PROBING PROTEIN STRUCTURE BY AMINO ACID-SPECIFIC COVALENT LABELING AND MASS SPECTROMETRY

Vanessa Leah Mendoza and Richard W. Vachet*
Department of Chemistry, University of Massachusetts Amherst, Amherst, MA

Received 7 April 2008; received (revised) 9 June 2008; accepted 9 June 2008

Published online 17 November 2008 in Wiley InterScience (www.interscience.wiley.com) DOI 10.1002/mas.20203

For many years, amino acid-specific covalent labeling has been a valuable tool to study protein structure and protein interactions, especially for systems that are difficult to study by other means. These covalent labeling methods typically map protein structure and interactions by measuring the differential reactivity of amino acid side chains. The reactivity of amino acids in proteins generally depends on the accessibility of the side chain to the reagent, the inherent reactivity of the label and the reactivity of the amino acid side chain. Peptide mass mapping with ESI- or MALDI-MS and peptide sequencing with tandem MS are typically employed to identify modification sites to provide site-specific structural information. In this review, we describe the reagents that are most commonly used in these residue-specific modification reactions, details about the proper use of these covalent labeling reagents, and information about the specific biochemical problems that have been addressed with covalent labeling strategies. © 2008 Wiley Periodicals, Inc., Mass Spec Rev 28:785–815, 2009

Keywords: covalent labeling; protein structure; protein–protein complexes; solvent accessibility

I. INTRODUCTION

Because of the link between structure and function, the determination of the molecular structure of proteins with high resolution will continue to be important in biology. NMR and X-ray crystallography are the most important techniques to obtain this high-resolution structural information, but an increasing number of protein systems are not amenable to these methods because of size, conformational flexibility, aggregation propensity, or limited sample amount. In addition, the prevalence and vast array of protein–protein and protein–ligand complexes that comprise cellular machinery make it clear that higher throughput methods are necessary to unravel the relationships between protein structural interactions and cellular function. Because of current technical limitations with NMR and crystallographic methods, mass spectrometry (MS) plays an ever-increasing role in protein structure determination due to its speed, sensitivity, and specificity.

Because MS measurements occur in the gas-phase, structural information about proteins in solution is acquired in an indirect way. Typically, this structural information is obtained by changing the mass of the protein or its proteolytic fragments in a structure-dependent manner. In other words, protein structural properties must be encoded in the mass-to-charge (m/z) ratios of the ions that are eventually measured by the mass spectrometer. To accomplish this, a variety of approaches have been used, including H/D exchange (Wales & Engen, 2006), intra- and intermolecular cross-linking (Sinz, 2006), noncovalent structural probing (Ly & Julian, 2006), and protein structural mapping via covalent labeling. Covalent labeling approaches use reagents that either modify specific amino acids or react generally with many different amino acids. In comparison to H/D exchange methods that use deuterium as a reagent, the possibilities of back-exchange and scrambling are essentially non-existent with covalent labeling reagents. In addition, covalent labeling approaches provide structural information about amino acid side chains that is generally not available with H/D exchange methods, making covalent labeling a complementary approach to H/D exchange. Due to the size of typical covalent modifications, however, protein structure is more likely to be perturbed with such reagents as opposed to when deuterium is used to probe structure. Also, modifications that affect a given amino acid’s chemical properties, such as its charge, hydrogen bonding capability or hydrophobicity, could also possibly perturb protein structure.

Cross-linking methods are similar to covalent labeling strategies in that the amino acid side chain modifications are not readily reversed. These methods create new intramolecular or intermolecular bonds that impose distance constraints on the location of two amino acid side chains, which allow one to deduce information about the three-dimensional structure of a single protein or protein complex. Key differences between covalent labeling methods and cross-linking approaches are the difficulty of the resulting MS analyses and the type of information gathered from each method. Because cross-linking approaches make new bonds between sites that are often distant from one another in the primary sequence, identifying the cross-linked sites with peptide mass mapping or tandem MS can be difficult. Analysis of covalently modified amino acids is usually more straightforward. As for information content, cross-linking provides distance constraints for two amino acids whereas covalent labeling approaches provide information about protein surface structure because the reactions are usually controlled by the accessibility of one or more amino acids to the covalent labeling reagent.

Covalent labeling is usually carried out with either non-specific labels or amino acid-specific labels. Approaches that use
non-specific labels almost exclusively rely on protein reactions with radicals (e.g., oxidative footprinting methods), and several reviews of these approaches and their applications have appeared recently (Maleknia & Downard, 2001; Guan & Chance, 2005; Downard, 2006; Takamoto & Chance, 2006; Xu & Chance, 2007). Amino acid-specific labels have been used quite extensively with MS to study protein structure and protein interactions, but to our knowledge no general review that has focused solely on this topic has appeared recently in the literature. Reviews of amino acid-specific tagging strategies for quantitation and selective fractionation of proteins for proteomics applications have appeared (Lundblad, 2004; Leitner & Lindner, 2006), but these reviews have not covered those labeling approaches that seek to elucidate protein secondary and tertiary structure. The current review, in contrast, will describe the wide range of amino acid-specific covalent labeling reagents and techniques that have been used with MS to study protein structure and protein interactions. We attempt to convey the scope of protein systems investigated with covalent labeling and MS by briefly describing most of the studies published on this topic in the last 16 years. Special attention will be paid to (1) the reagents that are used for amino acid-specific labeling, (2) to cautionary notes for the proper utilization of these labels, (3) to the protein structural problems that have been addressed with these labels, and (4) to the future prospects of such amino acid-specific covalent labeling strategies. We hope that this assessment of amino acid-specific labeling approaches coupled with MS for protein structural analysis will provide a good starting point for those interested in using these methods, and we will highlight the aspects of these approaches that require special care to be applied correctly.

II. BASIS OF COVALENT LABELING WITH MASS SPECTROMETRY

Obtaining structural information for proteins or protein–ligand complexes with a covalent labeling approach typically relies on the differential reactivity of amino acids upon exposure to a particular label. The implicit assumption in these experiments is that amino acids that are exposed to solvent and, therefore, accessible to a labeling reagent will be modified, whereas buried amino acids will be modified slowly or not at all. Protein conformational changes or protein–ligand binding will affect the extent to which certain amino acid(s) react with the added labeling reagent (Fig. 1). The amino acids that become more or less accessible to the reagent will react to greater or lesser extents; that difference indicates their involvement in a conformational change and/or their presence at a ligand-binding site. Amino acid-specific information about a protein’s structural changes is

![Analytical scheme showing covalent labeling with MS detection. The label modifies solvent accessible amino acids. ESI- or MALDI-MS can be used to calculate the extent of modification. The modified protein can then be subjected to proteolysis followed by either LC–MS/MS or MALDI-MS to determine the peptide fragments and/or the specific amino acids that have been modified. [Color figure can be viewed in the online issue, which is available at www.interscience.wiley.com.]

FIGURE 1. Analytical scheme showing covalent labeling with MS detection. The label modifies solvent accessible amino acids. ESI- or MALDI-MS can be used to calculate the extent of modification. The modified protein can then be subjected to proteolysis followed by either LC–MS/MS or MALDI-MS to determine the peptide fragments and/or the specific amino acids that have been modified. [Color figure can be viewed in the online issue, which is available at www.interscience.wiley.com.]
inferred from differential reactivity patterns of individual amino acids. This idea of using modification reagents to map protein structure has been around for at least 40 years, but the more recent ability of MS to quickly, sensitively, accurately, and precisely map protein modification sites has made covalent labeling approaches much more powerful methods to obtain protein structural information. In pre-MS days, the extents of modification were typically determined using amino acid analysis, biological activity assays, fluorescence, or absorbance measurements.

A. Analysis and Identification of Labeled Sites

Mass spectrometry can be used in several ways to provide a read-out of the covalent modifications that occur upon reaction with a labeling reagent. The simplest, but perhaps least informative means, is to monitor the number of reagent adducts that modify the protein of interest. Taking the mass spectrum of an intact protein after reaction with the label readily provides this information, and the maximum number of labels and/or the average number of labels can be useful "low resolution" information. This approach is most useful when comparing two protein states and when using a highly specific amino acid labeling reagent. A change in the number of labels in one state as compared to another can indicate whether a given type of amino acid is involved in a structural change or is near a ligand-binding site. For example, Leitner and Lindner (2005) recently used a combination of 2,3-butanedione and phenylboronic acid to modify Arg residues in a variety of proteins. In comparing myoglobin with apomyoglobin (i.e., without the heme group), electrospray ionization (ESI)-MS of the intact proteins indicated that the average number of modified Arg residues increased in the apo-protein; that increase is consistent with one or more of these residues being partially buried next to the heme group in the holoprotein. By monitoring the labeling reactions over time and closely inspecting the kinetics of these reactions, this approach allowed the actual number of exposed Arg residues to be calculated.

Whereas monitoring the modification extent of a whole protein by MS is quick and straightforward, it does not provide the same level of detail (or spatial resolution) that is possible when modified proteins are enzymatically digested and the resulting peptide fragments are analyzed by ESI-MS or matrix-assisted laser desorption/ionization (MALDI)-MS. Thus, the most common approaches to use chemical modifications to derive protein structure are peptide mass mapping with MALDI-MS or ESI-MS and peptide sequencing with tandem MS (MS/MS), usually in conjunction with reversed-phase liquid chromatography (RPLC). It should be noted that care must be taken when using MALDI-MS to monitor labeled peptides because the typical acidic matrices used in MALDI-MS can cause the dissociation of labels that are acid-labile.

The particular amino acids that have been labeled can usually be determined by proteolytic digestion of the protein after reaction with a labeling reagent. With peptide mass mapping (i.e., no MS/MS), the modified amino acids are identified by measuring the m/z ratios of the proteolytic fragments and finding the fragments whose m/z ratios differ, by the mass of label, from the predicted values (Fig. 2). If an amino acid-specific label is used, then the specific amino acid in a proteolytic fragment can usually be deduced; however, if the relative specificity of the reagent is low, then the identity of the modified amino acid often cannot be assigned with complete confidence. Nonetheless, identification of a particular modified proteolytic fragment still provides some spatial resolution so that more confident conclusions about a protein’s structure can be made. Tandem MS is required to identify the specific amino acids modified in a proteolytic fragment. In effect, MS/MS can increase the spatial resolution for a protein to the single amino acid level. Obtaining a complete peptide sequence by MS/MS is not always possible; however modified amino acids can typically be at least narrowed down to within two to three residues.

When the extent of protein modification is low or a mixture of proteolytic peptide fragments to be analyzed is complex, identification of the specific modification sites can be challenging. A powerful approach to address this problem is to use modification-specific product ions as markers of the modified peptides. The best example of this idea is found in the context of lysine acetylation experiments. Dissociation of peptides with acetylated lysine residues often results in the formation of abundant immonium ions of acetylated lysine at m/z 143, and if extracted ion chromatograms of m/z 143 are used, then peptides with modified lysine residues can be readily identified. Another common product ion of lysine-acetylated peptides at m/z 126, which corresponds to the loss of NH3 from the immonium ion, has also been used to facilitate the identification of peptides with acetylated lysines (Kim et al., 2002). Another means to identify modified peptides in a mixture of peptides is to use multiple, equally reactive labels that differ in mass. Recently, Gabant,
Augier, and Armengaud (2008) demonstrated that three lysine-specific reagents that differed in mass by 113 Da could be used in parallel. The reactivity of these reagents with the protein of interest is virtually identical, causing modified peptides to show up as a quartet of peaks. Finding the modified peptides was facilitated by identification of peptide fragments that had peaks at m/z that corresponded to M + 226, M + 339, and M + 452.

An alternative to enzymatic digestion followed by peptide mass mapping or MS/MS is top-down sequencing (Kelleher et al., 1999; Meng et al., 2005; McLafferty et al., 2007), which involves MS/MS analysis on the intact protein. Generally speaking, the key advantage of the top-down approach is that the protein sequence and any modification site(s) can be obtained potentially in a single MS/MS experiment. In addition, the long enzymatic digestion step can be avoided to allow the analysis to be done more quickly and to possibly avoid the loss of covalent labels that might not be stable during the relatively long digestion step. The main challenge associated with top-down sequencing is the large number of product ions that result from protein dissociation that can complicate spectral interpretation. Despite this challenge, top-down sequencing, especially on Fourier transform ion cyclotron resonance (FT-ICR) mass spectrometers has found more widespread usage (Meng et al., 2005; Zhai et al., 2005). The use of top-down sequencing to find chemical modification sites, however, presents additional challenges. Multiple potential sites of modification can lead to mixtures of protein isoforms. Upon dissociation, these isobaric protein ions generate even more complicated product ion spectra. Furthermore, for labels with relatively low specificity, narrowing down the modification site to individual amino acids can be difficult and effectively provides less spatial resolution on the protein structure. Nonetheless, top-down sequencing has been used effectively to map labeled sites so that 3D protein structural information can be obtained (Novak et al., 2004).

B. Extent of Modification

Determination of a particular amino acid modification site helps to identify the region(s) in a protein that underwent a structural change or formed an interface with another molecule; however, the degree of modification at this amino acid is also important to determine. The degree of modification usually reflects the extent to which a specific protein region is involved in the structural change. For example, the reactivity of an amino acid that is at the interface with another molecule is likely to decrease more than the reactivity of an amino acid that is at the periphery of the interface. Concentration effects are also important to consider as these can influence the observed extent of modification (Tong, Wren, & Konermann, 2007; Mendoza & Vachet, 2008). Higher protein concentration will result in lower labeling levels when low label:protein molar ratios are used. Thus, this effect must be accounted for to avoid ambiguities in data interpretation, particularly in comparative studies involving a protein and its binding partner.

Given the useful information contained in the degree of modification, several approaches have been developed and used to measure amino acid reactivities with modification reagents. The most common approach is to simply compare the relative ion abundances of a peptide fragment that contains the amino acid after the protein of interest is reacted under two different conditions—one condition in which the protein has its native structure, and another in which the protein undergoes the structural change of interest. The ratio of the modified peptide fragment’s ion abundance to the peptide fragment’s total ion abundance (i.e., unmodified + modified) allows one to easily identify the degree to which an amino acid underwent a reactivity change that coincided with the protein’s structural change. Although this approach is straightforward, it does have the potential to give misleading information. Some amino acid modifications can change the ionization efficiency of the measured peptide; consequently, the ion abundance ratio of the modified peptide does not accurately reflect its solution concentration (Cech & Enke, 2001). This problem is likely to be exacerbated if LC is used because modified and unmodified peptides will usually be separated from one another, and thus be ionized under different solvent conditions if solvent gradients are used. Furthermore, modification of residues such as lysine that are proteolytic cleavage sites could cause proteolytic enzymes such as trypsin to miss cleavage sites in modified copies of a protein but not in the unmodified copies of the protein. Such an occurrence could also lead to misleading ion abundance ratios if not properly accounted for.

