

# Methylene as a possible universal footprinting reagent that will include hydrophobic surface areas: Overview and feasibility: Properties of diazirine as a precursor

FREDERIC M. RICHARDS, RAPHAEL LAMED,<sup>1</sup> RICHARD WYNN,<sup>2</sup> DARSHAN PATEL,<sup>3</sup>  
AND GERARD OLACK

Department of Molecular Biophysics and Biochemistry, Yale University, New Haven, Connecticut 06520

(RECEIVED April 25, 2000; FINAL REVISION August 16, 2000; ACCEPTED August 18, 2000)

## Abstract

Methylene is one of, if not the, most reactive organic chemical known. It has a very low specificity, which makes it essentially useless for synthesis, but suggests a possible role in protein footprinting with special importance in labeling solvent accessible nonpolar areas, identifying ligand binding sites, and outlining interaction areas on protomers that form homo or hetero oligomers in cellular assemblies. The singlet species is easily and conveniently formed by photolysis of diazirine. The reactions of interest are insertion into C-H bonds and addition to multiple bonds, both forming strong covalent bonds and stable compounds. Reaction with proteins and peptides is reported even in aqueous solutions where the vast majority of the reagent is used up in forming methanol. Species containing up to 5 to 10 extra :CH<sub>2</sub> groups are easily detected by electrospray mass spectroscopy. In a mixture of a 14 *K<sub>d</sub>* protein and a noninteracting 1.7 *K<sub>d</sub>* peptide, the distribution of mass peaks in the electrospray spectra was close to that expected from random modification of the estimated solvent accessible area for the two molecules. For analysis at the single residue level, quantitation at labeling levels of one <sup>13</sup>CH<sub>2</sub> group per 10 to 20 kDa of protein appears to be possible with isotope ratio mass spectroscopy. In the absence of reactive solvents, photolysis of diazirine produces oily polymeric species that contain one or two nitrogen atoms, but not more, and are water soluble.

**Keywords:** diazirine, methylene, nonpolar, photochemical labeling, protein footprinting, surface area

Determination of atomic level structures of macromolecules is dominated by X-ray crystallography, solution NMR, and, on the near horizon, solid state NMR as the principal methods employed. Crystallography requires crystals, and NMR, either solution or solid state, is limited by molecular size and the need for specially labeled samples to surmount these limits. Starting at a larger dimensional scale increasingly sophisticated cryoelectron micros-

copy is working toward higher and higher resolution for macromolecular shape data but suffers from the lack of general experimental methods for fitting the linear sequences into the observed shapes. The interface between the surface of a macromolecule and the solvent represents the most difficult part of the structure determination by any of these techniques. The interfaces are, of course, the regions where biology may be said to begin.

At the first level, the appearance of function is most often associated with the binding of two components: a receptor and a ligand to form a complex, or two protomers to form a dimer. In each case, there is a change in the solvent-accessible surface of both partners in the resulting complex. If the surface regions can be labeled and the products recognized at the single residue level, then the differential labeling of the free and bound partners will identify the binding site and thus the nature of the interacting surfaces. The detailed results can be expected to depend on the reagents used and the chemistry of the labeling process. There may be additional complexities such as major conformational changes in domain structure on complex formation, or dynamic flexibility from random conformations in certain segments of the polypeptide chain. Such effects have to be addressed in each case individually. There is no general solution.

Reprint requests to: Frederic M. Richards, Department of Molecular Biophysics and Biochemistry, Yale University, P.O. Box 208114, New Haven, Connecticut 06520-8114; e-mail: fmrswr@csb.yale.edu.

<sup>1</sup>Present address: Department of Molecular Biology and Biotechnology, Tel Aviv University, Ramat Aviv 69978, Tel Aviv, Israel.

<sup>2</sup>Present address: Department of Applied Biotechnology, The DuPont Pharmaceutical Co., Experimental Station, P.O. Box 80336, E336/241B, Wilmington, Delaware 19880-0336.

<sup>3</sup>Present address: 1025 Walnut St., Box 576, Philadelphia, Pennsylvania 19107-5001.

**Abbreviations:** DZN, diazirine; SBW, slit bandwidth; ESMS, electrospray mass spectrometry; IRMS, isotope ratio mass spectrometry; MALDI, matrix-assisted laser desorption ionization (mass spectrometry); PTH, phenylthiohydantoin; MW<sub>avg</sub>, molecular weight for composition with average natural abundance of isotopes; UV, ultraviolet; DTT, dithiothreitol; CD, circular dichroism; HPLC, high-performance liquid chromatography.

The most commonly used procedure of this sort is probably the one developed by Tullius et al. (1989) to identify the binding sites on nucleic acids for selected binding proteins. The procedure has two steps: (1) hydroxyl radicals are generated in situ and react immediately with accessible pentose residues causing chain cleavage (the sugars are unique components restricted to the main chain, and a cleaved bond is the "label" in this case); (2) analysis consists of a run on a sequencing gel. Cleavage samples from both the free nucleic acid and the protein-loaded one can be run side by side on the same gel. The binding site on the nucleic acid is thus visualized immediately. An essential component of the simplicity of this analysis is that the hydroxyl radical reagent show no intrinsic cleavage selectivity in the extended polymer for either the base identity or sequence. Any difference in pattern will then reflect the influence of the tertiary structure of the protein-nucleic acid complex. Recently, the Tullius' laboratory has shown that the position of cleavage of the ribose ring was correlated to the accessible area of the hydrogen atoms on the DNA backbone (Balasubramanian et al., 1998). This information is more detailed than the residue level and confirms the surface nature of the modification, but it is restricted to the ribose and deoxyribose moieties. The surface accessibility of the bases, for example, is not defined.

The measurement of changes in the relative reactivity of modifiable functional groups on ligand binding has a long history in protein structure studies starting in the middle of the last century. Both chemical and enzymatic reactions have been developed for modification of proteins, but these were usually aimed at high specificity to ease subsequent sequence analysis and interpretation of functional changes (Means & Feeney, 1971). Reactions at low temperature, neutral pH, and aqueous solvents were sought so that the native structure of the protein would be maintained. Long before the first success of X-ray crystallographic procedures applied to proteins many general aspects of protein structure were laid out by such chemical and enzymatic studies. The position of both charged and uncharged polar groups could be inferred as well as the changes in the reactivity of these groups induced by ligand binding, but only occasionally at the single residue level. By default, the nonpolar residues were assumed to be "inside" because of their aversion to water. There were no thermochemical reagents that could be used in water to react with hydrocarbon side chains under conditions that would not destroy a protein. While a great deal of useful information was obtained, the relatively small number of reactive functional groups did not permit a full listing of the residues contributing to a binding site. More recently, the introduction of nitrenes and carbenes has provided a series of reagents that can react with hydrocarbons, but these have almost always been developed and used in the form of photoaffinity labels designed to bind very specifically and thus to give information about interesting, but limited, portions of the proteins (Tometsko & Richards, 1980).

With the advent of the X-ray structures of proteins starting with myoglobin in 1958 (Kendrew et al., 1958; Kendrew, 1963), the general ideas about proteins were confirmed, but in individual cases there had been misinterpretations based on unrecognized and unexpected chemical behavior, usually for unknown reasons, as pointed out by Glazer shortly after the X-ray data began to appear (Glazer, 1976, pp. 10–14). However, it also became clear that, using accessible surface area as measured on an atomic model, about one-half of the surface of the water-soluble proteins, whose structures were then known, was nonpolar (Lee & Richards, 1971) and that changes in nonpolar area made major contributions to the

binding of peptides and nucleotides to the relevant proteins (Richards et al., 1971).

Ideally a general footprinting reagent would define the surface of the protein independent of the chemical nature of the surface. Binding sites would then be located by differential labeling as is already done with more specific reagents. (For recent examples, see Buechler et al., 1989 and Hanai and Wang, 1994, where only the amino groups of the involved lysine residues were monitored.) The analytical procedures used to detect and quantify the labeling would also be at least at the single residue level. To include surface hydrocarbon area, the reagent would have to be either a free radical, a nitrene, or a carbene.

