VOLUME 121, NUMBER 20
MAY 26, 1999

JOURNAL OF THE AMERICAN CHEMICAL SOCIETY

Trifluoromethyldiazirinylphenyldiazenes: New Hemoprotein Active-Site Probes

Richard A. Tschirret-Guth, Katalin F. Medzihradszky, and Paul R. Ortiz de Montellano*

Contribution from the Department of Pharmaceutical Chemistry, University of California, San Francisco, California 94143-0446

Received February 4, 1999

Abstract: The photoactivatable trifluoromethyldiazirinylphenyldiazene probes **1a** and **2a** have been synthesized, and their utility in the mapping of hemoprotein active sites has been validated with myoglobin (Mb). Reaction of the probes with Mb yields Fe-aryl adducts. Photolysis of these adducts unmasks a carbene that cross-links to active-site protein residues. Migration of the aryl group from the iron to a porphyrin nitrogen then attaches the porphyrin chromophore to the labeled amino acid residue. Tryptic digestion of the labeled proteins followed by mass spectrometric analysis of the peptides has identified Leu-29, His-64, Ile-107, and Val-68 as active-site residues. Previous studies with an arylnitrene probe, which appears to only react with nucleophilic groups, identified His-64 as an active-site residue [Tschirret-Guth, R. A.; Medzihradszky, K. F.; Ortiz de Montellano, P. R. *J. Am. Chem. Soc.* **1998**, *120*, 7404−7410]. These studies have identified all but one of the active-site amino acids in contact with the probe. The present strategy not only labels active-site amino acids but also, because the probe is rigidly held in the active site, provides approximate locations for the labeled residues with respect to the heme iron atom. Validation of the strategy with myoglobin opens the way to use of the approach with hemoproteins of unknown active-site structure.

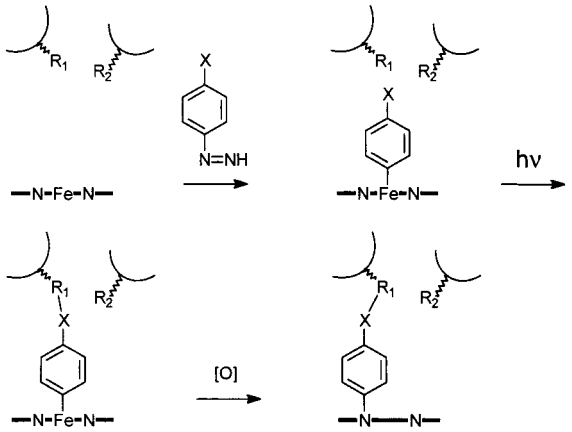
Cytochromes P450 are important enzymes involved in the biosynthetic and metabolic pathways of endo- and xenobiotics.[1,2] To date, more than 500 P450 isoforms have been identified, and many have been characterized with respect to their substrate specificity or function. Despite this overwhelming number of proteins, crystal structure coordinates are only available for five water-soluble bacterial P450 enzymes,[3] and information on the active-site structures of the mammalian membrane-bound isoforms has only been obtained by indirect methods such as homology modeling,[4,5] mechanism-based inactivation,[6−9] and photoaffinity labeling.[6,10−13] One technique developed by this

* To whom correspondence should be addressed at the School of Pharmacy. Phone: (415) 476-2903. Fax: (415) 502-4728. E-mail ortiz@cgl.ucsf.edu.

(1) Ortiz de Montellano, P. R. In *Cytochrome P450: Structure, Mechanism and Biochemistry*, 2nd ed.; Ortiz de Montellano, P. R., Ed.; Plenum Press: New York, 1995; pp 473−574.
(2) Guengerich, F. P. *Mammalian Cytochromes P450*, Vol. I and II; CRC Press: Boca Raton, FL, 1987.
(3) Poulos, T. L.; Cupp-Vickery, J. R.; Li, H. In *Cytochrome P450: Structure, Mechanism and Biochemistry*, 2nd ed.; Ortiz de Montellano, P. R., Ed.; Plenum Press: New York; pp 125−150.
(4) Poulos, T. L. *Methods Enzymol.* **1991**, *206*, 11−30.
(5) Modi, S.; Paine, M. J.; Sutcliffe, M. J.; Lian, L.-Y.; Primrose, W. U.; Wolf, C. R.; Roberts, G. C. K. *Biochemistry* **1996**, *35*, 4540−4550.
(6) Yun, C.-H.; Hammons, G. J.; Jones, G.; Martin, M. V.; Hopkins, N. E.; Alworth, W. L.; Guengerich, F. P. *Biochemistry* **1992**, *31*, 10556−10563.
(7) Roberts, E. S.; Hopkins, N. E.; Alworth, W. L.; Hollenberg, P. F. *Chem. Res. Toxicol.* **1993**, *6*, 470−479.
(8) Roberts, E. S.; Hopkins, N. E.; Zaluzec, E. J.; Gage, D. A.; Alworth, W. L.; Hollenberg, P. F. *Arch. Biochem. Biophys.* **1995**, *323*, 295−302.
(9) Koenigs, L. L.; Trager, W. F. *Biochemistry* **1998**, *37*, 13184−13193.
(10) Obach, R. S.; Spink, D. C.; Chen, N.; Kaminsky, L. S. *Arch. Biochem. Biophys.* **1992**, *294*, 215−222.
(11) Ohnishi, T.; Miura, S.; Ichikawa, Y. *Biochim. Biophys. Acta* **1993**, *1161*, 257−264.
(12) Miller, J. P.; White, R. E. *Biochemistry* **1994**, *33*, 807−817.

10.1021/ja990351h CCC: $18.00
Published on Web 05/06/1999

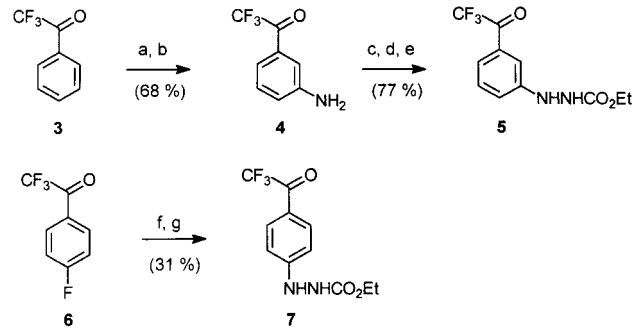
Scheme 1. Outline of the Strategy for Labeling of Active-Site Amino Acid Residues[a]



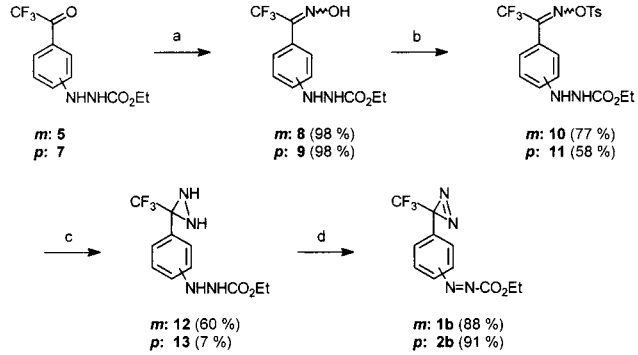
The heme is represented by N−Fe−N, and the side chains of active-site residues by R_1 and R_2. X is the photoactivatable group on the probe.

