientific, 3747 North Meridian Road, Rockford, Illinois 61101, United States

Supporting Information

ABSTRACT: Carbene chemistry has been used recently in structural mass spectrometry as a labeling method for mapping protein surfaces. The current study presents a method for quantitating label distribution at the amino acid level and explores the nature and basis for an earlier observation of labeling bias. With the use of a method based on liquid chromatography−tandem mass spectrometry (LC−MS/MS) applied to digests of holo-calmodulin, we developed a quantitation strategy to map site-specific incorporation of carbene, generated from photolysis of ionic label precursors 2-amino-4,4-azipentanoic acid and 4,4-azipentanoic acid. The approach provides reliable incorporation data for fragments generated by electron-transfer dissociation, whereas high-energy collisional dissociation leads to energy and sequence-dependent loss of the label as a neutral. However, both can produce data suitable for mapping residues in the interaction of holo-calmodulin with M13 peptide ligand. Site-specific labeling was monitored as a function of reagent, ionic strength, and temperature, demonstrating that electrostatic interactions at the protein surface can "steer" the distribution of label precursors to sites of surface charge and favor label insertion into residues in the vicinity of the surface charge. A further preference for insertion into carboxylates was observed, based on chemical reactivity. We suggest that decoupling surface partitioning from the chemistry of insertion offers a flexible, tunable labeling strategy for structural mass spectrometry that can be applied to a broad range of protein surface compositions and promotes the design of reagents to simplify the workflow.



Structural mass spectrometry (MS) has emerged as a complement to more established methods in structural biology, as it can be used to define contact surfaces between proteins and their ligands and to monitor protein dynamics.[1−5] This can be achieved through the use of chemical reagents for topographical mapping and employing the mass spectrometer to identify where in structure the reagents have distributed. Perturbations of these maps provide the data necessary for such purposes. For example, these data can be mined for distance restraints, useful to generate structural models for protein complexes that have an empirical basis.[6−8]

The utility of structural mass spectrometry relies on the properties of the protein chemistries used in the mapping exercise. The functional group specificity of the reagents and the kinetics of the labeling reactions are two important considerations. An ideal reagent might target all solvent-exposed amino acid residues equally, with reaction rates that exceed the dynamics of protein motion so that the folding and/or functional properties of proteins can be explored as well. Most structural mass spectrometry methods employ hydrogen/deuterium (H/D) exchange,[9] hydroxyl radical labeling,[10] or more selective chemical reagents.[4] These methods have permitted numerous useful contributions to structure analysis but possess limitations that drive the need for new solutions. For example, the H/D exchange of backbone amides has slow kinetics and the chemical label is transient. Residue-selective reagents that label side chains (e.g., succinimide esters targeting lysine) have even slower kinetics and together with their size can perturb the underlying structure and dynamics of the protein being probed.[4] Hydroxyl radical labeling is an appealing method that can label on the microsecond time scale and target about half the common amino acids,[10,11] although readily oxidized residues are strongly favored (e.g., methionine). The sensitivity of labeling to protein concentration is perhaps its most significant concern;[12] however, the power of mass spectrometry in protein analysis continues to drive new developments in these methods and stimulate ideas for new ones.

A substituted diazirine has been introduced recently as a highly soluble, potentially nonselective protein labeling reagent.[13] Upon photolysis of the diazirine, a reactive carbene is generated with the potential to rapidly insert into any X−H bond (where X can be C, O, N, or S).[14−16] Advantages to this type of reagent include protein concentration independence, fast reactions in the submicrosecond range, and no reactivity toward the protein prior to activation. The approach was

Received: January 12, 2012
Accepted: April 5, 2012
Published: April 5, 2012

   dx.doi.org/10.1021/ac300120z | *Anal. Chem.* 2012, 84, 4411−4418

demonstrated using L-2-amino-4,4-azipentanoic acid ("photoleucine") as a label source, which was converted to the corresponding carbene by pulsed laser photolysis at 355 nm. The study showed that the level of labeling is sensitive to changes in protein topography brought about by conformational change and ligand binding.

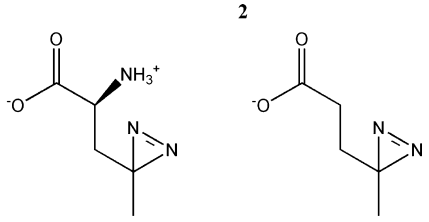
The approach was inspired by the use of diazirines for photoaffinity labeling[16] and more specifically by the photolysis of diazirine gas (CH_2N_2), which was proposed by Richards et al.[17] and Craig et al.[18] to be the ideal labeling agent. When bubbled into solution and photolyzed, reactive $:CH_2$ is generated that bears a geometric and size similarity to water. The protein reactivity that results is suggestive of indiscriminate residue labeling, at least based on a peptide-level analysis using tritiated methylene and scintillation detection.[19] Since a limitation of this method is the low solubility of the gas, requiring continual bubbling of reagent and photolysis of several minutes in duration,[20] we have continued to explore the use of highly soluble substituted diazirines that can be rapidly converted to MS measurable labeling yields with pulsed laser irradiation.

Although all amino acids are potential targets for carbene reactions, it has yet to be determined whether all amino acids are *equivalent* targets, particularly for substituted diazirines. Carbene diffusion should be limited in aqueous solution, since polar protic species have very high rates of reactivity with substituted or unsubstituted carbenes, and carbene lifetimes are extremely short.[14,16,21] Under this assumption, the distribution of labels should reflect the distribution of reagent across the protein surface prior to activation. Our previous study revealed a protein-dependent labeling yield, with incorporations both higher and lower than expected based on calculations of accessible surface area.[13] In the current work, we analyze the distribution of products at the peptide and residue level for two substituted diazirine reagents, to determine the feasibility of the method for higher-resolution structural studies and to study the nature and source of the labeling bias suggested in the earlier work.

