Available online at www.sciencedirect.com
ScienceDirect
Bioorganic & Medicinal Chemistry 14 (2006) 6213–6222
Bioorganic & Medicinal Chemistry

N-Terminal amino acid side-chain cleavage of chemically modified peptides in the gas phase: A mass spectrometry technique for N-terminus identification

Almary Chacon,[a] Douglas S. Masterson,[b] Huiyong Yin,[a] Daniel C. Liebler[c] and Ned A. Porter[a,*]

[a]*Department of Chemistry, Vanderbilt University, Nashville, TN 37235, USA*
[b]*Department of Chemistry and Biochemistry, University of Southern Mississippi, Hattiesburg, MS 39406, USA*
[c]*Department of Biochemistry and Mass Spectrometry Research Center, Vanderbilt University, USA*

Received 7 February 2006; accepted 30 May 2006
Available online 19 June 2006

Dedicated to a leader in the field of free radical chemistry, Bernd Giese, the 2005 Tetrahedron Prize winner for Creativity in Organic Chemistry.

Abstract—Although genome databases have become the key for proteomic analyses, de novo sequencing remains essential for the study of organisms whose genomes have not been completed. In addition, post-translational modifications present a challenge in database searching. Recognition of the *b* or *y*-ion series in a peptide MS/MS spectrum as well as identification of the b_1- and y_{n-1}-ions can facilitate de novo analyses. Therefore, it is valuable to identify either amino-acid terminus. In previous work, we have demonstrated that peptides modified at the ε-amino group of lysine as a *t*-butyl peroxycarbamate derivative undergo free radical promoted peptide backbone fragmentation under low-energy collision-induced dissociation (CID) conditions. Here we explore the chemistry of the N-terminal amino group modified as a *t*-butyl peroxycarbamate. The conversion of N-terminal amines to peroxycarbamates of simple amino acids and peptides was studied with aryl *t*-butyl peroxycarbonates. ESI-MS/MS analysis of the peroxycarbamate adducts gave evidence of a product ion corresponding to the neutral loss of the N-terminal side chain (R), thus identifying this residue. Further fragmentation (MS^3) of product ions formed by N-terminal residue side-chain loss (-R) exhibited an *m*/*z* shift of the *b*-ions equal to the neutral loss of R, therefore labeling the *b*-ion series. The study was extended to the analysis of a protein tryptic digest where the SALSA algorithm was used to identify spectra containing these neutral losses. The method for N-terminus identification presented here has the potential for improvement of de novo analyses as well as in constraining peptide mass mapping database searches.

1. Introduction

The rapid development and improvement of mass spectrometry, along with the availability of large genome databases, have made possible high-throughput characterization of complex protein mixtures.[1] Nevertheless, the presence of unexpected post-translational modifications often leads to unsuccessful or incorrect identifications in database searches. In addition, the study of proteins derived from organisms whose genomes are not complete is limited.[2] In such cases, more labor-intensive data analysis such as de novo sequencing must be used, where the peptide sequence is reconstructed from the MS/MS fragmentation data with the help of sequencing algorithms.[3–5] Due to the complexity of peptide MS/MS spectra, a number of methods have been developed to facilitate interpretation. N-terminal and C-terminal charged derivatives were developed in the late 1980s to be used with fast atom bombardment (FAB) analyses in order to simplify fragmentation spectra produced from high-energy collision-induced dissociation. In addition, these charged derivatives were useful in increasing signal intensity of peptide fragment ions in such experiments.[6,7] Keough and co-workers derivatized the N-termini of peptides with a sulfonic acid group.[8,9] Fragmentation of these peptides by collision-induced dissociation (CID) and postsource decay (PSD) matrix-assisted laser desorption ionization (MALDI) yielded selectively the *y*-ion series. Work by Geskell et al. investigated N-terminus derivatization using phenylthiocarbamoyl derivatives similar to those employed in

Keywords: Proteomics; Free radical; Peroxycarbonate; Backbone fragmentation; Gas phase digestion.
* Corresponding author. Tel.: +1 615 343 2693; fax: +1 615 343 1234; e-mail: n.porter@vanderbilt.edu

0968-0896/$ - see front matter © 2006 Elsevier Ltd. All rights reserved.
doi:10.1016/j.bmc.2006.05.060

Edman degradation.[10] CID experiments in this case produced predominantly b_1- and y_{n-1}-ions. James and co-workers carried out the modification of N-termini of peptides with nicotinyl-*N*-hydroxysuccinimide which resulted in suppression of the *y*-ion series in MALDI tandem time-of-flight (MALDI-TOF-TOF) analyses.[11] More recently, Reilly and co-workers investigated the effect of acetaminidation of N-termini on fragmentation patterns in CID spectra.[12] In most cases studied, there was an enhancement of y_{n-1}-ions, while diminishing other fragmentations, and in some cases contiguous series of *b*-NH_3 ions were observed. Another approach to improving de novo methods involves stable isotopic labeling. Esterification of the C-termini with deuterium labeled methanol results in a mass shift of the *y*-ions with respect to the unlabeled peptide ester thus identifying the *y*-ion series.[13] In a similar manner, O^{18} labeling of the C-termini of peptides provides information about the *y*-ion series.[14] Cannon and co-workers presented a computational method to obtain information from the natural isotopic distributions of peptide fragments in MS/MS spectra, in order to facilitate identification of *b*- and *y*-ions without prior derivatization.[15]