laboratory relies on the reaction of arylhydrazines and aryldiazenes with heme centers to give an iron-aryl (Fe-aryl) complex, followed by oxidation of the Fe-aryl complex to a mixture of the four possible N-arylprotoporphyrin-IX regioisomers.[14,15] The ratio of the four regioisomers thus obtained can be used to construct a low-resolution topological model of the active site. However, this technique does not provide information on the nature of the amino acids that form the active-site pocket. To overcome this shortcoming, we have recently introduced the use of azidoaryldiazenes as photoaffinity probes for the active sites of heme-containing proteins.[16,17] One advantage of this technique is that initial formation of the iron-aryl complex locks the probe in a defined geometry and ensures that subsequent labeling by the photoactivated azide group occurs in a defined region within the active site of the protein (Scheme 1). Using sperm whale myoglobin (Mb) as a test protein, both *m*- and *p*-azidophenyldiazene were found to cross-link to His64, which suggests that these probes are photoactivated to species that preferentially or exclusively react with nucleophilic active-site residues. To make this approach suitable for the study of hemoprotein active sites without strongly nucleophilic residues and to facilitate the labeling of multiple residues within a single active site, we have replaced the azido group by a function that is photoactivated to a more reactive species. The trifluoromethyldiazirine group has been chosen because it is relatively small and displays both remarkable photochemical properties and chemical stability.[18−19] It is activated by irradiation at 350 nm, a wavelength that is removed from both the protein (280 nm) and heme Soret (430−480 nm) absorbance maxima. In addition, the photogenerated carbene is long-lived and readily inserts into unactivated C−H bonds. We report here synthesis of the photoactivatable aryldiazene probes **1a** and **2a** and validation of their use to map the active sites of heme containing proteins by analysis of their reaction with sperm whale myoglobin.

Scheme 2. Conditions: (a) KNO_3, Concd H_2SO_4, 0 °C; (b) HCO_2H/Et_3N/(Pd/C), Acetonitrile, Reflux; (c) $NaNO_2$, 6 N HCl, 0 °C; (d) $SnCl_2$, 6 N HCl, 0 °C; (e) $ClCO_2Et$, Pyridine, Acetonitrile; (f) NH_2NH_2, DMSO, 90-100 °C, 30 min; (g) $ClCO_2Et$, Pyridine, Acetonitrile.

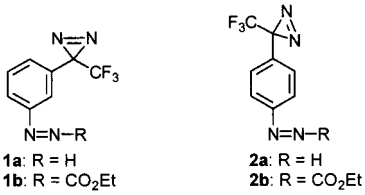


Scheme 3. Conditions: (a) $NH_2OH.HCl$, EtOH/Pyridine, Reflux; (b) Tosyl Chloride, NEt_3, Acetone; (c) NH_3(liq), Ether; (d) $Pb(OAc)_4$, $CHCl_3$.



Results

Synthesis. Since aryldiazenes are only moderately unstable,[20] **1a** and **2a** are best generated immediately before use by base



hydrolysis of the corresponding aryldiazenecarboxylate esters **1b** and **2b**. The strategies used to synthesize **1b** and **2b** require that introduction of the acylhydrazine function precede assembly of the trifluoromethyldiaziridinyl moiety (Schemes 2 and 3). Probe **1b** was prepared from 3-amino-α,α,α-trifluoroacetophenone (**4**), which in turn was obtained by nitration of trifluoroacetophenone (**3**) and reduction of the nitro group with formic acid/triethylamine/Pd/C.[21] Amine **4** was converted to a hydrazine by successive treatment with sodium nitrite and $SnCl_2$ in concentrated HCl, and the product was immediately derivatized to the acylhydrazine **5**. Sequential treatment with hydroxylamine, tosyl chloride, and ammonia gave the diaziridinyl acylhydrazine **9**. Standard treatment of oxime **8** with tosyl chloride in pyridine at reflux for 1−2 h failed to yield the desired product. However, tosylation in acetone/triethylamine at room temperature for 10

(13) Cvrk, T.; Hodek, P.; Strobel, H. W. *Arch. Biochem. Biophys.* **1996**, *330*, 142−152.
(14) Ortiz de Montellano, P. R. *Biochimie* **1995**, *77*, 581−593.
(15) Swanson, B. A.; Ortiz de Montellano, P. R. *J. Am. Chem. Soc.* **1991**, *113*, 8146−8153.
(16) Tschirret-Guth, R. A.; Medzihradszky, K. F.; Ortiz de Montellano, P. R. *J. Am. Chem. Soc.* **1998**, *120*, 7404−7410.
(17) Tschirret-Guth, R. A.; Ortiz de Montellano, P. R. *J. Org. Chem.* **1998**, *63*, 9711−9715.
(18) Brunner, J.; Senn, H.; Richards, F. M. *J. Biol. Chem.* **1980**, *255*, 3313−3318.
(19) Frey, A. B.; Kreibich, G.; Wadhera, A.; Clarke, L.; Waxman, D. J. *Biochemistry* **1986**, *25*, 4797−4803.
(20) Huang, P. C.; Kosower, E. M. *J. Am Chem. Soc.* **1968**, *90*, 2354−2362.
(21) Terpko, M. O.; Heck, R. F. *J. Org. Chem.* **1980**, *45*, 4992−4993.

min gave the tosyloxime **10** in 77% yield. Acylhydrazine **12** was oxidized to the final product **1b** using lead tetraacetate in $CHCl_3$ (or $CDCl_3$ when monitored by NMR). The reaction, which could be conveniently monitored by ^1H NMR (Figure 1), yielded a single product. The reaction is rapid and provides an excellent alternative to other diaziridine oxidation procedures. Probe **2b** was prepared from 4-fluoro-α,α,α-trifluoroacetophenone (**6**) and hydrazine in dimethyl sulfoxide. The hydrazine was immediately derivatized with ethylchloroformate to avoid further intermolecular reaction and hydrazone formation. Acyl hydrazine **7** was then converted to the final product (**2b**) as described above (Scheme 3). However, the yields were somewhat lower than those observed with **1b**. The final products **1b** and **2b** had UV absorption bands at 350 and 351 nm, respectively, positions typical for the trifluoromethyldiazyrinyl group.[18]

