analytical chemistry

Article

pubs.acs.org/ac

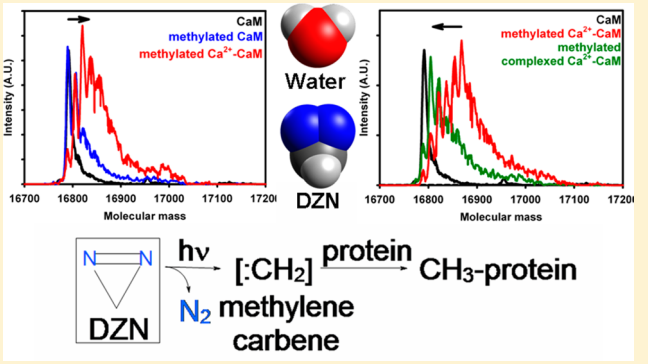
# Solvent Mimicry with Methylene Carbene to Probe Protein Topography

Gabriela Elena Gómez, José Luis E. Monti, Mariana Rocío Mundo, and José María Delfino*

Departamento de Química Biológica, Facultad de Farmacia y Bioquímica, Universidad de Buenos Aires e Instituto de Química y Fisicoquímica Biológicas (IQUIFIB-CONICET). Junín 956, C1113AAD Buenos Aires, Argentina

Supporting Information

**ABSTRACT:** The solvent accessible surface area (SASA) of the polypeptide chain plays a key role in protein folding, conformational change, and interaction. This fundamental biophysical parameter is elusive in experimental measurement. Our approach to this problem relies on the reaction of the minimal photochemical reagent diazirine (DZN) with polypeptides. This reagent (i) exerts solvent mimicry because its size is comparable to water and (ii) shows scant chemical selectivity because it generates extremely reactive methylene carbene. Methylation gives rise to the EM (extent of modification) signal, which is useful for scrutinizing the conformational change triggered by $Ca^{2+}$ binding to calmodulin (CaM). The increased EM observed for the full protein is dominated by the enhanced exposure of hydrophobic area in $Ca^{2+}$-CaM. Fragmentation allowed us to quantify the methylene incorporation at specific sites. Peptide 91−106 reveals a major reorganization around the calcium 151 binding site, resulting in local ordering and a greater exposure of the hydrophobic surface. Additionally, this technique shows a high sensitivity to probe recognition between CaM and melittin (Mel). The large decrease in EM indicates the occlusion of a significant hydrophobic area upon complexation. Protection from labeling reveals a larger involvement of the N-terminal and central regions of CaM in this interaction. Despite its smaller size, Mel's differential exposure can also be quantified. Moreover, MS/MS fragmentation realizes the goal of extending the resolution of labeled sites at the amino acid level. Overall, DZN labeling emerges as a useful footprinting method capable of shedding light on physiological conformational changes and interactions.




The solvent accessibility of the polypeptide chain underlies fundamental processes such as protein folding, conformational change, and interaction with partners. However, despite its great importance in protein science,[1] there is hardly any experimental method suitable for approaching a direct measurement of the solvent accessible surface area (SASA).

Among the techniques most successfully applied toward this aim is H/D exchange,[2−6] which addresses the accessibility of exchangeable protons with negligible modification in chemical nature. Despite its strong dependence on secondary structure and the labile nature of the label, protection factors derived from kinetic data have been extensively employed to describe the features of local environments. Other widely used footprinting techniques, based on hydroxyl radical (•OH) reactions, have proven valuable in exploring the structure and interactions of nucleic acids[7,8] and proteins.[9−11] To address the issue of chemical selectivity, current efforts counsel establishing a protection factor analysis based on reaction rate constants.[12] However, such results can be hampered by oxidative damage-induced conformational changes.[13]

Our approach to the problem of SASA measurement is based on the reaction of the polypeptide chain with the photochemical reagent diazirine (DZN).[14,15] DZN is a soluble gas (up to 20−25 mM in aqueous solvents) that generates the extremely reactive methylene carbene ($:CH_2$) upon irradiation at $\lambda > 300$ nm. Consequently, methylation of its molecular cage occurs, i.e., insertion into any X−H bond (where X = C, O, N, or S).[16,17] In this scenario, the technique has demonstrated minimal chemical selectivity at the peptide level with respect to the quality of the amino acid side chains involved at each target site.[18,19] Because of its similar size to water, unsubstituted DZN cannot only probe the outer solvent's exposed surface in atomic detail but also reveal inner surface components such as cavities, crevices, and channels, resulting in effective molecular mimicry of the aqueous solvent. Nevertheless, because of its mild apolar character, the protein solvation properties of DZN likely differ from those of water, conferring the molecule a bias toward labeling hydrophobic surfaces.[15,18−21] Another carbene labeling approach aimed at analyzing protein topography is based on the water-soluble amino acid "photoleucine" and its deaminated derivative.[22,23] The authors rightly discuss complexities related to issues such as charge, amphipathic character, and orientation effects. These factors influence the distributions of the parent

**Received:** July 20, 2015
**Accepted:** September 8, 2015





DOI: 10.1021/acs.analchem.5b02724
*Anal. Chem.* XXXX, XXX, XXX−XXX

reagent and the transient carbene intermediate, ultimately determining the labeling phenomenon. In our case, methylation of the target protein represents the least conceivable chemical perturbation, yielding products with a physicochemical character that is minimally different from that of their unreacted counterparts. With respect to MS analysis, equivalent ionization efficiency of modified and unmodified peptides would be expected. Introducing a stable methyl tag allows a wide range of analytical procedures to be used for subsequent product processing. In addition, DZN labeling is essentially free from damaging consequences to protein conformation that could give rise to artifacts or at least complicate the interpretation of the data.

In this work, we provide the very first description of the utility of the DZN labeling approach combined with mass spectrometry to probe ligand binding-induced conformational change and complex formation in a protein (bovine brain calmodulin). As will be presented, matrix-assisted laser desorption ionization-time-of-flight mass spectrometry (MALDI-TOF) analysis of methylated products (including MS/MS peptide fragmentation) can be profitably used to obtain relative quantitative estimates of the extent of reaction at individual site resolution. Therefore, the point of the analysis is its evaluation of particular sites across different conformational states or across the process of complexation. In addition, because of its scant dependence on amino acid composition, the comparison among different reactive sites in a given conformation or association state of a protein or peptide also creates an addressable issue. We believe that because DZN labeling has the ability both to probe solvent accessibility and to describe local environments, it adds its own value to the repertoire of solution methods designed to extract topographical information about proteins.