We report here derivatization of the N-terminal α-amino group of amino acids and peptides to give a reactive adduct which, under low-energy CID conditions, promotes cleavage of the N-terminal amino acid side chain (Scheme 1). This fragmentation provides information about the nature of the N-terminal residue and may also serve as a labeling technique for elucidating the *b*-ion series in de novo analyses. Thus, single amino acids and peptides were modified at the N-terminus with aryl peroxycarbonates, **1a**–**1c**, as shown in Scheme 1. Fragmentation of the resulting peroxycarbamates in an ESI triple quadrupole instrument gave evidence of a product ion corresponding to the neutral loss of the N-terminal side chain (R), thus identifying this residue.

2. Results

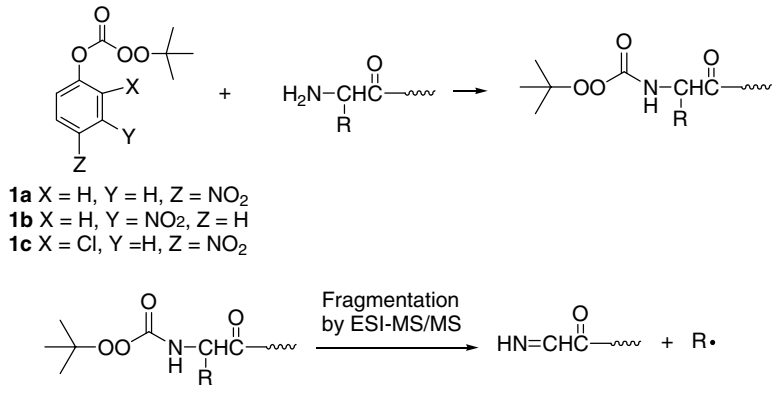
2.1. Single amino acids and peptides

CID spectra of the lithium adducts of modified amino acid ester derivatives in the form C_4H_9OO-$C(O)NHCH(R)CO_2$ R′ [$C_4H_9OOC(O)M$] displayed fragmentation involving the loss of *m*/*z* = 116, C_4H_8OO-C(O), to give the underivatized amino acid ester $NH_2CH(R)CO_2R'$, M([$C_4H_9OOC(O)M$—116]). In addition, the neutral loss of the *t*-butyl peroxycarbonate moiety, $C_4H_9OOC(O)$, plus the amino acid side chain, R, yielded the imine NH=CHCOOR′ ([C_4H_9OO C(O)M—117 –R]). A sample CID spectrum of modified valine methyl ester is shown in Figure 1. Lithium was used as the counter ion for ESI of the modified amino acids due to the unavailability of the N-terminus for protonation. Cleavage of the amino acid side chain was not observed in the CID spectra of the parent amino acid esters not bearing the peroxycarbamate moiety. In most cases [$C_4H_9OOC(O)M$—116] was the predominant fragment; however, the imine fragment was present in most spectra.

In the case of modified proline, the parent adduct ion was observed but the CID spectrum did not contain an imine fragment. In CID of the peroxycarbamate-modified glycine we did not observe an imine fragment which has a mass difference of 2 with respect to the unmodified amino acid, instead, we observed a fragment which matches the loss of only 1 hydrogen with respect to the unmodified amino acid [$C_4H_9OOC(O)M$—117] accompanied by the usual [$C_4H_9OOC(O)M$—116] fragment. Modified asparagines also did not lead to an imine fragment; however, an imine fragment was observed in the CID spectrum of modified peptides bearing asparagine at the N-terminus. A list of modified amino acids and corresponding CID fragmentations are presented in Table 1.

2.2. Peptide modifications

Peptide YRVRFLAKENVTQDAEDNC (entry **21** in Table 2) was modified under different pH conditions, ranging from 5.2 to 8.6, in order to determine the pH optimum for N-terminal modification over ε-N modification of Lys. Previous studies have shown that selectivity for reactions at the N-terminal α-amino group of peptides can be achieved by pH control, due to the lower pK_a of the N-terminus compared to the ε-amino group of lysine.[18–21] In addition to pH control, a less reactive reagent than **1a** (**1b**) offers higher selectivity for reaction



Scheme 1. Modification and fragmentation of derivatized N-termini.

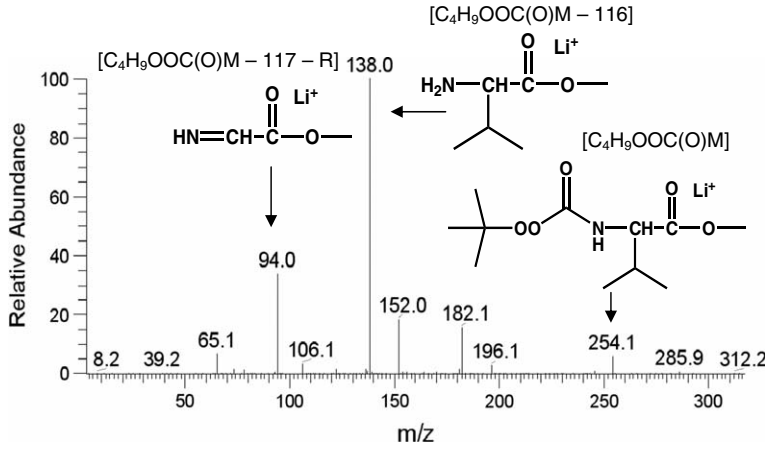


Figure 1. CID spectrum of N-terminal-modified valine methyl ester.