To date, there is no sequencing procedure for proteins that can match the speed and simplicity of the DNA methods. Because of the nature of the sample whose sequence is sought (mixtures of peptides with only a low level of modification, as discussed below), the required information is not likely to be revealed by the elegant mass spectroscopic procedures developed by Biemann (1992; fragmentation analysis) and Chait et al. (1993; peptide ladders from Edman cleavage) that are so effective for pure peptide samples. The closest equivalent is the main-chain amide proton exchange process (Roder & Wüthrich, 1986). The identification of the individual exchanging residues requires a fully identified, high resolution NMR spectrum, and study of the exchange kinetics of each peak. This procedure is very informative but the data only refer directly to the main-chain amide hydrogen atoms. Structural interpretation of the kinetic changes on ligand binding is not always straightforward. The hydroxyl radical may also be used to cleave peptide chains, but it has not yet been developed as a general footprinting reagent as it has with DNA. Rather it has been used in a form where there is a designed specificity in the cleavage pattern (Platis et al., 1993; Ermacora et al., 1994).<sup>4</sup>

In examining possible alternative reagents, it seemed as though methylene (:CH<sub>2</sub>) might be a candidate. This electron deficient compound is easily synthesized in the singlet state. It is the most reactive and nonselective of all of the carbenes. Its major reactions are the insertion into, or addition to, single or double bonds (see the major review by Kirmse, 1971). Although it does not have the hydrogen bonding properties of water, :CH<sub>2</sub> is roughly the size and shape of a water molecule, and thus, in the geometrical sense, it might be expected to sample the same surface seen by water.

In hydrocarbons, the singlet species of methylene reacts in the picosecond time range and thus does not have the opportunity to change to the triplet form, the intersystem crossing rate being several orders of magnitude slower (Turro et al., 1987). Thus, the complexities of radical chemistry can be largely avoided. With this rapid rate, one would expect that the reaction of :CH<sub>2</sub> occurs with one of the cage of atoms surrounding it in the liquid phase at the instant following its appearance. The covalent bond that is formed is stable, and thus many forms of product analysis are available that do not have the potential kinetic problem of a back reaction in

<sup>4</sup>After preparing this paper, papers by Goshe and Anderson (1999) and Goshe et al. (2000) were brought to our attention. This group has defined the conditions including N<sub>2</sub>O, D<sub>2</sub>O, DTT, and radiation-induced OH• radicals that will produce efficient conversion of CH groups to CD groups. This work clearly shows the potential for accessible area measurements similar to the one proposed in this paper. If in the future we can arrange some measurements on the same system by the two procedures, it will be of great interest to see if the inferred areas are the same. The chemistry of the two processes is significantly different but they are aimed at the same goal.

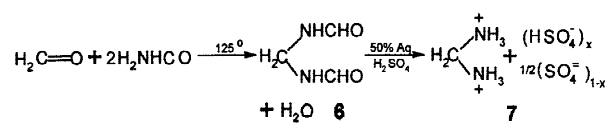
the sample that must be carefully avoided, or accounted for, in the proton exchange procedures. The reaction rates of  $\text{:CH}_2$  with primary, secondary, and tertiary carbons are nearly the same (Doering et al., 1956), an indication of its possible very low specificity for all insertion reactions. This lack of specificity has caused methylene to be described by Doering et al. (1956) as “the most indiscriminate reagent known in organic chemistry” and by March as “. . . virtually useless for synthetic purposes. . .” (March, 1977, p. 183). For synthesis this may indeed be true, but it is a characteristic which might make it ideal for a footprinting reagent.

Because of its high reactivity, methylene must be generated *in situ* in the sample. While there are many sources of methylene, photolysis of diazirine (DZN) appears to be the method of choice. DZN is the three member ring isomer of diazomethane and is a gas at room temperature. There are, however, a number of potential difficulties with this proposal:

1. It is anticipated that any system under study will require water to be the principal solvent where the water concentration will be about 55M or 110M in O-H bonds. The concentration of the solutes are likely to be millimolar at best. In aqueous solutions, the principal product will always be methanol. Will it be possible to get significant modification of the solute in the presence of this overwhelming solvent competitor?
2. The partition coefficient of diazirine between the gas and liquid phases differs with different solvents. The position of the DZN molecules at the protein/solvent interface just before photolysis may not mimic the distribution of water. Thus, the extent of labeling may not be simply related to the residue accessible area as normally defined.
3. There are many bonds other than C-H or C=C with which  $\text{:CH}_2$  might react. Not all of the relative rates, measured in water, are in the literature. Although  $\text{:CH}_2$  seems to be almost omnivorous, the reaction rates with various bonds are not identical, and labeling efficiency in every sample may need to be measured as a function of the extent of total reaction, thus unacceptably increasing the time and complexity of the analysis.
4. Each successful  $\text{:CH}_2$  addition/insertion will produce a new covalently linked methyl or methylene group that instantly becomes another potential target for further reaction. Thus, the extent of labeling must be kept low to avoid multiple hits on the same residue. This requirement is in conflict with the necessity to get enough label attached to provide sufficient signal for detection and assay.

This paper addresses some of these problems and provides some new data on the properties of diazirine whose distribution in a solution will have a large effect on the labeling pattern. A short meeting abstract on this work has appeared.<sup>5</sup> We particularly wish

<sup>5</sup>Abstracts of poster sessions on DZN labeling from the Delfino laboratory were distributed in the programs to attendees at two international meetings, and from the Richards laboratory at one meeting: Craig et al. (1998a, 1998b) and Olack et al. (1999). These meeting programs are not readily available to the general public. Anyone wishing copies of these abstracts should write to: Dr. Jose Delfino, Dept. de Quimica biologica e Instituto de quimica y Fisisoquimica Biologicas (IQUIFIB-CONICET), Facultad de Farmacia y Bioquimica, Universidad de Buenos Aires, Junin 956, 1113 Buenos Aires, Argentina for Craig et al. (1998a, 1998b); and to F.M. Richards, Dept. of Molecular Biophysics and Biochemistry, Yale University, P.O. Box 208114, New Haven, Connecticut 06520-8114 for Olack et al. (1999).



Scheme 1

to acknowledge receipt from Drs. Jose Delfino and Patricio Craig of an early draft of an unpublished manuscript on this same general topic. The work of their laboratory has been presented in poster sessions at two international meetings.<sup>5</sup>

## Results

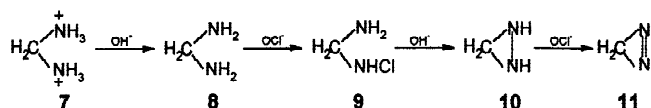
### Synthesis of diazirine

There are a number of synthetic schemes for making diazirine, but the procedure of Ohme and Schmitz (1964) is one of the simplest.

A 2:1 mixture of formamide and paraformaldehyde without solvent is heated to 120–130 °C for 12–18 h and then allowed to stand at room temperature for several days giving crystals of methylene bisformamide (**6**, in Scheme 1). For small scale runs with readily available, labeled paraformaldehyde, all the various combinations of isotopic products involving the methylene function with  $^{12}\text{C}$ ,  $^{13}\text{C}$ ,  $^{14}\text{C}$ ,  $^1\text{H}$ ,  $^2\text{H}$ ,  $^3\text{H}$  can be obtained with this same protocol. The  $^{13}\text{C}$  compound has been particularly useful in our study while P.O. Craig and J.M. Delfino (unpubl. data) have worked largely with the  $^{14}\text{C}$  form. In the usual labeling procedure, all of the nitrogen is given off as  $\text{N}_2$  and does not appear in the products. However, in the studies of photolysis in anhydrous conditions (see below), DZN prepared with  $^{15}\text{N}$  formamide, and thus containing  $^{15}\text{N}_2$ , has been useful.

When (**6**) is dissolved in 50% aqueous sulfuric acid at 4 °C and left for several days at that temperature, crystals of methylenediamine sulfate (**7**, in Scheme 1) appear in high yield. This salt may be kept anhydrous at room temperature for many months.

DZN (**11** in Scheme 2), a gas at room temperature and 1 atmosphere pressure, is formed by reaction of the salt, **7**, with aqueous sodium hypochlorite/NaOH in a single step. The presumed intermediates are **8**, **9**, and **10**, in Scheme 2. This is a two-phase system with at least two competing reactions: (1) direct reaction with water in the basic solution destroying the methylenediamine (**8**) and (2) the multistep reaction of (**8**) with hypochlorite to form DZN (**11**). The detailed steps in the synthesis, the collection and treatment of the DZN gas, and the photolysis of DZN, with a Hg arc source or a HeCd laser, to yield methylene for the labeling reaction are provided in Supplementary material in the Electronic Appendix (ESM). (Please note that we have had some flash burns while generating DZN. This has been reported before in the literature (Graham, 1962). See the ESM for more details and take appropriate precautions during the synthetic steps.)



Scheme 2

*Some properties of DZN*

The measurement methods and data for this section are given in the ESM. Some brief comments and/or the derived numbers are listed here.

*Near UV absorption of DZN in the gas phase*

Near UV absorbance in the gas phase is restricted to the region between 280 and 320 nm (as shown in ESM Fig. E1), which is similar to that given by Graham (1962). The wavelength region between 310 and 320 nm is a window in the UV absorption of most proteins and thus useful both for photolysis and for quantitative measurements of DZN. For analysis in the gas phase, we have used absorption measurements on the peak at 314.6 nm and a slit bandwidth of 0.5 or 1.0 nm.