## EXPERIMENTAL SECTION

**Materials.** Formaldehyde (37%, w/v) was purchased from E. Merck (Darmstadt, Germany). Formamide, urea, and freeze-dried bee venom were purchased from Sigma Chemical Co. (St. Louis, MO). TPCK trypsin from bovine pancreas was purchased from Worthington Biochemical Corporation (Lakewood, NJ). Calmodulin (CaM) and melittin (Mel) were purified from fresh bovine brain and bee venom, respectively (see the Supporting Information). Urea was recrystallized from ethanol before use. Acetonitrile was purchased from E. Merck (Darmstadt, Germany), and trifluoroacetic acid (TFA) was purchased from Riedel de Haën (Seelze-Hannover, Germany); both were HPLC grade. All other reagents and chemicals used were analytical grade.

**Photolabeling Procedure.** DZN was synthesized and quantified as previously described.[19] Briefly, DZN gas is generated into a 100 mL glass syringe from the precursor salt (methylene diammonium sulfate) mixed with an alkaline solution of sodium hypochlorite. DZN is (i) continuously infused into the stirred protein solution placed into a capped quartz cuvette at room temperature and (ii) simultaneously irradiated with light from an arc source (1000 W Hg/Xe lamp, Oriel code 6295) equipped with a cutoff filter (Oriel code 59044, $\lambda < 300$ nm). Consistently, neither infusing DZN into the sample nor the exposure to filtered UV light causes any deleterious effect on protein conformation, as ascertained by circular dichroism spectroscopy. The DZN concentration in aqueous solution can be accurately determined by measuring the absorbance of the dissolved gas at 320 nm using an extinction coefficient of $\varepsilon_{320} = 180\ M^{-1}\ cm^{-1}$.[15] See ref 19 for a full description of the photolysis procedure, including a diagram of the experimental setup and the equations describing the evolution of DZN concentration during infusion and photolysis.

**Processing of DZN Labeled Samples.** Labeled samples of apo and holo CaM were analyzed using electrospray ionization mass spectrometry (ESI-MS) with no further purification. To dissociate labeled samples of Mel-complexed CaM, both solid urea (up to 8 M) and EDTA (up to 5 mM) were added. The protein and peptide were separated using size-exclusion chromatography (SEC-FPLC) on a Superdex-75 column (GE Healthcare Life Sciences) run at a 0.60 mL/min flow rate in 50 mM Tris-HCl pH 8.0/4 M urea/100 mM NaCl/1 mM EDTA. The elution of the proteins was monitored by UV absorption at 280 nm. Isolated CaM samples were dialyzed against $(NH_4)HCO_3$ (0.07 mM, pH 7.5−8.0) and freeze-dried. Isolated Mel samples were run on a C4 column (Vydac 214TP54, 4.6 mm × 250 mm) at a 1 mL/min flow rate, with the gradient used for Mel purification (see the Supporting Information). For the sake of comparison, labeled samples of free CaM and free Mel were treated in the same fashion. After these procedures, the samples were analyzed using ESI-MS.

**ESI-MS.** DZN-reacted and control protein samples were routinely run on an LC−ESI-MS platform consisting of a Thermo Surveyor HPLC system (C8 column, Vydac 208TP5105, 50 mm × 1 mm) coupled to an electrospray mass spectrometer (Thermo Finnigan LCQ duo, equipped with an ion trap mass analyzer). All the ESI spectra shown have been deconvoluted to the zero-charge domain by the ProMass Deconvolution Software (Thermo Scientific) using a standard parameter set: average mass type, mass tolerance 0.02%, minimum tolerance 2 Da, relative impurity threshold 30%, input $m/z$ range 300−2000 units, adduct ion mass 1.0079, baseline removal set at 0.8, peak width 1, merge width 0.3, minimum score 2, normalized scores 1, comprehensive deconvolution set to on, smooth width 5, number of smooths 2, and a noise threshold set at 0.1% relative intensity.

**Tryptic Digestion.** Labeled and control CaM samples were completely digested with TPCK trypsin in $(NH_4)HCO_3$ buffer (0.1 M, pH 8.5), after 12−24 h at 37 °C, using a 2% (w/w) enzyme/substrate ratio. The tryptic mixture was desalted using a micro ZipTip μC18 (EMD Millipore, Merck KGaA, Darmstadt, Germany) and analyzed by MALDI-TOF.

**MALDI-TOF MS.** Peptide samples (0.5 μL) dissolved in 60% acetonitrile, 0.1% TFA in water were allowed to dry. Next, a saturated matrix solution (0.5 μL) of α cyano-4-hydroxycinnamic acid in acetonitrile:water (70:30 by volume, 0.1% TFA) was added to the spots and allowed to dry. The samples were ablated using a pulsed Nd:YAG laser (355 nm) at a nominal power of 3400 arbitrary units. All the MALDI spectra were recorded on an Applied Biosystems MALDI TOF/TOF 4800 Plus mass spectrometer operating in the reflector mode by applying an accelerating voltage of 20 kV and a delayed extraction time set to 375 ns. Each spectrum represents the average of 500 laser shots and is externally calibrated. The unambiguous identification of tryptic peptides was achieved after MALDI-MS/MS and postsource decay (PSD) fragmentation over selected ions at a laser power of 4400 arbitrary units. All the MS analyses were performed at our local protein facility (LANAIS-PROEM, UBA-CONICET).


DOI: 10.1021/acs.analchem.5b02724
Anal. Chem. XXXX, XXX, XXX−XXX

**Extent of Modification with Methylene Carbene (EM).** The extent of reaction of the target protein with methylene carbene (:$CH_2$) is quantified by the metric EM:[19]

$$\mathrm{EM} = \sum_{i=0}^{i=n} iI_i / \sum_{i=0}^{i=n} I_i \quad (1)$$

where $I_i$ stands for the intensity of the peak including $i$ methylene groups per protein molecule. The intensity of a given peak $i$ is approximated as the area under a Gaussian function centered at molecular mass $X_i$. Note that $X_0$ (= $M$) corresponds to the molecular mass of the unmodified species. To allow meaningful comparisons of results from independent experiments, EM values should be normalized by the load of DZN reagent[19] and expressed as moles of:$CH_2$ incorporated per mole of protein. One should note that if the peak intensity pattern obeys a Poisson distribution $P(k,\lambda) = \lambda^k\, e^{-\lambda}/k!$, EM equals $\lambda$, with the single parameter representing both the mean and the variance.