Because of these potential problems, several other approaches have been developed to measure the extent of amino acid modification. The first of these methods uses isotopic labeling to determine relative reactivities of amino acids and has been used with lysine acetylation (Glocker et al., 1994; Hochleitner et al., 2000). In this approach, the protein of interest is modified by a two-step process (Fig. 3): (1) the protein, under conditions of interest (e.g., bound to a ligand), is reacted with a low molar excess of the lysine-specific reagent; (2) the protein is then reacted with a high molar excess of an isotopically labeled form of the lysine-specific reagent so that all remaining unreacted lysines are modified. The relative reactivity of a given lysine residue can be determined from the ion abundance ratio between the unlabeled modified peptide and the isotopically labeled version of the modified peptide. Because an isotopically labeled version of the reagent is used, concerns over changes in ionization efficiency upon modification are minimized. Peptides that contain lysine residues that are not very exposed have low ion abundance ratios because very little isotopically unlabeled modified peptide is observed. In contrast, highly exposed lysine residues have high ion abundance ratios because most of these residues are modified by the isotopically unlabeled reagent during the first reaction step. Also, note that isotopic labeling can be helpful to identify modified peptides in a complex mixture.

A second way to obtain more reliable information about the extent of modification is to use an internal standard, which is usually another peptide that does not contain a modifiable residue (Glocker et al., 1996). Typically, the ion abundance of the unmodified version of the peptide that contains the modifiable residue of interest is compared to the ion abundances of peptides that do not contain modifiable residues. Measurement of the ion abundance of the unmodified peptide minimizes ionization efficiency concerns and can provide semi-quantitative measurements of the extent of modification.
A third experimentally involved, but more reliable method, is to measure the modification reaction kinetics of the protein or, better yet, its peptide fragments that contain the modified residues of interest (Suckau, Mak, & Przybylski, 1992; Leitner & Lindner, 2005; Gao & Wang, 2006; Mendoza & Vachet, 2008). Rate coefficients can be determined from the extent of modification over a range of reaction times or reagent concentrations (Fig. 4). These kinetic measurements can be performed by either measuring the increase in the ion abundance of the modified version of a peptide over time or the decrease in the ion abundance of the unmodified version of a peptide over time. The latter has the advantage that the rate coefficient is sensitive to only the unmodified population of the peptide (and the protein from which it comes), which minimizes any modification-induced changes to the reaction rate that could compromise the overall analysis. Measurement of modification kinetics has the additional potential advantage of more detailed structural information because most studies demonstrate that a good correlation exists between reaction rate coefficients and a residue’s solvent accessibility.

C. Ensuring a Protein’s Structural Integrity

Any protein structural information that is derived from covalent modification experiments is reliable only if the structural integrity of that protein is preserved during the reaction. The relatively large size of most amino acid-specific labels and their potential effect on an amino acid side chain’s charge or hydrophobicity make it quite likely that covalent modification to even a surface-exposed residue could perturb a protein’s structure so that any subsequent modification no longer reflects the protein’s initial structure. Given this serious concern, appropriate checks are required to ensure that the structural probe (i.e., the amino acid-specific reagent) does not distort the protein’s structure and thus provide incorrect structural information. Even though these checks are very important, we found during the preparation of this review that more than 60% of the studies that report the use of amino acid-specific labeling and MS did nothing to ensure the structural integrity of the studied protein, which brings into question the validity of these results. The remaining 40% used either circular dichroism (CD), fluorescence spectroscopy, activity assays, a limited amount of modification, reaction kinetics, or a combination of these checks to ensure protein structural integrity during the amino acid-specific labeling reaction. CD spectroscopy is the most common approach that researchers have used to check protein structure because it is sensitive to changes in a protein’s secondary structure. However, CD provides population-weight average properties of a protein, and is relatively insensitive to local changes in protein structure. As such, subtle or local structural changes caused by the modification reagent will usually be missed. In contrast, fluorescence spectroscopy can provide information about local regions of a protein’s structure. However, CD provides population-weight average properties of a protein, and is relatively insensitive to local changes in protein structure. As such, subtle or local structural changes caused by the modification reagent will usually be missed. In contrast, fluorescence spectroscopy can provide information about local regions of a protein’s structure. Unfortunately, because fluorescence typically only monitors the environment around tryptophan residues, it will likely miss any structural changes if they occur distant from tryptophan residues. Although activity assays are potentially powerful ways to monitor structural changes, often only perturbations to an enzyme’s active site lead to noticeable changes in activity.

Other approaches that use the extent of protein modification offer more reliable ways to ensure protein structural integrity. Perhaps the most reliable method is to limit the number of modifications to one per protein molecule. In this way, the
probe’s reactivity is only influenced by the unperturbed protein structure. Oftentimes, however, this limit means that a significant fraction of the protein will be unmodified. Furthermore, if multiple possible modification sites exist on the protein, the extent of modification at any given residue will be low and, therefore, difficult to pinpoint. Another approach, which was recently described by our group, is to measure the reaction kinetics for individual modification sites by monitoring the ion abundance of the peptide(s) that contain that residue (Mendoza & Vachet, 2008). The reactions of the modification reagent with the protein of interest give rise to second-order reaction kinetics; ensuring that this reaction order is maintained at low reagent doses can serve as the means by which protein structural integrity is assured (Fig. 5). Because the assumption is that second-order rate coefficients will be constant as long as the protein’s structure is unchanged, a second-order plot will remain linear. Deviations from linearity or changes in the plot’s slope indicate a variation in the reaction dynamics that are caused by changes in the microenvironment around the residue of interest. Measurements of the unmodified peptides (and not the modified peptides) further ensure that the initial protein state is monitored.

III. AMINO ACID-SPECIFIC LABELS AND EXAMPLES OF THEIR APPLICATIONS

Residue-specific modification reactions are usually carried out with reagents that utilize well-known organic reactions. Amino acid-specific reactions for only eight of the 20 naturally occurring amino acids have been used to monitor protein structure in conjunction with MS. In general, the reactivity of the different functional groups in a protein relies on the reactivity and the accessibility of the amino acid side chain to the reagent. In the sections that follow, we describe the amino acid-specific reactions that are most commonly used with a particular focus on (i) the reaction mechanisms; (ii) the range of reaction conditions that are accessible with these reactions; (iii) the side reactions that must be contended with; and (iv) specific biochemical problems that have been addressed.

A. Arginine

1. Reagents

   a. Reaction mechanisms. The modification of arginine residues is typically based on the reaction of vicinal dicarbonyl compounds to form cyclic adducts. Commonly used reagents are phenylglyoxal, p-hydroxyphenylglyoxal, 2,3-butanedione, 1,2-cyclohexanedione, and methylglyoxal. Aside from the vicinal dicarbonyl compounds, kethoxal is also used for arginine modification. The reactions with the most commonly used reagents are shown in Scheme 1.

   b. Reaction conditions. The arginine-specific reactions are most commonly carried out between room temperature and 37°C over a pH range of 7–10 with 10–200 mM buffer. The dicarbonyl labeling reagents are fairly soluble in water at concentrations up to 200 mM. Reagent concentrations typically vary from 0.1- to 3,000-fold molar excess relative to the protein of interest, and reaction times range from 5 min to 24 hr. For 2,3-butanedione, reactions are typically, but not universally, carried out in the dark to avoid possible photo-activation of this molecule, which could enhance nonspecific reactions with groups other than arginine (Fliss & Viswanatha, 1979; Riordan, 1979). Considered as a whole, the accessible reactions conditions are almost ideal to study proteins under native conditions. It has been noticed, however, that the adducts that are formed with the various dicarbonyls can be somewhat unstable, especially at pH below 7. Borate can stabilize the adducts and prevent the regeneration of arginine as shown in Scheme 2 (Riordan, 1973). Hence, arginine-specific labeling is most successful when borate is added along with the dicarbonyl of interest.

   c. Side reactions. Studies have shown that phenylglyoxal and 1,2-cyclohexanedione can have significant side reactions with amino groups in the absence of borate buffers (Takahashi, 1968; Patthy & Smith, 1975). Reaction of methylglyoxal with lysine can be considerable (Oya et al., 1999). The reactions of 2,3-butanedione with lysine and histidine residues are relatively minimal (Yankeelov, Mitchell, & Crawford, 1968) especially in borate buffers (Riordan, 1973).

2. Protein Surface Structure Analysis

Arginine specific labeling combined with MS has been actively used to probe protein structures. Arginine has the most basic pK_a of all amino acid side chains and thus is always protonated under physiological conditions. Because arginines are almost always found on protein surfaces, their value as targets for protein surface mapping would seem to be low. However, their ability to form intramolecular salt bridges with carboxylates, which could reduce their reactivity, can occasionally make them useful residues to probe. The use of 1,2 cyclohexanedione with 252-Cf

FIGURE 5. A second-order kinetics plot for the peptide fragment 49–60 from β-2-microglobulin being modified by sulfos-ν-hydroxysuccinamide acetate. Linear region (black) indicates unchanged protein structure. Deviation from linearity (gray) suggests a protein structural change due to the modification. [Color figure can be viewed in the online issue, which is available at www.interscience.wiley.com.]
SCHEME 1. Modification reactions of arginine by vicinal dicarbonyl compounds.

For phenylglyoxal, $X = H$
For p-hydroxyphenylglyoxal, $X = OH$

For $X = H$, $+116$ Da
For $X = OH$, $+132$ Da

For $X = H$, $+252$ Da
For $X = OH$, $+284$ Da

$+86$ Da

$+112$ Da

$+94$ Da
plasma desorption mass spectrometric peptide mapping was first reported by Przybylski and co-workers (Suckau, Mak, & Przybylski, 1992). Only 4 of the 11 arginine residues in hen egg white lysozyme were modified, and the reactivities that were observed correlated inversely with the solvent accessibilities of these arginine residues. The selective modification of only four of the arginines was ascribed to the presence of neighboring proton acceptor groups that led to intramolecular catalysis; these data indicated that this reaction is very sensitive to the residue’s local chemical environment. A similar inverse correlation of solvent accessibility and reactivity to 1,2-cyclohexanedione was observed by Pucci and co-workers in their characterization of the surface topology of Minibody, a de novo-designed β-protein (Zappacosta et al., 1997). The reactivity pattern of these arginine residues was used to support a structural model for the Minibody protein. Zou and co-workers compared the structural characteristics of the full-length native and hyperphosphorylated versions of the human replication protein A (RPA) with p-hydroxyphenylglyoxal (Liu et al., 2005). Their results revealed that 18 and 16 arginine residues were modified in native and hyperphosphorylated RPA, respectively. This drop in the number of modified arginines revealed a structural rearrangement that occurred upon phosphorylation that reduced the reactivity of these residues as obtained from crystal structures; that agreement indicated that the phenylboronic acid cap did not adversely affect the structural information gained from the labeling reaction. In addition, reactions with myoglobin, apomyoglobin, and ribonuclease A before and after disulfide reduction were consistent with the known structural changes that these proteins underwent.

Fabris and co-workers have demonstrated that reagents other than dicarboxyls can also selectively modify arginine residues. They demonstrated that kethoxal, which is an RNA footprinting reagent, can selectively modify the guanidinium group of arginine at neutral pH (Akinsiku, Yu, & Fabris, 2005). Susceptibility to kethoxal alkylation for two model proteins, ubiquitin and RNase A, correlated very well with the solvent accessibilities calculated from 3D structures.

3. Protein–Ligand Complex Analysis

When an electrostatic interaction is suspected to be a major contributor to a protein–ligand binding interaction, arginine-selective modification reactions can be a valuable probe because of arginine’s ability to form salt bridges. Arginine-specific reagents have been used to examine the surface topology of the quaternary structure-dependent heparin-binding region of bovine seminal plasma PDC-109; the reactivities of arginine residues of the free and the heparin-bound protein were compared (Calvete et al., 1999). Heparin is a highly negatively charged macromolecule whose interactions with proteins are expected to be electrostatic in nature. Peptide mapping with N-terminal sequencing and ESI-MS revealed that Arg57, Arg64, and Arg104 were protected from modification in the heparin-bound protein. These Arg residues are suspected to be a part of the cationic face in the PDC-109 oligomeric state that binds tightly to heparin. The ability of kethoxal modification and ESI-FTMS to probe nucleoprotein complexes was assessed with the well-characterized complexes formed by HIV-1 NC with RNA stem loops SL2 and SL3 (Akinsiku, Yu, & Fabris, 2005). Peptide mapping showed that all five arginine residues were alkylated in the free protein, but Arg7 and Arg10 were protected from modification in the NC-SL3 complex. For the NC-SL2 complex, only Arg32 was protected. These results closely matched the calculations of solvent accessibility based on 3D structures of the complexes. Carven and Stern (2005) employed hydroxyphenylglyoxal to
probe peptide-induced conformational changes in the human leukocyte antigen HLA-DR1. HLA-DR1, a class II major histocompatibility complex variant, binds peptides and presents them at the cell surface for inspection by T cells. UV–Vis absorption spectroscopy showed that 1.5 and 0.5 hydroxyphenylglyoxal molecules were incorporated into the empty and peptide-loaded proteins, respectively. However, MALDI-TOF MS data revealed that three arginine residues were modified in the ligand-free protein but no arginine residues were found modified in peptide-loaded HLA-DR1. This conflict between the UV and MS data was attributed to the imprecision associated with measuring small absorbances in the UV experiments.