Labeling of Sperm Whale Myoglobin with 1a. Probe **1a** was added to a sample of sperm whale Mb and the course of the reaction followed by UV spectroscopy (Figure 2A). The initial Soret maximum of ferric Mb at 408 nm was replaced by a much less intense band at 430 nm characteristic of histidine-ligated Fe-aryl complexes.[14] Complex formation occurred more slowly than previously observed with phenyldiazene or *m*-azidophenyldiazene, and a large excess of **1a** was required to take the reaction to completion. The reaction of Mb with **2a** was even slower, and only partial formation of the Fe-aryl complex was achieved (data not shown). To avoid nonspecific labeling by excess **1a**, the protein complex was purified by passage through a size exclusion PD-10 column before photolysis. The UV spectrum of the protein after elution from the column indicated that no loss of the Fe-aryl complex had occurred (Figure 2B). Irradiation of the complex with an unfiltered Hg pen lamp for 30 min occasionally caused a red-shift (with a concomitant narrowing) of the Fe-aryl absorbance from 430 to 422 nm (data not shown), suggesting that structural modification of the protein active site was caused by wavelengths below 280 nm. However, irradiation at 360 nm for 30 min with a Rayonnet reactor fitted with an additional Pyrex (<300 nm) filter resulted in only minor loss of the Fe-aryl absorbance (Figure 2B). After aerobic exposure to acid to promote iron-to-nitrogen shift of the aryl group, the protein was subjected to HPLC analysis with the eluent monitored at 214 and 416 nm (Figure 3). Panel A shows the elution profile of intact Mb. Under HPLC conditions, the heme dissociates from the protein core and elutes 2 min earlier than the protein. After complex formation with **1b** and oxidative decomposition in the absence of photolysis, the peak corresponding to free heme disappeared and was replaced by products that absorb at 416 nm and elute at 19−21 min. These products are attributed to the free N-trifluoromethyldiazirinylphenylprotoporphyrin IX isomers (Figure 3B). However, when the sample was photolyzed *prior* to denaturation, the protein was covalently attached to a chromophore that absorbs at 416 nm (panel C). A nonprotein-bound peak eluting at 16 min is attributed to free N-arylprotoporphyrin-IX isomers derived from photolyzed probe that did not cross-link to Mb. Residual 416 nm-absorbing peaks at ∼20 min suggest that photolysis was not complete.

Mass Spectrometric Analysis of Modified Tryptic Peptides of Labeled Mb. The labeled Mb was purified by HPLC and digested with trypsin. The resulting peptides were separated by HPLC with the detector set to monitor the effluent at 214 and 416 nm (Figure 4). The peaks absorbing at 416 nm were collected and analyzed by MALDI-TOF and electrospray mass spectrometry. Labeled species were identified by the 718 Da

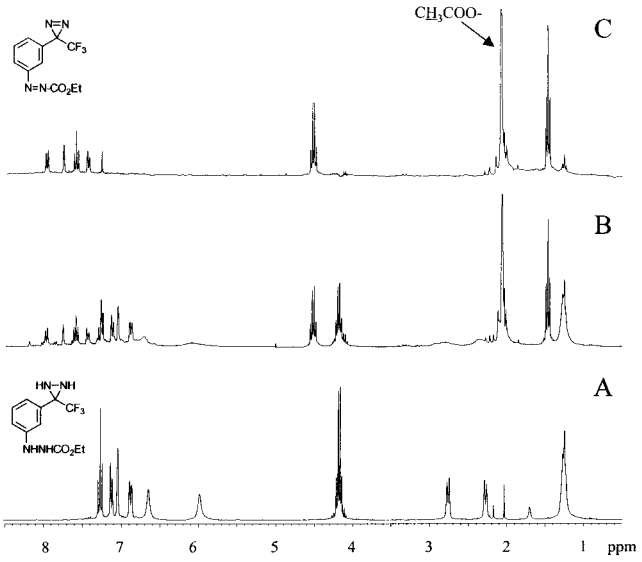


Figure 1. ^1H NMR of the oxidation of **9** to **1b** by lead tetraacetate; (A) no $Pb(OAc)_4$; (B) ∼1 equiv $Pb(OAc)_4$; (C) ∼2 equiv $Pb(OAc)_4$.

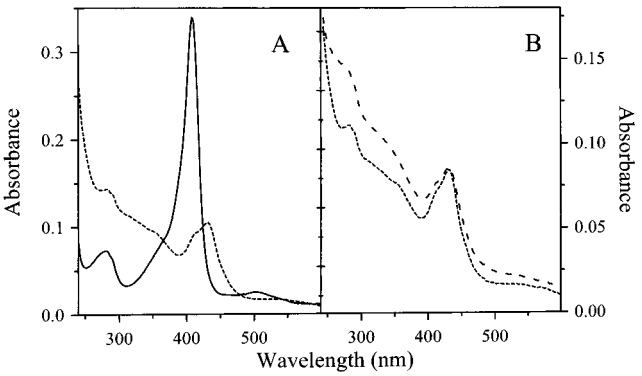


Figure 2. (A) (—) native Mb, (- - -) Fe-aryl complex; (B) (- - -) Fe-aryl complex after PD-10, (— — —), Fe-aryl complex after photolysis.

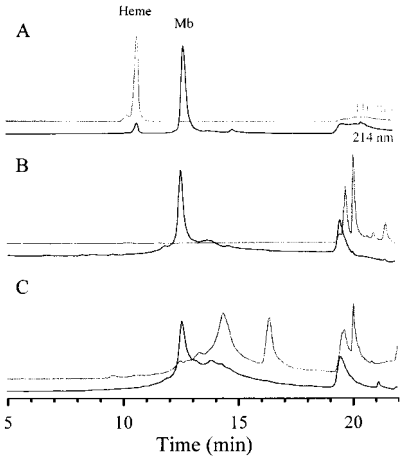


Figure 3. HPLC profile of (A) Mb, (B) Mb after complex formation with **1a** and oxidative shift, but without photolysis, (C) Mb after complex formation with **1a**, photolysis and oxidative shift.

mass increment (probe + heme) with respect to the mass predicted for the native tryptic peptides (Table 1).

The modified peptides eluting at 48−52 min were identified as derivatives of the tryptic peptide Val^{17}-Arg^{31}. The PSD spectrum of this peptide, $MH^+_{average}$ = 2312.5 (Figure 5A) clearly shows the presence of the heme (*m*/*z* 563), the heme-aromatic linker adduct (*m*/*z* 720), and an immonium ion at *m*/*z*

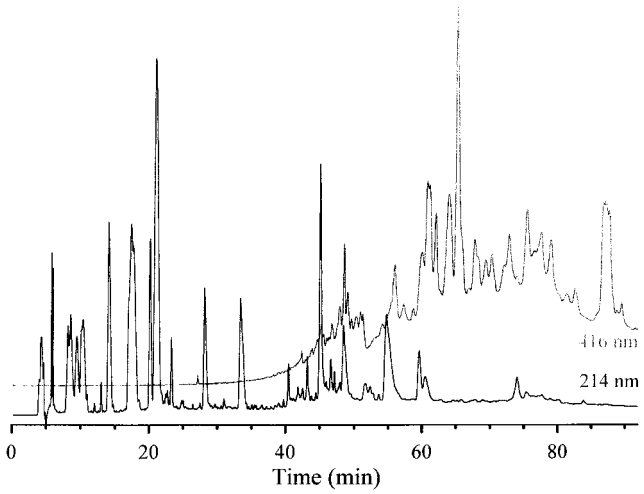


Figure 4. HPLC profile of the tryptic digest of Mb modified with **1a**.