Estimating EM for the full protein enables setting the modification regime to detect methylation products at the peptide level. Here, where only peaks corresponding to monomethylated species are expected, eq 1 readily yields the following:

$$\mathrm{EM}\,\% \approx 100 \times I_1/(I_0 + I_1) \quad (2)$$

**Analysis of Methylene Incorporation at Amino Acid Resolution.** MS/MS fragmentation (PSD in MALDI-TOF) of selected monomethylated tryptic peptides was conducted to identify sites of modification at amino acid resolution. Methylene insertion, mostly occurring at side-chain locations, as ascertained by recent NMR evidence (data not shown), is expected to exert a minimal effect on fragmentation yields. The basic tenet holds that REM, representing the observed relative probability of reaction of a given fragment of length $n$, equals the sum of independent probabilities $P_j$ of reaction at each contiguous amino acid site $j$:

$$\mathrm{REM}_{y_i} = \sum_{j=n-i+1}^{n} P_j \mathrm{REM}_{b_i} = \sum_{j=1}^{i} P_j i \le n \quad (3)$$

This assumption is warranted whenever a sufficiently low overall extent of reaction is achieved, meaning that the likelihood of having two neighboring sites modified in a given molecular entity becomes null, thus eliminating the possibility of mutual influence. In the absence of technical MS issues that could bias fragmentation yields, such as neutral losses (see discussion in ref 23), defined "stair" patterns are indeed observed for relative EM (REM) values along series of ions derived from isolated monomethylated species. To pinpoint labeling sites at a single amino acid resolution, we calculated REM values for each identifiable fragment along the y, b (or a) series of ions. Following a global fit analysis, individual $P_j$ values can be derived for the whole set of REM data by minimizing the overall sum of the squared deviations between the predicted and experimentally determined REM values for all the identified fragments. To compose a full map of methylene reactivity for the protein involving all available peptides, the REM values for any given peptide should be multiplied by its overall EM (cf. ref 23). An Excel spreadsheet executing a macro programmed in Visual BASIC was used for this purpose.

**Molecular Modeling.** Calculations of SASA (including polar and nonpolar surfaces) based on available high-resolution structures were performed with Surface Racer[24] using van der Waals atomic radii according to the parameter set defined by Richards[1] with a 1.4 Å probe radius selected for the water molecule. The structures used are the following: apo CaM (PDB code 1cfd[25]), holo CaM (1cll[26]), the complexes of CaM with CaM-dependent protein kinase II-$\alpha$ (CaM-CaMKII$\alpha$, 1cm1[27]) and CaM-binding domain of skeletal muscle myosin light chain kinase (CaM-MLCK, 2bbm[28]), the homo tetramer of Mel (2mlt[29]) and the mutant Mel with D-Pro at position 14 (1bh1[30]). In addition, the SASA of the models of fully exposed conformational ensembles of MLCK and Mel were calculated using ProtSA.[31]

## RESULTS AND DISCUSSION

**DZN Labeling Reveals Protein Conformational Change.** The purpose of this study is to develop a method with the ability to distinguish conformational states and interactions from the perspective of the protein surface exposed to aqueous solvent (SASA). Given the wealth of structural information available on CaM, our test case involved the change experienced by the protein in the presence of $Ca^{2+}$ (Figures S-1 and S-2). This work represents the first use of DZN labeling to sense conformational changes occurring under physiological conditions.

At the level of the full-length protein, ESI-MS allowed us to distinguish methylated products as individual $M + i \times 14$ peaks, where index $i$ points to the number of methylene groups incorporated per protein molecule. The spectrum of a DZN-modified sample of CaM (Figure 1), including all available protein species eluting after RP HPLC, reveals a pattern of peaks headed by one that corresponds to the unmodified species ($M$ = 16 791 Da) and is consistent with a separation of 14 amu.

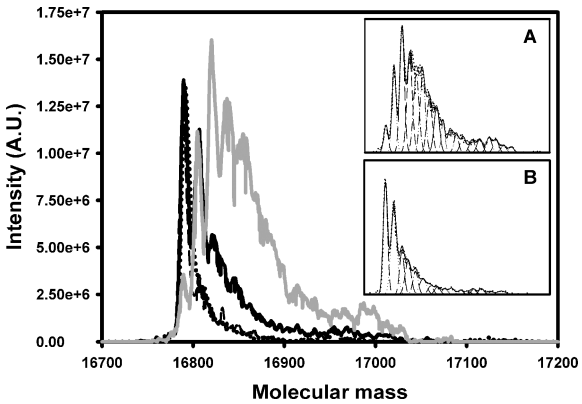



**Figure 1.** ESI spectra of CaM (60 $\mu$M) incubated with 1.5 mM $CaCl_2$ (gray line and inset A) or 0.1 mM EDTA (solid black line and inset B) in 20 mM Tris-HCl, pH 7.5 labeled with DZN for 6 min (simultaneous dissolution and photolysis of the gas at a flux of 1.0 mL/min). Controls, untreated CaM (dotted black line) and irradiated CaM in the absence of DZN (dashed black line), are also shown. Insets: a function (solid line) representing the sum of a set of Gaussian components (dashed lines) was fitted to the experimental data (dotted line). No significant data above the threshold value (set as zero in the graph) exist outside the range shown.

By integrating these signals, a robust quantitative metric can be derived that indicates the extent of modification (EM) with