Table 1. Amino acid adducts and their CID fragmentation

Amino acid[a]	$[M+Li^+]$ *m/z*	$[^tBuOOC(O)M+Li^+]$ *m/z*	Neutral loss[b]	R	R *m/z*
Gly	96.1	212.0	117.0	H	1
Ala	110.1	226.1	132.0	CH_3	15
Val	138.0	254.2	160.1	$CH(CH_3)_2$	43
Leu	152.1	268.2	174.1	$CH_2CH(CH_3)_2$	57
Ile	152.2	268.2	174.2	$CH(CH_3)CH_2CH_3$	57
Glu	210.0	326.0	204.1	$(CH_2)_2(CO)OC_2H_5$	101
Asp	168.0	283.9	190.0	$CH_2(CO)OCH_3$	73
Asn	194.4	311.1	—	$CH_2(CO)NH_2$	58
His	176.1	291.9	198.0	$CH_2C_3H_3N_2$	81
Met	170.1	285.9	192.1	$(CH_2)_2SCH_3$	75
Phe	186.0	302.0	208.1	$CH_2C_6H_5$	91
Ser	126.0	242.0	148.1	CH_2OH	31
Tyr	202.1	318.0	224.1	$CH_2C_6H_5O$	107
Trp	225.0	341.0	247.0	$CH_2C_8H_6N$	130
Thr	140.2	256.1	162.1	$CH(OH)CH_3$	46
Gln	208.3	325.1	189.0	$(CH_2)_2(CO)NH_2$	72
Lys	181.1	297.2	189.0	$(CH_2)_4NH_2$	72
Pro[c]	137.0	251.9	—	C_3H_7	43
Cys[d]	—	—	—	CH_2SH	47
Arg	189.0(H^+)	305.0	217	$(CH_2)_3NHC(NH_2)_2$	100

[a] Amino acids are methyl ester derivatives (M = $NH_2CH(R)CO_2CH_3$) with the exception of Asp and Glu which are the dimethyl ester and diethyl ester derivatives, respectively, Lys is the ethyl ester and Gln and Asn are the *t*-butyl ester derivatives.

[b] Neutral loss corresponds to 117 +R where 117 is the *m/z* of the $C_4H_9OOC(O)$ group +1 hydrogen and R is the amino acid side chain.

[c] Cleavage of the side chain was not observed for Pro, only loss of the $C_4H_9OOC(O)$ group was seen.

[d] The MS spectrum of modified Cys was complex and did not contain a peak corresponding to $[C_4H_9OOC(O)M]$, suggesting decomposition of the amino acid may occur during the modification reaction. Modified Asn was also an exception where no side-chain loss was observed.

at the N-terminus. LC/MS analysis of the modified peptide revealed that under basic conditions (pH 7.5–8.5) and with reagent **1a**, modification occurs at both amino group sites, but with greater efficiency on the lysine side chain. Slightly acidic conditions (pH 5.5) and reagent **1b** gave better selectivity for the N-terminus. Figure 2 shows the LC/MS trace of the modified peptide at pH 8.6 and 5.5. Peak **1** in the chromatogram corresponds to *m/z* = 758.2 ($[M+3H]^{3+}$) where M = YRVRFLAKENVTQDAEDNC. Peaks **2** and **3** indicate a single modification with *m/z* = 796.8 ($[C_4H_9OOC(O)M+3H]^{3+}$). Peak **2** corresponds to modification at the N-terminal group, whereas peak **3** corresponds to α-N modification of lysine. This was confirmed by the CID fragmentation of peak **2**, also shown in Fig. 2, where the ion with *m/z* = 722.2 ($[C_4H_9OOC(O)M—117–107+3H]^{3+}$) indicates the neutral loss of $C_4H_9OOC(O)$ and the N- terminal tyrosine side chain (–CH_2–C_6H_4OH, *m/z* = 107). This fragment was absent in the CID spectrum of peak **3**. At lower pH and with the use of reagent **1b**, peak 3 is diminished, which implies a preference for N-terminal modification under these conditions.

Several peptides containing different N-terminal amino acids were modified using reagent **1b** at pH 5.5 and the N-termini were identified in CID experiments (Table 2). Thus, CID spectra of peptide peroxycarbamates bearing alanines at the N-terminus give product ions that indicate the loss of CH_3 ($[C_4H_9OOC(O)M—117–15]$), whereas terminal valines show a similar loss of CH_3CHCH_3. In cases where cysteine and proline were the terminal residues, the CID spectra did not exhibit side-chain cleavage product ions. Instead, as seen for