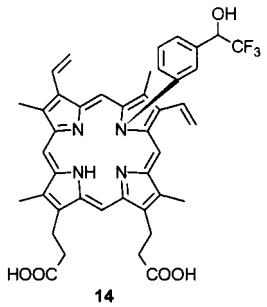
805 corresponding to a modified leucine or isoleucine. The peptide also produced a series of relatively abundant a-, b-, and (b+H_2O)-type internal fragments, all bearing the porphyrin ring and all containing a Leu and an adjacent Ile residue. The presence of a modified (C-terminal) y_4 and (N-terminal) a_{13} in the PSD spectrum indicated that cross-linking occurred either at Ile 28 or Leu 29. On the basis of other PSD spectra of porphyrin-*cross*-linked peptides (see below and ref *16*), we observed that the first (of lowest mass) fragment ion detected of the N-terminal a-type ion series is formed by a cleavage at the modified residue. All other a-ions result from cleavage at positions beyond the modified residue. Thus, since the a_{12} fragment was absent from the spectrum, we concluded that the modification had occurred at Leu-29.

Fractions eluting between 59 and 67 min contained the modified peptides His^{64}-Lys^{77} and His^{64}-Lys^{78} or Lys^{63}-Lys^{77} with masses at 2112.5 and 2239.2, respectively (Table 1). Figure 5B shows the PSD spectrum of the modified peptide His^{64}-Lys^{77}. The most abundant fragments in the spectrum were the protonated porphyrin ring at *m*/*z* 563 and a modified valine immonium ion at *m*/*z* 791. Modified N-terminal fragment ions from a_5 to b_{10} and modified internal fragments at *m*/*z* 903 (a-type VL) and *m*/*z* 1005 (a-type TVL or VLT) pinpointed Val-68 as the labeled amino acid. The other peptide is due to a missed cleavage by trypsin at one terminus of this peptide. PSD analysis (data not shown) revealed the additional Lys residue at the C-terminus, and Val-68 as the modified amino acid.

The modified peptides Ala^{57}-Lys^{77} and Ala^{57}-Lys^{78}, with MH^+ at 2882.62 and 3010.62, respectively, were also identified. The PSD spectrum of the shortest peptide (Figure 5C) gave an abundant modified His immonium ion at *m*/*z* 828 (unmodified would be *m*/*z* 110) and a series of modified fragment ions with charge-retention at the N-terminus (a_8-a_{13}), the first member of which was formed by cleavage between the His-64 and Gly-65 residues. In addition, a modified C-terminal fragment ion, y_{14}, formed via peptide bond cleavage between Lys-63 and His-64, was formed, indicating that the peptide was modified at His-64. The longer peptide resulted form a missed cleavage at Lys-77 due to the poor exoproteolytic properties of trypsin.

The MALDI-TOF spectra of the fraction containing peptide Ala^{57}-Lys^{77} also showed the presence of a peptide with MH^+ at *m*/*z* 2801.63 (data not shown). This mass corresponded to tryptic peptide His^{97}-Arg^{118} + 158 Da. The observed mass increment indicated the presence of the aromatic structure of the probe without the porphyrin ring, i.e., a labeling reaction in which the oxidative iron to nitrogen shift did not occur. The peptide was subjected to ESIMS analysis by nanospray ionization performed on a hybrid quadrupole-TOF mass spectrometer. The 5+ ion of this species (*m*/*z* 561.13) was subjected to low-energy CID analysis (data not shown). Low-energy CID data revealed that the peptide was His^{97}-Arg^{118} (Table 1). The fragment ions y_4, y_6, y_8, y_9, y_{11}, and b_{10} were observed unmodified and showed that Ile-107 was the site of the modification. An abundant ion at *m*/*z* 244, corresponding to a modified Ile immonium ion, confirmed this finding. A minor component was observed with MH^+ *m*/*z* 3361.84. This mass corresponds to a modified (heme containing) ^{97}His-^{118}Arg peptide (MH^+ calculated = 3361.79, −15 ppm). Unfortunately, it did not produce an abundant enough ion for further (PSD or CID) analysis by any of the ionization techniques.

MS analysis of the set of three products eluting at 87−90 min gave a protonated molecular mass of 737.3. Since they also had an absorbance maximum at 416 nm, the products were identified as the N-arylprotoporphyrin IX isomers **14**.



14

Discussion

The novel hemoprotein active-site probes, 3- and 4-(trifluoromethyldiazirinyl)phenyl diazenes **1a** and **2a,** are readily synthesized from commercially available starting materials (Schemes 1 and 2), and their acyl precursors (**1b** and **2b**) are stable for months when kept in the dark at −20 °C. When tested against sperm whale Mb, we found that the rates of Fe-aryl formation of **1a** or **2a** with Mb were slower than the rates for their azido counterpart. This difference is probably due to the constricted nature of the Mb active site and the larger size of the trifluoromethyldiazirine than azido group, although there may be an electronic contribution to the rate difference due to the differences in the electron-withdrawing effects of the two functional groups. It can be expected that the steric contribution to this difference in rates will be less important with proteins such as the cytochrome P450 enzymes, which generally have larger active sites than that of Mb.

Photolysis of the Fe-(3−diazirinylphenyl) complex of sperm whale Mb followed by oxidative shift of the aryl group to the porphyrin nitrogens resulted in covalent attachment of the heme group to the protein via the cross-linking probe (Figure 3). The levels of heme labeled Mb that are obtained are lower than those observed with *m*-azidophenyldiazene.[16] Several factors may contribute to this difference: (a) incomplete photolysis, (b) less efficient cross-linking by the photoactivated function with respect to deactivating reactions, (c) insertion into a larger number of active-site residues, and/or (d) incomplete iron to nitrogen shift, yielding a protein that is covalently attached to the probe but not to the heme chromophore. Our data suggest that all these factors may come into play: (a) peaks between 19 and 21 min in panel C of Figure 3 indicate that not all of the probe was photolyzed and that a photolysis time longer than 30 min might be advantageous, (b) the presence of an important

Table 1. Modified Peptides Isolated from Tryptic Digest of Labeled Mb

$MH^+_{observed}$[a]	peptide	$MH^+_{calculated}$	deviation (ppm)	sequence[b]
2310.96	$Val^{17}-Arg^{31}$	2311.12	−70	VEADVAGHGQDI**L**IR
2111.24	$His^{64}-Lys^{77}$	2111.13	+52	HGVT**V**LTALGAILK
2239.32	$His^{64}-Lys^{78}$	2239.23	+38	HGVT**V**LTALGAILKK
2882.62	$Ala^{57}-Lys^{77}$	2882.55	+24	ASEDLKK**H**GVTVLTALGAILK
3010.47	$Ala^{57}-Lys^{78}$	3010.64	−56	ASEDLKK**H**GVTVLTALGAILKK
2801.31	$His^{97}-Arg^{118}$	2801.55	−78	HKIPIKYLEF**I**SEAIIHVLHSR[c]

[a] Masses obtained by MALDI-TOF mass spectrometry, in delayed extraction, reflectron mode, with two point external calibration. [b] The modified residue is indicated in bold. [c] Modified with the probe only.