DOI: 10.1021/acs.analchem.5b02724
Anal. Chem. XXXX, XXX, XXX−XXX

methylene carbene, expressed in units of moles of $CH_2$ incorporated per mol of protein, normalized to a given concentration of DZN (see Experimental Section). The intensity pattern of molecular species satisfies a Poisson-like distribution, as expected for a general methylation phenomenon. Consistently, the evolution of this distribution as the load of DZN increases clearly shows not only the shift of the maximum but also asymmetric broadening (data not shown). In the range assayed, EM exhibits a linear dependence with the load of the reagent,[15,19,20] thus warranting a straightforward comparison among samples in a given condition for chemical modification. No detectable concurrent photolytic damage to the protein is occurring because the spectrum of a control sample irradiated in the absence of DZN is essentially identical to the spectrum of an untreated sample. The overall intensity profile of peaks is consistent with a general modification of the polypeptide chain.[14] However, there is a striking difference between CaM samples incubated with $CaCl_2$ or EDTA and labeled with an equal load of DZN. From that comparison, a large shift toward heavier mass peaks is observed in the holo form (with $Ca^{2+}$) of the protein. Label incorporation is linearly dependent on the SASA of proteins or peptides (Figures 3 and 4 in ref 18), indicating the lesser role played by chemical selectivity in the labeling event. The EM for the apo or holo forms of CaM yields values of 1.8 or 4.1 mol of $CH_2$ per mol of protein, respectively. Indeed, the total SASA values for the CaM forms differ slightly: 10 674 (1cfd, apo) and 9685 $Å^2$ (1cll, holo), representing a 9% relative decrease. Therefore, the change in the signal cannot be treated as a measure solely dependent on the extent of the geometrical static surface. Here, the labeling phenomenon is likely dominated by the greatly increased exposure of the external hydrophobic area in CaM: 3965 and 5163 $Å^2$ for the apo and holo forms, respectively, a 30% relative increase. Indeed, this characteristic is a well-known characteristic of the holo form, as has been amply demonstrated by its ability to interact both with a variety of amphipathic peptide ligands[32,33] and with hydrophobic matrixes.[34] Therefore, any difference in the labeling yield will be attributed to a change in the extent and/or nature of the SASA. Interestingly, the DZN function grafted into an amino acid scaffold, as in the case of "photoleu,"[22,23] causes a labeling shift amply favoring the apo form, most likely resulting from the heavier weight of electrostatics and orientation effects in the phenomenon.

Next, the analysis of the DZN labeling pattern along the polypeptide chain at the level of small peptides enabled us to illuminate the conformational change at higher resolution. To this end, DZN-modified samples of apo and holo CaM were analyzed by MALDI-TOF after complete digestion with trypsin. To obtain an appropriate signal, the experimental conditions for the modification reaction had to be more exhaustive than the conditions used for the full protein. Indeed, the ability to detect modified species is illustrated for peptides 91−106 and 107−126 (Figure 2 and Figure S-3).

First, as shown by comparing panels A,B and C,D, the methylated species derived from each peptide occur exclusively among DZN-treated samples. At the level of peptides of this size, only peaks corresponding to monomethylated species appear in the mass spectra. This fact is consistent with the expectation for a general labeling phenomenon to give rise to a Poisson-like distribution of products. The relative proportion of monomethylated species with respect to the unmodified counterpart provides sufficient information to calculate EM eq 2).

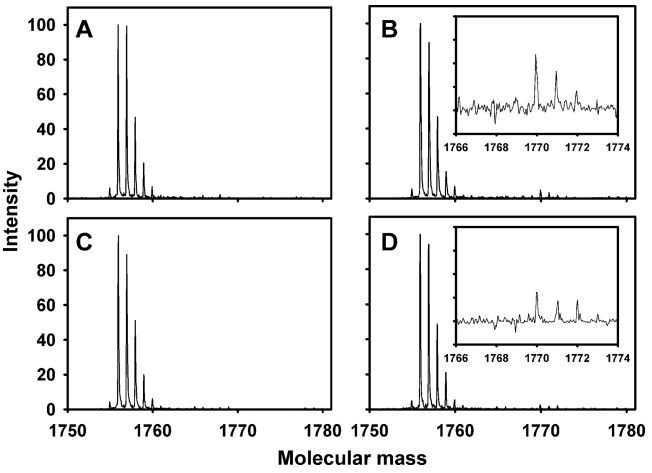



**Figure 2.** MALDI-TOF spectra of tryptic peptide ($V_{91}$FDKDGNGYISAAELR$_{106}$; $MH^+$ 1754.8) derived from holo (B) or apo CaM (D) labeled with DZN for 20 min (simultaneous dissolution and photolysis of the gas at a flux of 1.0 mL/min). Panels A and C show CaM irradiated in the absence of DZN in the holo or apo forms, respectively. The insets correspond to zooms around the range in which monomethylated species occur, plotted as the differences between samples irradiated in the presence or absence of DZN. Values of the extent of modification (EM %) were calculated from these spectra and expressed as mol of $CH_2$ incorporated per 100 mol of protein, normalized by the load of DZN reagent. EM % values are as follows: 4.1 (holo) and 2.2 (apo), yielding a holo/apo ratio of 1.86 ± 0.44.

For peptide 91−106, a holo/apo EM% ratio of 1.86 ± 0.44 is observed (Figure 2, a value that is substantially higher than expected from a static SASA calculation (0.78). This paradoxical finding points to the involvement of this peptide in building one of the binding sites for calcium (Figure 3).

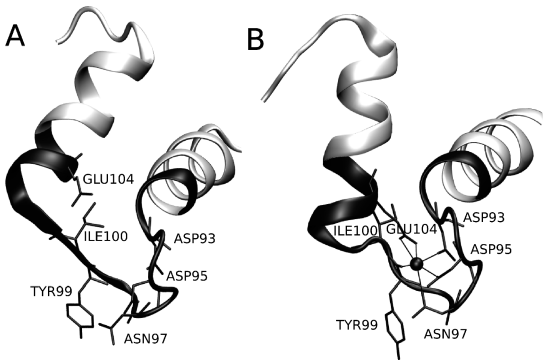



**Figure 3.** Involvement of peptide 91−106 (in black) in one of the binding sites for calcium. For structures 1cfd (apo CaM, A) and 1cll (holo CaM, B), the overall exposure of amino acid residues D93, D95, N97, and E104 changes from 14, 120, 92, and 63 $Å^2$ in the former to 7, 96, 86, and 18 $Å^2$ in the latter. In addition, upon binding of calcium 151, the relative hydrophobic exposure (hydrophobic SASA/total SASA per residue) of these same amino acids changes from 2, 29, 14, and 39% to 6, 47, 39, and 81%. The figure was rendered using VMD.[35]

A sizable number of amino acids (D93, D95, N97, Y99, I100, and E104) belonging to this peptide compose the environment around calcium 151. It is well documented that this region undergoes a major reorganization upon calcium binding. Carboxylate and carbamido groups belonging to the side chains of polar amino acids D93, D95, N97, and E104, which


DOI: 10.1021/acs.analchem.5b02724
Anal. Chem. XXXX, XXX, XXX−XXX

are exposed in the apo form, adopt an inward orientation in the holo form because they participate directly in calcium coordination. In contrast, Y99 does not change the general orientation of its side chain because only its backbone participates in calcium chelation. Consequently, in the holo form, the hydrophobic moieties of the side chains appear exposed to the solvent. Despite the somewhat lower overall exposure of this section in the holo form (approximately a 20% decrease in total SASA), this peptide appears more heavily labeled because of the increased hydrophobicity at this location. In addition, the comparison of mobility in this region, as attested to by the change in crystallographic temperature factors ($B$) and deviations from the average atomic positions measured by NMR (Figure S-4), stresses that this peptide adopts a more organized structure in the holo form, thus causing a higher persistence of hydrophobic surface exposed to the aqueous solvent. In contrast, the adjacent peptide 107−126 shows a holo/apo EM % ratio of 0.84 ± 0.16 (Figure S-3), a value consistent with the expectation based on the SASA calculation (0.75).