EXPERIMENTAL SECTION

Materials. Lyophilized calmodulin (CaM)[22] (heavy and light) was provided by Dr. Hans Vogel of the Department of Biological Sciences at the University of Calgary (Calgary, AB). L-2-Amino-4,4-azipentanoic acid (**1**) and 4,4-azipentanoic acid (**2**) were synthesized by variations to methods described by Suchanek et al.,[23] and the structures are shown in Scheme 1. Reagent **1** is commercially available, and reagent **2** can be obtained through a custom synthesis request (both from Thermo Fisher Scientific). Other materials used in the research are listed in the Supporting Information.

Scheme 1. Structures of 2-Amino-4,4-azipentanoic Acid (1) and 4,4-Azipentanoic Acid (2)



Labeling Protocol. A solution of protein (100 μM CaM or CaM−M13 complex) was equilibrated with 100 mM reagent **1** or 100 mM reagent **2** in Tris-buffered saline (10 mM Tris, 100 mM KCl, pH 7.5) for 15 min at room temperature. A 5 μL droplet was placed in the center of a small glass vial optically isolated from its surroundings, except for the vial mouth. Covalent modification was initiated by photolysis of the droplet at a wavelength of 355 nm, using the third harmonic of a Nd:YAG pulsed laser (Spectra Physics, 20 mJ/pulse, pulse width of 8−10 ns fwhm). The beam width was adjusted to be slightly greater than the droplet radius, to ensure the entire droplet was irradiated. Approximately 100 pulses were accumulated per sample (at 10 Hz), sufficient for the conversion of 200 μmol of diazirine (data not shown), or 4 times the fluence necessary for complete photolysis of diazirine in the droplet. Samples were generated in triplicate for subsequent analysis by mass spectrometry.

Diffusion Control. Aliquots of buffered protein equilibrated with reagent **1** were subjected to various temperature conditions prior to labeling. For room-temperature labeling (293 K), samples were equilibrated on ice for 15 min. Room-temperature samples were also rapidly cooled in dry ice (195 K) and equilibrated for a minimum of 2 min or plunge-frozen in liquid nitrogen (77 K) and equilibrated for a minimum of 2 min. The vials were removed from equilibrated conditions 5 s prior to photolysis and irradiated for 10 s as described, to ensure complete diazirine conversion.

Proteolytic Digestion. Each 100 μM aliquot of labeled protein was diluted to 10 μM with trypsin in a 20:1 (protein/enzyme) mass ratio and digested at room temperature for 16 h. Aliquots of an unlabeled control followed the same procedure. Digests were further diluted to 1 μM with an aqueous solution containing 0.2% formic acid and 3% acetonitrile, prior to liquid chromatography−tandem mass spectrometry (LC−MS/MS) analysis.

High-Performance Liquid Chromatography. The separation of labeled and unlabeled peptides was carried out on a Dionex Ultimate 3000 HPLC system (Sunnyvale, CA) in a nanoflow configuration (C18 trap, 75 μm i.d. × 2 cm, 3 μm particle diameter, and C18 column, 75 μm i.d. × 15 cm, 2 μm particle diameter, Dionex). Peptides were eluted using 0.2% formic acid that increased from 3% to 90% acetonitrile over 10 min at a flow rate of 0.3 μL/min.

Mass Spectrometry. All mass spectra of labeled and unlabeled samples were acquired in Orbitrap mode on a Thermo Scientific LTQ Orbitrap Velos mass spectrometer (San Jose, CA) equipped with a capillary or nanoflow electrospray ionization source as appropriate, with nebulization. Spray voltages were 5 and 2.9 kV for capillary and nanoflow, respectively, and desolvation temperature was 275 °C. Peptide spectra were acquired in positive ion mode for 300−2000 m/z. To obtain MS/MS data of labeled peptides, ions were fragmented by high-energy collisional dissociation (HCD) or electron-transfer dissociation (ETD) in separate experiments using inclusion lists, followed by detection at a nominal resolution setting of 60 000. MS/MS spectra were acquired for 100−2000 m/z. HCD-based precursor activation was performed with an activation time of 0.1 ms under collision energies optimized as described later in the article. ETD-based precursor activation was carried out for 200 ms, including charge-state-dependent supplemental activation. Precursor ions were isolated with width of 2 Th. Prior to the analysis of labeled peptides, unlabeled peptide sequences were identified using

dx.doi.org/10.1021/ac300120z | Anal. Chem. 2012, 84, 4411−4418

fragment fingerprinting methods standard in proteomics, and spectra were manually assessed to identify all observable sequence ions (i.e., y-ions for HCD and z-ions for ETD).

Peptide-Level Data Analysis. The extent of modification for each peptide was determined using the ratio between the chromatographic peak areas (PA) for all singly modified forms of the peptide to the sum of the modified and unmodified forms, resulting in the fraction modified of a given peptide (eq 1):

$$P = \frac{\mathrm{PA}_{\mathrm{modified}}}{\mathrm{PA}_{\mathrm{modified}} + \mathrm{PA}_{\mathrm{unmodified}}} \quad (1)$$

where P denotes the fractional amount of peptide modified. The quantitation method assumes equivalent ionization efficiencies between the modified and unmodified peptides, an assumption that may not always hold. The values are equivalent to an average number of labels/peptide for a singly modified peptide.

Residue-Level Data Analysis. A labeled peptide with the label distributed throughout the sequence can generate both labeled and unlabeled sequence ions in its MS/MS spectrum. For the MS/MS analysis of singly labeled peptides, we define the fractional modification of a given sequence ion as the ratio between the modified sequence ion intensity to the total sequence ion intensity (modified and unmodified). For example, for HCD spectra generating a y-ion series, eq 2 describes this fraction:

$$f(\mathrm{y}_i) = \frac{I(\mathrm{y}_i)_{\mathrm{modified}}}{I(\mathrm{y}_i)_{\mathrm{modified}} + I(\mathrm{y}_i)_{\mathrm{unmodified}}} \quad (2)$$

where $f(\mathrm{y}_i)$ denotes the fractional modification of y-ion i and $I(\mathrm{y}_i)$ the denotes the intensity of the y-ion of interest, whether in the modified or unmodified form. Finally, the absolute level of labeling for a given amino acid residue i is based on both the fractional amount of peptide modified and the fractional modification of the corresponding sequence ions. This is shown in eq 3:

$$\mathrm{labels/residue}_i = P[f(\mathrm{y}_i) - f(\mathrm{y}_{i-1})] \quad (3)$$

where P is derived from eq 1 and the term in square brackets is the difference between the fractional modification of adjacent residues represented by y_i and y_{i-1}. A similar approach is followed for z-ions.