4. Protein–Protein Complex Analysis

Arginine is the second most prevalent amino acid at protein–protein interfaces (Bogan & Thorn, 1998). This is probably due to its ability to have multiple interactions with binding partners. It can form up to five hydrogen bonds, form salt bridges, have hydrophobic interactions via the three methylene carbons on its side chain, and interact with aromatic groups via its guanidinium group. Consequently, arginine-specific modification reactions combined with MS can be a useful tool to characterize protein–protein complexes. A few examples of arginine-selective reagents have been reported. Przybylski and co-workers investigated hen egg white lysozyme and its complex with a monoclonal IgM-type antibody with 1,2-cyclohexanedione along with peptide mapping by MALDI-MS and plasma desorption MS (Fiedler et al., 1998). Their results, that Arg14 was protected from modification in the complex, indicated that it is part of the epitope on native lysozyme that is recognized by IgM. Tomer and co-workers have used hydroxyphenylglyoxal modification to investigate the complex formed by glycosylated full-length gp120 protein from HIV strain SF2 with its human receptor CD4 (Hager-Braun & Tomer, 2002). Results revealed that Arg59 but not Arg58 on CD4 was shielded from hydroxyphenylglyoxal and, therefore, was considered to be part of the interaction site with gp120. These results are in agreement with the crystal structure of CD4 bound to a truncated and deglycosylated form of gp120, which showed the main interactions sites on CD4 to be Phe43 and Arg59 (Fig. 6). This work by Tomer and co-workers illustrates the complementary information that such covalent labeling experiments can provide when the structure of a full-length protein is very difficult to obtain by X-ray crystallography due to the heterogeneity that is inherent to protein glycosylation. Covalent labeling/MS results can potentially provide definitive information about protein–protein interfaces when they can be interpreted in light of other higher-resolution structural data.

Arginine modification with 1,2-cyclohexanedione, limited proteolysis, and alanine mutagenesis was recently used to determine conformational changes and binding sites associated with the interaction between the hepatitis C virus E2 envelope glycoprotein and the CBH-5 antibody (Jacob et al., 2008). Arg587, Arg630, and Arg651 were modified in the free E2 protein, but were protected from modification upon antibody binding. These residues are located near domain A of the E2 protein, which is required for recognition by all monoclonal antibodies. This protection from modification can arise from direct contact between these residues and the monoclonal antibody or from a conformational change induced by the antibody. When the MS-based data were combined with alanine scanning mutagenesis results, the authors concluded that a significant conformational change in the region of residues 579–644 occurs upon CBH-5 binding.

B. Carboxylic Acids

1. Reagents

Carbodiimides such as 1-ethyl-3-(3-dimethylaminopropyl)carbodiimide hydrochloride (EDC) are used to modify carboxyl groups in proteins. However, addition of nucleophiles like glycine ethyl ester (GEE) and glycineamide (GA) is required to establish the extent of modification because the initial products from the reaction with EDC are inherently unstable. The reactions with these reagents are shown in Scheme 3.

b. Reaction conditions. Reactions with carbodiimides are mostly carried out at room temperature in the pH range of 4.5–6 with moderate (50 mM) buffer concentrations. EDC and the nucleophiles needed to stabilize the initial reaction product are very soluble in water, and reagent amounts vary from 40- to 2,000-fold molar excess for EDC and 200- to 500-fold molar excess of the nucleophiles relative to the protein of interest. Reaction times are typically 30–120 min. The relatively acidic pH conditions typically used for the reactions with EDC make it difficult to use this reagent to probe protein structure under native conditions.

c. Side reactions. Carbodiimides can significantly react with sulphydryls, whereas reactions with tyrosine residues to form o-arylisourea occur more slowly (Hoare & Koshland, 1966; Carraway & Koshland, 1972).
2. Protein Surface Structure Analysis

Given the difficulty to selectively modify carboxylates under native conditions, relatively few studies have appeared in which carboxylate-specific reagents were used to map protein surfaces. In addition, just like arginine residues, aspartic and glutamic acid residues are almost always found on protein surfaces; that hydrophilicity makes these amino acids less useful targets for mapping protein structure. Nonetheless, EDC and glycineamide–HCl have been used to characterize the structural difference between the inactive (Ras\(^{\text{GDP}}\)) and active (Ras\(^{\text{GTP}}\)) forms of the Ras protein (Akashi et al., 1997). MALDI-TOF MS and ESI-MS analyses revealed that the effector region of the Ras protein, which included several aspartate and glutamate residues, is more accessible to solvent in the active form. In another study, Kominami and co-workers investigated the cell membrane topology of guinea pig cytochrome P450 17\(\alpha\) with EDC, glycineamide, and MALDI-TOF MS (Izumi et al., 2003). Peptide mapping showed that seven acidic residues were modified in the detergent-solubilized state. However, in proteoliposomes three aspartate residues were protected from modification, indicating that these amino acids are at or near the membrane-binding domains of P450 17\(\alpha\).

3. Protein–Protein Complex Analysis

Aspartate is the fifth most frequently found residues at protein–protein interfaces (Bogan & Thorn, 1998); thus, it could be a reasonable target for understanding protein–protein interactions. Glutamate is less frequently found at protein interfaces. Sanderson and Mosbaugh (1996) used carboxylate-specific reactions to identify the carboxylic acid residues of the bacteriophage PBS2 uracil-DNA glycosylase inhibitor (Ugi) that is required to form a complex with uracil-DNA glycosylase (Ung). Amino acid sequence analysis and MALDI-TOF MS revealed that glutamic acid residues at positions 28 and 31 are important in forming a stable irreversible Ung:Ugi complex. These findings are supported by the crystal structures of human and herpes simplex virus type-1 uracil-DNA glycosylase and Ung complexes as shown in Figure 7 (Mol et al., 1995). Takio and co-workers studied the Ras protein in the presence of one of its target proteins, Raf-1 RBD, to better differentiate between the inactive and active forms (Akashi et al., 1997). Their results showed that the extent of protein modification was less for the active form in the presence of Raf-1 RBD; that difference indicated that carboxylate groups in the effector region are involved in the interaction with a target protein.

C. Cysteine

1. Reagents

   a. Reaction mechanisms. Because thiol groups are inherently very reactive, a large number of reagents can be used to modify cysteine residues. Iodoacetamide and its derivatives, iodoacetic acid, and \(N\)-alkylmaleimides are extensively used. 5-((((2-Iodoacetyl)amino)ethyl)amino)naphthalene-1-sulfonic acid (IAEDANS), chloroacetamide, iodoethanol, arsonous acids, 4-vinylpyridine, and 2-nitro-5-thiocyanobenzoic acid (NTCB) are other examples of reagents that react with thiols. Also, labeling with fluorescent reagents like acrylodan and 5-iodoacetamide fluorescein (5-IAF) has been used. The reactions with the more commonly used labels are shown in Scheme 4.
b. Reaction conditions. The reactions with iodoacetamide (and its derivatives) and N-alkylmaleimides are mostly carried out at pH 6.9–8.7 with 10–300 mM buffer concentrations. The reactions are effective at temperature from 0 to 37°C. Typically, reagent concentrations vary from 1- to 1,000-fold molar excess, and reaction times can be as short as 15 min or as long as 8 hr. For iodoacetamide, reactions are carried out in the dark because this reagent is photolabile. Iodoacetic acid is also best used in the dark, and its reactions are most efficient at pH 8.5. Reagent excesses of 5- to 300-fold are typically used with reaction times from 20 min to 4 hr.

The reactions with the other labels are usually carried out at room temperature or 37°C, at pH between 7 and 9, and with 10–300 mM buffer. These other labels are typically used in a 2- to 300-fold molar excess, and are allowed to react with the protein of interest for times between 10 min and 12 hr.

In general, most of the cysteine-specific reagents can be used to study proteins under native conditions; however, because the labels have limited solubility in water, a small percentage of ethanol, acetonitrile, or dimethyl sulfoxide must be used. Thus, care must be taken to ensure that the amount of organic solvent added will not disrupt the protein structure. The percentage of organic solvent is usually kept as low as possible; the exact amount that may be used will depend on the protein and the organic solvent.

c. Side reactions. Acrylodan can also modify lysine residues (Yem et al., 1992). Iodoacetamide and iodoacetate can modify histidine and lysine residues at pH values above 5 and 7, respectively, to a lesser extent compared to thiol groups, and reactions with tyrosyl groups have been observed infrequently (Heinrikson et al., 1965; Gurd, 1967; Whitehurst et al., 2007). In contrast, the reaction of these two labels with exposed methionine residues is fast and is independent of pH (Gurd, 1967).

d. Other details of reactivity. Reactions with N-alkylmaleimides proceed by nucleophilic addition. These reactions are more specific and faster compared to thioether formation with iodoacetamide, which occurs via nucleophilic displacement (Britto, Knipling, & Wolff, 2002). Iodoacetamide forms a very stable thioether linkage, which is well-suited for LC–MS analysis, whereas adducts formed with N-alkylmaleimides undergo partial ring hydrolysis to form an isomeric mixture of maleamic acid adducts (Whitehurst et al., 2007). In addition, the chain length of the alkyl substituents in N-alkylmaleimides has a non-uniform effect on MALDI-TOF MS ion abundance (Apuy et al., 2001b).

2. Protein Surface Structure Analysis

Free cysteine residues (i.e., residues that are not involved in disulfide bonds) are rarely found on the surface of proteins because thiol groups are very reactive. Consequently, cysteine-specific chemical modifications in conjunction with MS can be a very valuable way to monitor the changes in protein surface topology. In an early example of covalent labeling with MS, the fluorescent label acrylodan and FAB-MS were used to examine the accessibility of the two cysteine residues of recombinant interleukin-1β (Yem et al., 1992). Results indicated that Cys8 was modified; that result was consistent with other studies that probed the accessibility and reactivity of cysteine residues of interleukin-1β. A single lysine residue (Lys103) was also labeled by acrylodan. The authors concluded that the basic pH reaction medium (pH 9.0) was favorable for labeling some lysine residues, and they also suggested that the presence of Lys103 at the bottom of a hydrophobic pocket facilitated the reaction because of the hydrophobic character of acrylodan. If this latter chemistry was found to be more general, then such environment-specific reactivity would expand the utility of amino acid-specific labels. Baldwin and co-workers employed pulsed-alkylation MS to investigate the structural differences between the native heterodimeric form of bacterial luciferase and a folding intermediate that is well-populated in 2 M urea (Apuy et al., 2001b). Pulsed-alkylation MS involves selective ratiometric chemical modification, proteolysis, and MALDI-TOF MS. In these experiments, reactions are initiated by the addition of d5-N-ethylmaleimide, and aliquots are taken at various times followed by addition of H2N-N-ethylmaleimide to completely modify all remaining cysteine residues. MALDI-TOF MS analysis revealed that cysteine residues in the β subunit were not modified in 2 M urea, but that cysteine residues at positions 307, 324, and 325 in the α subunit were modified in the presence of 2 M urea. The data suggested an unfolding of the C-terminal region of the α subunit in the folding intermediate.
Iodoacetic acid, iodoacetamide, and derivatives of these reagents have been used extensively to map cysteine residues. Balaram and co-workers used iodoacetic acid and iodoacetamide to determine the differential reactivities of cysteine residues in *Plasmodium falciparum* triose-phosphate isomerase (Maithal et al., 2002). ESI-MS analysis showed that the rates of modification correlate well with the calculated solvent accessibilities. Whitehurst et al. determined the label-accessible Cys residues in the Sindbis virus E1 and E2 glycoproteins by incubation of the native virus particles in the presence of iodoacetamide (Whitehurst et al., 2007). ESI-MS analysis revealed that three and 5 of the 17 cysteine residues in E1 and E2, respectively, are solvent accessible. They also showed that the conformational changes in the two glycoproteins that were induced by a more acidic pH did not alter the location of the free Cys residues. Patapoutian and co-workers used iodoacetamide and 2-(aminoethyl)methanethiosulphonate to modify the cysteine residues of TRPA1 (Macpherson et al., 2007), which is a member of the transient receptor potential (TRP) family of ion channels, and is expressed in nociceptive neurons. ESI-MS analysis revealed that 14 Cys residues on the cytosolic side of this channel are modified, three of which are required for normal channel function.

Wolff and co-workers investigated the reactivity of the 20 cysteine residues in the tubulin dimer with four different sulfhydryl reagents (Britto, Knipling, & Wolff, 2002), and concluded that solvent accessibility is not the only factor that controlled Cys reactivity. MALDI-TOF MS and N-terminal sequence analysis revealed that five and two cysteine residues in the a and b subunits of tubulin, respectively, were very reactive with iodoacetamide, ethylmaleimide, and IAEDANS. The 3D structure of tubulin indicates that the reactive Cys residues are located within 6.5 Å of positively charged residues that tend to stabilize the thiolate anion and, thus, lower the pKₐ of Cys (Fig. 8A). Some of the unreactive Cys residues were in the vicinity of one or more carboxylate groups that tend to suppress thiol dissociation and raise the pKₐ of Cys residues (Fig. 8B). This study indicated that the local electrostatic environment of Cys residues might be more important than solvent accessibility in determining the reactivity of cysteine residues. The most reactive of all the Cys residues in tubulin was difficult to determine because the labels were so reactive. To overcome this difficulty, the tubulin dimer was modified with less reactive chloroacetamide. These reactions revealed that Cys347 of a-tubulin was the most reactive residue.

Extending the idea further that cysteine reactivity could be used to experimentally determine the pKₐ values of individual Cys residues, Poole and co-workers very recently used an isotope-coded, iodoacetamide-based reagent along with MALDI-TOF MS analysis (Nelson et al., 2008). The modification rates of cysteine residues with isotope-coded N-phenyl iodoacetamide (iodoacetanilide) were measured over a range of pH values. Proteolytic digestion and MALDI–TOF MS were used to monitor the reaction extent of specific cysteinyl residues at various time points. This approach demonstrated that the pKₐ values of the two cysteine residues in Escherichia coli thioredoxin were 6.5 (Cys32) and greater than 10.0 (Cys35), which are in good agreement with previous reports that used chemical modification approaches (Kallis & Holmgren, 1980).

FIGURE 8. Tubulin cysteines in positively charged and negatively charged environments (PDB code 1JFF). A: Cys241 and Cys246 are within 5 Å of Arg320. B: Cys129 is within 6 Å of Glu3. [Color figure can be viewed in the online issue, which is available at www.interscience.wiley.com.]