Further fragmentation of the tryptic peptides allowed us to evaluate sites of modification at amino acid resolution. For peptide 107−126, we advanced into the precise locale of labeling after isolating the monomethylated species ($m/z$ 2415.1) in MALDI experiments and subjecting it to PSD fragmentation (Figure S-5). The basis for this analysis lies in the quantitative evaluation of each monomethylated species along the sequence with respect to its corresponding parent (unmodified) ion (i.e., calculating EM % for each fragment), from which the probability of reaction at any individual (amino acid) site can be estimated (see Experimental Section). After a global fitting procedure, for both the apo and holo forms of CaM, the occurrence of methylated spots appears to be associated with the C-terminal third of the peptide. Remarkably, no bias toward a particular chemical selectivity is observed because this stretch of the peptide includes functional groups of very different natures: hydrophobic (I, M, V) and both negatively (D, E) and positively (R) charged amino acids.

**DZN Labeling Uncovers a Protein−Peptide Interaction.** CaM interacts with Mel to form a high affinity complex ($K_d \approx 3$ nM[36]) that has been extensively studied.[37−41] For our purpose, experimental conditions were selected that favor complex formation while avoiding the tetramerization of Mel, as verified by far-UV CD and Trp fluorescence (Figure S-6). We then undertook the DZN labeling of both the free and Mel-bound forms of CaM (Figure 4), and we found EM values corresponding to the CaM moiety of 8.5 and 4.5 mol of $CH_2$ per mol protein, respectively. This reduction (47 ± 5%) in the extent of modification likely results from the expected decrease in the geometrical SASA compounded with the occlusion of a significant amount of hydrophobic area.

Any interpretation of these results requires accounting not only for the interaction phenomenon but also for the conformational change associated with complexation. The total calculated SASA yields the following values: 9695 (1cll) and 8490 Å$^2$ (1cm1) for CaM in the free or complexed forms, respectively, amounting to a 12% decrease. In addition, nonpolar SASA also decreases significantly (18%): from 5163 (1cll) to 4213 Å$^2$ (1cm1). Overall, it becomes clear from these results that occlusion of the hydrophobic area acquires a substantial weight in the labeling phenomenon. As shown above in the case of apo and holo CaM, this effect can even override small geometrical differences in the opposite direction.

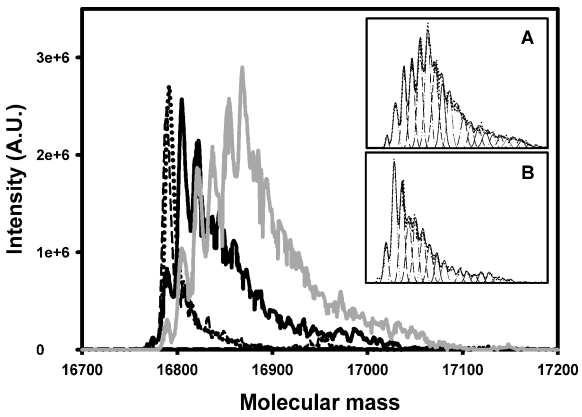



**Figure 4.** ESI spectra of free (gray line and inset A) or complexed (solid black line and inset B) CaM (50 $\mu$M) in 50 mM Tris-HCl, pH 8.0/2 mM $CaCl_2$ labeled with DZN for 20 min (simultaneous dissolution and photolysis of the gas at a flux of 1.0 mL/min). The protein/peptide ratio was 1:1. Untreated CaM (dotted black line) and CaM irradiated in the absence of DZN (dashed black line) are also shown. Insets: a function (solid line) representing the sum of a set of Gaussian components (dashed lines) was fitted to the experimental data (dotted line). No significant data above the threshold value (set as zero in the graph) exist outside the range shown.

Caution should be exerted in interpreting results exclusively based on the purely static picture provided by the time-averaged structures. This analysis must necessarily include differences that might arise from the dynamics of the structures involved. In this regard, our laboratory's previous work, which addresses the labeling of the molten globule states of *B. licheniformis* $\beta$-lactamase[20] and bovine $\alpha$-lactalbumin,[18] underscores the importance of the appearance of liquid-like permeable hydrophobic phases as a major factor boosting the measured signal.

Next, DZN-modified samples of CaM in the free or complexed forms were analyzed using MALDI-TOF after digestion with trypsin. In principle, because of the quasi-random (Poisson-like) character of the DZN modification reaction, the EM metric is expected to depend linearly on the size of the polypeptide. This trend can be inferred from the correlation between the extent of methylene labeling and the molecular weight of peptides (Figure S-7). A similar trend was previously observed both for a series of proteins of differing molecular weight[19] and for a set of tryptic peptides derived from bovine $\alpha$-lactalbumin.[18] Deviations from this general behavior become informative with respect to the particular features of the tertiary structure or quaternary association of the target protein. The ability to detect modified species at this peptide level is illustrated in Figures S-8 (1−13), S-9 (91−106), and S-10 (107−126). Table 1 summarizes the EM % values for all the recovered tryptic peptides.