Table 2. Peptides analyzed by CID for N-terminal side-chain loss

Entry	Sequence	M[a] *m/z*	$[C_4H_9\ OOC(O)M]$[b] *m/z*	Imine fragment[c] *m/z*	R	R *m/z*
1	VYIHPF	775.4($1H^+$)	891.5($1H^+$)	731.1($1H^+$)	$CH(CH_3)_2$	43
2	HPKRPWIL	523.4($2H^+$)	581.8($2H^+$)	483.1($2H^+$)	$CH_2C_3H_3N_2$	81
3	MEHFRW	962.0($1H^+$)	1078.0($1H^+$)	887.2($1H^+$)	$(CH_2)_2SCH_3$	75
4	SRVSRRSR	335.2($3H^+$)	373.8(3H+)	324.5($3H^+$)	CH_2OH	31
5	NRVYVHPF	516.4($2H^+$)	574.4($2H^+$)	486.9($2H^+$)	$CH_2(CO)NH_2$	58
6	FHPKRPWIL	597.4($2H^+$)	655.4($2H^+$)	551.9($2H^+$)	$CH_2C_6H_5$	91
7	KNPYIL	768.4($1Na^+$)	885.2($1Na^+$)	679.7($1Na^+$)	$(CH_2)_4NH_2$	72
9	GFLRRI	761.4($1H^+$)	877.5($1H^+$)	760.3($1H^+$)	H	1
10	WHWLQL	882.5($1H^+$)	998.5($1H^+$)	752.1($1H^+$)	$CH_2C_8H_6N$	130
11	QRPRLSHKGPMPF	518.0($3H^+$)	556.4($3H^+$)	592.6($3H^+$)	$(CH_2)_2(CO)NH_2$	72
12	LSRLFDNAl	468.5($2H^+$)	526.5($2H^+$)	877.7($1H^+$)	$CH_2CH(CH_3)_2$	57
13	ISRPPGFSPFR	630.9($2H^+$)	689.0($2H^+$)	602.0($2H^+$)	$CH(CH_3)CH_2CH_3$	57
14	EDKDLD	795.2($1Na^+$)	911.5($1Na^+$)	722.0($1Na^+$)	$(CH_2)_2(CO)OH$	73
15	YGSLPQKSQRSQDEN	868.6($2H^+$)	926.8($2H^+$)	815.11($2H^+$)	$CH_2C_6H_4OH$	107
16	TPRK	251.1($2H^+$)	308.9($2H^+$)	205.8($2H^+$)	$CH(OH)CH_3$	46
17	APRLRFYSL	561.9($2H^+$)	619.8($2H^+$)	553.1($2H^+$)	CH_3	15
18	PPGFSPFR	904.3($1H^+$)	1020.4($1H^+$)	—	C_3H_7	43
19	RPKPQQF	900.3($1H^+$)	1016. 5($1H^+$)	801.2($1H^+$)	$(CH_2)_3\ NHC(NH_2)_2$	100
20	CNLAVAAASHIYQNQFQ	660.1($3H^+$)	—	—	CH_2SH	47
21	YRVRFLAKENVTQDEDC	758.2($3H^+$)	796.8($3H^+$)	722.2($3H^+$)	$CH_2C_6H_4OH$	107
22	DRVYIHPF	523.5($2H^+$)	582.1($2H^+$)	493.8($2H^+$)	$CH_2(CO)OH$	59

[a] M = *m/z* of unmodified peptide.
[b] *m/z* of peptide modified with reagent **1b** at pH 5.5.
[c] $[C_4H_9OOC(O)M—117 –R]$ fragment or imine fragment in CID spectrum resulting from the neutral loss of the$C_4H_9OOC(O)$ group +1 hydrogen (117) and the N-terminal amino acid side chain (R).