*Quantum yield for the loss of UV absorption*

Although sharp by solution standards, the peaks in the absorption spectrum are not as sharp as one might expect for a gas phase sample, thus suggesting a high quantum yield for the photolysis induced by the absorbed beam. A direct measurement of the quantum yield for the disappearance of DZN, described in the ESM, was  $1.0 \pm 0.1$  (Murov, 1973).

*Estimation of the molar extinction coefficient in the gas phase*

The extinction coefficients ( $\epsilon$ ) in the gas and liquid phases are related by the equilibrium absorbance ratios in the two phases and the partition coefficients between them. When absolute numbers are not required, all liquid extinction coefficients can be expressed as ratios to the gas value without actually knowing what that value is. In ratios of the values of  $\epsilon$  for two different solvents, the value of  $\epsilon_{gas}$  will cancel out. However, to get actual concentrations, this latter number must be known. Our best estimate for  $\epsilon_{gas}$  at this time at  $l = 314.6$  nm is 340 L/mole/cm with SBW = 0.5 nm or 240 L/mole/cm with SBW = 1.0 nm. The estimated error in these numbers is  $\pm 10\%$ . The details of the measurements of this quantity are given in the ESM (see Williams, 1984 for analysis of the principal photolysis product methanol).

*Partitioning between the gas phase and some pure liquids*

With the absorbance high enough for easy measurement, it is convenient to use the gas phase over a solution to measure DZN concentration. The gas phase has been used as the reference for all other measurements. The output of the synthesis appears to contain no gases that absorb in the same region as DZN, thus no absorption corrections are required. Given the gas phase extinction coefficient, one only needs the partition coefficients to determine the actual concentrations in liquid phases. The partition coefficients for the various solvents are listed in Table 1. These can be measured without knowledge of the extinction coefficients. The spectra of DZN dissolved in various solvents (ESM, Fig. E2), as well as the equations showing the calculation of the coefficients from the measured data are shown in the ESM.

*Self-polymerization on photolysis of anhydrous DZN*

The extreme reactivity of methylene ensures that in most situations it will react with the components of the solution in which it is dissolved. However, it does not insert into C-F bonds, and thus does not react with any of the saturated perfluoro hydrocarbons, nor does it react with the noble gases. Thus in gas mixtures with argon or in liquid solution in perfluorohexane, for example, diaz-

**Table 1.** Some properties of diazirine dissolved in various solvents<sup>a</sup>

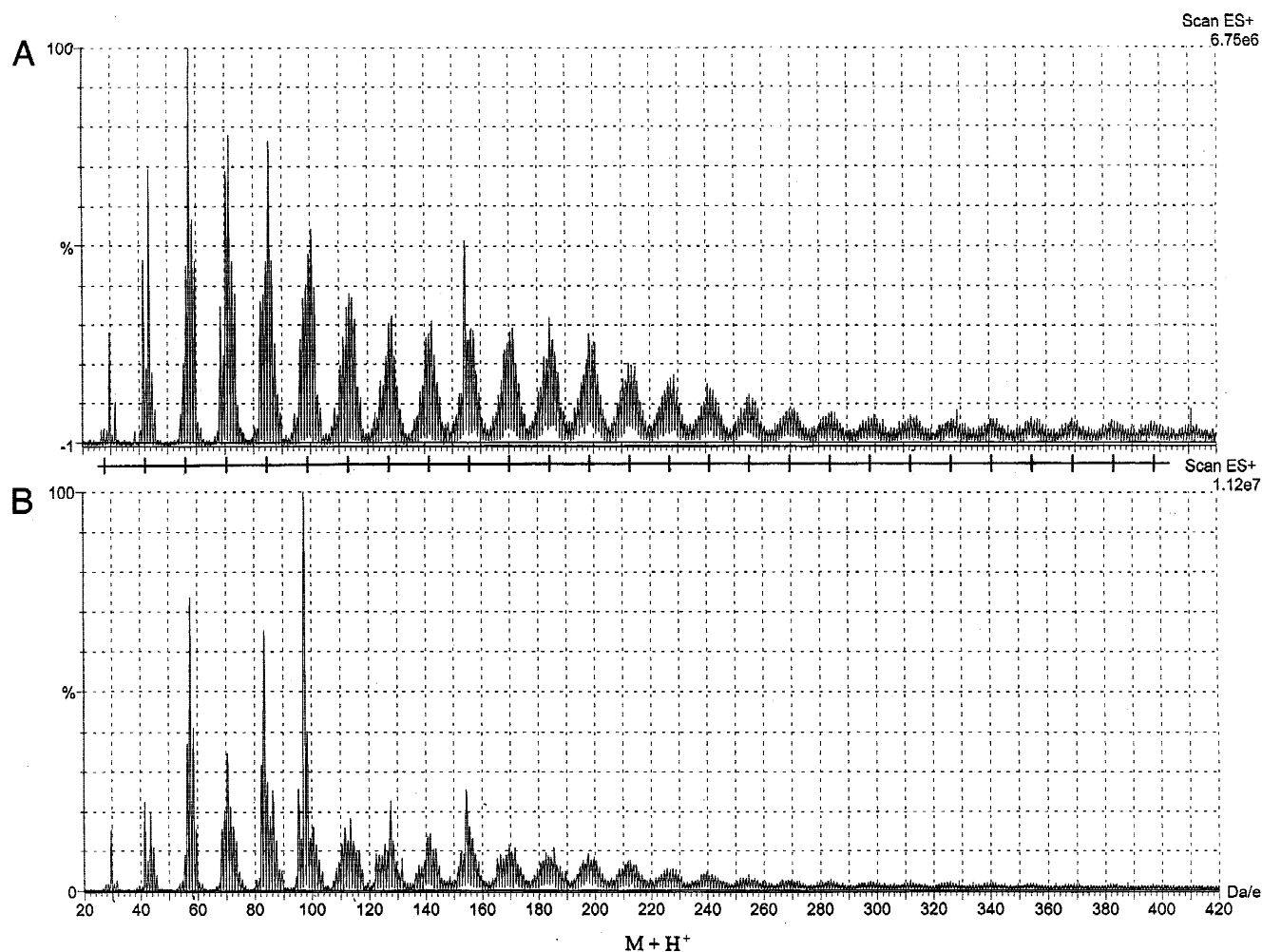
Solvent	Partition $l^{max}$ in			
	Coefficient $K_i$	Solvent	$A_{li}^{l^{max}}/A_g^{314.5}$	$\epsilon_{li}^{l^{max}}/\epsilon_g^{314.5}$
Water	3.7	320	1.46	0.39
	7.7 (5°)		2.60	0.34
NaCl 2M	2.8	319.7	0.9	0.52
SDS 15% aq	3.2	318.0	1.3	0.41
Methanol	18.1	320.2	12.4	0.68
Ethanol	12.3	317.8	9.3	0.76
Butanol-1	9.9	318.4	5.6	0.57
Pentanol-1	9.3	318.2	7.5	0.81
Octanol-1	7.0	318.2	4.9	0.70
Perfluorohexane	4.6	315.9	4.2	0.92
Hexane	10.9	317.4	6.5	0.60
Heptane	10.0	317.7	8.5	0.85
Cyclohexane	10.4	317.5	6.0	0.58
Benzene	24.5	320.0	19.2	0.78

<sup>a</sup>The temperature of the experiments was  $23 \pm 1^\circ\text{C}$  unless otherwise indicated. The reproducibility of the partition coefficients is on the order of 10% of the listed value.

irine will yield  $:\text{CH}_2$  as usual, but there are only other  $:\text{CH}_2$ ,  $\text{N}_2$  or unphotolyzed DZN molecules with which the methylene can initially react. When such photolyses are carried out, oily material rapidly appears as a cloudy suspension in perfluorohexane or a cloudy covering on the inside of the gas-containing cuvette as noted by Kirkbride and Norrish (1933). The bulk of the material is assumed to be polymethylene either linear or branched. Surprisingly, however, all of the material is readily soluble in water implying the presence of some nitrogen in every product molecule. The ESMS of the two water solutions are curious and are shown in Fig. 1.

Beyond masses of 100 or 120, there is a peak at every mass number out to at least 400 to 500 Da. The peaks appear in groups that have a similar bell shape beyond mass 120 and a repetition interval of  $\sim 14$  mass units. Below 100 Da, the peaks are still grouped, but are incomplete in that not all the mass positions are represented. (Below mass 90, there are certain mass values that cannot be provided by any combination of C, H, and N atoms.) The higher mass groups would seem to be formed by the addition of  $:\text{CH}_2$  groups to each member of the lower mass group and without any rate discrimination since the shape of the group envelope is not altered. This is easily imagined if the subsequent reactions beyond mass 100 are all C-H insertions into methyl or methylene groups formed in the earlier compounds in the series.