3. Protein–Ligand Complex Analysis

A number of studies have employed chemical modification of sulfhydryl groups in combination with MS to identify ligand-binding sites. Cesura and co-workers examined the reactivity of cysteine residues in recombinant human catechol O-methyltransferase (rhCOMT) in the presence and absence of S-adenosyl-L-methionine (AdoMet) with iodoacetamido fluorescein (5-IAF) as the labeling reagent and ESI-MS (Vilbois et al., 1994). Results showed that four of the seven cysteine residues were modified in free rhCOMT, whereas the reactivity of Cys68 and Cys94 were dramatically reduced in the presence of AdoMet; these data indicate that these residues are at or near the AdoMet binding site. These findings are in agreement with X-ray crystallographic data (Fig. 9) and are another example of how covalent labeling can be interpreted in light of high-resolution structural data to provide more definitive information about protein interactions in solution. Witkowska and co-workers employed iodoacetic acid for carboxymethylation of the human estrogen receptor ligand-binding domain (hER LBD) in the free and estradiol-bound forms (Hegy et al., 1996). Of the four cysteine residues in hER LBD, MALDI-TOF and ESI-MS analysis showed that only Cys447 was protected from modification in the estradiol complex, suggesting a role for this amino acid in estradiol binding.
acid in the interaction of ER LBD with estradiol. Iodoacetate and MALDI-TOF MS have also been used to investigate the roles of cysteine residues in the activity of glycine reductase, an enzyme responsible for the synthesis of acetyl phosphate, from *Eubacterium acidaminophilum* (Kohlstock et al., 2001). The protection of Cys359 from carboxymethylation in the presence of acetyl phosphate indicated its presence in the active site. N-Ethylmaleimide has been used to probe the ligand-induced structural changes of the A and B subunits of the cyclic nucleotide-gated (CNG) channel of retinal rod photoreceptor cells (Bauer & Krause, 2005). CNG was reacted in the presence and absence of one of the activators of this channel, 8-bromo-cyclic guanosine monophosphate (8-Br-cGMP). Peptide mapping revealed that all cysteine residues in the C-terminal domain of the protein were alkylated with N-ethylmaleimide when 8-Br-cGMP was absent. The reactivity of Cys505 from the A subunit and Cys1104 from the B subunit, which are residues located at the cyclic nucleotide-binding sites, were significantly reduced in the presence of 8-Br-cGMP. Another residue, Cys949 of the B subunit, was only accessible in the nonliganded state of CNG, indicating that this part of the protein experiences major structural changes upon channel activation.

Labeling with N-ethylmaleimide and N-methylmaleimide was also used to study ligand-induced conformational and solvent accessibility changes of translutaminase Factor XIII as a function of activation state (Turner et al., 2004). Cysteine residues in the non-activated, thrombin-activated, and calcium-activated forms were monitored with MALDI-TOF on proteolytic fragments of the labeled protein. Results revealed that only one out of nine cysteine residues, Cys188, was modified in the non-activated state. Three additional cysteine residues (Cys314, Cys409, and Cys695) were modified in the activated state regardless of the way in which the enzyme was activated. These data suggested that the region containing Cys188 either did not undergo any conformational change or maintained its accessibility after activation. Furthermore, the authors concluded that the regions of the protein near the other three cysteine residues underwent significant conformational changes upon activation.

4. Metal-Binding Sites

The thiolate of cysteine is a very common ligand for metal ions in metalloproteins, and when bound to a metal, the thiolate’s reactivity with covalent labels markedly decreases. Therefore, a few studies have used chemical modification and MS to characterize metal-binding sites in proteins. Forest and co-workers used cysteine alkylation by iodoacetamide to investigate the ligands in the zinc binding site of the metalloregulatory Fur protein (Gonzalez de Peredo et al., 1999). The kinetics of Fur alkylation monitored with MALDI-MS showed that one Cys residue was highly reactive, whereas two others had very low reactivities. In the presence of ethylenediaminotetraacetic acid (EDTA), which stripped Zn from the protein, the two slow-reacting Cys residues were as reactive as the other Cys residues; this indicated that the two slow-reacting Cys were bound to zinc. Peptide mapping identified Cys132 as the fastest reacting Cys residue; this is in agreement with secondary structure predictions in which Cys132 is located in an exposed region of the protein. This work by Forest and co-workers illustrates how covalent labeling can provide corroborating evidence for secondary structure predictions. Giedroc and co-workers used pulsed alkylation MS with N-ethylmaleimide to investigate the relative stabilities of the six zinc fingers in the metal-response element (MRE)-binding transcription factor-1 (MTF-1) (Apuy et al., 2001a). The relative cysteine thiolate reactivity they determined, $F_5 > F_6 \geq F_1 > F_2 \approx F_3 \approx F_4$, suggested that the Zn binding stabilities of the F5 and F6 domains were lower than in the other four domains. These observations are consistent with a previous study that showed F5 and F6 are the two domains that most readily lose Zn(II) upon extensive dialysis (Chen, Agarwal, & Giedroc, 1998; Chen, Chu, & Giedroc, 1999).

5. Disulfide Mapping

Disulfide bonds are critical to how a protein folds, and thus their presence and location are important to define protein structure. In recent years, cysteine-specific labels and MS have been used to map disulfides in proteins. The implicit assumption in these experiments is that Cys residues are unreactive when in a disulfide form; however, special control experiments are necessary to distinguish between Cys residues that form disulfides and those that are simply buried in the protein’s interior. Glocker and co-workers used melarsen oxide, an arsonous acid derivative, to selectively bridge bis-cysteinyl residues to characterize the tertiary structure of partially reduced bovine pancreatic trypsin inhibitor (Happersberger, Przybylski, & Glocker, 1998). MALDI-TOF MS peptide mapping revealed that only one of the three disulfide bonds, Cys14-Cys38, was predominantly reduced. The same group employed iodoacetamide to investigate the structural differences between reduced and activated heat shock protein Hsp33 (Barbirz, Jakob, & Glocker, 2000). Peptide mapping showed that all six cysteine residues were alkylated in reduced Hsp33. A peptide with Cys232 and Cys234 as well as a peptide with Cys265 and Cys268 were found unmodified in oxidized Hsp33; these data indicated that these cysteines were linked with a disulfide bridge. In addition, protection of Cys239 from modification in the oxidized form suggested it became buried in the oxidized protein. Macher and co-workers used a biotinylated form of iodoacetamide and ESI-MS to determine whether the two
cysteine residues in α-1,3 galactosyltransferase form a disulfide bond (Shetterly et al., 2001). The positive reactivity of both cysteine residues indicated that the protein does not contain a disulfide bond. Codina et al. (2001) used specific chemical cleavage and cysteine modification with 2-nitro-5-thiocyanobenzozoic acid (NTCB) and 4-vinylpyridine, respectively, to locate the disulfide bonds in mature pea α-L-fucosidase. The first strategy involved reaction of the denatured protein with NTCB to selectively cyanylate free Cys residues, which led to protein cleavage at these sites upon exposure to alkaline conditions. 4-Vinylpyridine simply modified free Cys residues without any subsequent cleavage of the protein. MALDI-TOF MS analysis revealed that reaction of the protein with either NTCB or 4-vinylpyridine resulted in the modification of only one of the five Cys residues. From their experiments, they concluded that there are two disulfide bonds and one free thiol group in the protein. Brown and co-workers used iodoacetamide to probe the nature of the network of disulfide bonds in the Sindbis virus E1 and E2 glycoproteins (Whitehurst et al., 2007). Edwards and co-workers investigated the regulation of the enzyme activity of soybean Glycine max protein tyrosine phosphatase (GmPTP) with glutathione and iodoacetamide (Dixon, Fordham-Skelton, & Edwards, 2005). ESI-MS peptide mapping revealed that Cys266 was resistant to thiolation by glutathione. In contrast, Cys78 and Cys176 were readily thiolated and/or alkylated by iodoacetamide. These results, together with kinetic studies and homology modeling, were used to generate a model for the redox regulation of GmPTP. In their model, GSSG rapidly glutathionylates cysteines 176 and 78 of GmPTP. A slow rearrangement follows, and an intramolecular disulfide forms between Cys78 and the active site Cys266. Franz and co-workers used iodoacetamide to determine the disulfide pattern of Salmon Egg Lectin 24K from the Chinook salmon Oncorhynchus tshawytscha (Yu et al., 2007). Reduction and mass spectral analysis of tryptic peptides showed that there were seven disulfide bonds, but that there were also two free cysteines not involved in a disulfide bond. These results were only partially consistent with disulfide patterns generated by several automated algorithms. From all of the above studies, it is evident that covalent labeling along with MS can offer some insight into the location of disulfide bonds, but the presence of buried thiols introduces some level of ambiguity in the disulfide assignments.

Finally, Fridovich and co-workers used chemical modification to determine whether the dimer of human Cu/Zn superoxide dismutase 1 (hSOD1), which has a solvent-exposed thiol at Cys111, forms an intersubunit trisulfide bond (Okado-Matsumoto, Guan, & Fridovich, 2006). Previous conflicting reports in the literature existed about the modification status of Cys111. hSOD1 contains four cysteines per subunit: two are involved in an intrasubunit disulfide bond, one is buried, and another is solvent-exposed (Cys111). Fridovich and co-workers found that a cysteine residue from hSOD1 was reactive with N-ethylmaleimide, 4-vinylpyridine or 5,5′-dithiobis(2-nitrobenzoic acid) (DTNB); this reactivity indicated that Cys111 existed as a free thiol. In contrast, commercially obtained hSOD1 did not react with the reagents and had other spectral features that were consistent with a trisulfide linkage. The authors finally concluded that, although Cys111 does not have an endogenous trisulfide linkage, the reason for its existence in the commercially obtained sample was unclear.

D. Histidine

1. Reagents

a. Reaction mechanisms. Diethylpyrocarbonate (DEPC) has almost exclusively been used to modify histidine residues. A single modification predominates at low DEPC concentrations, but higher concentrations of the reagent can lead to a second carbethoxylolation of histidine. The reactions of DEPC with histidine residues are shown in Scheme 5.
b. Reaction conditions. The reactions are mostly carried out at room temperature and 37°C in the pH range of 5.5–7.5. DEPC is insoluble in water, but it does dissolve in water-miscible solvents such as ethanol and acetonitrile. Typically, reagent amounts between 0.5- and 1,000-fold molar excess are used with reaction times from 1 to 120 min. These conditions are suitable to study proteins under native conditions; however, because DEPC is not water-soluble, care must be taken to ensure that the amount of organic solvent added does not disrupt the protein structure. The percentage of organic solvent is usually kept as low as possible; the exact amount that may be used will depend on the protein and the organic solvent.

c. Side reactions. Studies have shown that reactions of DEPC with lysine, tyrosine, cysteine, arginine, serine, and threonine residues can be common in some cases (Miles, 1977; Mendoza & Vachet, 2008). The amount of lysine modification increases over the pH range of 6–8 (Dage, Sun, & Halsall, 1998).

d. Other details of reactivity. DEPC hydrolysis is known to occur with a reaction half-life of 9 min at 25°C and pH 7, and the hydrolysis reaction is significantly faster at pH above 7 (Lundblad & Noyes, 1984). Mono-modification of histidine residues is reversible under acidic and alkaline conditions, and in the presence of nucleophiles such as hydroxylamine and tris(hydroxymethyl)aminomethane (TRIS) (Miles, 1977). The half-life of N-carbethoxyimidazole is approximately 55 hr at pH 7, 2 hr at pH 2, and 18 min at pH 10 (Melchoir & Fahrney, 1970). Bis-modification of histidine residues, however, is irreversible. Modification of lysine and tyrosine residues are irreversible and reversible, respectively (Miles, 1977), and it has been recently shown that modifications of Ser and Thr residues are reversible over 20 hr (Mendoza & Vachet, 2008). Finally, the extent of histidine’s reactivity with DEPC increases in the presence of acetate buffers in the pH range of 6–7 (Kalkum, Przybylski, & Glocker, 1998).

2. Protein Surface Structure Analysis

Histidine is the third least frequently occurring amino acid in proteins, yet its aromaticity, moderate basicity, H-bonding capacity, and ability to bind divalent transition metals causes it to be quite commonly involved in protein biochemistry. This fact, and the ease with which DEPC can be used, make histidine a relatively common target to study protein surface structure. Most studies indicate that histidine’s reactivity is controlled by its solvent accessibility and protonation state. In an early study, recombinant human macrophage colony-stimulating factor β (rhM-CSFβ) was investigated with DEPC modification and MALDI-TOF MS (Glocker et al., 1996). Peptide mapping revealed that solvent-accessible histidines were selectively modified at low DEPC:rhM-CSFβ molar ratios (<50:1). Modification of all five histidine residues as well as nonspecific modification of Tyr and Lys residues occurred when high DEPC:rhM-CSFβ molar ratios (>50:1) were used. Halsall and co-workers used MALDI-TOF and ESI-MS to determine the location and extent of DEPC modification of z1-acid glycoprotein (Dage, Sun, & Halsall, 1998). Results indicated that His97 was modified at a faster rate than His100; that difference was presumed to reflect differences in either the ionization state of the imidazole or its accessibility to DEPC. This study also observed DEPC-modified lysine residues, and the amount of lysine modification increased from pH 6 to 8. The reactivity of insulin with DEPC has also been investigated (Kalkum, Przybylski, & Glocker, 1998). Up to six carboxy groups per insulin molecule were detected. At DEPC:insulin molar ratios of 50:1, the products consisted of mixtures of the monomodified (CET), formyl-CET (FCEt)-modified, and urethane-CET (UCEt)-modified histidine. Peptide mass mapping indicated that carboxylation occurred at His5, His10, and N termini of both subunits. However, only biscarboxylation (FCEt and UCEt) of His10 was observed. The ε-nitrogen (NE) and the δ-nitrogen (ND) of His10 both have high solvent accessibility in both subunits, whereas His5 was only accessible on one side of the imidazolyl ring (Fig. 10). These findings reveal that reactivity correlates well with the calculated solvent accessibilities of the imidazolyl nitrogen atoms. Finally, the location and extent of DEPC modifications on cytochrome b561 have been studied (Tsubaki et al., 2000). Cytochrome b561 transports electron equivalents across vesicle membranes to convert intravesicular monodehydroascorbate radicals to ascorbate. DEPC modification reactions of cytochrome b561 were examined to elucidate the mechanism of the transmembrane electron transfer. A map of the tryptic peptides indicated that there were three major modification sites, Lys85, His161, and His88, and three minor modification sites, His92, Tyr218, and Tyr192. All of the DEPC modification sites, except for Tyr192, are fully conserved, and 3D structures showed that these sites are located at a very narrow region on the extraspecific surface of the protein. These conserved residues are found in cytochrome b561 from multiple organisms, which indicates the structural and/or functional importance of these residues in the protein’s ability to accept electrons from ascorbate. An important observation from the work described above together with the work from other groups (Glocker et al., 1996; Dage, Sun, & Halsall, 1998; Kalkum, Przybylski, & Glocker, 1998; Mendoza & Vachet, 2008) is that competing reactions of DEPC with residues other than histidine are fairly common. As a result, peptide mass mapping is not
sufficient to confirm the identity of modified residues; tandem MS must be used. The relatively promiscuous reactivity of DEPC with not only histidine residues but also Lys, Tyr, Ser, Thr, Cys, and occasionally Arg residues also presents an opportunity. These seven amino acids cover approximately 32% of the sequence of the average protein (Trinquier & Sanejouand, 1998). This fact makes DEPC a reasonable standalone reagent for protein surface mapping.