Only those peptides are included that could be properly isolated and identified and for which a reliable quantitative measurement of intensity above noise could be achieved. Overall, this peptide collection represents a ∼73% sequence coverage of the full-length protein. EM % ratio values between complexed and uncomplexed forms (C/F) of less than one indicate protected regions occurring because of complex formation. In general, most peptides show values less than or close to one, as expected from the known closure of the structure of CaM in a horseshoe fashion around the target peptide. However, certain defined regions appear more affected


DOI: 10.1021/acs.analchem.5b02724
Anal. Chem. XXXX, XXX, XXX−XXX

Table 1. MALDI-TOF Analysis of Tryptic Peptides Derived from CaM Labeled in the Free Form or Complexed with Mel[a]

| peptide sequence[b] | $m/z$ (monoisotopic) measured for $[M + H]^+$ | $m/z$ (monoisotopic) measured for $[M + H + CH_2]^+$ | EM % (mol of $CH_2$/mol peptide)[c] F | C | C/F (±SD)[d] |
|---|---|---|---|---|---|
| $A_1$DQLTEEQIAEFK$_{13}$ | 1563.7[f] | 1577.7 | 1.9 | 1.5 | 0.79 ± 0.19 (0.76)[e] |
| $S_{38}$LGQNPTEAELQDMINEVDADGNGTIDFPEFLTMMAR$_{74}$ | 4069.8 | 4083.8 | 7.9 | 19.0 | 2.40 ± 0.96 (0.92)[e] |
| $V_{91}$FDKDGNGYISAAELR$_{106}$ | 1754.8 | 1768.8 | 3.4 | 2.7 | 0.79 ± 0.16 (0.82)[e] |
| $H_{107}$VMTNLGEKLTDEEVDEMIR$_{126}$ | 2401.1[g] | 2415.1 | 3.8 | 4.0 | 1.05 ± 0.21 (0.66)[e] |
| $E_{127}$ADIDGDGQVNYEEFVQMMTAK$_{148}$ | 2490.0 | 2504.0 | 5.8 | 5.3 | 0.91 ± 0.18 (0.83)[e] |

[a]The reaction conditions for modification were as follows: DZN inflow rate 1.0 mL/min, an input-plus-photolysis phase lasting 20 min, steady-state DZN concentration ($C_0$) ~ 0.9 AU. After digestion with trypsin, peptides derived from CaM were analyzed by MALDI-TOF. For details, see the Experimental Section. [b]The reported mass for each [M + H]+ ion differs in less than 0.1 amu from the theoretical monoisotopic mass (calculated with PeptideMass: web.expasy.org). [c]The extent of modification (EM%) is expressed as mol of $CH_2$ incorporated per 100 mol of protein, normalized by the load of DZN reagent. A typical data set is shown. For all the peptides, only the unlabeled and monomethylated species exist in the mixture (see Figures S-8–S-10). [d]The labeling C/F ratio (±SD) is calculated as the EM % of a peptide sample derived from complexed CaM (C) relative to the value measured for the free form (F). [e]Values between brackets indicate the SASA ratio, which represents the static solvent exposure of each peptide in the complex (1cm1) relative to the free form (1cll), as calculated using Surface Racer.[24] [f]$m/z$ values correspond to the N-terminal acetylated form of the peptide. [g]$m/z$ values correspond to the peptide triply methylated in $K_{115}$.

by complexation than others. To frame the following discussion, Figure 5 includes several descriptors of local environments.

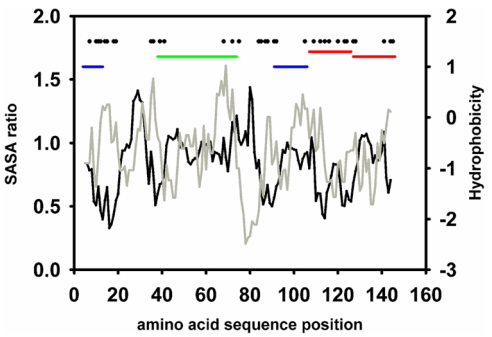



**Figure 5.** Local solvent exposure and hydrophobicity of CaM. The SASA was calculated with Surface Racer[24] using the van der Waals atomic radii set defined by Richards[1] and a probe radius of 1.4 Å on structures 1cm1 (C, complex) and 1cll (F, free). Individual amino-acid accessibility values were averaged along the sequence over a moving pentapeptide window, and the SASA ratio between the two forms (C/F) was then calculated (black line). The hydrophobicity plot uses both the Kyte and Doolittle scale[42] and a nonapeptide averaging window (gray line). Dots indicate those amino acid residues in complexed CaM that lie less than 4.5 Å from Mel. Horizontal bars show the location of tryptic peptides shaded according to the EM % C/F ratio: < 1 (blue), ~1 (red), and >1 (green).

As expected, it is clear that low SASA ratios map to the spots with the highest density of neighboring Mel amino-acid residues. Our results are in general agreement with the findings reported using a different footprinting technique based on the hydroxyl radical oxidation of the polypeptide chain and detection by ESI-MS.[43] These authors observed the highest protection values in the very same regions in CaM upon complexation to Mel.

In particular, peptide 1–13 shows one of the lowest C/F values (0.79 ± 0.19). This result can be readily explained by the high number of amino acids involved at the interface with Mel, consequently displaying a low SASA ratio. The other low C/F value (0.79 ± 0.16) corresponds to peptide 91–106. In contrast to the former peptide, this peptide displays an intermediate level of occlusion, as indicated by the SASA ratio and the number of neighbors. Nevertheless, our results agree with the results of this system's oxidative challenge, which points to the protection of this region at a level similar to that of the N-terminal domain. In particular, Wong et al.[43] attribute this effect to the movement of the Y99 side chain into the complex and, therefore, away from the aqueous solvent. With respect to those peptides that map to the C-terminal region (107–126 and 127–148), we do not observe significant differential labeling (C/F ratios close to one). Similarly, the authors cited above observe that the C-terminal region of CaM appears not to be protected in any significant measure because it forms the complex with Mel. Our results follow this same trend, which might be explained by the more distant relative location of the C-terminus (in comparison to the N-terminus) from the Mel helix.

From an exclusively geometric standpoint, peptide 38–74 presents a SASA ratio close to one. Because of its considerable length and because it is obtained in low yield, this peptide was more difficult to quantify, giving rise to a more uncertain estimate of EM %. Nevertheless, it appears to yield an anomalously high EM % ratio (2.40 ± 0.96). Notably, it is not unthinkable to measure "reverse" ratios (>1) as the consequence of peptides' participation in the vicinity of cavities and crevices, which are actually inner surface components with the ability to lodge DZN.[15,18–21] Peptide 38–74 folds into a loop–helix–loop–helix motif in both free and complexed CaM, essentially maintaining an almost identical relative orientation of helices and lying close to (but not in direct contact with) the N-terminal end of Mel. However, the environment around this peptide changes dramatically. It stands next to the "hinge" region in complexed CaM, which is a situation quite different from the central, fully solvent-exposed $\alpha$-helix in free CaM. Thus, in the complex, a mobile crevice with the ability to preferentially lodge DZN might develop. Importantly, the C-terminal stretch of peptide 38–74 includes


DOI: 10.1021/acs.analchem.5b02724
Anal. Chem. XXXX, XXX, XXX–XXX

the sequence FLTMMA, the most hydrophobic region in the protein, which lies next to the most hydrophilic region, located around the center of the sequence (see the hydropathy plot in Figure 5).