Solvent Accessibility Calculation. Reagent accessibility was approximated using solvent accessibility calculations in Pymol (Schrödinger LLC) applied to the appropriate Protein Data Bank (PDB) coordinate files for holo-CaM (1PRW) and CaM−M13 (2O5G). Dot solvent was set to 1, and dot density was set to 3.

RESULTS AND DISCUSSION

Detection and Quantification of Modified Peptides. We have previously shown that moderate protein labeling could be achieved by photolysis of buffers doped with reagent **1**, governed primarily by the concentration of reagent **1** and the rate of energy deposition into solution.[13] This was established on the basis of whole-protein analysis by mass spectrometry. For the current study, we generated tryptic digestions of labeled proteins in order to investigate if labeling at higher structural resolution could be achieved, using conventional "bottom-up" style protein analysis methods. Under our chosen conditions, calcium-bound bovine brain calmodulin (holo-CaM) was labeled until reagent **1** was fully depleted, resulting in an average of approximately two insertions per molecule (data not shown). A tryptic map of unlabeled holo-CaM identified 12 peptides (Supporting Information Figure S1), 10 of which were selected to cover the complete protein sequence. These peptides and their carbene-modified forms were then quantified using LC−MS.

The extracted ion chromatograms (XICs) for the unlabeled and labeled peptides reveal differences in elution characteristics (Supporting Information Figure S2). Peak splitting in the XIC of the labeled peptide suggests the existence of unique congeners, pointing to at least a partial distribution of label across the peptide. The distribution in chromatography could be a means of monitoring individual label sites. However, it cannot be assumed that a unique peptide congener would be chromatographically differentiated from all others. Surveying the XICs for the singly modified forms of all 10 peptides separated by nanoLC, we observed that peak splitting was highly variable and not correlated with the label distribution. For the purposes of this study, we therefore implemented a quantitation strategy that did not depend on congener separation.

The level of modification was calculated as a ratio of modified to total peptide XIC peak areas, for a given peptide. The accuracy of this approach assumes that the labeled peptide has the same ionization efficiency as the unlabeled form, which we recognize as a possible limitation, but high precision can be maintained (Supporting Information Figure S3). The digest of labeled holo-CaM with M13 (a high-affinity peptide ligand) was compared to the corresponding digest of labeled holo-CaM alone. The M13 binding site and the conformational distortion created upon binding represents a large fraction of the total protein, and as expected, most peptides show a significant decrease in carbene labeling. We confirmed this using an isotope dilution method and tracking only the residual unlabeled peptide. For this experiment, holo-CaM was produced in ^{13}C- and ^{15}N-enriched media (heavy form) or normal media (light form). We measured the change in heavy and light forms for unlabeled peptides only and found general agreement between the two quantitation strategies (see Supporting Information Figure S4 for details).

Residue-Level Label Quantitation. Our approach to residue-level quantitation involved the generation of an averaged MS/MS spectrum across all chromatographic features for a modified peptide. The relative intensities of labeled and unlabeled fragments in an ion series were then measured and plotted as a function of sequence ion number (eq 2). A peptide with an equal probability of labeling at every residue would generate fractional modification levels that are linearly dependent on the sequence ion number (Figure 1A, open circles). Deviations from a simple linear dependence would be due to the effects of local structure and/or sequence-dependent reaction rates. The slope between two adjacent sequence ions ($n - 1$ and n) multiplied by the average number of labels in the total peptide (MS-level data) determines the average number of labels at the nth residue.

To illustrate, the modification ratios for residues in LTDEEVDEMIR (117−127) are shown in Figure 1A (filled circles). No appreciable labeling occurred on LTD and MIR based on slopes of zero between z_{11}−z_9 and z_3−z_1, respectively. All the labeling is found internally at z_8−z_4, or EEVDE. Scaling these relative intensities by the number of labels per peptide

dx.doi.org/10.1021/ac300120z | Anal. Chem. 2012, 84, 4411−4418

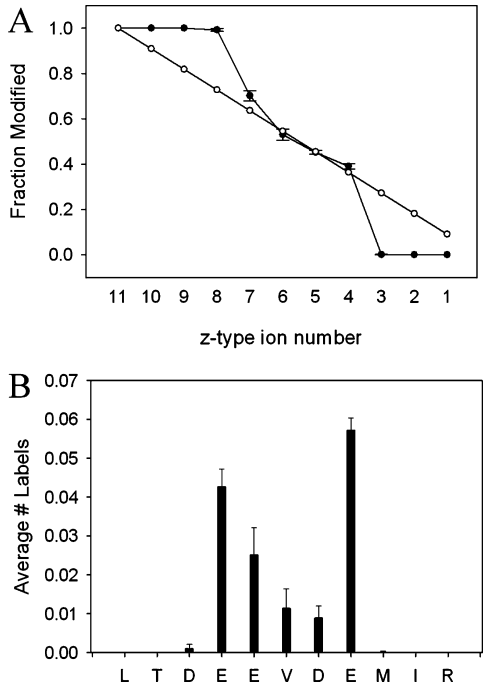


Figure 1. Distribution of carbene labels arising from the photolysis of holo-CaM in the presence of reagent **1** (2-amino-4,4-azipentanoic acid) for peptide LTDEEVDEMIR (117−127), determined using ETD fragmentation. (A) The fractional distribution of label (filled circles) relative to a theoretical distribution that assumes all residues are equally reactive and accessible to the reagent (open circles). (B) The corresponding absolute label incorporation for the peptide, calculated according to eq 3. Error bars represent ±1 SD.

generates a plot of the number of labels per residue (Figure 1B). Most of the labeling is found on three glutamates, but measurable levels are also observed for an aspartate and a valine. All residues with the possible exception of M are surface-exposed in the holo-CaM structure, so the lack of labeling on LTD and IR is noteworthy.