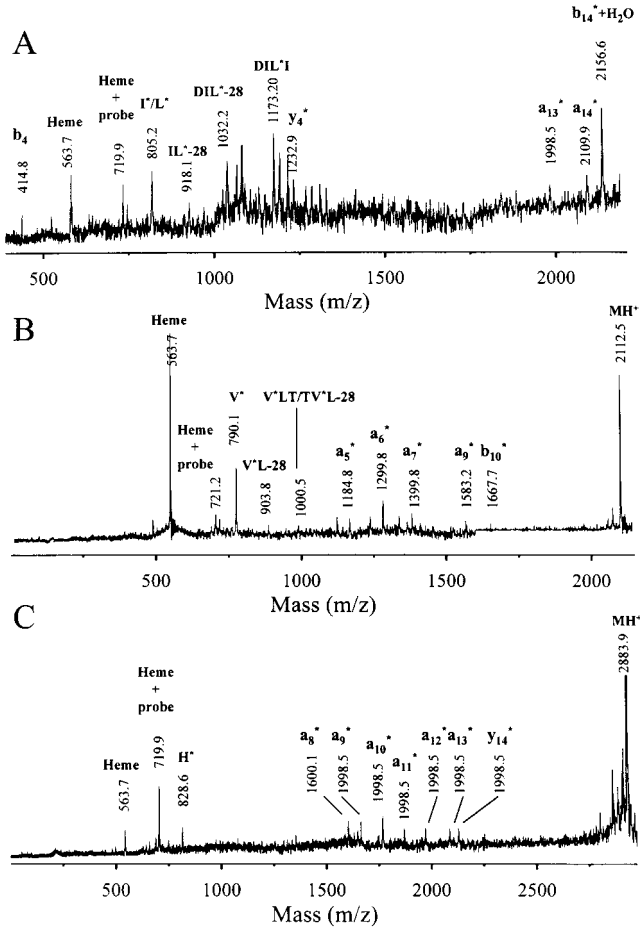


Figure 5. Post source decay (PSD) mass spectrum of labeled peptides (A) $Val^{17}-Arg^{31}$, (B) $His^{64}-Lys^{77}$, (C) $Ala^{57}-Lys^{77}$; the modified fragments are labeled with an asterisk.

nonprotein peak absorbing at 416 nm that elutes between 16 and 17 min (Figure 3, panel C) indicates the formation of free N-arylprotoporphyrin IX adducts derived from photolyzed complex that did not cross-link to the protein but underwent the iron to nitrogen shift, and (c) the isolation of a peptide modified at Ile-107 that does not bear the heme suggests that covalent attachment of the probe to the protein may sometimes hinder the iron to nitrogen shift (see below also).

Tryptic digestion and mass spectrometric analysis of the HPLC-purified peptides indicated labeling of at least four different amino acid side chains within the Mb active site (Table 1, Figure 6). This result clearly confirms the superior labeling properties of the trifluoromethyldiazirine over the azido group in the context of an active-site mapping strategy. Three out of the four residues identified by the former probe are nonnucleophilic and are located no more than 2−3 Å away from the estimated position of the photogenerated carbene (Figure 6).

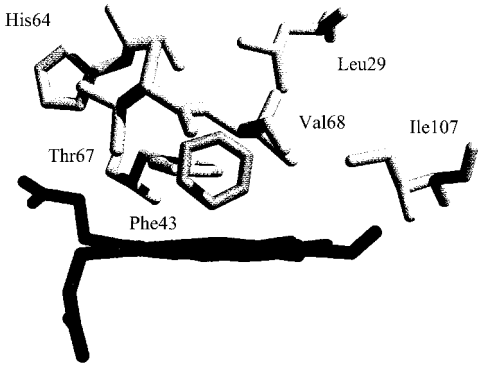


Figure 6. Side view of the distal side of the active site of the Mb-phenyl complex. The phenyl group bound to the iron is shown, and the residues are labeled.

This result illustrates one of the strengths of the present approach, since the relative position of the labeled residues with respect to the iron of the heme group can be approximated by inspection of the geometry of the probe. Examination of the crystal structure of the Mb iron-phenyl complex revealed that Phe-43 is the only residue in proximity of the probe for which a cross-linked peptide was not isolated. This observation is not surprising, as the phenylalanine side chain is unfavorably positioned with respect to the probe, but we cannot exclude the possibility that the phenylalanine was actually labeled but the iron to nitrogen shift did not occur and the adduct escaped undetected (Figure 6).

The peptides were obtained as isomeric mixtures. This is not unexpected given that cross-linking may occur to multiple points on the same residue (i.e., Leu-29 can be modified at C-β, C-γ, or C-δ) and that the iron to nitrogen shift can produce up to four different N-arylporphyrin regioisomers. The retention times of these closely related peptides are expected to vary only slightly, so the peptides elute as an isomeric cluster. We have been unable to detect differences in the mass spectrometric fragmentation patterns of these isomeric peaks.

Ile-107 was found covalently bound to the aryl probe but not to the heme, which suggests that the iron to nitrogen shift was inhibited or suppressed by covalent attachment of the probe to this residue. Covalent attachment of the probe to Ile-107 may interfere with the motion necessary for the shift to occur or may otherwise interfere with the shift, resulting in decomposition of the complex rather than the iron to nitrogen shift. The detection of this peptide was fortuitous because it coeluted with a heme-bearing peptide. It would not have been detected otherwise since it lacks the heme chromophore used to monitor the position of the labeled peptides. This shortcoming can be circumvented, if desired, by the use of radiolabeled probes that would render covalent attachment of the heme to the probe unnecessary.

The presence of the free N-arylprotoporphyrin IX isomers **14**, or closely related structures derived from the precursor

carbene, in the tryptic digest of modified Mb was unexpected since the protein had been purified by HPLC prior to tryptic digestion in order to remove heme products that were not covalently bound to the protein. It is possible that a fraction of the probe formed an unstable adduct with Mb that decomposed, releasing the modified N-arylporphyrins, during the tryptic digestion at pH 8.3 but not at the highly acidic pH values employed for the oxidative shift and the protein purification (both at pH <1). A candidate for such a labile alkylation site is Thr-67 (Figure 6). It is located close to the probe, and its hydroxyl group is optimally situated in proximity of the photogenerated carbene to allow trapping of the carbene by the alcohol group. It is not clear, however, whether the resulting trifluoromethyl-substituted ether adduct would be sufficiently unstable under basic conditions to be cleaved during the tryptic digestion.

In summary, the use of probes that bind to the heme iron and are then photoactivated to unmask a carbene functionality complements and greatly expands the previously reported use of related probes in which photoactivation unmasked a nitrene function.[16] The nitrene function primarily reacts with nucleophilic residues in the hemoprotein active site, whereas the carbene function successfully inserts into a range of unactivated amino acids in the same active site. The nitrene probes thus provide information on reactive amino acids, but the carbene probes make possible a more complete mapping of the active-site residues. Furthermore, as the distance and angle of the photoactivated moiety with respect to the heme iron atom and the heme plane is fixed, it is possible to assign approximate positions to the labeled amino acids within the active site. The topological information provided by this approach is thus much firmer than that obtainable by normal photoaffinity labeling or the use of mechanism-based inactivating agents, for which it is necessary to question whether the labeling even occurs in the active site. The demonstration that the probes can be used to map all but one of the accessible active-site residues in sperm whale Mb confirms the validity of the method and opens the way to the use of these probes to map unknown hemoprotein active sites.