To investigate this situation a little further, diazirine was prepared from formamide containing 98+%  $^{15}\text{N}$ . The  $^{15}\text{DZN}$  then has a mass of 44 rather than 42 although the methylene produced is still mass 14. The gas phase form of the reagent was produced and photolyzed as usual. The water-soluble oil was obtained, and an enlarged section of its ES spectrum along with the same mass range from the normal  $^{14}\text{DZN}$  reagent is shown in Fig. 2. The most obvious changes are that every other peak in the  $^{14}\text{N}$  spectrum is very low or missing in the  $^{15}\text{N}$  spectrum and that the centroid of the envelope of each group of peaks has shifted about 2 units to higher mass. The repetition interval between peak groups remains



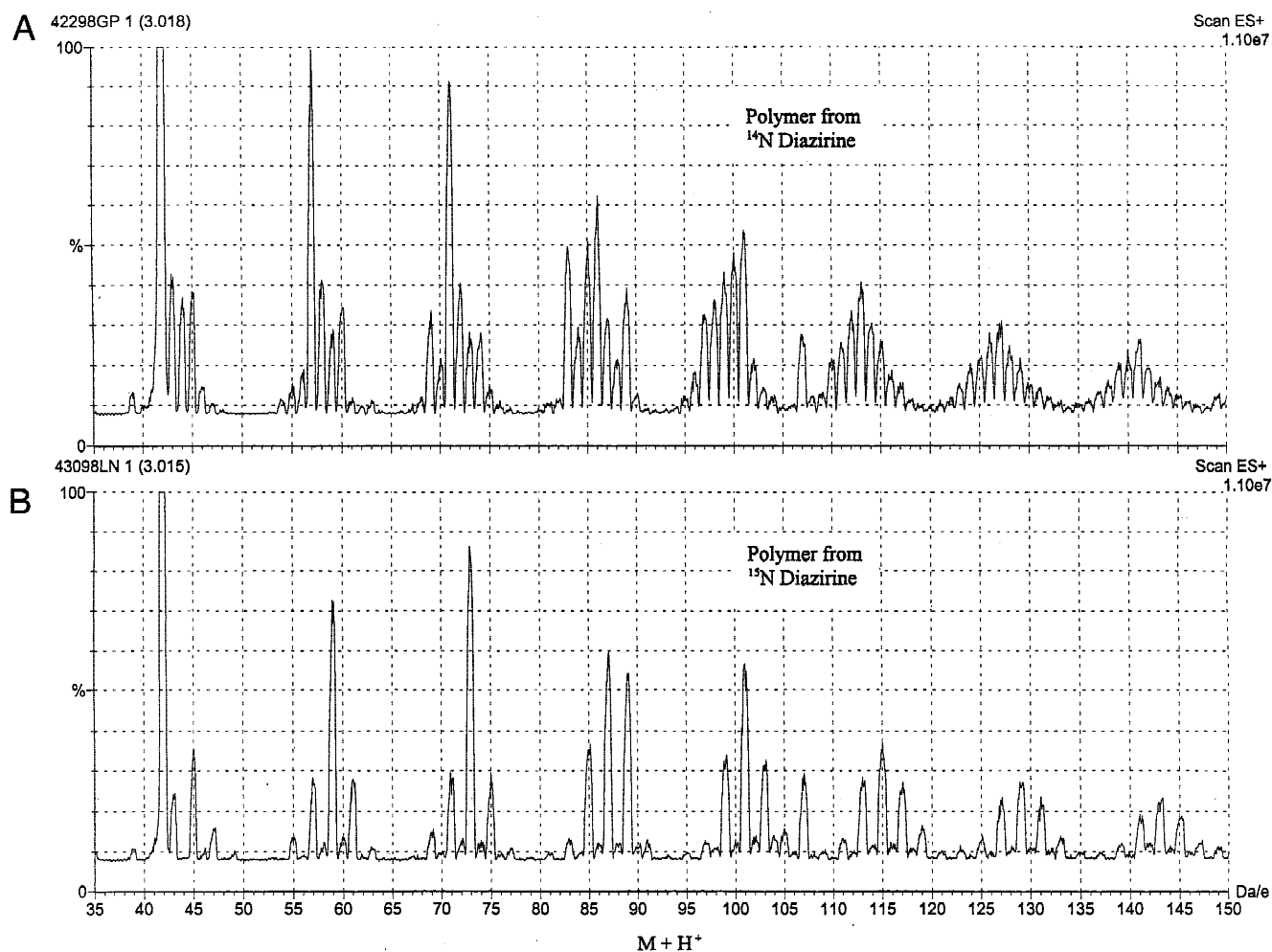
**Fig. 1.** Electro spray mass spectra of the aqueous solutions of the products of photolysis of diazine in the absence of other reactive species. **A:** Diazine dissolved in neat perfluorohexane, previously purged with argon, and photolyzed as a solution. The products of photolysis, soluble and insoluble, were extracted from the fluorocarbon with water. All oily material dissolved. The aqueous solution was adjusted to 20% acetonitrile: 0.1% trifluoroacetic acid and injected for mass analysis without any further treatment. The thick line as second abscissa has tick marks at intervals of 14 au adjusted to start at mass 28. Only molecules with a +1 charge were observed in the relatively low mass range of the components in these samples. **B:** DZN photolyzed in the gas phase with no solvent present. Again the oily material dissolved in water, but any product existing in the gas phase was probably lost.

at 14 au. The only explanation that we have come up with is that all species produced by the photolysis contain either 1 or 2 nitrogen atoms and that they appear alternately in the mass spectrum. Any compound with 2 nitrogen atoms will increase 2 units in mass, with 1 nitrogen atom the mass increase will be 1 unit. If the  $^{14}\text{N}$  compounds are arranged  $\text{N}_2\text{-N}_1$  in order of increasing mass, on conversion to the  $^{15}\text{N}$  forms the  $\text{N}_2$  species will move 2 units higher in mass, the  $\text{N}_1$  will move 1 unit higher and both  $^{15}\text{N}$  containing peaks will overlap. The original  $^{14}\text{N}_1$  position will be empty thus generating the spectrum in Fig. 2B.

We have no definitive information on what the nitrogen containing section of these molecules actually is, but some possibilities for neutral masses 41 through 46 would be:  $\text{C}_2\text{H}_3\text{N}$  (41) acetonitrile;  $\text{CH}_2\text{N}_2$  (42) diazine or diazomethane;  $\text{C}_2\text{H}_5\text{N}$  (43) aziridine;  $\text{CH}_4\text{N}_2$  (44) diaziridine;  $\text{C}_2\text{H}_7\text{N}$  (45) dimethyl amine;  $\text{CH}_6\text{N}_2$  (46) methyl hydrazine. Larger units would start to include pyrrole (67), pyrazole (68), pyrrolidine (71), and other ring compounds as suggested much earlier (Kirkbride & Norrish, 1933).

Each peak may represent multiple compounds with the same elementary composition but different structures. Once each  $\text{N}_1$  or  $\text{N}_2$  containing compound appears, it continues to add  $\text{CH}_2$  groups and thus fairly soon all unit mass positions are filled and each peak may contain a collection of positional isomers with the same mass. At the higher mass levels, the accurately repeating pattern of 14 au also suggests that the CH insertion process is the only one occurring under these conditions.

Attempts in this study to examine the methylene labeling of dry or only slightly moist samples of proteins or peptides to reduce the competition from water has resulted in the appearance of some of this polymeric material. We have not found any evidence in the mass spectra of cross-linking of the larger polymeric units. Thus all the photochemically activated species appear truly monofunctional as expected and are most likely to be methylene itself. The lack of cross-linking means the polymer will not be covalently bonded to proteins in the mixture. Its water solubility would suggest that, if necessary, it could be removed by procedures used for



**Fig. 2.** Spectra of polymers obtained from photolysis of diazirine in the gas phase. **A:** A portion of a spectrum similar to that shown in Figure 1B, but with an expanded mass scale. The diazirine was from a standard preparation with the nitrogen present as the  $^{14}\text{N}$  isotope. For the spectrum in **B**, the diazirine was prepared with the  $^{15}\text{N}$  isotope present in both N atoms at 98+ % enrichment. The samples were prepared, photolyzed, and analyzed as described for Figure 3. There were small peaks in the spectrum out to at least mass 400, but not shown in this figure.

other detergents. The polymeric material is unlikely to be found in dilute solutions in water or hydrocarbon solvents.