Impact of Fragmentation Chemistry on Site Determination. Under the right conditions, electron-based fragmentation methods such as ETD can be superior to collision-induced dissociation (CID)-based methods for preserving sites of modification and generating long sequence reads.[24,25] ETD was selected as the fragmentation method of choice for these reasons, based partly on speculation that carbene labeling of certain side chains may generate labile modifications. To test if CID fragmentation would report similar internal label distributions, peptides were fragmented by both ETD and HCD. The latter generates CID-like fragmentations spectra with the potential for sequence reads longer than conventional trap-based CID.[26] The internal distributions of two peptides are shown in Figure 2, parts A and B. HCD generates distributions that follow the trends established by ETD, but there are notable deviations. For example, in Figure 3A, the labeling at y_{10} is lower than the labeling at y_9, pointing to label loss that is dependent on the site of fragmentation. MS/MS spectra generated by HCD contained neutral loss peaks of variable intensity, with a mass of $115/x$ (x = peptide charge state). Neutral loss could be controlled in part by the collision energy used in the experiment. Reducing the collision energy to minimize label loss returned internal distribution plots similar to ETD data, but the intensity of sequence ions was also generally reduced. All experiments using HCD fragmentation

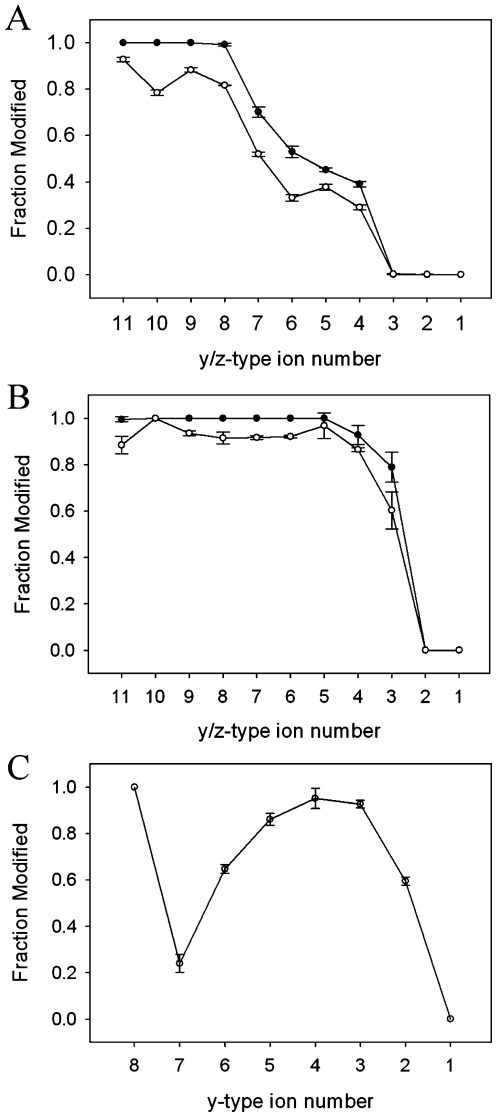


Figure 2. Fractional distribution of carbene label derived from reagent **1** (2-amino-4,4-azipentanoic acid), as determined by ETD (filled circles) and HCD (open circles) for (A) peptide LTDEEVDEMIR (117−127), (B) MKDTDSEEEIR (77−87), and (C) EAFSLFDK (15−22). ETD was unsuccessful for EAFSLFDK. Error bars represent ±1 SD.

were therefore first optimized for collision energy on this basis (Supporting Information Figure S5).

Neutral loss represents a loss of information. This is illustrated by the label distribution of EAFSLFDK (15−22) fragmented with HCD, as shown in Figure 2C. The label on D21 (y_2) and F20 (y_3) accounts for ∼90% of label incorporation, but fragmentation at successively higher y-ion numbers erodes all but ∼25% of the label. Interestingly, the intact peptide does not display neutral loss (y_8), highlighting that the gas-phase ion conformation and/or the location of the mobile proton are critical factors governing the rate of loss. Site-specific label information can be lost in such situations, where the rate of loss is not constant across the sequence. We hypothesize that neutral loss is occurring primarily from labile esters formed by the insertion of the carbene into the carboxylic acid of aspartate and glutamate residues. The distribution data shown in Figures 1 and 2 demonstrate that labeling is weighted toward these residues, and although carbenes can insert into

dx.doi.org/10.1021/ac300120z | *Anal. Chem.* 2012, 84, 4411−4418

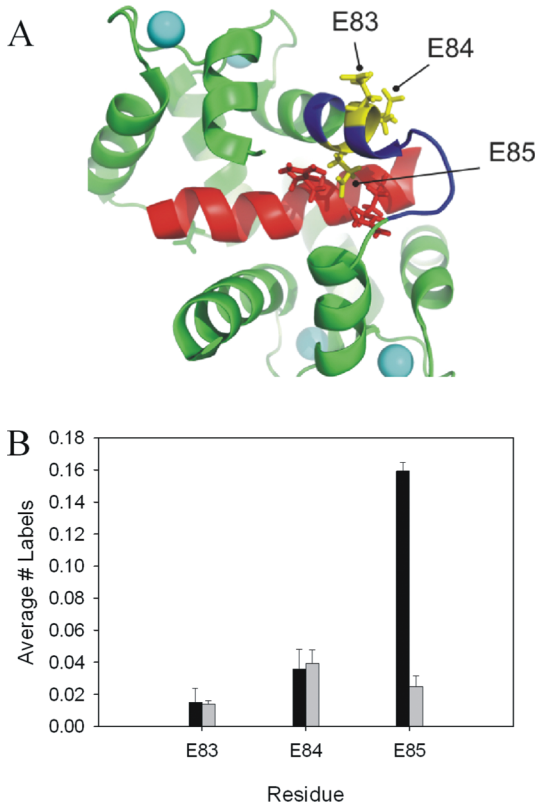


Figure 3. Labeling of holo-CaM and holo-CaM/M13 using reagent **1** (2-amino-4,4-azipentanoic acid). (A) Structure highlighting E85 in the binding site, along with E83 and E84 that remain solvent-exposed upon binding (shown in yellow). Holo-CaM is shown in green, M13 is shown in red, and the tryptic peptide MKDTDSEEEIR (77−87) is shown in blue. Bound Ca^{2+} ions are colored cyan. Side chains for two M13 arginines flanking E85 are also shown. (B) Absolute label incorporation of holo-CaM (black bars) and the M13/holo-CaM complex (gray bars), based on ETD fragmentation of MKDTDSEEEIR. Error bars represent ±1 SD.