3. Protein–Ligand Complex Analysis

Several studies have also used DEPC modification to characterize protein–ligand interactions. These studies involved modification of histidine residues followed by assessment of the protein’s activity or the reactivity of a protein’s histidine residues with DEPC in the presence and absence of a ligand. In an example of the former, the role of histidine residues in the ligand–receptor interaction of rhM-CSFβ (Glocker et al., 1996) was studied. An increase in the extent of modification of His9 and His15 corresponded with a decrease in rhM-CSFβ bioactivity. This suggested that the A helix of rhM-CSFβ, which contains His9 and His15, is involved in the ligand–receptor interaction. Several examples exist in which protein–ligand interaction sites have been deduced from the DEPC reactivity of a protein in the presence and absence of a ligand. Hondal et al. (2001) investigated the role of histidine residues in heparin binding of rat selenoprotein P. Amino acid sequencing and MALDI-TOF MS revealed that His834 and His547 were carbethoxylated upon reaction with DEPC. The extracellular binding of 4,4'-dinitrostilbene-2,2'-disulfonic acid (DNDS) protected His834, but not His547, from carbethoxylation. DNDS is an extracellular stilbene compound that competitively inhibits anion transport, but binds preferentially to the outward configuration of the protein. The His and Lys data both clearly indicate that the region in selenoprotein P from residues 80 to 91 is part of the heparin-binding region. Hamasaki and co-workers investigated the possible roles of histidine residues in the conformational changes that occur during human erythrocyte band 3-mediated anion exchange (Jin et al., 2003). ESI-MS revealed that His834 and His547 were carbethoxylated in the absence of Cu(II). Upon Cu(II) addition, fluorescence results indicated that COMMD1 bound Cu(II) with a stoichiometry of 1:1, but did not bind with other divalent metals. MS analysis indicated that four histidine residues were carboxethoxylated in the absence of Cu(II). Upon Cu(II) addition, His101 and His134 were protected from modification. As a whole, the data indicated that the copper binding site is located in the region that encompasses residues 61–154.

4. Metal-Binding Sites

An important role of histidine residues in proteins is to bind transition metals, and thus DEPC modification has been used to map metal-binding sites in proteins. A particularly noteworthy example involved an assessment of the Cu(II) binding sites of the prion protein (PrP) (Qin et al., 2002). MALDI-MS data showed that incubation of human PrP with Cu(II) resulted in addition of up to four Cu(II) ions; this number is in agreement with previous reports that there are four Cu(II) binding sites in the N-terminal fragment (HuPrP23-98) of the human protein. However, when the protein was incubated with Zn(II), Ni(II), Mg(II), Mn(II), or Fe(II), no metal-HuPrP23-98 complexes were detected. These results validated the copper-binding specificity of the prion protein. MALDI-MS peptide mapping also indicated the histidine residues within HuPrP23-98 that were protected by Cu(II) from DEPC modification. It was found that five histidine residues were protected from modification when a 10-fold molar excess of Cu was used. In addition, the peptide mass mapping data obtained in the presence of Zn(II) and Ni(II) suggested that one or two histidines were partially protected from DEPC modification. This same study investigated the mouse protein (MoPrP23-231), which contained almost the entire sequence of the prion protein, and the data showed that five histidine residues were also protected from modification by Cu(II) coordination. Post-source decay data of chymotryptic peptides confirmed DEPC modification on His60, His68, His76, and His84 within the four PHGGGSWGQ octarepeat units and on His95 within the related sequence GGGTHNQ of the full length MoPrP23-231.

Sarkar and co-workers used DEPC modification together with EPR, UV–Vis, and fluorescence spectroscopies to characterize the copper-binding properties of the human copper metabolism gene MURR1 domain (COMMD1), which belongs to a family of multifunctional proteins that regulate the nuclear function of the transcription factor nuclear factor-kappa B (NF-κB) (Narindrasorasak et al., 2007). EPR, UV–Vis, and fluorescence results indicated that COMMD1 bound Cu(II) with a stoichiometry of 1:1, but did not bind with other divalent metals. MS analysis indicated that four histidine residues were carboxethoxylated in the absence of Cu(II). Upon Cu(II) addition, His101 and His134 were protected from modification. As a whole, the data indicated that the copper binding site is located in the region that encompasses residues 61–154.

E. Lysine

1. Reagents

a. Reaction mechanisms. Amino groups such as the ε-NH2 of lysine residues and the N-terminal ω-NH2 are most commonly modified by acylation with various organic acid anhydrides such as acetic anhydride, maleic anhydride, and succinic anhydride, as well as N-hydroxysuccinimide derivatives. Biotin N-hydroxysuccinimide derivatives are commonly used to modify amino groups, but the biotin moiety in these studies is not commonly used for purification or isolation purposes. Other examples of amino group-specific reagents are imido esters (e.g., methyl acetimidate and S-methylthioacetimidate), 2,4,6-trinitrobenzenesulfonic acid (TNBS), and S-[(fluorosulfonyl)benzoyl]adenosine (FSBA). The reactions with the most commonly used labels are shown in Scheme 6.

b. Reaction conditions. The reactions are typically carried out between 20 and 37°C and at pH 5–9. Most of the labels are sparingly soluble in water, and thus must be dissolved in organic solvents like acetonitrile, ethanol, dimethyl formamide, and dimethyl sulfoxide. Reagent amounts vary from 1- to 5,000-fold
molar excess, and reaction times range from 2 min to at least 2 hr. For reactions with acetic anhydride, the addition of a base or a well-buffered solution is required to prevent the pH decrease associated with anhydride hydrolysis. The above reaction conditions are mostly suitable to study proteins under native conditions; however, because most of the labels are not water-soluble, care must be taken to ensure that the added organic solvent does not disrupt the protein structure. The percentage of organic solvent is usually kept as low as possible; the exact amount that may be used will depend on the protein and the organic solvent. An interesting aspect of some of the labels such as succinic anhydride and sulfo-N-hydroxysuccinimamide acetate is their ability to efficiently modify lysine residues at acidic pH, which allows these covalent labels to be used under acid-unfolded conditions.

c. Side reactions. Acetic anhydride readily acetylates tyrosine residues, but the O-acytyrosine product is unstable in the presence of acetate or at pH C 9 (Riordan & Vallee, 1967; Scholten et al., 2006). Although succinic and maleic anhydrides are specific for amino groups, reactions with tyrosines are frequent; acetylation of sulfhydryl groups is less common (Klapper & Klotz, 1972). Serine and threonine residues are apparently not modified by these labels because the hydroxyl groups of these amino acids are not as nucleophilic as the phenol group of tyrosine. S-Methylthioacetimidate can modify cysteine residues because the leaving group of this label readily forms a disulfide bond with sulfhydryl groups (Janecki, Beardsley, & Reilly, 2005).

d. Other details of reactivity. Modification of lysine residues with succinic anhydride results in a charge reversal at neutral pH. In contrast, the positive charge of the lysine residue is maintained after modification with methyl acetimidate (or S-methylthioacetimidate). The reactions of amino groups with citraconic anhydride are reversible at pH C 4 (Kadlik, Strohalm, & Kodicke, 2003) or upon treatment with hydroxylamine at pH 10 (Habeeb & Atassi, 1970).

2. Protein Surface Structure Analysis

Lysine is one of the most common amino acids found in proteins, and it is almost always found on protein surfaces. Nonetheless, modification of lysine residues has been the most commonly used covalent labeling strategy to probe protein surface structure. In one of the earliest examples of using covalent labeling and MS, Knock et al. (1991) determined the relative reactivities of amine groups in Aplysia egg-laying hormone with the N-hydroxysuccinimide ester of biotin. Sequence analysis and FAB-MS revealed that the ε-amine group at the N-terminus was the least reactive. Bioactivity assays together with the modification results suggested that the N-terminal amino acid plays a role in the protein’s function. In one of the first examples in which MS was exclusively used to determine specific lysine modification sites, Przybylski and co-workers used acetic anhydride to probe the surface topology of hen egg white lysozyme (Suckau, Mak, & Przybylski, 1992). In this study, 252Cf plasma desorption (PD) and MS peptide mapping revealed the order of reactivity of the lysine residues, which were found to correlate well with the calculated solvent accessibilities from crystal structure data. The same group later employed acetylation and succinylation of amino groups to characterize the surface topology of three other model proteins (Glocker et al., 1994). A two-step acetic anhydride acetylation/trideuteroacetylation was developed to facilitate an estimation of the relative reactivities of each amino group. Again, good correlations between the surface accessibilities determined from crystal structures and the reactivity trends were found. These early results by Przybylski and co-workers were very influential in establishing the potential of covalent labeling strategies when combined with MS.

This initial work with covalent labeling of lysine residues set the stage for the numerous examples that have since appeared in the literature. Several examples, included chronologically below, are summarized to demonstrate the scope of the protein systems that have been studied with lysine covalent labeling. Welfle and co-workers used maleic anhydride to determine the relative reactivities of lysine residues in the HIV-1 capsid protein p24...
(Ehrhard et al., 1996). Results from the labeling experiments along with antibody binding, conformation, and stability assays were used to conclude that an epitope with low-affinity to antibodies is hidden within the p24 and becomes exposed upon conformational changes that were induced by the low levels of maleic anhydride modification. The structure of general-diffusion porins from *Rhodobacter capsulatus* was characterized with lysine modification with succinic anhydride, X-ray crystallography, and MALDI-TOF MS (Przybylski et al., 1996). Mass spectrometric and X-ray crystallographic data revealed that the N-terminus and three lysine residues were succinylated. All of these residues were solvent-accessible and localized on the surface of the inner channel. Single-channel conductance experiments of the labeled porins showed that a change of the positive charge of the lysine residues in the inner channel to a negative charge resulted in an increased cation selectivity and single-channel conductance. Pucci and co-workers characterized the surface topology of Minibody with lysine modification with acetic anhydride (Zappacosta et al., 1997) to test the structural model for the Minibody. Their results were consistent with the existing model except for the lack of reactivity of Lys31, which was attributed to either shielding by a nearby tyrosine residue or the formation of a salt bridge with a nearby aspartic acid. These observations are intriguing because they suggested that covalent labeling of lysine residues might be sensitive to this residue’s involvement in salt bridge structures; that sensitivity is not always obvious from other structural techniques.

Sampieri and co-workers investigated the role of lysine residues on the activity of scorpion Toxin VII (Tsyγ) using a combination of chemical modification and toxicity assays. Sulfo-N-acetate was used to acetylate lysine residues (Hassani et al., 1999), and it was found that only Lys12 was acetylated. Subsequent toxicity and binding assays revealed that a modification of Lys12 led to a large reduction in the toxicity of Tsyγ; this indicated the obvious importance of this residue for the protein’s toxic activity and binding properties. Lysine acetylation with acetic anhydride has also been used to investigate the capsid mobility of the flock house virus (FHV) in solution (Bothner et al., 1999). MALDI-TOF MS of proteolytic and acetylated fragments from six types of FHV particles revealed differences in the degrees of digestion and acetylation. These results indicated that virus particles with identical high-resolution crystal structures can exhibit significant differences in their conformational changes that were induced by the low levels of maleic anhydride modification. The structure of general-diffusion porins from *Rhodobacter capsulatus* was characterized with lysine modification with succinic anhydride, X-ray crystallography, and MALDI-TOF MS (Przybylski et al., 1996). Mass spectrometric and X-ray crystallographic data revealed that the N-terminus and three lysine residues were succinylated. All of these residues were solvent-accessible and localized on the surface of the inner channel. Single-channel conductance experiments of the labeled porins showed that a change of the positive charge of the lysine residues in the inner channel to a negative charge resulted in an increased cation selectivity and single-channel conductance. Pucci and co-workers characterized the surface topology of Minibody with lysine modification with acetic anhydride (Zappacosta et al., 1997) to test the structural model for the Minibody. Their results were consistent with the existing model except for the lack of reactivity of Lys31, which was attributed to either shielding by a nearby tyrosine residue or the formation of a salt bridge with a nearby aspartic acid. These observations are intriguing because they suggested that covalent labeling of lysine residues might be sensitive to this residue’s involvement in salt bridge structures; that sensitivity is not always obvious from other structural techniques.

Sampieri and co-workers investigated the role of lysine residues on the activity of scorpion Toxin VII (Tsyγ) using a combination of chemical modification and toxicity assays. Sulfo-N-acetate was used to acetylate lysine residues (Hassani et al., 1999), and it was found that only Lys12 was acetylated. Subsequent toxicity and binding assays revealed that a modification of Lys12 led to a large reduction in the toxicity of Tsyγ; this indicated the obvious importance of this residue for the protein’s toxic activity and binding properties. Lysine acetylation with acetic anhydride has also been used to investigate the capsid mobility of the flock house virus (FHV) in solution (Bothner et al., 1999). MALDI-TOF MS of proteolytic and acetylated fragments from six types of FHV particles revealed differences in the degrees of digestion and acetylation. These results indicated that virus particles with identical high-resolution crystal structures can exhibit significant differences in their conformational changes in solution.

The membrane topology of guinea pig cytochrome P450 (P450) was characterized by a combination of selective chemical modification and mass spectrometry (Izumi et al., 2003). MALDI-TOF MS peptide mapping revealed that 11 lysine residues were acetylated with acetic anhydride in the detergent-solubilized state. In contrast, Lys29, Lys50, Lys490, and Lys492 were not acetylated in the proteoliposomes, and thus were protected from acetylation by incorporation of the enzyme into liposome membranes. These results suggest that the lysine residues not modified in proteoliposomes are at or near the membrane-binding domains of P450. D’Ambrosio et al. (2003) used acetylation with acetic anhydride to investigate the solution structure of porcine aminoacylase 1 (ACY1), a zinc-binding metalloenzyme. MALDI-MS peptide mapping and limited proteolysis showed that eight out of the 17 lysine residues were acetylated, indicating that these residues are solvent-exposed. Based on these results together with tyrosine modification and cross-linking experiments, an ACY1 structural model was generated wherein the monomer consists of two domains, a catalytic and a dimerization domain.