#### The methylene labeling reaction in water

##### Products from 1-butanol

With the exception of the perfluorocarbons, methylene will react with any solvent molecules in addition to solutes of principal interest. In general the volume concentration of the solute will be small compared to that of the solvent. As a result, most of the methylene will be "lost" in reactions of no immediate interest. By itself this might just cause a waste of reagent, but there is the potential for the nature of the solvation shell around the solute to alter the distribution of the diazirine prior to photolysis and thus affect the methylene labeling quantitatively, and perhaps qualitatively, for the same solute in different solvents. Some preliminary experiments have been carried out to see how serious this might be.

Butanol-1 was chosen as an example. There are only five expected insertion products: **12**) pentanol-1, **13**) pentanol-2, **14**) 2-methyl-butanol-1, **15**) 3-methyl-butanol-1, and **16**) methyl butyl ether. Butanol has 10 hydrogen atoms. The above list assumes that X-H insertion is the only type of reaction. If the insertion rate into all of the bonds is the same, then the ratio of the amounts of the above products expected would be: **12:13:14:15:16** = 3:2:2:2:1.

Three reactions were carried out: the first in neat butanol-1 served as both solute and solvent; the second in butanol-1 dissolved in  $\text{H}_2\text{O}$  at a concentration of 6 vol%; the third in butanol dissolved in  $\text{D}_2\text{O}$  also at 6%. The latter was to check whether or not the expected isotope effect of slower insertion into a D-O bond was sufficiently large to be useful in reducing the water:methylene reaction rate. It was not.

Standard GCMS (gas chromatograph/mass spectrometer) calibration runs were made over a concentration range from 0.5 to 5.0 vol% for each of the products with butanol-1 as the solvent. For analysis of the samples reacted with DZN, following photolysis, hexanol-1 was added as an internal standard for the GC runs. The

hexanol peak in the chromatogram came just after butanol and all of the product peaks. The neat butanol samples were diluted 1:100 with methanol. Aliquots of 1 mL were used for injection.

For photolysis, a DZN/argon mixture was pumped through the sample-containing cuvette. The differential UV absorption of the input and output gas streams was used to estimate the amount of DZN converted by photolysis. For the neat butanol samples,  $72 \pm 11\%$  of the photolyzed DZN could be accounted for with the sum of the products identified in the GCMS peaks (products **12,13,14+15,16**).

For the 6 vol% samples in H<sub>2</sub>O and D<sub>2</sub>O, the products of photolysis were extracted with diethyl ether, dried with anhydrous Na<sub>2</sub>SO<sub>4</sub>, and used directly for injection. To generate enough of the products for accurate quantitation, a large volume of the argon/DZN mixture had to be pumped through the system. Unfortunately, that resulted in the loss of the butyl methyl ether product (**16**) that is not water soluble and is also volatile. Blank runs on a stock solution of (**16**) showed that bubbling argon through the solution as in a regular experiment was removing all of the ether.

We were able to get an estimate of the ether production by carrying out the reaction in a sealed cuvette where the volatile products could not escape, in place of the flow system usually used. With the setup available, we could only photolyze a much smaller amount of DZN with a corresponding diminution of the amounts of the products. The background and peaks in the GC runs were much noisier, but they were good enough to provide reasonable, but less accurate, integrated areas. The ether was indeed present and in a larger amount than originally predicted, as we had seen in the neat butanol run.

Products **14** and **15** were not adequately separated for individual quantization on the column used. The summed area of these peaks was recorded. All products had the correct mass to be pentanol derivatives as expected. The total yield of products in the aqueous samples was only about 1% of the DZN converted by photolysis. The results are summarized in Table 2.

In the neat sample, the amount of the methylbutylether(**16**) was substantially larger than expected indicating that the O-H bond reacted faster than any of the C-H bonds. For the water samples and the flow system where the ether was lost and recorded as a "0" yield, the areas of peaks **12, 13, 14+15** were normalized to sum to 9.0. The H<sub>2</sub>O and D<sub>2</sub>O areas agree quite well with each other and with the expected yield ratios for equivalent reaction rates. There is more variation in the neat sample, but still it is qualitatively similar. Thus the reaction behavior of the C-H and O-H bonds does not seem to be affected in a major way by the shift from a hydrocarbon to an aqueous solvent.

#### Products from a mixture of lysozyme and peptide S-15

A mixture of hen egg white lysozyme, a very stable protein of MW<sub>avg</sub> 14,314 with a completely native structure, was made with a 15 residue peptide, MW<sub>avg</sub> 1,751 representing the first 15 residues from ribonuclease-S peptide. By CD the latter has no demonstrable secondary structure in solution at room temperature. This mixture was run through the DZN photolysis procedure. The enzyme and the peptide do not interact, but are simply two independent solutes being simultaneously modified under identical conditions. The entire mixture was then analyzed by mass spectrometry. The octanol used to prevent foaming was extracted with

**Table 2.** Product ratios in methylene labeling of butanol-1 relative GCMS peak areas

Product	Expected if reaction rates equal	Found in butanol-1						
		Neat		6% in H <sub>2</sub> O	6% in D <sub>2</sub> O	6% in H <sub>2</sub> O		
A	B <sup>a</sup>	C <sup>b</sup>	D <sup>c</sup>	E <sup>d</sup>	F <sup>e</sup>	G <sup>f</sup>	H <sup>g</sup>	I <sup>h</sup>
<b>12</b> ) Pentanol-1	3	2.2	3.1	3.2	3.2	2.1	2.5	2.5
<b>13</b> ) Pentanol-2	2	0.8	1.1	1.8	1.7	1.3	1.6	1.6
<b>14</b> ) 2-Methyl-butanol-1	2	3.4	4.8	4.0	4.0	4.0	4.8	4.8
<b>15</b> ) 3-Methyl-butanol-1	2							
<b>16</b> ) Methylbutyl ether	1	3.6	(5.1)	nd <sup>i</sup>	nd	2.6	(0)	(3.2)
Totals	10	10.0	14.1	9.0	9.0	10.0	9.0	12.2
Sum of identified products as % of total DZN photolyzed		72 ± 10		1 ± 0.1	1 ± 0.1			

<sup>a</sup>If the 9 C-H + 1 O-H = 10 bonds in butanol are assigned the same probability of reaction then the expected yields, as estimated by the GCMS product areas, would be given by the numbers in column B normalized to a sum of 10. The peaks for products **3** and **4** were not separated sufficiently to measure separately so the areas were summed before normalization to the indicated totals.

<sup>b</sup>Areas normalized to a sum of 10.

<sup>c</sup>Area sum for **12 + 13 + 14 + 15** was normalized to 9. The area figure for the ether (**16**) was adjusted to what it would be if the sum of the first 4 was correct.

<sup>d,e</sup>No ether (**16**) was found in these two runs as explained in the text. The areas were normalized to a sum of 9 and should be directly compared to the comparable values in columns B, C, and D.

<sup>f,g,h</sup>The experimental numbers in column G were obtained from the run in a sealed cuvette for the DZN reaction. The ether was found in this case where it was lost in runs E and F. The areas in G were summed to 10. In H, the ether was omitted and the other 4 normalized to 9 for direct comparison with E and F. The numbers in I were calculated as in D and should be compared with that column.

<sup>i</sup>nd, not determined.

ethyl acetate, the solution dried, redissolved in water, and diluted in a methanol-formic acid solution for analysis. The molecular weights are sufficiently different that there is no overlap between the two electrospray envelopes. The transformed data are shown in Figures 3 and 4.

In a simulation of what might be expected under ideal conditions, the following assumptions were made: all reaction rates of methylene with accessible surface areas are the same; there is no interference between the two solutes, which are thus labeled independently; the total accessible surface area of lysozyme in the native state is  $\sim 6,700 \text{ \AA}^2$  (Lee & Richards, 1971), and is divided into 100 equal patches of  $67 \text{ \AA}^2$  each, a not unreasonable value for an average surface residue; the unstructured peptide is 15 residues long and is assumed to have 15 patches of the same size; the probability  $p$  of a successful reaction is related to time in an actual experiment; since the assumed extent of reaction is very low, no account was taken of possible multiple hits on a given patch. The binomial theorem was then applied to the two samples, one with  $n = 15$  and one with  $n = 100$ . The compositions of the solutions are expressed as the fraction of each sample with a given number of methylene groups added. The results are shown in Figure 5.