C−H bonds, this would not generate a cleavable modification at the vibrational energies achievable in HCD. This apparent bias is explored below.

ETD is the preferred fragmentation method for label distribution analysis because of the issues associated with neutral loss in HCD. Even with ETD, however, the accuracy of the site-specific quantitation strategy requires that labeling of a peptide does not alter its rate of fragmentation. Although this cannot be proven with the data shown thus far, all peptides fragmented by ETD have well-behaved internal distributions of the label. No negative slopes were observed, which might be expected to occur with some frequency if labeling altered the distribution of radicals on the peptide backbone. The CID-based methods retain utility if neutral loss is minimal, or perhaps when applied in a differential manner to map induced changes in labeling. Such data should obviously be interpreted with caution.

Interaction Mapping at Higher Resolution. We then used ETD to determine if reliable interaction mapping could be achieved, keeping in mind an apparent bias toward acidic residues in the labeling. The complexation of M13 peptide to holo-CaM involves the peptide's insertion into an inducible channel that is predominantly hydrophobic, but with partial electrostatic character.[27] Three glutamate residues are found within the tryptic fragment MKDTDSEEEIR (77−87), only one of which participates in the M13 binding site (E85). E83 and E84 remain solvent-exposed as do residues D79 and D81 (Figure 3A). Holo-CaM with and without saturating amounts of M13 was labeled using reagent **1**, and the label distribution was measured for this tryptic fragment in both cases (Figure 3B). Labeling was reduced only at E85, consistent with protection of this residue in the structural data for the complex. Similar results were obtained with tryptic peptide HVMTNLGEK (108−116), where an HCD-based analysis localized a change in labeling to E115. This residue forms a salt bridge with a lysine in the M13 peptide ligand (Supporting Information Figure S6). These examples accumulate additional observations of an apparent residue bias, in favor of selective labeling at glutamate residues. The aspartate residues in MKDTDSEEEIR are not labeled at all even though they are solvent-exposed, and E115 contributes ∼80% of labeling in HVMTNLGEK (108−116), even though most of this peptide is solvent-exposed.

Labeling Site Distribution: Effect of Probe Chemistry. It was noted in our earlier study that labeling yields for holo-CaM using reagent **1** were higher than anticipated, based on a simple statistical model of reagent distribution at the protein surface.[13] These calculations avoided any surface partitioning terms, and carbene diffusion through solution was disallowed. Experimental labeling was ∼4 times greater than the level calculated by this method. On the basis of the observations in Figures 1 and 2, a bias toward aspartate and particularly glutamate residues could rationalize higher labeling, as 25% of the protein sequence is composed of acidic residues. We hypothesized that electrostatic interactions between protein carboxylic acids and reagent amino groups were responsible for reagent preconcentration at the protein surface, as the former will be negatively charged and the latter zwitterionic under neutral pH conditions. To test this we synthesized 4,4-azipentanoic acid (**2**), which is negatively charged at neutral pH, and labeled holo-CaM under otherwise identical conditions. Labeling was ∼1.5 labels per protein, somewhat lower than the 2 labels per protein using reagent **1**. Label distribution on the protein was markedly different, when viewed at a peptide-level resolution (Figure 4A). VFDKDGNGYISAAELR (92−107) is singly labeled at a high level (0.46 ± 0.01 labels/peptide). This peptide and two others were sequenced using ETD to localize the label, and the data were compared to label levels using reagent **1** (Figure 4B). Full sequence reads were not possible, but it is clear that reagent **2** does not label glutamates nearly as well as reagent **1**. Conversely, higher labeling is seen at lysine and tyrosine using reagent **2**.

The results in Figure 4 are consistent with the idea that electrostatic interactions at the protein surface influence both yield and distribution of the label, although the high yield on Y100 was initially puzzling. We inspected the structural properties of holo-CaM in the vicinity of Y100 and noted that this residue falls between two Ca^{2+} coordination sites, defined by the EF3 and EF4 hands in the C-domain of the protein.[28,29] The bipyramidal coordination sphere around Ca^{2+} in the EF3 hand is known to involve the oxygen of Y100 at one position, although structurally it could also coordinate the Ca^{2+} in the EF4 hand (Figure 4C). It is conceivable that reagent **2** displaces Y100 at the EF3 coordination site, or is perhaps concentrated in the region by coordinating Ca^{2+} in the EF4 hand, bringing the diazirine functional group very close to Y100 prior to photolysis. A mechanism based on electrostatically