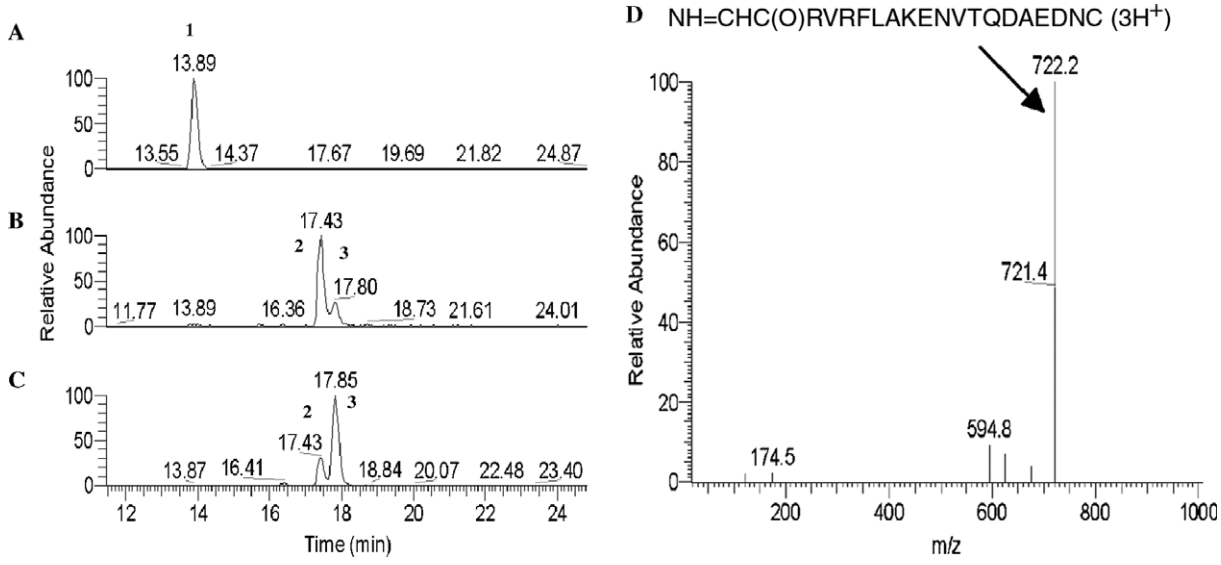


Figure 2. LC/MS of modified peptide YRVRFLAKENVTQDAEDNC. (A) Selected Ion Monitoring (SIM)*m/z* = 758.2 ($[M+3H]^{3+}$) where M is the unmodified peptide YRVRFLAKENVTQDAEDNC. (B) SIM *m/z* = 796.8 ($[C_4H_9OOC(O)M+3H]^{3+}$) where M was modified with reagent **1b** at pH 5.5. Peaks **2** and **3** correspond to N-terminus and lysine modifications, respectively. (C) SIM *m/z* = 796.8 ($[C_4H_9OOC(O)M+3H]^{3+}$) where M was modified with reagent **1a** at pH 8.6. (D) CID spectrum of peak **2,** parent m/z = 796.8 ($[C_4H_9OOC(O)M+3H]^{3+}$). The product ion *m/z* = 722.2 corresponds to the loss of $C_4H_9OOC(O)$ and the N-terminal tyrosine side chain, $CH_2C_6H_4OH$ ($[C_4H_9OOC(O)M—117–107+3H]^{3+}$).

the single amino acids, modified cysteine gave a complex MS spectrum and the CID spectrum of modified proline only showed the loss of the $C_4H_9OOC(O)$. Nevertheless, the presence of these amino acids anywhere else in the sequence did not affect the fragmentation chemistry at the N-terminus of modified peptides. CID of a modified peptide bearing glycine at the N-terminus yielded similar fragmentation to the CID of the modified single amino acid, where fragment $[C_4H_9OOC(O)M—117]$ was observed instead of the expected fragment $[C_4H_9OOC(O)M—117–1]$. Results from the CID of a modified peptide bearing Asn at the N-terminus (entry **5** in Table 2) indicated loss of the Asn side chain, which is in contrast to results obtained with the modified Asn amino acid derivative.

When a mixture of the two isobaric peptides HPKRPWIL and DRVYIHPF (entries **2** and **22** in Table 2) was modified at pH 5.5 using reagent **1b**, the two peptides were distinguishable in LC–MS selected reaction

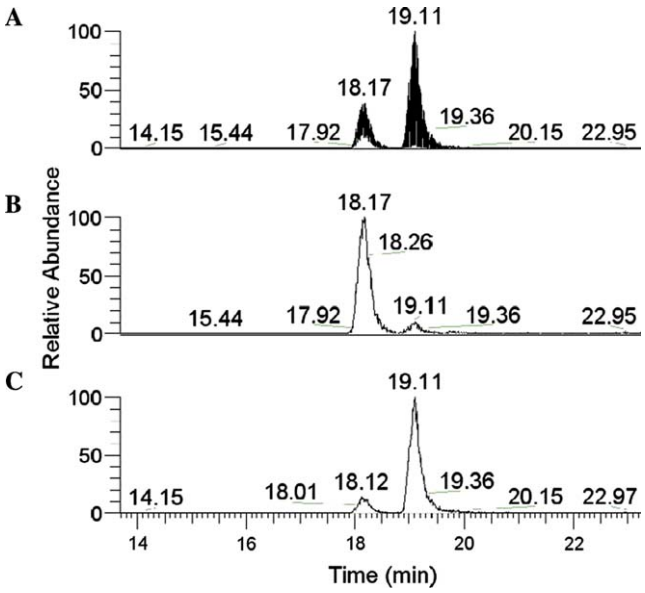


Figure 3. LC-SRM trace of peptide mixture HPKRPWIL and DRVYIHPF modified with reagent **1b** at pH 5.5. (A) SIM*m/z* = 582 ($[C_4H_9OOC(O)M+2H]^{2+}$) where M is either peptide. (B) SRM of *m/z* = 582 ($[C_4H_9OOC(O)M+2H]^{2+}$)→*m/z* = 483 ($[C_4H_9OOC(O)M—117–81+2H]^{2+}$) where M is HPKRPWIL, and the histidine side chain (CH_2-$C_3H_3N_2$) has *m/z* = 81.(**C**) SRM of *m/z* = 581 ($[C_4H_9OOC(O)M—117–81+2H]^{2+}$) → *m/z* = 494 ($[C_4H_9OOC(O)M—117–59+2H]^{2+}$) where M is DRVYIHPF and the aspartic acid side chain (–CH_2COOH) has *m/z* = 59.

monitoring (SRM) and neutral loss experiments by N-terminal side-chain cleavage (Fig. 3). Thus, the peptide bearing an N-terminal histidine can be distinguished from the peptide bearing aspartic acid without further sequencing. Figure 3(B) shows the SRM of *m/z* = 582 ($[C_4H_9OOC(O)M+2H]^{2+}$) → *m/z* = 483 ($[C_4H_9OOC(O)M—117–81+2H]^{2+}$) where M is HPKRPWIL and the side chain of histidine (CH_2–$C_3H_3N_2$) has *m/z* = 81. Figure 3(C) shows the SRM of *m/z* = 582 ($[C_4H_9OOC(O)M+2H]^{2+}$) → *m/z* = 494 ($[C_4H_9OOC(O)$-M—117–59+2H$]^{2+}$) where M is DRVYIHPF and the side chain of aspartic acid (–CH_2COOH) has *m/z* = 59.