Novak et al. (2004) used top-down Fourier transform ion cyclotron resonance mass spectrometry (FTICR-MS) combined with acetylation with N-hydroxysuccinimidy acetate to study reactivities of the amino groups of ubiquitin under native and denaturing conditions. Under denaturing conditions, ubiquitin was fully modified with a 21-fold molar excess of the label. Under native conditions, not all of the primary amino groups were modified even with an 81-fold molar excess. FTICR-MS data revealed the order of reactivities to be: Met1 > Lys6 > Lys48 > Lys63 > Lys33 > Lys11 > Lys27, Lys27, Lys29. These data are in good agreement with the solvent accessibility of residues obtained from the crystal structure of ubiquitin.

Reilly and co-workers explored the use of S-methyl thioacetimidate and propionimidate as lysine-modification reagents to probe the surface topology of three model proteins: ubiquitin, carbonic anhydrase and hemoglobin (Janecki, Beardsley, & Reilly, 2005). The authors used these reagents because amidination by these reagents is perceived to have several advantages over previously investigated lysine modification reagents. For example, acylation (e.g., using acetic anhydride) or alkylation by trinitrobenzenesulfonylic acid alters both the size and the charge of the lysine side chain. These structural changes might cause partial denaturation and lead to incorrect conclusions about solvent accessibility. In addition, these side chain modifications might also lead to lower MS detection efficiency because acylation and alkylation can reduce the ionization efficiency of the modified peptides in MALDI or ESI. Experiments with S-methyl thioacetimidate and propionimidate revealed that the reactivities of amino groups in the three model proteins correlated well with the solvent accessibilities calculated from the protein crystal structures. Unlike most other lysine-modification reagents, their studies revealed that these labels can also modify cysteine residues. Using the same reagents, the structure of the *Caulobacter crescentus* ribosome was characterized (Beardsley, Running, & Reilly, 2006). The ribosomal proteins were amidinated before and after ribosome disassembly. ESI-TOF MS revealed that the extent of labeling was consistently lower for the intact ribosome. However, the ribosome stalk proteins, which are known to extend from the core of the organelle and be relatively solvent-accessible, were labeled quite extensively in the intact and disassembled forms. MS data also revealed that the N-terminal region of the stalk protein L12 is not very solvent-accessible. These results are consistent with previous reports that indicated that it is the N-terminal domain that anchors it to the ribosome. Taken as a whole, the labeling results were found to be mostly in agreement with the calculated solvent accessibility in the crystal structures of the 50S subunit of D. radiodurans and the 30S subunit of *E. coli*; that agreement suggested that the overall conformation of the ribosome for the three species is similar. These experiments on the *Caulobacter crescentus* ribosome show the potential of covalent labeling strategies to
study such large protein complexes, which contain more than 50 proteins.

Reilly and co-workers have also used amidination with S-methylisothiourea to probe the structure and activity of trypsin (Liu, Broshears, & Reilly, 2007). MALDI-TOF and LC-ESI-MS revealed that 14 of the 15 amino groups were modified in trypsin. The N-terminus was the lone unmodified site; that finding is consistent with a three-dimensional model of trypsin derived from X-ray diffraction experiments. In this structure, the N-terminal residue is buried on the inside of the folded protein, whereas all of the lysines are located on the surface. Intriguingly, amidinated trypsin was found to be an active proteolytic enzyme and autolysis was minimized.

Zou and co-workers compared the reactivities of lysine residues of full-length native and hyperphosphorylated human replication protein A (RPA) after modification with N-hydroxysuccinimide biotin (Liu et al., 2005). MALDI-TOF MS and ESI-MS analysis revealed that Lys343 was protected from modification in the hyperphosphorylated RPA, indicating that a structural rearrangement occurred upon phosphorylation. Alter and co-workers used the solvent accessibility of lysine residues to evaluate the relevance of three-dimensional models of native RPA generated by remote-homology-based modeling procedures to the native solution-state structure of the protein (Nuss, Sweeney, & Alter, 2006). The reaction rate constants of tryptic fragments with sulfo-N-hydroxysuccinimide were used to monitor the microenvironment of the lysine residues. These reactivities were compared to the anticipated reactivities of candidate structural models. Results showed that this approach could assess structural models and can be a basis for selecting the most relevant model.

Tomer and co-workers used sulfosuccinimidobiotin to probe the surface topology of the E1 and E2 glycoproteins that constitute the protein shell of the Sindbis virus (Sharp et al., 2006). MS analysis identified seven sites of modification in the E2 glycoprotein, whereas only one modification site was identified in the E1 glycoprotein. These results confirm that the E1 protein is almost completely buried in the virus structure. Curruthers and co-workers employed biochemical assays, lysine modification with sulfo-NHS-LC-biotin, and MS to characterize the structural basis of the inhibition of the human erythrocyte glucose transport protein (GLUT1) by cytoplasmic ATP (Blodgett et al., 2007). GLUT1 is a prototypic member of the facilitative glucose transporter family that, when complexed with ATP, displays reduced glucose import capacity but increased affinity for sugar. Their analyses revealed that the labeling of lysine residues 245, 255, and 256 in loop 6–7 is inhibited by ATP. Protection of these residues might reflect altered conformational accessibility of loop 6–7 upon ATP binding. MS data and biochemical analyses suggested that ATP binding to GLUT1 caused the GLUT1 carboxyl terminus to interact with GLUT1 cytoplasmic loop 6–7; that interaction inhibits transport. Sulfosuccinimidyl-6-(biotinamido) hexanoate, another biotin derivative, was used to investigate the surface topology of band 3 anion exchanger, a cell-surface protein, from the human erythrocyte (RBC) and the malaria-infected form (iRBC) (Azim-Zadeh et al., 2007). Three Lys residues (Lys826, Lys829, and Lys892) were modified in band 3 from RBC and only two (Lys829 and Lys892) were modified in band 3 from the iRBC. These results suggested that, upon infection, a conformational change takes place in band 3 from the RBC only in the region around Lys826.

Other biotin-labeled reagents (Sulfo-NHS-biotin, Sulfo-NHS-LC-biotin, and Sulfo-NHS-LC-LC-biotin) have also been used recently in parallel together with MALDI-TOF MS to assess the solvent accessibility of amino acids (Gabant, Augier, & Armengaud, 2008). Data from these parallel reactions showed that the reactivity of these three reagents with the THUMPz protein at four reagent/polypeptide molar ratios did not differ. Assignment of the modified peptides was more accurate due to the triple redundancy caused by specific increments in monoisotopic mass. In addition to lysines, the hydroxyl groups of serine and tyrosine residues were also found labeled. An interesting observation from the many studies that have used biotin-labeled reagents is that very few researchers take advantage of the biotin moiety for selective purification of the modified peptides. In some cases, using avidin columns to selectively purify biotin-labeled peptides might allow modified peptides with low abundance to be more readily identified. The drawback to this approach, however, is that information about the relative reactivity of different protein sites would be lost because the unmodified peptides would not be detected.

3. Protein–Ligand Complex Analysis

With pK_a values typically close to 10.5, more than 99% of lysine residues are positively charged under physiological conditions and are commonly involved in interactions between proteins and negatively charged binding partners. Consequently, covalent labeling of lysine residues along with MS detection has been used effectively to study protein-nucleic acid interactions, which are particularly important to regulate DNA replication and transcription. The surface topology of the thyroid transcription factor 1 homeodomain (TTF1HD)–DNA complex was characterized by Scaloni et al. (1999) who used a combination of limited proteolysis, selective acetylation with acetic anhydride, and MS analysis. MALDI-TOF MS and limited proteolysis revealed that the N-terminus, Lys4, and Lys25 were acetylated in the isolated protein but not in the TTF1HD–oligonucleotide complex; that difference indicated that there is a structural change in these regions upon interaction with DNA. Lys46 in the recognition helix was protected in the complex and, therefore, interacted directly with the oligonucleotide. In addition, Lys64 became accessible to acetylation upon interaction with DNA suggesting that the C-terminal region adopts a more flexible conformation in the complex.

Le Grice and co-workers used covalent labeling and MALDI-TOF MS to identify lysine residues in the heterodimer of HIV-1 reverse transcriptase (RT), which helps form the nucleoprotein complex with viral RNA (vRNA) and the tRNA primer (Kvaratskhelia et al., 2002). Analysis of the free RT revealed that 12 and 15 Lys residues were modified in the p51 and p66 subdomains, respectively. Acetylation of RT was subsequently carried out in the presence of vRNA:tRNA and duplex DNA. In the RT-DNA:DNA complex, two Lys residues in p51 and five Lys residues in p66 were considerably protected from acetylation; those numbers are in good agreement with
crystal structures. The modification pattern for RT-vRNA:tRNA was similar to RT-DNA:DNA, but two additional Lys residues in p66 were protected. These data showed that, like in the RT-DNA:DNA complex, the majority of the RT contacts to vRNA:tRNA is located in the primer-binding cleft.

Zou and co-workers used N-hydroxysuccinimide biotin to identify the lysine residues present in the DNA-binding domains (DBD) A, B, and C located in the p70 and DBD-D of the p32 subunits of human replication protein A (hRPA) (Shell et al., 2005). Lysine residues of hRPA were modified in the presence and absence of the single-stranded (dT)$_{30}$ oligonucleotide. MS analyses revealed that seven lysine residues in the p70 subunit and no residues from the p32 and p14 subunits were protected from modification by the single-stranded DNA (Fig. 11). Residues readily biotinylated in the free RPA protein but protected from modification in the nucleoprotein complex include Lys263 from DBD-A, Lys343 from DBD-B, and Lys489, Lys577, and Lys588 of DBD-C. All these residues are located in the binding cleft of DBD, and the data indicated that these five residues interact with the ssDNA. Lys183 and Lys259, which are in the DBD but not in the binding cleft of the DBD, were also protected from modification. The reduced reactivity of these residues was attributed to a conformational change within the p70 subunit upon ssDNA binding. These observations again highlight the utility of combining covalent labeling data with structural data (e.g., crystal structures) because, without such complementary data, assignment of residues as being directly involved in binding is difficult with only the covalent labeling data.

Lysine modification in conjunction with MS has also been used to probe complexes between proteins and other types of ligands. Remacle and co-workers investigated the amino acid residues that might be involved in the substrate binding and catalysis of l-alanine dehydrogenase from Bacillus subtilis with chemical modification and kinetic studies (Delforge et al., 1997). Amino groups were modified by reactions with 2,4,6-trinitrobenzenesulfonic acid (TNBS), N-succinimidyl 3-(2-pyridyldithio)propionate (SPDP) and 5′-(p-(fluorosulfonyl)benzoyl) adenosine (FSBA). Amino acid sequencing and ESI-MS showed that Lys74, which is part of the protein’s active site, was modified in the free protein, but not when the protein was bound simultaneously to the reduced form of nicotinamide adenine dinucleotide (NADH) and pyruvate. Interestingly, their results showed that little protection was afforded by NADH or pyruvate alone; the necessity of the ternary complex for activity was suggested. Sanz and co-workers examined the role of lysine residues in the heparin-binding bovine seminal plasma protein PDC-109 (Calvete et al., 1999). Peptide mapping with N-terminal sequencing and ESI-MS revealed that, in the heparin-bound protein, Lys34, Lys59, and Lys68 were protected from modification, indicating that these basic residues are present in the heparin-binding region of PDC-109.

The ligand-induced conformational changes in transglutaminase Factor XIII were investigated by Maurer and co-workers with acetylation with acetic anhydride and MS detection (Turner et al., 2004). Twelve of the 38 lysine residues in the protein were modified by acetic anhydride in the non-activated and activated forms. This indicated that the regions containing these residues either do not undergo any conformational change or maintain their accessibility after any structural change. When the enzyme was activated by calcium, three new lysine residues were modified: Lys221, Lys677 (or Lys678), and Lys73. Similarly, when FXIII was activated by thrombin, three new lysine residues were modified, but in this case Lys156 was acetylated instead of Lys73. Clearly, these data indicate that FXIII undergoes significant and different conformational changes when activated by calcium or thrombin. A similar modification approach was employed by Liu and co-workers to investigate the heparin sulfate (HS)-induced conformational changes in heparan sulfate 3-O-sulfotransferase-1 (3-OST-1) (Edavettal et al., 2004). Lysine residues of 3-OST-1 were modified with acetic anhydride and hexadeuteroacetic anhydride in the absence and presence of HS. The relative reactivities of the amino groups were measured from the MALDI-TOF MS abundance ratio of acetylated and deuteriacetylated peptides. The data suggested that HS induces a conformational change in 3-OST-1, particularly in the C-terminal region of the protein. Lysine modification combined with MS was used to characterize the conformational changes induced by

![FIGURE 11. Crystal structures of the p70 subunit of human replication protein A. A: Structure of DBD-AB (PDB code 1JMC). Lys263 and Lys343 are found in the binding clefts of DBD-A and DBD-B, respectively, and are protected from modification by direct contact with the single-stranded (ss) piece of DNA. Lys183 and Lys259 are also protected from modification when ssDNA is present, even though they are not bound to the ssDNA. B: Structure of DBD-C (PDB code 1L10). Lys489, Lys577, and Lys588 are located in the binding cleft of DBD-C and are protected from modification when ssDNA is present. [Color figure can be viewed in the online issue, which is available at www.interscience.wiley.com.]](image-url)
peptide binding to HLA-DR1, which is a class II major histocompatibility complex (MHC) variant (Carven & Stern, 2005). Four lysine residues were modified in the empty and peptide-loaded forms. However, modifications at Lys67 in the α subunit and Lys98 in the β subunit, which are not in the peptide-binding region, were only observed for the empty protein. These results indicated that there is a conformational change upon peptide binding in the region around these Lys residues.