dx.doi.org/10.1021/ac300120z | Anal. Chem. 2012, 84, 4411−4418

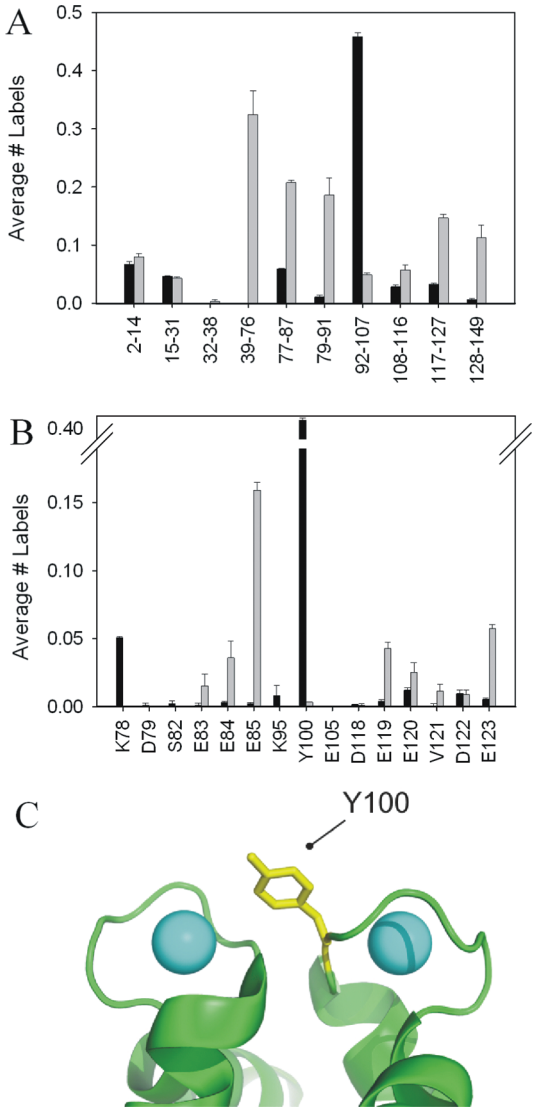


Figure 4. (A) Labeling levels using reagent **1** (2-amino-4,4-azipentanoic acid, gray bars) and reagent **2** (4,4-azipentanoic acid, black bars) for the set of 10 CaM tryptic peptides. (B) Residue-level reagent incorporation for select residues, based on ETD fragmentation of MKDTDSEEEIR (77−87), VFDKDGNGYISAAELR (92−107), and LTDEEVDEMIR (117−127) with **1** (gray bars) and **2** (black bars). Error bars are ±1 SD. (C) Location of Y100 (yellow) between two Ca^{2+} ions (cyan spheres).

driven preconcentration of the reagent would therefore seem consistent with the observation of high labeling yields at this residue. However, this interpretation does imply that the labeling of "nearest neighbor" residues is possible, without requiring a direct interaction with the reagent and further suggests that solvating a protein with a reagent could induce a structural or dynamical perturbation, depending on reagent type and concentration. The labeling of V122 using reagent **1** (Figure 1B) is also consistent with a nearest neighbor interpretation. This hydrophobic residue sits in a strongly acidic region of structure, which could permit preconcentration of the reagent based on interactions with its amino group.

To further test the role of electrostatics in preconcentrating the reagent at the protein surface, we determined the labeling yields at low and high ionic strength (Supporting Information Figure S7). Labeling was reduced significantly at high ionic strength and throughout the protein structure, with some exceptions. For example, the interdomain helical region (partially represented by peptides 77−87 and 79−91) is strongly acidic and labeling is suppressed by over 75%. However, a peptide with high labeling at E115 in HVMTNLGEK (108−116) actually increased slightly in yield in the high ionic strength experiment, which indicates that reagent partitioning governed by electrostatics provides only a partial basis for an understanding of reagent distribution.

Labeling Site Distribution: Carbene Reactivity. We next considered selectivity in carbene reactivity, as a possible explanation for a bias toward glutamate. Carbenes are short-lived intermediates, where the spin state of the unshared electrons (singlet or triplet) plays a role in the carbene lifetime, the chemistry of insertion and ultimately the distribution of products.[15,16] For alkyl carbenes of the type generated in this study, the difference in free energy between states is minimal[30] and reactivity is dominated by the singlet state, even though the ground state is triplet.[16,31] The singlet state has a preference for polar protic bond insertion over C−H insertion; therefore, it is conceivable that preferred insertion into the carboxylate group contributes to the bias we observe.

We formed low-temperature matrixes by plunge-freezing buffered solutions of reagent **1** and holo-CaM to control the diffusion of the carbene upon its generation. The frozen samples remained transparent to 355 nm radiation, allowing for protein labeling in the frozen state (Figure 5A). Yields at reduced temperature were comparable to those at room temperature in many cases, although some regions show increased labeling (e.g., peptide 92−107) or decreased labeling (e.g., peptide 79−91). The distribution was also affected. LTDEEVDEMIR (117−127) was labeled at 195 K to approximately the same level as room temperature, but labeling was higher at D119 and V122 and lower at E124 (Figure 5B). The distribution was again different at 77 K, most notably at V122. The label data for LTDEEVDEMIR (117−127) was plotted as a function of solvent-accessible surface area for the three temperatures (Figure 5C). On the basis of goodness of fit, a linear relation holds only at the lowest labeling temperature. We chose to highlight this phenomenon with LTDEEVDEMIR as it presents good ETD fragment data, and the segment sits in a stable region of structure. The correlation will be explored more fully in a future study using a wider set of proteins and reagents.

The retention of labeling and its redistribution under low-temperature conditions points to effects at or very near the surface of the protein. If we assume that the average protein conformation in this region of structure is minimally perturbed upon rapid freezing (which is reasonable based on the success of cryoelectron microscopy for structure determination), then the altered distribution at low temperature reflects reduced diffusion of the labeling agent, that is, the carbene generated upon photolysis. In other words, the half-life of the carbene must be sufficiently long to permit diffusion and capture at sites of higher reactivity under room-temperature conditions. This is not inconsistent with the ultrafast reaction rates of singlet alkyl carbenes, which are on the order of 10^7−10^9 M^{-1} s^{-1}.[16] Such rates approach the diffusion limit, but similarly high rate constants for DNA−protein interactions have been measured, even though the protein must "search" the sequence for its target binding site.[32] We therefore propose that the carbenes generated upon photolysis are sufficiently long-lived to sample their immediate environment. The high reactivity with