2.3. Analysis of modified protein tryptic digest

A myoglobin tryptic digest standard was modified using reagent **1c** at pH 5.5. LC–MS/MS analysis was done on the unmodified as well as the modified digest. Visual inspection of the MS data identified 12 most abundant peptides, 7 of which were modified (Table 3). Identification of the peptides from MS/MS spectra was confirmed by database search algorithm SEQUEST[22] and the P-Mod algorithm.[23] N-terminal side-chain loss was observed in data-dependent MS^2 experiments for entries **2**, **4**, **5**, **7**, and **8**. MS^3 gave a product ion spectrum of the [$C_4H_9OOC(O)$-M—117 –R] fragment for entry **2**. This MS^3 spectrum exhibited a labeled *b*-ion series indicating the loss of the N-terminal valine side chain + 1 hydrogen with respect to the umodified peptide (mass shift = −44), while the *y*-ion series remained unchanged (Table 4).

The SALSA algorithm[17] was used to search through tandem MS data for spectra indicating the N-terminal side-chain losses. The SALSA algorithm was designed to find characteristics in the MS/MS spectra such as charged and neutral losses, a particular product ion, or a sequence motif. It can also search for a combination of two or more of these characteristics and it does not require a peptide sequence input. SALSA scores do not indicate the probability of false positives, but instead rank spectra based on intensity and number of ions found that match a search criteria. We searched our modified myoglobin digest data for the neutral loss of 116, which leads to the fragment [$C_4H_9OOC(O)M$—116] or simply [M]. This generated a list of precursor ions that match the *m/z* of modified peptides in the digest. To find peptides bearing a certain amino acid at the N-terminus, we added a secondary search characteristic corresponding to the neutral loss of 117 + R, which leads to the side-chain loss fragment [$C_4H_9OOC(O)$—117 –R]. Thus, ions corresponding to (neutral loss of 117 + R) were scored only in CID spectra also containing the primary loss of 116. This scoring scheme reduced the number of false positive matches. For example, a SALSA search for MS/MS spectra of peptides bearing histidine at the N-terminus scored a primary neutral loss

Table 3. Peptides detected by LC–MS from a modified myoglobin digest

Entry[a]	Peptide sequence	Unmodified *m/z*	Modified[b] *m/z*	Mass shift[c]
1	HLKTEAEMK	371.8($4H^+$)	—	—
2	ALELFR	374.8($2H^+$)	864.3($1H^+$)	116
3	YKELGFQG	472.3($2H^+$)	—	—
4	LFTGHPETLEK	424.8($3H^+$)	694.5($2H^+$)	116
5	HGTVVLTAIGGILK	460.4($3H^+$)	748.0($2H^+$)	116
6	HPGDFGADAQGAMTK	752.0($2H^+$)	810.0($2H^+$)	116
7	VEAKIAGHGQEVLIR	536.6($3H^+$)	862.1($2H^+$)	116
8	GLSDGEWQQVLNVWGK	908.6($2H^+$)	966.5($2H^+$)	116
9	GHHEAELKPLAQSHATK	464.5($4H^+$)	—	—
10	YLEFISDAIIHVLHSK	943.3($2H^+$)	—	—
11	LFTGHPETLEKFDK	416.2($4H^+$)	—	—
12	ELGFQG	650.2($1H^+$)	766.1($1H^+$)	116

[a] Entries in blue are peptides for which modification at the N-terminus is observed.
[b] *m/z* of modified peptides ([$C_4H_9OOC(O)M$]).
[c] Mass shift is adjusted for singly charged species.

Table 4. *b*- and *y*-ions from MS^2 of unmodified peptide VEAKIAGHGQEVLIR compared to *b*- and *y*-ions from MS^3 of the modified peptide([$C_4H_9OOC(O)M$])

	b-ions			*y*-ions		
Position	MS^2 unmodified peptide[a]	MS^3 modified peptide[b]	Mass shift[c]	MS^2 unmodified peptide[a]	MS^3 modified peptide[b]	Mass shift
2	228.79	—	—	—	—	—
3	299.92	—	—	401.14	401.40	0.26
4	414.86	371.26	−43.60	500.16	500.45	0.29
5	528.21	483.72	−44.49	629.14	629.56	0.42
6	599.25	—	—	757.22	757.54	0.32
7	656.19	612.14	−44.05	814.34	814.34	0.00
8	793.06	749.19	−43.87	951.60	951.52	−0.08
9	850.39	806.07	−44.32	1008.45	1008.49	0.04
10	978.37	933.46	−44.91	1079.51	1079.59	0.08
11	1108.16	1063.40	−44.76	1192.67	1192.40	−0.27
12	1206.36	1162.33	−44.03	1308.66	1307.56	−1.10
13	1319.99	1276.47	−43.52	1378.14	—	—
14	1432.56	1388.61	−43.95	—	—	—

[a] Parent ion $m/z = 804.1(M+2H)^{2+}$ where M is the peptide VEAKIAGHGQEVLIR.
[b] Parent ion $m/z = 782.1([C_4H_9OOC(O)M—117–43+2H]^{2+})$, where valine ($–CH(CH_3)_2$) has $m/z = 43$.
[c] The difference between the *b*-ions corresponds to the valine side chain +1 hydrogen.