Further fragmentation of tryptic peptides allowed us to evaluate sites of modification at amino acid resolution. As a case in point, methylated sites in peptide $V_{91}$FDKDGNGYISAAELR$_{106}$ were identified (Figure 6).

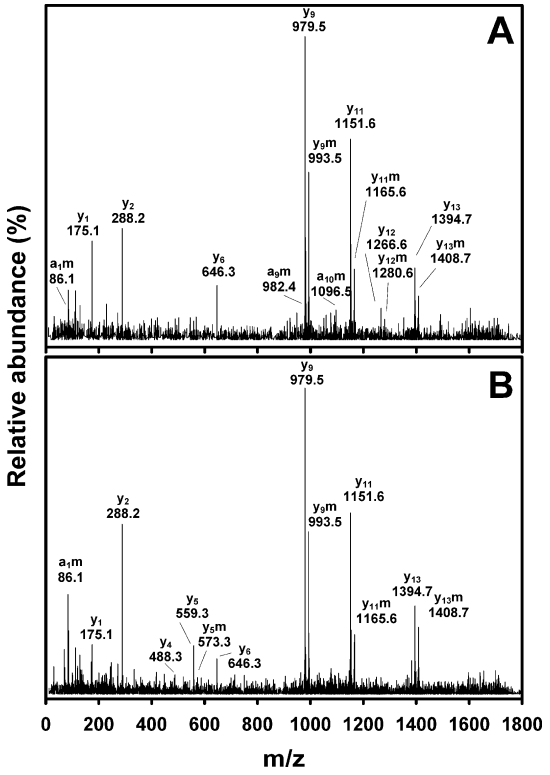



**Figure 6.** MS/MS spectra corresponding to the fragmentation of monomethylated species [$m/z$ 1768.8] of peptide $V_{91}$FDKDGNGYISAAELR$_{106}$ derived from free (A) or complexed (B) $Ca^{2+}$-CaM labeled with DZN for 20 min (simultaneous dissolution and photolysis of the gas at a flux of 1.0 mL/min). Methylated products are marked with the index *m*. Each spectrum represents the average of two scans.

Again, we compared the intensities of each fragment ion with its methylated counterpart after MS/MS analysis of the monomethylated species obtained from free or complexed CaM. In both spectra, the y series of fragments is readily apparent because of its higher yield, whereas the b series of ions is absent. After analyzing the whole pattern for site-specific modification (according to the formalism expressed in eq 3), the probability of methylation occurrence appears to be higher between amino acids 94 and 98. Nevertheless, given the poor recovery of fragments $y_3$ to $y_8$ and the presence of $a_{1m}$, one cannot exclude the possibility of some methyl incorporation between positions 91 and 104.

Observing the CaM−Mel interaction from the side of the peptide partner also reveals a decrease in EM upon complex formation: EM of free or complexed Mel is 15.9 or 10.8 mol of $CH_2$ per mol peptide, respectively (Figure S-11). This reduction in EM (32 ± 8%) might result from the expected decrease in SASA upon complex formation. All the calculated differences between free and complexed forms yield surface occlusion values ranging between 50 and 56% (see the Supporting Information). This result illustrates that the DZN method can detect differences in the labeling signal, even in cases in which one of the partners involved in the interaction (Mel) is considerably smaller than the other (CaM).

## CONCLUSIONS AND PERSPECTIVES

The results presented in this paper demonstrate the worth of combining a mild photochemical method with the superior advantage provided by modern MS techniques to probe the protein surface at solvent size resolution. In this fashion, the method proved to be well suited not only to dissecting aspects of an overall physiological conformational change but also to advancing the understanding of an interface relevant to protein-peptide recognition. The unobtrusive nature of the covalent tag (methyl group) attached to the protein not only permits the conformation to be preserved but also enables it to fully exploit the similarities to the unmodified protein in the subsequent analytical processing of samples. Evaluating the new metric EM on the peptide set derived from labeled samples allowed us to collect meaningful information about solvent accessibility. Furthermore, MS/MS fragmentation achieves the goal of attaining sufficient detail to differentiate among labeled sites at amino acid level resolution. Current efforts include the use of high-power UV LEDs tuned at 315 nm for irradiation. This approach paves the way for miniaturization and automation to reduce sample volume, DZN load, and photolysis time. Given its versatility and technical simplicity, DZN labeling holds promise as a particularly relevant footprinting method in the context of current proteomic and interactomic efforts.

## ASSOCIATED CONTENT

### Supporting Information

The Supporting Information is available free of charge on the ACS Publications website at DOI: 10.1021/acs.analchem.5b02724.

Additional experimental details and data (PDF)

## AUTHOR INFORMATION

### Corresponding Author

*Phone: 54 11 4964 8291, ext. 116. Fax: 54 11 4962 5457. E-mail: delfino@qb.ffyb.uba.ar.

### Notes

The authors declare no competing financial interest.

## ACKNOWLEDGMENTS

For the interpretation of mass spectra, the expert advice of Ms. Susana Linskens and Mr. Carlos Paván (LANAIS-PROEM, UBA-CONICET) is gratefully acknowledged. The authors thank Dr. Mariano González Lebrero for his help in rendering Figure 3 and Figure S-1. G.E.G. and J.M.D. hold teaching positions at the Universidad de Buenos Aires (UBA) and are career researchers at the Consejo Nacional de Investigaciones Científicas y Técnicas (CONICET). This research has been supported by grants to J.M.D. from UBACyT, CONICET, and ANPCyT. G.E.G. is the recipient of a starting research grant from ANPCyT.

## REFERENCES

(1) Richards, F. M. *Annu. Rev. Biophys. Bioeng.* **1977**, *6*, 151−176.
(2) Englander, S. W. *J. Am. Soc. Mass Spectrom.* **2006**, *17*, 1481−1489.