dx.doi.org/10.1021/ac300120z | Anal. Chem. 2012, 84, 4411−4418

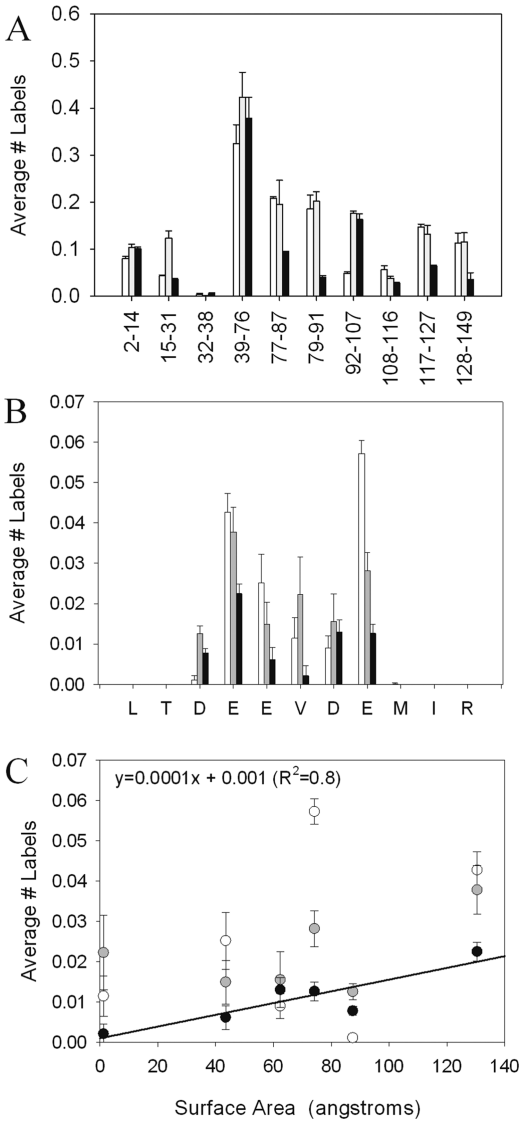


Figure 5. Effect of temperature on the labeling of holo-CaM with reagent **1** (2-amino-4,4-azipentanoic acid) showing (A) the labeling yield measured for the tryptic peptides at three different temperatures, (B) the label distribution for LTDEEVDEMIR (117−127), and (C) the corresponding relation to solvent accessible surface area in holo-CaM for the six residues in this peptide with measurable yields. All white bars and symbols denote 273 K, all gray bars and symbols denote 195 K, and all black bars and symbols denote 77 K. A straight line was fit to the 77 K data in panel C. Error bars represent ±1 SD.

glutamate at room temperature must still be a combination of preferred association and intrinsic reactivity, however. Both are required to rationalize why aspartate is not labeled as extensively as glutamate. Extrusion of nitrogen from the reagent would appear to generate a carbene with an affinity for glutamate that is even higher than the reagent itself (see labeling yields for D119 and E124 in Figure 5B, for example).

Only when the "molecular cage" of the carbene is limited in size by temperature reduction does it appear that labeling yields match surface accessibility (Figure 5C). We suggest that, under equilibrium conditions, reagent distribution is dominated by an electrostatic interaction term that does not discriminate between a carboxylate from aspartate or glutamate, and therefore the distribution of labeling on these residues could be governed primarily by solvent accessibility. At sufficiently low temperature, the indiscriminate nature of the singlet carbene will permit a reaction in the local structural environment, whatever the residue happens to be. When carbene diffusion is permitted at higher temperature, it redistributes to sites of higher reactivity, eroding a correlation with surface accessibility as we observe.

As one further test, we explored the effect of both increased ionic strength and temperature reduction on the internal distribution of label for peptide 108−116 (HVMTNLGEK). Labeling is dominated by residue E115 at low ionic strength, which could not be suppressed at high ionic strength as noted above. Reducing the temperature to 77 K at low ionic strength did not alter the distribution either. However, labeling at both low temperature and high ionic strength strongly altered the distribution (Supporting Information Figure S8). It approached the limit described by unbiased labeling of residues for an unstructured peptide, at least for residues TNLGE. This is consistent with the high solvent exposure of this region in holo-CaM structure and confirms that reagent preconcentration and diffusion of the carbene to sites of higher reactivity both influence label distribution.

■ CONCLUSIONS AND IMPLICATIONS

The highly reactive carbene labels can be effectively localized to individual residues within a protein, using a simple quantitation strategy that integrates across all congeners of singly labeled peptides. The approach was found to be reliable in the current example, generating label distributions that suggest a minimal effect of label position on peptide fragmentation efficiency, provided that ETD is used. Collisional activation leads to sequence-dependent neutral loss, which may be attributed to a labeling bias toward carboxylate groups, principally on glutamate. This bias results from a complex relationship involving the distribution of the "preactivated" reagent (the diazirine) and a redistribution of the "active" reagent (the carbene). The bias is readily tuned by chemically altering the surface-targeting properties of the diazirine. The apparent preference of singlet carbene for carboxylates can be controlled by temperature at the very least, and probably through substituent effects as well. The two sources of labeling bias presented in this study may not be fully uncoupled obviously, as redistribution of the carbene would also be sensitive to local noncovalent interactions. Within the bias defined in this study, carbenes can support typical applications of protein footprinting, as an analysis of the CaM−M13 interaction demonstrates.

A set of diazirine reagents with different protein surface activities should be an effective strategy for addressing labeling bias for the most part and permit high-yield mapping of all solvent-accessible residues. Reagent design could also simplify differential labeling experiments, for example, through the incorporation of isotope coding strategies. The diffusible nature of the carbene represents the most interesting feature of this footprinting strategy, and perhaps the most significant limitation to higher-resolution mapping of proteins and their interactions. If the preferred surface chemistry for the diazirine reagent differs from the preferred site of carbene reactivity, then carbene diffusion will scramble diazirine surface populations. This diffusion must be limited to the immediate surface of the protein, because any excursion into bulk solvent should quench the carbene, given its preference for polar protic bonds. (In our view, a similar concern must also be raised for any footprinting strategy where the active agent is generated from a distinct

dx.doi.org/10.1021/ac300120z | Anal. Chem. 2012, 84, 4411−4418

chemical species, such as in hydroxyl radical footprinting by synchrotron radiation or laser photolysis.) Rapid sample freezing can limit this effect, but it is unclear if this will present a general solution to the problem. It certainly raises the possibility of measuring distances between sites, provided that the half-life of the carbene in aqueous solution is known. Nevertheless, the high degree of control over label distribution demonstrates that carbene chemistry is a very useful addition to the collection of protein chemistries in the emerging field of structural mass spectrometry.