Experimental Procedures

General Procedures. The ^{1}H and ^{13}C spectra were recorded on a General Electric GE 300 NMR spectrometer, FTIR spectra on an Impact 400 Nicolet infrared spectrophotometer, and UV−vis spectra on an HP 8452 diode-array spectrophotometer. Mass spectra were obtained at the University of California, San Francisco Mass Spectrometry Facility. Reagents were from Aldrich or Fisher and were used without further purification. The chromatographic purification of synthetic organic compounds was carried out on Merck 60 (230−600 mesh) silica gel. *m*-Nitrotrifluoroacetophenone was prepared according to the method of Steward and Vander Linden[22] and reduced to *m*-aminotrifluoroacetophenone with triethylammonium formate.[21] *p*-Fluorotrifluoroacetophenone was prepared as previously reported.[23] Melting points are uncorrected. Sperm whale Mb was purchased from Sigma and sequencing grade modified trypsin was obtained from Promega.

Ethyl 2-(3-Trifluoroacetylphenyl)hydrazinecarboxylate (5). *m*-Aminotrifluoroacetophenone (1.7 g, 9.0 mmol) was dissolved in 30 mL of 6 N HCl and cooled to −5 °C. $NaNO_2$ (0.68 g, 9.8 mmol) in 1 mL of water was added, and the mixture was stirred for 12 min. $SnCl_2$ (3.91 g, 20.6 mmol) in 3 mL of 12 N HCl was added, and stirring was continued for another 12 min. The mixture was neutralized with Na_2CO_3 and extracted with AcOEt. The organic layer was washed with brine, dried over anhydrous $MgSO_4$, and evaporated to dryness. The oily residue was redissolved in 20 mL of MeCN before pyridine (0.92 mL) and ethyl chloroformate (1.08 mL) were sequentially added. After 10 min, the solvent was removed under vacuum and the residue was purified by chromatography with hexanes/AcOEt 60:40 as the eluting solvent. This process yielded 1.73 g (70%) of pale yellow crystals, mp 107−109 °C: ^{1}H NMR δ ($CDCl_3$) 1.283 (t, J = 7.2 Hz, 3H), 4.209 (q, J = 7.2 Hz, 2H), 6.120 (s, 1H), 6.678 (s, 1H), 7.143 (dd, J = 1.8, 8.1 Hz, 1H), 7.389 (t, J = 8.1 Hz, 1H), 7.494 (s, 1H), and 7.565 ppm (d, J = 8.1 Hz, 1H); ^{13}C NMR δ ($CDCl_3$) 14.448, 62.273, 113.452, 116.620 (q, J_{C-F} = 1156 Hz), 120.027, 122.651, 129.831, 130.757, 148.883, 157.005, and 180.458 ppm (q, J_{C-F} = 137 Hz); IR ν (KBr) 3284, 1722, 1696, 1291, 1249, 1204, 1163, 1140 cm^{-1}; HRMS (EI) calcd for $C_{11}H_{11}N_2O_3F_3$ (M^+) 276.0722, found 276.0716.

Ethyl 2-(4-Trifluoroacetylphenyl)hydrazinecarboxylate (7). *p*-Fluorotrifluoroacetophenone (1.02 g, 5.33 mmol) was dissolved in 20 mL of dimethyl sulfoxide. Hydrazine (0.18 mL, 5.73 mmol) was added, and the mixture was heated to 90 °C for 1 h before the mixture was cooled, poured into 200 mL of water, and extracted with AcOEt. The organic layer was washed with brine, dried over anhydrous $MgSO_4$, and evaporated to dryness. The residue was dissolved in MeCN (20 mL) before pyridine (0.5 mL) and ethylchloroformate (0.55 mL) were sequentially added. After 10 min, the solvent was removed under vacuum and the residue purified by chromatography (hexanes/AcOEt, 70:30), yielding a white solid (0.46 g, 31%), mp 138−140 °C: ^{1}H NMR δ ($CDCl_3$) 1.293 (t, J = 6.9 Hz, 3H), 4.229 (q, J = 6.9 Hz, 2H), 6.390 (s, 1H), 6.623 (s, 1H), 6.861 (d, J = 8.7 Hz, 2H), and 7.971 ppm (d, 8.7 Hz, 2H); ^{13}C NMR δ (acetone-d_6) 14.891, 62.216, 112.258, 118.280 (q, J_{C-F} = 1156 Hz), 121.070, 133.238, 156.666, 157.693, and 178.475 ppm (q, J_{C-F} = 133 Hz), IR ν (KBr) 3337, 3298, 1693, 1601 cm^{-1}; HRMS (EI) calcd for $C_{11}H_{11}N_2O_3F_3$ (M^+) 276.0722, found 276.0724.

Oximes. The trifluoromethyl ketone **5** or **7** (1−2 mmol) was dissolved in 20 mL of EtOH containing 0.5 mL of pyridine. Hydroxylamine hydrochloride (1.2 eq) was then added, and the mixture was refluxed for 2 h before the solvent was removed under vacuum and the residue was purified by chromatography (hexanes/AcOEt 50:50). The oxime products **8** and **9** are thus obtained as a mixture of isomers that can be separated by flash chromatography (hexanes/AcOEt 70:30).

Ethyl 2-(3-Trifluoroacetyloximephenyl)hydrazinecarboxylate, "Early Fraction" (8a). White solid, mp 139−141 °C: ^{1}H NMR δ (acetone-δ_6) 1.203 (s, 3H), 4.102 (q, J = 6.9 Hz, 2H), 6.890 (d, J = 7.8 Hz, 1H), 6.949 (d, J = 7.8 Hz, 1H), 6.961 (s, 1H), 7.194 (s. 1H), 7.309, (t, J = 7.8 Hz, 1H), 8.266 (s, 1H), and 11.627 ppm (s, 1H); ^{13}C NMR δ (CD_3OD) 15.047, 62.652, 113.834, 115.134, 120.901, 122.622 (q, J_{C-F} = 1083 Hz), 129.901, 130.136, 147.487 (q, J_{C-F} = 129 Hz), 150.672, and 159.926 ppm; IR ν (KBr) 3350, 3309, 1715, 1127 cm^{-1}; HRMS (EI) calcd for $C_{11}H_{12}N_3O_3F_3$ (M^+) 291.0831, found 291.0831.