The Turchi group used chemical modification with biotin-labeled reagents along with limited proteolysis to determine the influence of DNA binding on the structure of the protein Ku (Lehman, Hoelz, & Turchi, 2008). Ku is a heterodimeric protein composed of 70 (Ku70) and 80 (Ku80) kDa subunits that binds broken DNA ends and initiates the nonhomologous end joining (NHEJ) repair of DNA double-strand breaks induced by exogenous sources. Sites in the C-terminal region of Ku70 were differentially modified in the absence and presence of DNA, whereas no detectable changes in the modification extent of the C-terminal region of Ku80 were observed. Limited proteolysis experiments of free and DNA-bound Ku, however, revealed different proteolysis patterns that suggested changes in the C-terminal domains of Ku70 and Ku80. The m22G10-methyltransferase protein (TrmG10)–tRNAAsp complex was also characterized with biotin-labeled reagents (Sulfo-NHS-biotin, Sulfo-NHS-LC-biotin, and Sulfo-NHS-LC-LC-biotin) to confirm the mode of interaction between the protein and its nucleic acid substrate (Gabant, Augier, & Armengaud, 2008). In the absence of RNA, 13 residues in TrmG10 were labeled. Protection of Lys90 and Lys153 from modification in the nucleoprotein complex indicated direct involvement of these residues with the nucleic acid.

Finally, selective acetylation of lysine residues with acetic anhydride in combination with PLIMSTEX (protein–ligand interactions by mass spectrometry, titration, and H/D exchange) was used to study the interaction of double-stranded telomeric DNA with the DNA-binding domain of human telomeric repeat binding factor 2 (hTRF2) (Sperry et al., 2008). This protein plays an important role in capturing human telomeres to protect them from DNA damage repair systems. Amide H/D exchange revealed portions of the protein that have contacts with the phosphate backbone of DNA, whereas acetylation reactions showed that there was a decrease in solvent accessibility of regions containing Lys447 and Lys488, which are likely involved in the interactions with the DNA major and minor grooves. Combining amide H/D exchange and selective side chain modification is an approach that much potential to map protein structures, protein–ligand interactions, and protein–protein interactions because of the complementary information available with these two methods.

4. Protein–Protein Complex Analysis

Lysine residues are very commonly found on the surface of proteins, but they are not as commonly found at protein–protein interfaces, presumably due to the energetic cost of burying its positively charged side chain. Nonetheless, the frequency of lysine residues on protein surfaces still make them useful probes of protein–protein interactions, and numerous examples of lysine-selective modifications to study these binding inter-

actions have appeared over the last 15 years. In early work, Ohguro et al. (1994) probed the conformational changes and regions of the protein arrestin that interact with photoactivated and phosphorylated rhodopsin (P-Rho+) with acetic anhydride and deuterated acetic anhydride. MALDI-TOF MS analysis showed that 15 lysine residues were protected from modification, and three lysine residues in the C-terminal region of arrestin demonstrated increased levels of modification in the presence of P-Rho+. The changes in acetylation were attributed to steric protection by P-Rho+ and conformational changes within arrestin that alter the reactivity of individual lysine residues. Przybylski and co-workers investigated the interactions of hen egg white lysozyme with a monoclonal IgM-type antibody with acetylation with acetic anhydride (Fiedler et al., 1998). MALDI-MS and PD-MS peptide mapping of free and antibody-bound protein showed that Lys13 and Lys96 were protected from modification in the complex to indicate that it is part of the epitope surface area of native lysozyme. The same group characterized the complexes of the elongation factor proteins EF-Ts and EF-Tu-GDP from Thermus thermophilus (Glocker et al., 1998). Acetylation reactions of EF-Ts, EF-Tu-GDP, and EF-Ts EF-Tu with acetic anhydride were monitored with ESI-MS. Results indicated that there are structural changes in the effector loop of EF-Tu and in a central helix-turn-helix region of EF-Ts upon complex formation, and that the N-terminal region of EF-Ts is involved in its interaction with EF-Tu. Tomer and co-workers used lysine covalent modification to characterize an epitope of the HIV core protein p24 (Hochleitner et al., 2000). Their data revealed that the relative reactivities of five lysine residues of the free protein and monoclonal antibody 5E2.A3-bound protein were similar, but that the N-terminal proline residue showed reduced reactivity in the complex to indicate that Pro1 is part of the epitope on HIV-p24. Wang et al. (2004) characterized the interactions of rhodopsin and the α-subunit of transducin (Gt) with acetylation of lysine residues with sulfo succinimidyl acetate. Membrane preparations of unacti-

vated (Rh) and light-activated rhodopsin (Rh*) in the presence and absence of a synthetic peptide that corresponded to the C-terminal residues of the α-subunit of Gt, Gt(340–350), were acetylated. Results showed that the modification extents of the lysine residues located in the cytoplasmic C1 and C2 loops and the C-terminal tail of rhodopsin were decreased upon light activation; that decrease indicated that these regions undergo conformational changes. In the presence of Gt(340–350), acetylation sites on cytoplasmic loops 1, 2, and 4 of Rh* were protected, suggesting that these sites are involved in the interaction with Gt. The data also indicated an interaction between the end of the C-terminal tail of rhodopsin and Gt in the unactivated state.

Heck and co-workers used a combination of lysine acetylation with N-acetyl-succinimide and nanoLC-MALDI MS to probe the surface interactions of the DNase domain of Colicin E9 (E9) with its immunity protein Im9 (Scholten et al., 2006). This system was used to test if chemical modification and nanoLC-MALDI MS, together with data filtering with immunon marker ions, could be used to characterize protein–protein interactions. Experiments with unbound E9 revealed that all 20 lysine residues and the N-terminus were accessible to modification; those data are in good agreement with the
calculated solvent accessibilities from the 3D structure. In the presence of Im9, Lys55, Lys63, Lys76, Lys81, Lys89, and Lys97 of E9 were protected from modification (Fig. 12). Based on the crystal structure of the E9:Im9 complex, Lys89 and Lys97 are involved directly in the interaction site. Lys76 and Lys81 are located in a α-helical region near the interaction surface, and Lys55 and Lys63 do not interact with Im9. These data suggest that Im9 binding to E9 induces conformational changes in some amino acid residues. These results again illustrated that chemical modification approaches can be used to probe protein–protein interactions and provide complementary data to structural models from X-ray crystallography. The findings also indicated some of the differences in the solution–phase structure of the E9:Im9 complex compared with the crystal structure. These differences could be attributed to the absence of crystal lattice effects in the solution–phase complex that is probed by covalent labeling. Lysine acetylation with acetic anhydride was used by Kominami and co-workers to study the surface interactions of cytochrome P450 17α (P450) with NADPH-cytochrome P450 reductase (CPR) (Nikfarjam et al., 2006). MALDI-TOF MS revealed dual acetylation at Lys326 and/or Lys327 of P450 in the absence of CPR, which were modifications not detected in the presence of CPR. These data indicate that residues Lys326 and/or Lys327 of the J-helix play critical roles in the interaction of P450 with CPR. The same approach was used by Tomer and co-workers to probe the dimer interface of yeast MutLα, a heterodimer of MLH1 and PMS1 (Cutalo et al., 2006). MALDI and ESI-MS revealed that 34 out of the 43 lysine residues in MLH1 were acetylated. Three residues (Lys665, Lys675, and Lys704) were acetylated in the monomer MLH1 but were protected from acetylation upon formation of the MLH1-PMS1 dimer. These data indicate that these three residues are at or near the dimer interface. The authors used these results to refine secondary structure predictions and to develop homology models for the N- and C-terminal regions of MLH1.

Norris and co-workers used limited proteolysis and chemical modification to study the surface topology of the HIV core protein, p24, which is recognized by two different monoclonal human antibodies (mAbs) isolated from HIV + patients (Williams et al., 2006). The results from limited proteolysis experiments provided a rough estimation of the regions of p24 recognized by each of the mAbs. Lysine modification with acetic anhydride allowed for more specific localization of the binding interfaces between the two different mAbs and p24. The epitopes that these two mAbs recognize appear to be adjacent or possibly overlap because both mAbs protected two nearby lysine residues, Lys131 and Lys140, from acetylation.

Finally, because NADPH-cytochrome P450 reductase (CPR) and biliverdin reductase (BVR) are thought to compete with each other in binding to heme oxygenase-1 (HO-1), it was suggested that BVR’s and CPR’s binding sites on HO-1 partially overlap (Wang & Ortiz de Montellano, 2003). Acetylation of the lysine residues of rat HO-1 by acetic anhydride in the absence and presence of CPR or BVR were carried out to test this idea (Higashimoto et al., 2008). The presence of CPR hindered the acetylation of Lys149 and Lys153 on HO-1, which are located in the F-helix. In contrast, the presence of BVR did not hinder acetylation of any lysine residues on HO-1. These results indicate that the interaction mechanism of BVR with HO-1 is somewhat different from that of CPR with HO-1.

F. Tryptophan

1. Reagents

a. Reaction mechanisms. Tryptophan residues are most commonly modified by 2-hydroxy-5-nitrobenzyl bromide, which is known as Koshland’s reagent, N-bromosuccinimide, and o-nitrophenylsulfenyl chloride. The reactions of these reagents with tryptophan are shown in Scheme 7.

b. Reaction conditions. The reactions of all labels are typically carried out at room temperature. Koshland’s reagent can be used over a pH range of 4–7.6, but its specificity for tryptophan is greatest at a value between 4 and 5. The label is soluble in dry acetonitrile or dimethyl sulfoxide. Reagent amounts typically vary from 1- to 500-fold molar excess, and reaction times range from 5 to 30 min. For N-bromosuccinimide, a pH range of 5–7 is typically used, but its specificity for tryptophan is greatest at the pH closer to 5. Reagent amounts vary from 5- to 200-fold molar excess, and reaction times range from 1 to 5 min. Unlike Koshland’s reagent and N-bromosuccinimide, the reactions of o-nitrophenylsulfenyl chloride are carried out at acidic pH where it is more soluble. Equimolar amounts of the label can be used with reaction times around 30 min.

The above reagents work best at acidic pH, which makes them less suitable to study proteins under native conditions; however, they can be readily used to study acid-unfolded proteins. The labels are also not water-soluble, and care must be taken to ensure that the amount of organic solvent used to solubilize them does not disrupt protein structure. The percentage
of organic solvent is usually kept as low as possible; the exact amount that may be used will depend on the protein and the organic solvent.

c. Interferences. For reactions with N-bromosuccinimide, the presence of halides in the solvent must be avoided because these ions are readily oxidized by the reagent to their elemental form, which can have damaging and irreproducible effects on the studied protein (Horton & Koshland, 1984).

d. Side reactions. Koshland’s reagent readily reacts with cysteine and tyrosine residues at acidic (pH < 4) and alkaline pH, respectively (Horton & Koshland, 1965). Side reactions of N-bromosuccinimide with arginyl, tyrosyl, histidyl, amino and sulfhydryl groups are encountered occasionally (Spande & Witkop, 1967). However, the modification of tryptophan residues decreases at more basic pH (Inokuchi et al., 1982). Sulfenylation of cysteine residues occurs to the same extent as modification of tryptophan (Fontana & Scoffone, 1972).

e. Other details of reactivity. Koshland’s reagent is extremely sensitive to hydrolysis. The reagent is also photosensitive, so reactions with this compound are carried out in the dark to minimize loss of label. A precipitate of 2-hydroxy-5-nitrobenzyl alcohol can form from the hydrolytic product of Koshland’s reagent. Tryptophan oxidation by N-bromosuccinimide is reversible (Carven & Stern, 2005).

2. Protein Surface Structure Analysis

Tryptophan is the least abundant amino acid in proteins. This fact, and tryptophan’s hydrophobicity cause it to be rarely found on the surface of proteins. Hence, there are only a few studies that use chemical modification of tryptophan to characterize protein surface structures. One of these studies was the investigation of the role of tryptophan residues in the activity of scorpion Toxin VII (Ts7) using a combination of chemical modification, mass spectrometry, and toxicity assays (Hassani et al., 1999). Tryptophan residues on Ts7 were sulfenylated with o-nitrophenylsulfenyl chloride, and MALDI-MS peptide mapping identified Trp39, Trp50, and Trp54 as the sites of modification. Toxicity and binding assays revealed that modification of Trp39 and Trp54 led to a large reduction in the toxicity of the protein in mice and in the binding affinity for rat brain and cockroach synaptosomal preparations. Modification of Trp50 also decreased the protein’s toxicity in mice, but only moderately affected its binding properties. The authors thus concluded that only Trp39 and Trp54 are essential for the protein’s toxic activity and binding properties. Furthermore, homology modeling starting from the 3D structure coordinates of either the variant 3 of Centruroides sculpturatus or a-type toxin Androctonus (AahII) showed that these two tryptophan residues were clustered in the hydrophobic region of the protein and were presumed to be part of the docking site of the toxin. These two structures were chosen because the primary structure of Ts7 is homologous to the non-toxic “variant 3” protein from the Centruroides sculpturatus scorpion and AahII. Kodíček and co-workers investigated the potential of Koshland’s reagent to assess the surface accessibility of tryptophan residues in four model proteins: cytochrome c, human serum albumin, myoglobin, and lysozyme (Strohalm et al., 2004). The results showed that, under native conditions, only the surface-accessible tryptophan side chains were modified, whereas in the denatured proteins all the tryptophan residues were modified. These promising results suggested that Koshland’s reagent can be used effectively to probe the presence of tryptophan residues on a protein’s surface. Finally, a recent study by Edwards and co-workers investigated the use of halo compounds such as chloroform, 2,2,2-trichloroethanol (TCE), 2,2,2-trichlorooacetate (TCA), and 3-bromo-1-propanol (BP) under UV irradiation at 280 nm to probe the accessibility of tryptophan residues of carbonic anhydrase (Ladner, Turner, & Edwards, 2007). The light-driven reactions attach a formyl (chloroform), hydroxyethanone (TCE), carboxylic acid (TCA), or propanol group (BP) onto the indole ring of Trp, depending on the reagent used. MALDI-TOF MS analyses revealed that the reactivity of the Trp residues correlated well with their solvent accessibilities.

3. Protein–Ligand Complex Analysis

Only two studies have appeared in which tryptophan modification has been used to investigate protein–ligand binding. In one example, Takahashi et al. (1999) used N-bromosuccinimide to investigate the amino acid residues essential for the inhibitory activity of the α-amylase inhibitor (PHA-I) of Phaseolus vulgaris, which has a tetrameric structure (αβ)4. Two out of eight tryptophan residues were modified in free PHA-I—one tryptophan residue (Trp188) on each β subunit. In contrast, no modifications were detected when PHA-I was complexed with two porcine pancreatic α-amylase molecules. These results indicated that one tryptophan residue is located on each of the two active sites of PHA-I, and both are essential for activity.
In another example, the role of tryptophan residues in the interaction of DNA with the ethanol regulon transcription factor AlcR from *Aspergillus nidulans* was investigated (Marie et al., 2001). ESI-MS of the intact protein revealed that an average of 1.75 tryptophan residues in free AlcR was modified by the Koshland’s reagent, whereas an average of 0.58 residues was modified in the AlcR-DNA complex. Peptide mass mapping revealed that the rate of Trp45 modification decreased from 70% to 4% in the presence of DNA to indicate that this residue is located at the AlcR–DNA interface. The decrease in the rate of modification of Trp53 from 24% to 5% suggests that either this residue is involved in DNA recognition or the C-terminal region becomes more ordered upon complex formation. In contrast, Trp36 becomes more solvent exposed upon DNA binding as shown by the increase in its modification rate.