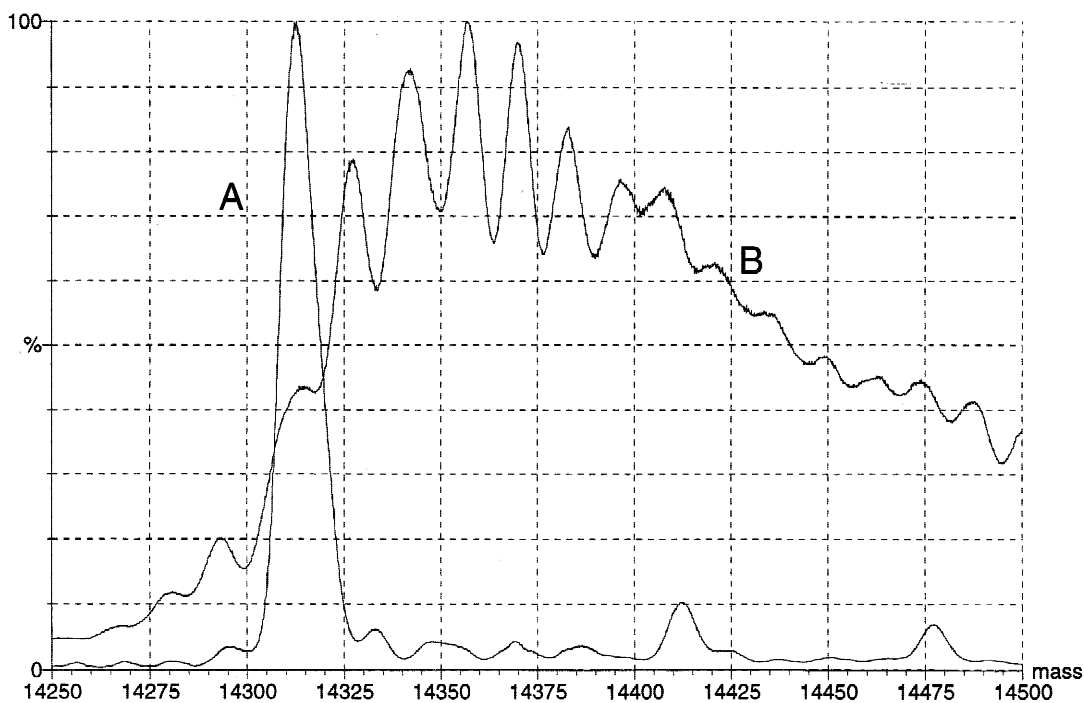
### Discussion

The purpose of this study is to investigate the feasibility of developing a general labeling procedure that would provide a measure of accessible area including the nonpolar regions. Such a method would be useful for differential protein footprinting in defining

ligand binding areas and protomer association regions in oligomeric complexes. (NOTE: There is some ambiguity about what the probe is in this system. DZN is the molecule that wanders around and tests the surfaces; however, it is not the reactant. The photochemically produced methylene is the reactant but it has a chance to survey only the immediately available atoms making up its cage and does not "search" as a solvent water molecule does. Until shown otherwise, the chemical behavior of both DZN and  $:\text{CH}_2$  must be considered in interpreting labeling results.) The principal focus of this particular paper is on the four questions listed at the end of the introduction.

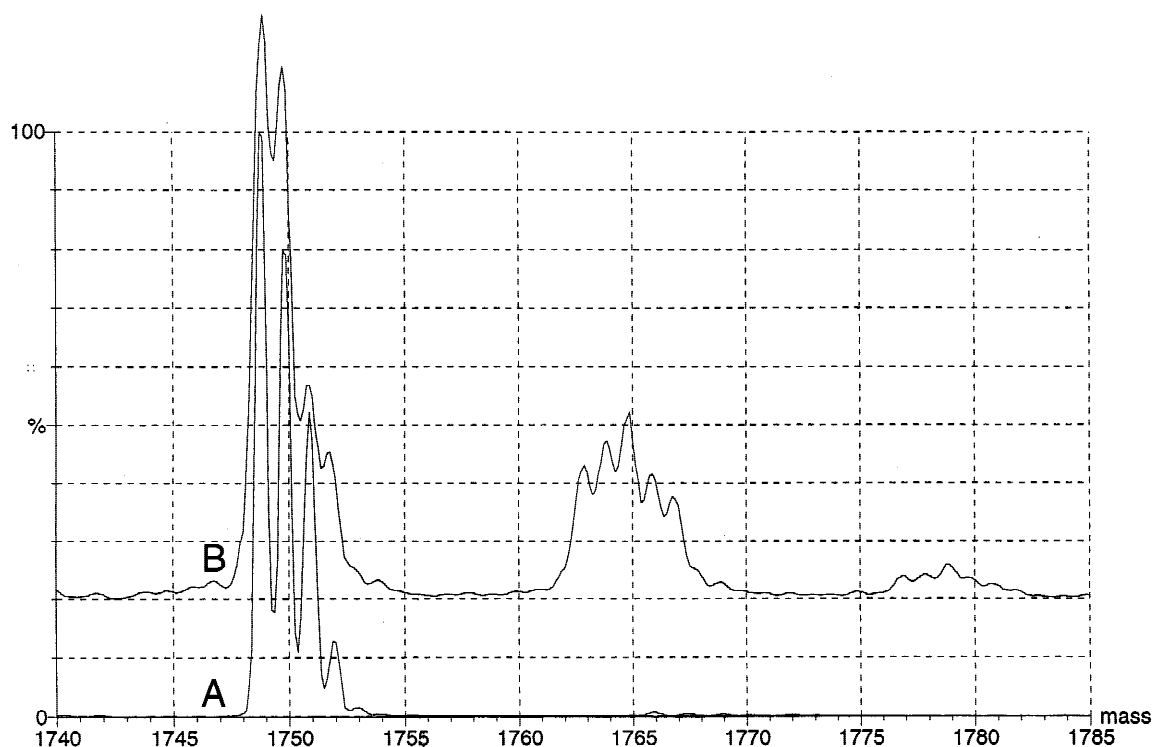
*First question:* Can one actually detect the reaction of  $:\text{CH}_2$  with a target solute in the presence of enormous concentrations of highly reactive solvents? This is especially a problem if the O-H reaction is as fast or faster than that of the C-H bond as appears to be the case. The data shown for butanol and for the lysozyme/S-15 mixture indicate that the answer is *yes* albeit the yield based on DZN is very low.

*Second question:* Is the solute-solvent interface sufficiently variable that differences in DZN partitioning will seriously affect the labeling pattern of the various types of solvent interface area? There is insufficient data at this moment to answer this question in general. It has been shown here that the product ratios for the labeling of the different carbon atoms and the hydroxyl group in 1-butanol are very similar when comparing butanol and water as solvents. The overall yields, of course, are very different. This result is perhaps not surprising. Highly polar substances could be very different. Much further testing of the relation of labeling to



**Fig. 3.** The transformed ES mass spectra of only the hen egg white lysozyme component of the mixture of this enzyme and the peptide S-15. The methylene labeling was carried out on a mixture of the peptide and protein in a single experiment. The large single peak at the left of the diagram marked **A** is the lysozyme component ( $MW_{\text{avg}} 14,315$ ) of this mixture before modification. The multippeak trace marked **B** shows the protein products of reaction with methylene. For the protein portion of each spectrum, the largest peak was adjusted to 100. In absolute terms, the highest peak in **B** is about one-quarter that of the highest peak in **A**. The integrated peak areas are thus not equivalent. The peaks in the actual data used in the transformation had charges +8 to +12 for peak **A** and +6 to +12 for **B**.





**Fig. 4.** The peptide portion of the spectra of the samples described in Figure 3. The charges on the peaks before transformation were +3 to +5. Again **A** is the peptide from the unmodified sample and the peaks labeled **B** from the same modified mixture as Figure 3B. In this case, the natural abundance  $^{13}\text{C}$  peaks are clearly visible. The **B** spectrum has been displaced up 20% to avoid overlap. As in Figure 3, the adjustment of the maximum peak heights interferes with a comparison of integrated areas.

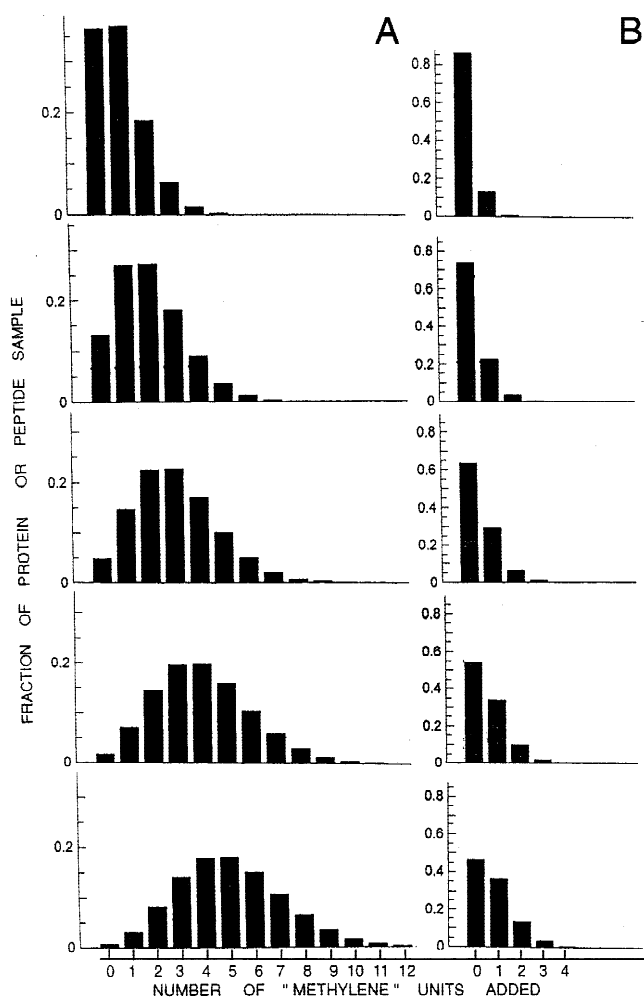
geometrical surface area will be needed on a variety of both small molecules and proteins.