Ethyl 2-(3-Trifluoroacetyloximephenyl)hydrazinecarboxylate, "Late Fraction" (8b). White solid, mp 140−142 °C: ^{1}H NMR δ (acetone-d_6) 1.205 (s, 3H), 4.102 (q, J = 6.9 Hz, 2H), 6.898 (d, J = 7.8 Hz, 1H), 6.944 (d, J = 7.8 Hz, 1H), 6.957 (s, 1H), 7.175 (s. 1H), 7.260, (t, J = 7.8 Hz, 1H), 8.271 (s, 1H), and 11.977 ppm (s, 1H); ^{13}C NMR δ (CD_3OD) 15.0048, 62.652, 113.390, 120.295 (q, J_{C-F} = 1119 Hz), 120.693, 130.101, 133.177, 147.259 (q, J_{C-F} = 115 Hz), 150.680, and 159.923 ppm; IR ν (KBr) 3340, 3278, 1698, 1143 cm^{-1}; HRMS (EI) calcd for $C_{11}H_{12}N_3O_3F_3$ (M^+) 291.0831, found 291.0842.

Ethyl 2-(4-Trifluoroacetyloximephenyl)hydrazinecarboxylate, "Early Fraction" (9a). White solid, mp 138−139 °C: ^{1}H NMR δ (acetone-d_6) 1.218 (t, J = 7.2 Hz, 3H), 4.119 (q, J = 7.2 Hz, 2H), 6.901 (d, J = 8.7 Hz, 2H), 7.403 (s, 1H), 7.467 (d, 8.7 Hz, 2H), 8.322 (s, 1H), and 11.527 ppm (s, 1H); ^{13}C NMR δ (acetone-d_6) 15.004, 61.750, 112.440, 118.036, 122.571 (q, J_{C-F} = 1081 Hz), 130.845, 146.581 (q, J_{C-F} = 126 Hz), 152.057, and 157.869 ppm; IR ν (KBr) 3319, 1667, 1173, 1122 cm^{-1}; HRMS (EI) calcd for $C_{11}H_{12}N_3O_3F_3$ (M^+) 291.0831, found 291.0835.

Ethyl 2-(4-Trifluoroacetyloximephenyl)hydrazinecarboxylate, "Late Fraction" (9b). White solid mp 156−158 °C: ^{1}H NMR δ (acetone-d_6) 1.214 (t, J = 7.2 Hz, 3H), 4.109 (q, J = 7.2 Hz, 2H), 6.861 (d, J = 8.7 Hz, 2H), 7.345 (d, 8.7 Hz, 2H), 8.310 (s, 1H), and 11.778 ppm (s, 1H); ^{13}C NMR δ (acetone-d_6) 15.010, 61.699, 112.680, 120.183 (q, J_{C-F} = 1120 Hz), 122.259, 130.183, 146.763 (q, J_{C-F} = 106 Hz),

(22) Steward, R.; Vander Linden, R. *Can. J. Chem.* **1960**, *38*, 399.
(23) Herkes, F. E.; Burton, D. J. *J. Org. Chem.* **1967**, *32*, 1311.

152.130, and 157.826 ppm; IR ν (KBr) 3388, 3329, 1715, 1209, 1127 cm^{-1}; HRMS (EI) calcd for $C_{11}H_{12}N_3O_3F_3$ (M^+) 291.0831, found 291.0833.

Tosyl Oxime Preparation. The oximes prepared above (2−4 mmol) were treated with 4-toluenesulfonyl chloride (1.1 equiv) in acetone (20 mL) and triethylamine (1 mL). After 10 min the solutions were filtered, evaporated to dryness, and purified by flash chromatography (hexanes/AcOEt, 80/20). In the case of **10**, the product is obtained as a mixture of isomers that can be separated by chromatography.

Ethyl 2-(3-Trifluoroacetyltosyloximephenyl)hydrazinecarboxylate, "Early Fraction" (10a). Yellow oil: ^{1}H NMR δ ($CDCl_3$) 1.252 (t, J = 7.2 Hz, 3H), 2.465 (q, J = 7.2 Hz, 2H), 4.197 (s, 3H), 6.690 (s, 1H), 6.809 (s, 1H), 6.857, (d, J = 7.8 Hz, 1H), 6.935 (dd, J = 1.8, 7.8 Hz, 1H), 7.282 (t, J = 7.8 Hz), 7.372 (d, J = 8.1 Hz, 2H), and 7.868 ppm (d, J = 8.1 Hz, 2H); ^{13}C NMR δ ($CDCl_3$) 14.382, 21.712, 62.156, 112.286, 116.018, 119.500 (q, J_{C-F} = 1102 Hz), 120.443, 125.269, 129.200, 129.656, 129.830, 131.122, 146.089, 148.419, 154.071 (q, J_{C-F} = 134 Hz), and 157.027 ppm; IR ν (KBr) 3401, 1720, 1386, 1196 cm^{-1}; HRMS (EI) calcd for $C_{18}H_{18}N_3O_5SF_3$ (M^+) 445.0919, found 445.0918.

Ethyl 2-(3-Trifluoroacetyltosyloximephenyl)hydrazinecarboxylate, "Late Fraction" (10b). Yellow oil: ^{1}H NMR δ ($CDCl_3$) 1.269 (t, J = 7.2 Hz, 3H), 2.451 (q, J = 7.2 Hz, 2H), 4.197 (s, 3H), 6.611 (s, 1H), 6.868 (s, 1H), 6.922, (d, J = 7.8 Hz, 1H), 6.950 (dd, J = 1.8, 7.8 Hz, 1H), 7.248 (t, J = 7.8 Hz), 7.361 (d, J = 8.1 Hz, 2H), and 7.860 ppm (d, J = 8.1 Hz, 2H); ^{13}C NMR δ ($CDCl_3$) 14.433, 21.718, 62.177, 112.923, 116.193, 117.245 (q, J_{C-F} = 1126 Hz), 121.230, 128.505, 129.074, 129.540, 129.874, 131.412, 145.973, 148.387, 153.484 (q, J_{C-F} = 128 Hz), and 156.991 ppm; IR ν (KBr) 3367, 1720, 1386, 1195 cm^{-1}; HRMS (EI) calcd for $C_{18}H_{18}N_3O_5SF_3$ (M^+) 445.0919, found 445.0916.

Ethyl 2-(4-Trifluoroacetyltosyloximephenyl)hydrazinecarboxylate (11). Yellow oil: ^{1}H NMR δ ($CDCl_3$) 1.241 (t, J = 7.2 Hz, 3H), 2.433 (s, 3H), 4.177 (q J = 7.2 Hz, 2H), 6.471 (s, 1H), 6.782 (d, J = 8.4 Hz, 2H), 6.947 (s, 1H) 7.346 (d, J = 8.1 Hz, 2H), 7.360 (d, J = 8.4 Hz, 2H), and 7.862 ppm (d, J = 8.1 Hz, 2H); ^{13}C NMR δ ($CDCl_3$) 14.306, 21.557, 62.111, 111.895, 115.333, 119.800 (q, J_{C-F} = 1104 Hz), 129.719, 130.423, 131.055, 145.997, 150.940, 153.013 (q, J_{C-F} = 129 Hz), and 157.070 ppm; IR ν (KBr) 3365, 1724, 1608, 1195 cm^{-1}; HRMS (EI) calcd for $C_{18}H_{18}N_3O_5SF_3$ (M^+) 445.0919, found 445.0917.