DOI: 10.1021/acs.analchem.5b02724
*Anal. Chem.* XXXX, XXX, XXX−XXX

(3) Englander, S. W.; Mayne, L.; Krishna, M. M. *Q. Rev. Biophys.* **2007**, *40*, 287−326.
(4) Zhu, M. M.; Rempel, D. L.; Du, Z.; Gross, M. L. *J. Am. Chem. Soc.* **2003**, *125*, 5252−5253.
(5) Percy, A. J.; Rey, M.; Burns, K. M.; Schriemer, D. C. *Anal. Chim. Acta* **2012**, *721*, 7−21.
(6) Truhlar, S. M.; Croy, C. H.; Torpey, J. W.; Koeppe, J. R.; Komives, E. A. *J. Am. Soc. Mass Spectrom.* **2006**, *17*, 1490−1497.
(7) Tullius, T. D.; Dombroski, B. A. *Science* **1985**, *230*, 679−681.
(8) Tullius, T. D.; Dombroski, B. A.; Churchill, M. E. A.; Kam, L. In *Recombinant DNA Methodology*; Wu, R.; Grossman, L.; Moldave, K., Eds.; Academic Press: San Diego, CA, 1989.
(9) Xu, G.; Chance, M. R. *Chem. Rev.* **2007**, *107*, 3514−3543.
(10) Wang, L.; Chance, M. R. *Anal. Chem.* **2011**, *83*, 7234−7241.
(11) Maleknia, S. D.; Downard, K. M. *Chem. Soc. Rev.* **2014**, *43*, 3244−3258.
(12) Huang, W.; Ravikumar, K. M.; Chance, M. R.; Yang, S. *Biophys. J.* **2015**, *108*, 107−115.
(13) Sharp, J. S.; Tomer, K. B. *Biophys. J.* **2007**, *92*, 1682−1692.
(14) Richards, F. M.; Lamed, R.; Wynn, R.; Patel, D.; Olack, G. *Protein Sci.* **2000**, *9*, 2506−2517.
(15) Craig, P. O.; Ureta, D. B.; Delfino, J. M. *Protein Sci.* **2002**, *11*, 1353−1366.
(16) Frey, H. M. *Adv. Photochem.* **1966**, *4*, 225−256.
(17) Turro, N. J.; Cha, Y.; Gould, I. R. *J. Am. Chem. Soc.* **1987**, *109*, 2101−2107.
(18) Craig, P. O.; Gómez, G. E.; Ureta, D. B.; Caramelo, J. J.; Delfino, J. M. *J. Mol. Biol.* **2009**, *394*, 982−993.
(19) Gómez, G. E.; Mundo, M. R.; Craig, P. O.; Delfino, J. M. *J. Am. Soc. Mass Spectrom.* **2012**, *23*, 30−42.
(20) Ureta, D. B.; Craig, P. O.; Gómez, G. E.; Delfino, J. M. *Biochemistry* **2007**, *46*, 14567−14577.
(21) Gómez, G. E.; Cauerhff, A.; Craig, P. O.; Goldbaum, F. A.; Delfino, J. M. *Protein Sci.* **2006**, *15*, 744−752.
(22) Jumper, C. C.; Schriemer, D. C. *Anal. Chem.* **2011**, *83*, 2913−2920.
(23) Jumper, C. C.; Bomgarden, R.; Rogers, J.; Etienne, C.; Schriemer, D. C. *Anal. Chem.* **2012**, *84*, 4411−4418.
(24) Tsodikov, O. V.; Record, M. T., Jr.; Sergeev, Y. V. *J. Comput. Chem.* **2002**, *23*, 600−609.
(25) Kuboniwa, H.; Tjandra, N.; Grzesiek, S.; Ren, H.; Klee, C. B.; Bax, A. *Nat. Struct. Biol.* **1995**, *2*, 768−776.
(26) Chattopadhyaya, R.; Meador, W. E.; Means, A. R.; Quiocho, F. A. *J. Mol. Biol.* **1992**, *228*, 1177−1192.
(27) Wall, M. E.; Clarage, J. B.; Phillips, G. N., Jr. *Structure* **1997**, *5*, 1599−1612.
(28) Ikura, M.; Clore, G. M.; Gronenborn, A. M.; Zhu, G.; Klee, C. B.; Bax, A. *Science* **1992**, *256*, 632−638.
(29) Gribskov, M.; Wesson, L.; Eisenberg, D. (to be published, as it appears in pdb).
(30) Hewish, D. R.; Barnham, K. J.; Werkmeister, J. A.; Kirkpatrick, A.; Bartone, N.; Liu, S. T.; Norton, R. S.; Curtain, C.; Rivetta, D. E. *J. Protein Chem.* **2002**, *21*, 243−253.
(31) Bernadó, P.; Blackledge, M.; Sancho, J. *Biophys. J.* **2006**, *91*, 4536−4543.
(32) Anderson, S. R.; Malencik, D. A. *Calcium Cell Funct.* **1986**, *6*, 1−42.
(33) Clore, G. M.; Bax, A.; Ikura, M.; Gronenborn, A. M. *Curr. Opin. Struct. Biol.* **1993**, *3*, 838−845.
(34) Gopalakrishna, R.; Anderson, W. B. *Biochim. Biophys. Acta, Mol. Cell Res.* **1985**, *844*, 265−269.
(35) Humphrey, W.; Dalke, A.; Schulten, K. *J. Mol. Graphics* **1996**, *14*, 33−38.
(36) Comte, M.; Maulet, Y.; Cox, J. A. *Biochem. J.* **1983**, *209*, 269−272.
(37) Seeholzer, S. H.; Cohn, M.; Putkey, J. A.; Means, A. R.; Crespi, H. L. *Proc. Natl. Acad. Sci. U. S. A.* **1986**, *83*, 3634−3638.
(38) Steiner, R. F.; Albaugh, S.; Fenselau, C.; Murphy, C.; Vestling, M. *Anal. Biochem.* **1991**, *196*, 120−125.
(39) Scaloni, A.; Miraglia, N.; Orru, S.; Amodeo, P.; Motta, A.; Marino, G.; Pucci, P. *J. Mol. Biol.* **1998**, *277*, 945−958.
(40) Schulz, D. M.; Ihling, C.; Clore, G. M.; Sinz, A. *Biochemistry* **2004**, *43*, 4703−4715.
(41) Zhang, H.; Gau, B. C.; Jones, L. M.; Vidavsky, I.; Gross, M. L. *Anal. Chem.* **2011**, *83*, 311−318.
(42) Kyte, J.; Doolittle, R. F. *J. Mol. Biol.* **1982**, *157*, 105−132.
(43) Wong, J. W. H.; Maleknia, S. D.; Downard, K. M. *J. Am. Soc. Mass Spectrom.* **2005**, *16*, 225−233.


DOI: 10.1021/acs.analchem.5b02724
*Anal. Chem.* XXXX, XXX, XXX−XXX