*Third question:* Are the possible reactions of  $:\text{CH}_2$  other than insertion important? For example, an extensively reported reaction is the addition of  $:\text{CH}_2$  to  $\text{C}=\text{C}$  alkyl double bonds. This results in cyclopropane derivatives. For the labeling purposes described in this paper, this would be a perfectly useful reaction. The product is stable and does not tend to undergo further reactions under the conditions used in protein experiments. The standard 20 protein amino acids have no such bonds, but they do occur in lipids where this probe might also be useful. The amino acids do have aromatic rings, and the partial double bond character of the ring bonds provides a known target. In this case, the initial product is a fused ring derivative that usually will convert by ring expansion and produce a cycloheptatriene, again an acceptable reaction. Some preliminary experiments have been started on 2-phenylethanol as a test small molecule. All the possible insertions and additions will produce a series of compounds that will all have the same mass and the same number of added carbon atoms. This might still allow a measure of accessible area without the necessity of separately measuring all of the individual compounds.

There is the possibility of adding to  $\text{C}=\text{O}$  and  $\text{C}=\text{N}$  bonds as well. These would form epoxides and cyclic imines that are chemically reactive and have a rich chemistry of their own. In this instance, depending on what is in their immediate environment, the final stable species produced by further reaction may, or may not, interfere with the analysis and its subsequent interpretation. If these thermochemical reactions are indeed playing a role, the pattern should be simplified by using scavengers or by working at

much lower temperatures. Due to the very high reactivity of the photochemically produced  $:\text{CH}_2$ , the insertion or addition reactions in the very first step of any modification are unlikely to be affected. Methylene reactions in solid matrices are observed at liquid helium temperatures.

*Fourth question:* Concerns the conflict in the desired level of modification: to get enough sample to satisfy the needed signal to noise ratio for the analyses while keeping the level low enough to avoid multiple modifications on the same residue. The promising part of the lysozyme/S-15 experiments is the comparison of Figures 3 and 4 with Figure 5. The latter is a calculation of the idealized reaction and what might be expected in the mass spectra of the products. The two noninteracting materials of very different size were present in the same solution and subject to precisely the same reaction conditions during the period of labeling. There is a surprising degree of agreement between third panel pair in Figure 5 and the experimental spectra shown in Figures 3B and 4B. Note that the calculations for Figure 5 do not take account of the isotope distribution of the naturally occurring  $^{13}\text{C}$ . These are not resolved in the protein spectra, but are in those of the peptide. The peak clusters for the latter should be summed into single peaks to make the comparison with Figure 5. Normalizing such sums from the peaks shown in Figure 4B to 1.00 gives the values 0.65/0.31/0.04 to be compared in Figure 5 showing the comparable values 0.62/0.29/0.07 in the  $p = 0.03$  panel. Because of the distortions in the higher mass part of Figure 3, and uncertainty about the proper base line, the best comparisons for the protein are in the first few peaks, especially the comparison of the remaining unmodified material with the first and second modified peaks in both Fig-



**Fig. 5.** Models of expected mass spectra for random labeling of noninteracting, identical, sites with a "methylene" type reagent. Column A is for a "protein" with 100 sites. Column B for a "peptide" with 15 such sites. The assumed uniform probability of reaction of any one site is listed on each panel. This probability is related to the overall extent of reaction. The standard binomial distribution was assumed. The diagrams should be compared to Figures 3 and 4. The best match for both **A** and **B** occurs in the vicinity of  $p = 0.03$ . The individual peaks showing the mass distribution due to  $^{13}\text{C}$  natural abundance is not shown explicitly in this figure. The  $^{13}\text{C}$  peaks may be thought of as lumped together within each bar.

ures 3B and 5 (panel pair at  $p = 0.3$ ). The assumptions made in constructing Figure 5 may not be wildly out of line, i.e., random unrelated reactions with comparable rates for the protein and peptide molecules.

The extent of modification could be increased by further treatment of such samples. Some earlier labeling attempts with the 20 residue ribonuclease S-peptide provided samples that had no observable unmodified material in the HPLC trace. However, since the covalent structure of the protein or peptide is being altered, one would really like the number of methylenes added to be as small as possible not larger.

An unrelated question is raised in Figure 4B by the two peak clusters of mass 1,763–1,767 and 1,777–1,781. These clusters do not have the intensity distributions expected for the expected natural abundance  $^{13}\text{C}$  series in a peptide of this size. With the major

peaks at 1,765 and 1,779, it would seem to indicate that there had been a partial reaction in the first modification step that resulted in the inclusion of an atom of oxygen (+16 au). There is a methionine residue in the S-15 peptide that could readily form a sulfoxide. We have seen these +16 peaks before in the modified peak clusters, but irregularly. We do not yet know for sure the origin of the oxygen atom (free oxygen gas, water, or the peptide itself are the only candidates) or the chemistry involved in its activation. The unmodified starting peptide appears to be reasonably clean since the principal peak has the correct mass and only traces of an M+16 component are occasionally seen. Even if this does occur, it should not interfere with the isotope ratio analysis discussed below a procedure that will report solely on the carbon insertion/addition reactions that have occurred. We hope to find the source of this nuisance and possible ways to prevent it.

#### *The next major problem*

The structural payoff in this approach will be to measure the extent of modification at the single residue level. This will require accurate measurements of labeling in each residue at very low levels of overall labeling. The problem arises from the central assumption of this whole proposal, *random modification based solely on surface accessibility*. We may actually be approaching such a protein preparation in the work described above.

For discussion, let us assume that we have a sample where each protein molecule has just a single  $^{13}\text{CH}_2$  group added. An ESMS measurement will easily show that the molecular weight is increased by 15 au. If this represented a good affinity label, we would simply proceed, using ESMS or MALDI procedures with the sequence study knowing that we were going to find all of the label in one or two peptides and eventually identify the altered amino acid(s). However, the sample of present interest is a large ensemble, not at all a pure compound. There will be many different residues modified in the collection of protein molecules. Further, the residues of a given type may yield a variety of products depending on the extent and position of the area accessible to the methylene. With only a single  $^{13}\text{CH}_2$  group added, all of these protein molecules will have the same mass, M+15. Each accessible residue will be modified to a variable but small extent. The chances of any two protein molecules in the sample having the same covalent structure is vanishingly small. To simplify the further discussion, we have assumed that each residue has the same probability of being modified.

If the protein were pancreatic ribonuclease in a true random coil form with 124 residues and a molecular weight of 13,670, the individual residue would be labeled to the extent of 1/124 or 0.008 atoms of  $^{13}\text{C}$  in the average residue of weight 110. The analytical sequencing procedure should be designed so that this level of labeling in a given position in the chain could be measured ideally with 1% accuracy or  $\pm 0.008\%$  of the total signal present in each protein molecule in the ensemble. ESMS is not well suited to make accurate quantitative measurements, and certainly not on peaks that differ in height by a factor of 100. It may be hard to even identify the small peak reliably above the background. An alternate approach is provided by isotope ratio mass spectrometry (IRMS). (See the recent review by Criss, 1999.) In this type of analysis, an organic sample is destroyed completely by combustion to  $\text{CO}_2$ . An IRMS spectrometer then measures with high accuracy and sensitivity the  $^{13}\text{C}/^{12}\text{C}$  ratio in the  $\text{CO}_2$ .

As an example, bovine pancreatic ribonuclease has 575 carbon atoms per molecule. With an average natural abundance of about 1.11%  $^{13}\text{C}$ , there will be 6.4 atoms of  $^{13}\text{C}$  and 568.6 atoms of  $^{12}\text{C}$ . To this molecule, we add 1 atom of  $^{13}\text{C}$  in the methylene label. Since we will only be interested in differences, we will use the unmodified material as the "standard." Let  $R$  be the mole ratio of  $^{13}\text{C}/^{12}\text{C}$ , then  $R_{\text{std}} = 6.4/568.6$ ;  $R_{\text{sample}} = (6.4 + 1.0)/568.6$ . The difference of these ratios as a fraction of  $R_{\text{std}} = 0.1562$ . This number is usually expressed on a per mil basis and called the  $d$  function. In this example,  $\delta = 156.20$ ;  $\delta$  can usually be measured with a precision of 0.02–0.05. Assuming uniform labeling, this ratio is independent of the size of the sample used and thus would apply to a single residue as well as the whole protein. If lowered accessibility reduced the labeling of a particular residue to 1% of the mean value, the measured value of  $\delta$  for that residue would be 1.56. For this latter value, a precision of 3–5% should be attainable.