On a final note, although tryptophan-specific reactions have not been used very much to study protein–ligand or protein–protein complexes, there is reason to think that such experiments could be very valuable. An analysis of several protein complexes by Bogan and Thorn (1998) indicated that tryptophan is likely to be a very important residue at the interface of many protein–protein complexes.

### G. Tyrosine

#### 1. Reagents

**a. Reaction mechanisms.** The typical reagents used to modify tyrosine residues are tetranitromethane, iodine, and N-acetylimidazole. The reactions involving these reagents are shown in Scheme 8.

**b. Reaction conditions.** The reactions of all labels are typically carried out at room temperature to 37°C and in a pH range of 6.5–9. Tetranitromethane and N-acetylimidazole are soluble in ethanol and water, respectively. Iodine, when mixed with KI to produce I⁻�, is soluble in water. Iodination with chloramine T and NaI is also readily carried out in water. The amount of tetranitromethane that is used varies from 2- to 5,000-fold molar excess, whereas the amount typically used for the other reagents varies from 1- to 150-fold molar excess. These reaction conditions are suitable to study proteins under native conditions; however, for the labels that are not as water-soluble, care must be taken to ensure that the amount of organic solvent added does not disrupt the protein’s structure. The percentage of organic solvent is usually kept as low as possible; the exact amount that may be used will depend on the protein and the organic solvent. In addition, it is important to be aware of the oxidizing capability of the elemental halides.

**c. Side reactions.** N-Acetylimidazole reacts with lysine and serine residues less readily (Houston & Walsh, 1970; Riordan & Vallee, 1972b). At high molar excess and pH ≥ 8, side reactions of tetranitromethane with histidine, methionine, and tryptophan can be significant (Sokolovsky, Harell, & Riordan, 1969). Iodination with histidyl residues to a significant extent; it oxidizes methionine, cysteine, and tryptophan considerably (Koshland et al., 1963; Filmer & Koshland, 1964).

**d. Other details of reactivity.** When reactions with tetranitromethane are performed at acidic pH, covalent cross-linking of tyrosyl residues can occur to produce inter- and intramolecular associations (Riordan & Vallee, 1972a). At alkaline pH, the o-nitrotyrosine product is relatively stable. In contrast, o-acetyltyrosine, which is formed upon reaction with N-acetylimidazole, is unstable under mildly alkaline conditions (Lundblad & Noyes, 1984). The presence of nucleophiles, such as common buffers like TRIS, can also decrease the stability of o-acetyltyrosine.

#### 2. Protein Surface Structure Analysis

Tyrosine is a fairly useful residue to probe protein surface structures because it can either be buried or exposed. In one
example, Pucci and co-workers used \textit{N}-acetylimidazole and tetranitromethane to probe the surface topology of Minibody (Zappacosta et al., 1997). ESI-MS analysis revealed that three tyrosine residues were modified by both labels, whereas only Tyr24 and Tyr47 were modified by tetranitromethane. The five tyrosine residues modified by tetranitromethane are consistent with the model for Minibody. The inability of Tyr24 and Tyr47 to react with \textit{N}-acetylimidazole was attributed to hydrogen bonding of the phenolic group of tyrosine with other amino acids, which impairs the reaction with \textit{N}-acetylimidazole. If the subtle differences in reactivity of tetranitromethane and \textit{N}-acetylimidazole were also observed in other protein systems, then the differential reactivity of these two labels could provide insight into the microenvironment of tyrosine residues. Leite and Cascio (2002) used tetranitromethane reactivity and MS to probe the topology of the human glycine receptor (GlyR) to determine which topological model is a better description of GlyR. ESI-MS data showed that six out of the 16 Tyr residues were nitrated. All modified Tyr residues, except two, were located at the hypothesized \textit{N}-terminal domain of GlyR, and the reactivities are in agreement with their solvent accessibilities calculated from one of the topological models. Modification of Tyr161 is also consistent with a proposed model wherein this residue is near the \textit{N}-terminal region of the ligand binding site.

Despite observations that tyrosine reactivity with different labels reflects the residue’s solvent accessibility, two other studies question the simplicity of this connection. Kodı´cˇek and co-workers investigated the dependence of tyrosine reactivity on solvent accessibility for three model proteins (cytochrome \textit{c}, lysozyme, human serum albumin) with tetranitromethane and iodine (Santrucek et al., 2004). The reactivities obtained for the three proteins did not have an apparent correlation with solvent accessibility. For example, Tyr148 and Tyr341 of human serum albumin showed high reactivity despite their low solvent accessibility, and several tyrosine residues, such as Tyr263 and Tyr497, have high solvent accessibility but showed poor reactivity (Fig. 13). In addition, some tyrosine residues like those at positions 161 and 411 exhibited different reactivity towards iodine and tetranitromethane. These data suggest that other factors such as the presence of charged residues near a tyrosine residue or the conformational flexibility of the protein in solution affect tyrosine reactivity.

McGuirl and co-workers also found a weak correlation between reactivity and solvent accessibility when peroxynitrite and tetranitromethane were used to nitrate tyrosine residues in two conformational isomers of the recombinant hamster prion protein (residues 90-232; PrP90-232) (Lennon et al., 2007). Six Tyr residues of the normal cellular isoform of PrP90-232 were nitrated by peroxynitrite, but the degree of nitration did not correlate well with the surface accessibility of these residues. Both nitrating agents did show that the most solvent-exposed tyrosines, Tyr225 and Tyr226, in the C-terminal region of the normal cellular isoform were the most reactive; that reactivity allowed for some conclusions to be drawn for this region of the protein. In the \textit{b}-isoform of PrP90-232, the degree of nitration at Tyr225 and Tyr226 was much lower, which indicates a change in the local environment around these residues. Two additional Tyr residues, Tyr149 and Tyr150, were modified in the \textit{b}-isoform, which were not modified in the normal cellular isoform. These results indicated that this region of the protein also underwent a change upon conversion to a \textit{b}-rich isoform.

3. Protein–Ligand Complex Analysis

Wallace and co-workers employed chloramine \textit{T}-mediated iodination of tyrosine residues to characterize the insulin-like
growth factor (IGF) binding sites of bovine insulin-like growth factor protein-2 (biGFP-2) (Hobba et al., 1996). Their experiments revealed that five of the six tyrosine residues of biGFP-2 were iodinated. Di-iodotyrosyl derivatives were detected for all except one of the tyrosine residues, indicating that the other five tyrosine residues were very solvent accessible. The extent of iodination of Tyr60 from biGFP-2 was reduced fivefold when complexed with IGF. Also, the reactivity of Tyr71 increased in the biGFP-2:IGF complex. These results were interpreted to indicate that Tyr60 either directly interacts with IGF or lies in a region that undergoes a conformational change upon IGF binding. The role of tyrosine residues in the inhibitory activity of the α-amylase inhibitor (PHA-I) of Phaseolus vulgaris was investigated with N-acetylimidazole and tetranitromethane (Takahashi et al., 1999). Two tyrosine residues were modified with N-acetylimidazole with a concomitant loss in activity; however, upon analysis to identify the modified residues, the resulting O-acetyltyrosines were hydrolyzed. Hence, tetranitromethane was used to identify the reactive tyrosine residues. Results revealed an addition of up to 2.1 nitro groups in the free protein, whereas no modification was observed when PHA-I was complexed with two porcine pancreatic α-amylase molecules. Peptide mapping and sequencing identified Tyr186 of the β-subunit as the site of nitration, which indicated the importance of this residue for the inhibitory activity of PHA-I.

Mailfait et al. (2000) probed the role of tyrosine residues in the binding of the retinoic acid receptor α (RARα) to retinoic acid (RA) with tetranitromethane. The data revealed that only Tyr277 of the three Tyr residues in the ligand-binding domain (LBD) of RARα was not modified. From these observations, the researchers concluded that Tyr277 is directly involved in RA binding. Selective chemical modification with tetranitromethane combined with MS was used to characterize the conformational changes of HLA-DR1 induced by peptide binding (Carven & Stern, 2005). MALDI-MS peptide mapping of tryptic digests revealed that five tyrosine residues were nitrated in empty and peptide-loaded conformations; that similarity indicated that the local environments that surround these residues are the same in both forms.

4. Protein–Protein Complex Analysis

Covalent labeling of tyrosine has great potential to identify protein–protein interface sites because this residue is the third most important residue at protein–protein interfaces (Bogan & Thorn, 1998). Even so, only two studies have appeared in which tyrosine reactivity has been used to obtain such information. The first is a study of the interaction between urokinase-type plasminogen activator (uPA) and its glycoplid-anchored receptor (uPAR) (Ploug et al., 1995). The receptor binding site of uPA is within its isolated growth factor-like module, residues 4-43 (GFD). Tyrosine modification with tetranitromethane was performed on the uncomplexed uPAR and GFD and on the uPAR-GFD complex. One and six tyrosine residues were nitrated in GFD and uPAR, respectively. In the complex, because Tyr57 of uPAR and Tyr24 of GFD were protected from modification, it was suggested that these residues are directly involved in the formation of the complex. In a second study, Przybylski and co-workers investigated hen egg white lysozyme and its complex with a monoclonal IgM-type antibody with iodination (Fiedler et al., 1998). MALDI-MS showed that because the level of modification of Tyr20 and Tyr23 in the free and antibody-bound forms were the same, these residues are not part of the epitope structure.

IV. FUTURE PROSPECTS

Selective covalent labeling in conjunction with MS has been extensively used in recent years to study protein structure and interactions. As we have shown in this review, the information obtained from this methodology is complementary to other biophysical techniques. Further improvements in this approach are possible, and could offer even greater insight into protein structure in solution. One of the needed improvements is the development of new amino acid-specific labels. Although many reagents are available and have been successfully used, the chemistry used to modify proteins is dominated by electrophilic reagents that target nucleophilic functional groups. There is a need for alternative reagents that can target different amino acid side chains with similar levels of selectivity and yield. For example, new labels that can modify the carboxylate groups of aspartate and glutamate are needed. The reaction conditions (i.e., low pH for EDC) for the labels currently available are not readily amenable probing protein structure under physiological conditions.

Even though most investigations so far have found reasonable correlations between reagent reactivity and side chain solvent accessibility, another potential improvement would be the development of reagents or combinations of reagents that could provide more specific information about the local chemical environment of amino acids. For instance, two of the reagents used to modify tyrosine side chains, N-acetylimidazole and tetranitromethane, target different sites on the tyrosine side chain, the hydroxyl group and the aromatic ring, respectively. Because these functionalities can often be found in different microenvironments within a protein structure, a combination of N-acetylimidazole and tetranitromethane might provide more detailed information on the position of tyrosine side chains in a given protein. Other combinations of reagents or even multifunctional reagents (e.g., with charged and residue-specific reactive functionalities) might offer more specific reactivity that allows the chemical environment around a given residue to be probed more precisely.

Right now, most covalent labeling data are used in conjunction with structural data, but a way in which covalent labeling could be used as the sole experimental data set is to combine it with computer modeling. Several groups have used selective covalent labeling data with homology modeling to generate structural models (Fiedler et al., 1998; Hassani et al., 1999; D’Ambrosio et al., 2003; Dixon, Fordham-Skelton, & Edwards, 2005; Cutalo et al., 2006). To date, this approach has required three-dimensional structural information (i.e., crystal structures or NMR structures) of related proteins with high sequence homology to the protein of interest. Another group has employed chemical modification to assess computational structural models, and they have shown that covalent labeling
data can be used to select the most relevant model (Nuss, Sweeney, & Alter, 2006). With the explosion of computing power and the development of new algorithms to predict protein structure, the correct covalent labeling data could provide the necessary constraints to predict a protein’s structure. The amount and type of labeling data that is necessary to provide the appropriate structural constraints for a given computational model needs to be explored. Questions like the following must be addressed. What is the minimal number of amino acid sites that need to be probed? Is the extent of an amino acid’s reactivity important, or is binary information (i.e., reactive or not) sufficient? Is information about solvent-accessible amino acids adequate or is information about buried residues also necessary?

In many studies, the identification of the modified fragments can often be difficult and, therefore, time consuming. One way of resolving these difficulties is by covalent modification of proteins with isotopically encoded reagents, which are usually a pair of reagents in light and heavy isotope forms. Through their characteristic isotope mass shifts, the peptides with covalent labels can be readily distinguished from other peptides in the mass spectra. Several groups have used deuterated and non-deuterated forms of N-alkylmaleimides, iodoacetanilide, and acrylamide to facilitate easy identification of labeled cysteine residues (Apuy et al., 2001a; Codina et al., 2001; Nelson et al., 2008). In a similar manner, deuterated and non-deuterated forms of acetic anhydride have been used to label lysine residues (Glocker et al., 1994; Hochleitner et al., 2000; Edavettal et al., 2004; Williams et al., 2006). Another approach is to utilize equal amounts of differentially mass-encoded reagents. Apuy et al. (2001b) employed N-alkylmaleimides that differ chemically by one methylene group to make easier the identification of labeled cysteine residues. Armengaud and co-workers used in parallel three lysine-specific reagents that differed in mass by 113 Da to easily locate modified fragments (Gabant, Augier, & Armengaud, 2008). Similar approaches with other amino acid-selective reagents could also be developed to facilitate modification site identifications.

Lastly, covalent labeling strategies are complementary to techniques such as H/D exchange because they provide information about amino acid side chains whereas H/D exchange only provides backbone information. Relatively few groups have combined such techniques (Turner & Maurer, 2002; Turner et al., 2004; Ohguro et al., 2004; Sperry et al., 2008), but there appears to be great potential in using side chain and backbone labels in concert. An intriguing prospect would be to combine these labeling approaches simultaneously instead of on separate samples.

ACKNOWLEDGMENTS

This work was supported by a grant from the National Institutes of Health (RO1 GM075092). V. M. also acknowledges support via a National Research Service Award T32 GM08515 from the National Institutes of Health.

REFERENCES