■ ASSOCIATED CONTENT

Supporting Information

Additional information as noted in text. This material is available free of charge via the Internet at http://pubs.acs.org.

■ AUTHOR INFORMATION

Corresponding Author

*Phone: 403-210-3811. Fax: 403-283-8727. E-mail: dschriem@ucalgary.ca.

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

We thank Dr. Yujun Shi (Department of Chemistry, University of Calgary) for access to the Nd:YAG laser. This work was funded by CFI (Canada Foundation for Innovation) and NSERC (Natural Sciences and Engineering Research Council of Canada). D.C.S. is supported by a Canada Research Chair in Chemical Biology and a Senior Heritage scholarship from Alberta Ingenuity—Health Solutions.

■ REFERENCES

(1) Kiselar, J. G.; Mahaffy, R.; Pollard, T. D.; Almo, S. C.; Chance, M. R. *Proc. Natl. Acad. Sci. U.S.A.* **2007**, *104*, 1552−1557.

(2) Sharon, M. *J. Am. Soc. Mass Spectrom.* **2010**, *21*, 487−500.

(3) Ben-Nissan, G.; Sharon, M. *Chem. Soc. Rev.* **2011**, *40*, 3627−3637.

(4) Mendoza, V. L.; Vachet, R. W. *Mass Spectrom. Rev.* **2009**, *28*, 785−815.

(5) West, G. M.; Tang, L.; Fitzgerald, M. C. *Anal. Chem.* **2008**, *80*, 4175−4185.

(6) D'Ambrosio, C.; Talamo, F.; Vitale, R. M.; Amodeo, P.; Tell, G.; Ferrara, L.; Scaloni, A. *Biochemistry* **2003**, *42*, 4430−4443.

(7) Kamal, J. K. A.; Chance, M. R. *Protein Sci.* **2008**, *17*, 79−94.

(8) Nuss, J. E.; Sweeney, D. J.; Alter, G. M. *Biochemistry* **2009**, *48*, 7892−7905.

(9) Konermann, L.; Pan, J. X.; Liu, Y. H. *Chem. Soc. Rev.* **2011**, *40*, 1224−1234.

(10) Xu, G.; Chance, M. R. *Chem. Rev.* **2007**, *107*, 3514−3543.

(11) Gau, B. C.; Sharp, J. S.; Rempel, D. L.; Gross, M. L. *Anal. Chem.* **2009**, *81*, 6563−6571.

(12) Tong, X.; Wren, J. C.; Konermann, L. *Anal. Chem.* **2007**, *79*, 6376−6382.

(13) Jumper, C. C.; Schriemer, D. C. *Anal. Chem.* **2011**, *83*, 2913−2920.

(14) Griller, D.; Nazran, A. S.; Scaiano, J. C. *Acc. Chem. Res.* **1984**, *17*, 283−289.

(15) Turro, N. J.; Cha, Y.; Gould, I. R. *J. Am. Chem. Soc.* **1987**, *109*, 2101−2107.

(16) Das, J. *Chem. Rev.* **2011**, *111*, 4405−4417.

(17) Richards, F. M.; Lamed, R.; Wynn, R.; Patel, D.; Olack, G. *Protein Sci.* **2000**, *9*, 2506−2517.

(18) Craig, P. O.; Ureta, D. B.; Delfino, J. M. *Protein Sci.* **2002**, *11*, 1353−1366.

(19) Ureta, D. B.; Craig, P. O.; Gomez, G. E.; Delfino, J. M. *Biochemistry* **2007**, *46*, 14567−14577.

(20) Gomez, G. E.; Mundo, M. R.; Craig, P. O.; Delfino, J. M. *J. Am. Soc. Mass Spectrom.* **2012**, *23*, 30−42 . DOI: DOI: 10.1007/s13361-011-0266-x.

(21) Toscano, J. P.; Platz, M. S.; Nikolaev, V. *J. Am. Chem. Soc.* **1995**, *117*, 4712−4713.

(22) Medzihradszky, K. F.; Campbell, J. M.; Baldwin, M. A.; Falick, A. M.; Juhasz, P.; Vestal, M. L.; Burlingame, A. L. *Anal. Chem.* **2000**, *72*, 552−558.

(23) Suchanek, M.; Radzikowska, A.; Thiele, C. *Nat. Methods* **2005**, *2*, 261−267.

(24) Syka, J. E.; Coon, J. J.; Schroeder, M. J.; Shabanowitz, J.; Hunt, D. F. *Proc. Natl. Acad. Sci. U.S.A.* **2004**, *101*, 9528−9533.

(25) Good, D. M.; Wirtala, M.; McAlister, G. C.; Coon, J. J. *Mol. Cell. Proteomics* **2007**, *6*, 1942−1951.

(26) Olsen, J. V.; Macek, B.; Lange, O.; Makarov, A.; Horning, S.; Mann, M. *Nat. Methods* **2007**, *4*, 709−712.

(27) Clore, G. M.; Bax, A.; Ikura, M.; Gronenborn, A. M. *Curr. Opin. Struct. Biol.* **1993**, *3*, 838−845.

(28) Chattopadhyaya, R.; Meador, W. E.; Means, A. R.; Quiocho, F. A. *J. Mol. Biol.* **1992**, *228*, 1177−1192.

(29) Fallon, J. L.; Quiocho, F. A. *Structure* **2003**, *11*, 1303−U1307.

(30) Nemirowski, A.; Schreiner, P. R. *J. Org. Chem.* **2007**, *72*, 9533−9540.

(31) Sander, W. W.; Patyk, A.; Bucher, G. *J. Mol. Struct.* **1990**, *222*, 21−31.

(32) Halford, S. E. *Biochem. Soc. Trans.* **2009**, *37*, 343−348.