We do not know of a better procedure than the standard Edman method for sequential production of single residue samples of isotope analysis. A portion of the completed cleavage reaction would be checked by HPLC: (1) to get the PTH derivatives for all amino acids, original and modified, at that sequence position; (2) to determine the fraction of material being carried forward from previous Edman cycles; and, equally important for this analysis, (3) the amount of carbon-containing reagent products such as diphenylurea that will dilute the  $^{13}\text{C}$  in the PTH amino acids and thus require correction of the IRMS analysis. An effort needs to be made to lower the concentration of these diluting substances before the IRMS measurements as far as possible without causing isotope fractionation of the various PTH molecules in the sample.

In an experiment just completed, a  $^{13}\text{C}$  modified sample of ribonuclease-A intended to provide an average composition of at least one added carbon atom per protein molecule was run through the IRMS analysis. The value of  $\delta$  obtained was  $\sim 420$ , a value outside the upper end of the current calibration range of the instrument. Thus the protein was modified more heavily (2.5–3.0 atoms of  $^{13}\text{C}$  per protein molecule) than was targeted. With this result, we feel that the measurement of accessible area at the single residue level is a viable goal and that a suitable protocol can be developed.

### Supplementary material in the Electronic Appendix

Details are provided for all the experimental methods, materials, and instruments and for data on diazirine, the  $\text{CH}_2$  precursor: quantum yield for UV absorption loss; UV absorption spectra, extinction coefficients, and partition coefficients for the gas phase and solutions in various solvents.

### Acknowledgments

This project has proceeded by fits and starts over a number of years. We have had a great deal of help and advice during the entire period. Our special thanks go to M. Ermacora for introduction to HPLC; David Schweitzguth, Joel Tolman, Melanie Cocco, and James Prestegard for help with NMR; The W.M. Keck Foundation Biotechnology Resource Laboratory, Kathy Stone, Jim Elliott, Walt McMurray; Myron Crawford, Introduction to mass spectroscopy and peptide synthesis; Chemical Instrumentation Center, Department of Chemistry, Susan Degala for GCMS; University of Illinois, Mass Spectroscopy Center, high resolution mass spectra; Peter Moore, Donald Crothers, sympathy when required; P.H. Vaccaro, Kurt Zilm, J.A. Berson, K.A. Wiberg, F.W. McLafferty, N. Turro, M.S. Platz, N.C. Craig, general advice on carbenes, precursors, and spectroscopy, mass and optical; Danny Rye and John Kingston, introduction to

and help with the IRMS procedures. We wish also to thank one of the reviewers for the comments on probable oxidation and the interpretation of Figure 4.

This work was supported by the National Institute of General Medical Sciences Grant #5P01 GM22778.

### References

- Balasubramanian B, Pogozelski WK, Tullius TD. 1998. DNA strand breaking by the hydroxyl radical is governed by the accessible surface area of the hydrogen atoms of the DNA backbone. *Proc Nat Acad Sci USA* 95:9738–9743.
- Biemann K. 1992. Mass spectrometry of peptides and proteins. *Annu Rev Biochem* 61:977–1010.
- Buechler JA, Vedvick TA, Taylor SS. 1989. Differential labeling of the catalytic subunit of cAMP-dependent protein kinase with acetic anhydride: substrate-induced conformational changes. *Biochemistry* 28:3018–3024.
- Chait BT, Wang R, Beavis RC, Kent SBH. 1993. Protein ladder sequencing. *Science* 262:89–92.
- Craig PO, Kleinman CL, Delfino JM. 1998a. Probing the conformation of proteins with diazirine, a minimal photochemical reagent suitable for labeling cavities. *On protein condensation: In honor of Gregorio Weber. Biophys J* 75:418.
- Craig PO, Kleinman CL, Delfino J. 1998b. Conformation dependent labeling of proteins with diazirine, a generally reactive photoprobe. *Protein Sci* 7 (Suppl 1):104.
- Criss RE. 1999. *Principles of stable isotope distribution*. New York, Oxford: Oxford University Press.
- Doering WE, Buttery RG, Laughlin RG, Chaudhuri N. 1956. Indiscriminate reaction of methylene with the carbon-hydrogen bond. *J Am Chem Soc* 78:3224.
- Ermacora MR, Ledman DW, Hellinga HW, Hsu GW, Fox RO. 1994. Mapping staphylococcal nuclease conformation using an EDTA-Fe derivative attached to genetically engineered cysteine residues. *Biochemistry* 33:13625–13641.
- Glazer AN. 1976. The chemical modification of proteins by group-specific and site-specific reagents. In: Neurath H, Hill RL, eds. *The proteins*. New York: Academic Press. pp 1–103.
- Goshe MB, Anderson VE. 1999. Hydroxyl radical-induced hydrogen/deuterium exchange in amino acid carbon-hydrogen bonds. *Radiat Res* 151:50–58.
- Goshe MB, Chen YH, Anderson VE. 2000. Identification of the sites of hydroxyl radical reaction with peptides by hydrogen/deuterium exchange: Prevalence of reactions with the side chains. *Biochemistry* 39:1761–1770.
- Graham WH. 1962. A direct method of preparation of diazirine. *J Am Chem Soc* 84:1063–1064.
- Hanai R, Wang JC. 1994. Protein footprinting by the combined use of reversible and irreversible lysine modification. *Proc Nat Acad Sci USA* 91:11904–11908.
- Kendrew JC. 1963. Myoglobin and the structure of proteins. *Science* 139:1259–1266.
- Kendrew JC, Bodo G, Dintzis HM, Parrish RG, Wyckoff H, Phillips DC. 1958. A three-dimensional model of the myoglobin molecule obtained by X-ray analysis. *Nature (London)* 181:662–666.
- Kirkbride FW, Norrish RGW. 1933. Primary photochemical processes. Part II. The absorption spectrum and photochemical decomposition of diazomethane. *J Chem Soc* 136:119–126.
- Kirmse W. 1971. *Carbene chemistry*. New York: Academic Press.
- Lee B, Richards FM. 1971. The interpretation of protein structures: Estimation of static accessibility. *J Mol Biol* 55:379–400.
- March J. 1977. *Advanced organic chemistry*. New York: McGraw-Hill Book Company.
- Means GE, Feeney RE. 1971. *Chemical modification of proteins*. San Francisco: Holden-Day.
- Murov SL. 1973. *Handbook of photochemistry*. New York: Marcel Dekker.
- Ohme R, Schmitz E. 1964. Notiz über eine einfache Synthese des Cyclo-diazomethans. *Chem Ber* 97:297–298.
- Olack G, Patel D, Wynn R, Lamed R, Richards FM. 1999. Seeing the surface: Analysis of covalent modification of proteins using diazirine. *Protein Sci* 8 (Suppl 1):163.
- Platis IE, Ermacora MR, Fox RO. 1993. Oxidative polypeptide cleavage mediated by EDTA-Fe covalently linked to cysteine residues. *Biochemistry* 32:12761–12767.
- Richards FM, Wyckoff HW, Carlson WD, Allewell NM, Lee B, Mitsui Y. 1971. Protein structure, ribonuclease-S and nucleotide interactions. *Structure and function of proteins at the three-dimensional level*. Cold Spring Harbor, New York: Cold Spring Harbor Laboratory. pp 35–43.

- Roder H, Wüthrich K. 1986. Protein folding kinetics by combined use of rapid mixing techniques and NMR observation of individual amide protons. *Proteins Struct Funct Genet* 1:34–42.
- Tometsko AM, Richards FM. 1980. *Applications of photochemistry in probing biological targets*. New York: New York Academy of Sciences.
- Tullius TD, Dombroski BA, Churchill MEA, Kam L. 1989. Hydroxyl radical footprinting: A high-resolution method for mapping protein-DNA contacts. In: Wu R, Grossman L, Moldave K, eds. *Recombinant DNA methodology*. San Diego, California: Academic Press.
- Turro NJ, Cha Y, Gould IR. 1987. Reactivity and intersystem crossing of singlet methylene in solution. *J Am Chem Soc* 109:2101–2107.
- Williams S, ed. 1984. Analysis of methanol. *Official methods of analysis*. Arlington, Virginia: Association of Official Analytical Chemists, Inc. p 186.