Table 5. First 6 results from SALSA search for peptides with histidine at the N-terminus in a myoglobin digest

Entry	Score	Precursor ion m/z $(2H^+)$[a]	Retention time (min)	Assigned modified peptide (^{t}BuOO(CO)-M)	Loss = 116[b]	Loss = 198[c]
1	19.25	747.75	27.01	HGTVVLTALGGILK	689.93	—
2	14.46	748.09	27.13	HGTVVLTALGGILK	689.92	648.71
3	7	966.63	27.70	GLSDGEWQQVLNVWGK	908.54	867.27
4	6.11	601.27	22.80	HLKTEAEMK	542.79	502.26
5	4.56	739.04	27.95	ALELFRNDIAAK	681.11	—
6	4.4	817.95	21.14	HPGDFGADAQGAMTK	759.87	718.79

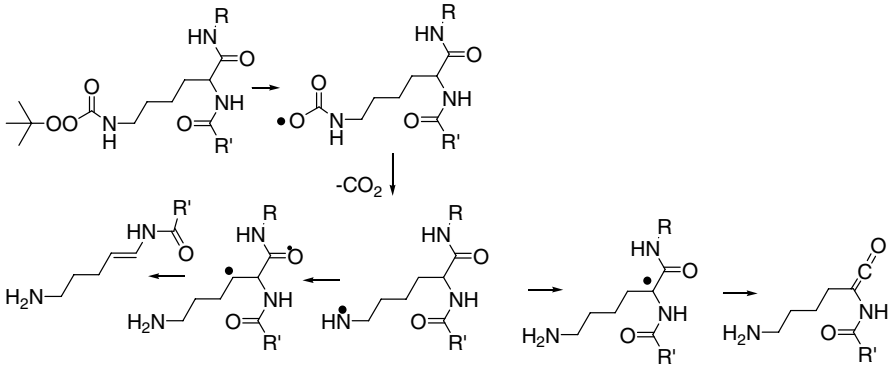
[a] Doubly charged precursor ions which lead to product ions in the MS/MS spectra matching a neutral loss of 116 (loss of $C_4H_9OOC(O)$) alone or accompanied by the neutral loss of 198 (loss of $C_4H_9OOC(O)$ +1 hydrogen –R of histidine).
[b] Doubly charged product ions corresponding to the neutral loss of 116 with respect to the precursor ion.
[c] Doubly charged product ions corresponding to the neutral loss of 198 with respect to the precursor ion.

of 116 linked to a secondary neutral loss of 198. To improve matching accuracy, we also specified doubly charged ions, since these dominate the spectra.

The first 6 of 24 results in the SALSA output for this search are shown in Table 5. Given the scoring scheme described above, the highest scoring spectra met both the primary and secondary criteria. The top score matched the m/z of modified peptide $C_4H_9OOC(O)$HGTVVLTALGGILK (entry **2** in Table 5). This result was validated by visual inspection of the MS/MS data. Other peptides bearing histidine at the N-terminus (entries **4** and **6**) exhibited low scores due to lower intensity of product ions corresponding to the scored losses. The peptide in entry **3** does not contain histidine at the N-terminus; however, this precursor ion leads to an abundant [$C_4H_9OOC(O)M$—116] product ion (neutral loss of 116), therefore it scores higher than entries **4** and **6**. It addition, this precursor ion also leads to a neutral loss of 198, which is unrelated to the N-terminal side-chain loss; however, this ion is present in low abundance. Entry **5** has a higher score than entry **6** even though the peptide does not bear histidine at the N-terminus, because this precursor ion also gives an abundant [$C_4H_9OOC(O)M$—116] product ion; however, it does not match the neutral loss of 198.

3. Discussion

The chemistry of peroxycarbamates is based on previous work from our laboratory on the derivatization of lysine residues.[16,24] In this study, α-*N*-acetyl lysine methyl ester was converted at the ε-amino group to its peroxycarbamate derivative by reaction with 4-nitrophenyl-*t*-butyl peroxycarbonate (**1a**). CID experiments on the lysine derivative as the lithium adduct yielded a unique fragmentation, which included amide bond cleavage at the C-terminus of lysine. Similar results were obtained when peptides containing one or more lysine residues were modified under the same conditions. The mechanism for this fragmentation is thought to be initiated by the dissociation of the *t*-butyl peroxy group to form an aminyl radical intermediate, followed by remote hydrogen abstraction and β-fragmentation of the resulting carbon radicals, see Scheme 2. Introduction of the labile peroxide bond provides a low-energy fragmentation pathway to free radicals that ultimately affect the peptide fragmentation patterns. Recently, Beauchamp and co-workers[25] made use of reagents that modified peptide lysines with an azo functional group. Decomposition of the labile azo linkage led to a protocol they dubbed 'Free Radical Initiated Peptide Sequencing (FRIPS)'. Again,



Scheme 2. Free radical fragmentation of lysine ε-amino peroxycarbamates.

the azo linkage provides a low-energy pathway to free radicals that affect peptide fragmentation patterns.