Diaziridinyl Hydrazine. The tosyloxime (**10** or **11**) was dissolved in diethyl ether (3−4 mL) and added to 4−5 mL of NH_3(l) at −78 °C. The mixture was stirred at −78 °C for 4 h and brought to room-temperature overnight. The mixture was then evaporated to dryness and purified by flash chromatography (hexanes:AcOEt 70:30).

Ethyl 2-(3-Trifluoromethyldiaziridinylphenyl)hydrazinecarboxylate (12). White solid (60%), mp 112−114 °C: ^{1}H NMR δ ($CDCl_3$) 1.254 (t, J = 7.2 Hz, 3H), 2.296 (d, J = 8.4 Hz, 1H), 2.770 (d, J = 8.4 Hz, 1H), 4.173 (q, J = 7.2 Hz, 2H), 6.002 (s, 1H), 6.678 (s, 1H), 6.861 (d, J = 7.5 Hz, 1H), 7.111 (s, 1H), and 7.253 ppm (d, 7.5 Hz, 1H); ^{13}C NMR δ ($CDCl_3$) 14.160, 57.971 (q, J_{C-F} = 141 Hz), 61.841, 112.145, 114.097, 119.926, 123.334 (q, J_{C-F} = 1105 Hz), 129.329, 132.292, 148.426, and 157.331 ppm; IR ν (KBr) 3278, 3261, 3209, 1728, 1266, 1155 cm^{-1}; HRMS (EI) calcd for $C_{11}H_{13}N_4O_2F_3$ (M^+) 290.0991, found 290.1005.

Ethyl 2-(4-Trifluoromethyldiaziridinylphenyl)hydrazinecarboxylate (13). Yellow oil (7%): ^{1}H NMR δ ($CDCl_3$) 1.254 (t, J = 7.2 Hz, 3H), 2.198 (d, J = 9.0 Hz, 1H), 2.746 (d, J = 9.0 Hz, 1H), 4.190 (q, J = 7.2 Hz, 2H), 5.986 (s, 1H), 6.656 (s, 1H), 6.810 (d, J = 8.4 Hz, 2H), and 7.448 ppm (d, J = 8.4 Hz, 2H): ^{13}C NMR δ ($CDCl_3$) 14.455, 57.654 (q, J_{C-F} = 139 Hz), 62.128, 112.680, 123.598 (q, J_{C-F} = 1108 Hz), 123.607, 129.198, 149.492, and 157.046 ppm; IR ν (KBr) 3322, 1719, 1155 cm^{-1}; HRMS (EI) calcd for $C_{11}H_{12}N_3O_3F_3$ (M^+) 291.0831, found 291.0831.

Diazirinyl Diazene. The diazirinyl hydrazine (**12** or **13**) was dissolved in $CHCl_3$ or $CDCl_3$. Lead tetraacetate was added stepwise and the reaction monitored by TLC (CH_2Cl_2) or ^{1}H NMR until completion. The final product was purified by chromatography (CH_2Cl_2) and obtained in ∼90% yield.

Ethyl 2-(3-Trifluoromethyldiazirinylphenyl)diazenecarboxylate (1). Orange liquid: ^{1}H NMR δ ($CDCl_3$) 1.474 (t, J = 7.2 Hz, 3H), 4.532 (q, J = 7.2 Hz, 2H), 7.434 (d, J = 7.8 Hz, 1H), 7.588 (t, J = 7.8 Hz, 1H), 7.758 (s, 1H), and 7.965 ppm (d, J = 7.8 Hz, 1H); ^{13}C NMR δ ($CDCl_3$) 14.026, 28.204 (q, J_{C-F} = 156 Hz), 64.642, 121.809, 121.853 (q, J_{C-F} = 1090 Hz), 124.468, 129.976, 130.724, 131.014, 151.515, and 161.818 ppm; IR ν (KBr) 1762, 1252, 1234, 1146 cm^{-1}; UV λ_{max} (MeOH) 350, 282 nm; HRMS (EI) calcd for $C_8H_4N_4F_3$ ($M^+ - CO_2CH_2CH_3$) 213.0388, found 213.0389.

Ethyl 2-(3-Trifluoromethyldiazirinylphenyl)diazenecarboxylate (2). Orange liquid: ^{1}H NMR δ ($CDCl_3$) 1.474 (t, J = 7.2 Hz, 3H), 4.529 (q, J = 7.2 Hz, 2H), 7.345 (d, J = 8.4 Hz, 2H), and 7.948 ppm (d, J = 8.4 Hz, 2H); ^{13}C NMR δ ($CDCl_3$) 14.131, 28.453 (q, J_{C-F} = 164 Hz), 64.714, 121.794, (q, J_{C-F} = 1096 Hz), 123.856, 127.352, 134.403, 151.665, and 161.860 ppm; IR ν (KBr) 1761, 1246, 1155 cm^{-1}; UV λmax (MeOH) 351, 288 nm; HRMS (EI) calcd for $C_8H_4N_4F_3$ ($M^+ - CO_2CH_2CH_3$) 213.0388, found 213.0398.

Sperm Whale Myoglobin Labeling. The aryl-iron complexes were formed according to a previously published procedure.[16] Prior to photolysis, the iron-aryl complex was run through a PD-10 (Pharmacia) column. Irradiation of the complex was performed at 25 °C for 30 min using a Rayonet photoreactor at 350 nm. The sample was held 2 cm from the source and an additional Pyrex filter was placed between the lamp and the samples to avoid protein modification. The iron to nitrogen shift of the probe was induced by pouring the mixture into acetonitrile/trifluoroacetic acid (96:4) and keeping the mixture at 4 °C for 16 h. HPLC analysis and purification of labeled Mb was performed according to a reported procedure.[16] The labeled Mb was lyophilized and resuspended in 200 μL of 50 mM ammonium carbonate. Sequencing grade trypsin (Promega) was added at a ratio 2/100 (w/w) and the digestion carried out at 37 °C for 3 h. HPLC purification of the tryptic peptides was performed on a Vydac C-18 RP column eluted at a flow rate of 0.8 mL/min with a gradient of solvent B (acetonitrile/trifluoroacetic acid 0.085%) into solvent A (water/trifluoroacetic acid 0.1%): 0−1 min 5% B, 1−40 min 5−35% B, 40−80 min 35−45% B, and 80−90 min 45−80% B. Fractions were collected every minute.

Mass Spectrometric Characterization of the Isolated Tryptic Fractions. MALDI (matrix-assisted laser desorption ionization) and MALDI-post source decay (PSD) experiments were carried out as previously reported.[16] Low-energy CID data were obtained on a QSTAR mass spectrometer (PE Sciex) equipped with an electrospray source. This instrument afforded sufficient resolution of the fragment ions allowing for their charge state determination. The peak nomenclature was taken from the literature.[24]

Acknowledgment. We thank PE Sciex for the opportunity of obtaining data on a QSTAR mass spectrometer. This research was supported by National Institutes of Health Grant GM25515. Mass spectrometric data were obtained at the Mass Spectrometry Facility (A. Burlingame, director) of the University of California at San Francisco supported by National Institutes of Health Grants RR 01614 and Liver Core Center 5 P30 DK26743.

JA990351H

(24) Biemann, K. *Methods Enzymol.* **1990**, *193*, 886−887.