3.1. N-Terminal amino acid side-chain loss by CID

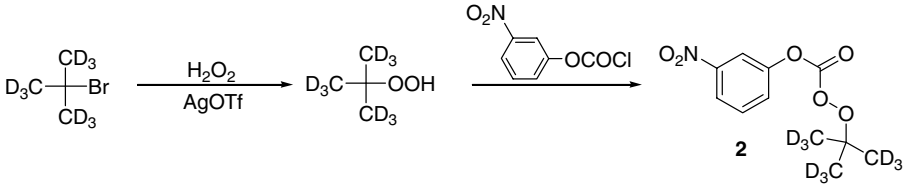
ESI-MS spectra of N-terminal peroxycarbamate modified single amino acids and peptides showed a molecular ion, $[C_4H_9OOC(O)M]$, where M is the amino acid or peptide and exhibited a fragment whose *m*/*z* is equal to the *m*/*z* of the unmodified species $[C_4H_9OOC(O)M—116]$. This suggests that gas phase decomposition of the *t*-butyl peroxycarbamate group is an important process. In most cases $[C_4H_9OOC(O)M—116]$ is the dominant fragment and it is accompanied by another fragment corresponding to the neutral loss of the N-terminal amino acid side chain $[C_4H_9OOC(O)M—117 –R]$. MS/MS of the $[C_4H_9OOC(O)M—116]$ fragment obtained in Q1 (by in-source CID) does not lead to side-chain fragmentation (data not shown). Instead, this ion generates fragment ions normally observed for the unmodified amino acid or peptide. This suggests that the $[C_4H_9OOC(O)M—116]$ fragment cannot be an intermediate in the side-chain loss, $[C_4H_9OOC(O)M—117 –R]$. Two mechanisms must be in play, with one leading to formation of M from $C_4H_9OOC(O)M$ and the other giving side-chain loss.

It is unclear how the peroxycarbamate decomposes by CID to provide the unmodified peptide or amino acid and to address this question we prepared compound **2** shown in Scheme 3. We suspect that the *t*-butyl group of the peroxycarbamate is a possible source of an H atom during the CID process leading to the peptide or amino acid fragment, $[C_4H_9OOC(O)M] \rightarrow M$. Compound **2** was prepared from the commercially available 2-bromo-2-methylpropane-d_9 by reaction with anhydrous hydrogen peroxide in the presence of silver triflate.[26] The resulting hydroperoxide was treated with 3-nitrophenyl chloroformate providing **2**. The chloroformate was prepared from 3-nitrophenol as described in Section 5.

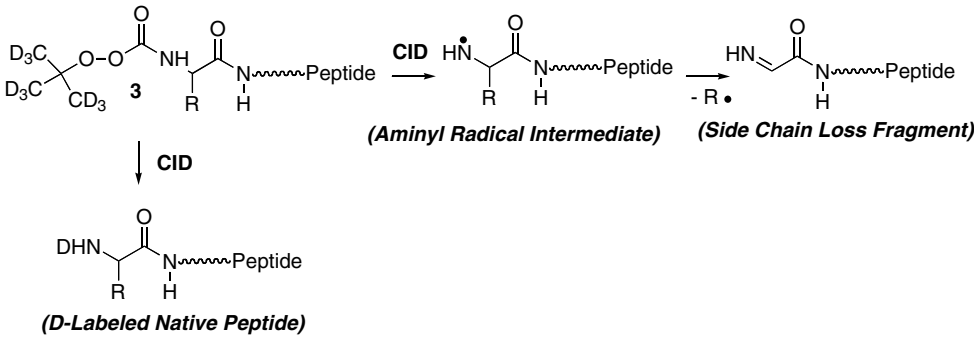
Compound **2** was then used to modify the N-terminus of the peptide WHWLQL (entry **10** in Table 2) using conditions previously described for modifications with the reagent **1b**. The deuterium labeled peroxycarbamate (C_4H_9-d_9OO(CO)M, **3** in Scheme 4) was subjected to CID and the resulting product ions included the N-terminal side-chain loss fragment, and a fragment corresponding to M+1, suggesting that deuterium transfer to the peptide attends peroxycarbamate dissociation. This indicates that the decomposition of the peroxycarbamate **3** follows a pathway in which a hydrogen (deuterium) atom is transferred from the *t*-butyl group to the α-amino of the N-terminus resulting in a CID fragment identical to the unmodified peptide (Scheme 4). See supporting information for details in the preparation of **2** and CID spectra of the modified peptide. The mechanistic details of this hydrogen (deuterium) atom transfer remain unclear, although an intermediate carbamic acid could be involved, as could either radical or ion pair species.

A reasonable mechanism for loss of R, the N-terminal amino acid side chain, proceeds through the aminyl radical, which is formed by homolysis of the weak O–O bond and loss of CO_2. Loss of R by β-fragmentation of the aminyl radical appears to be mechanistically reasonable.

Side-chain loss fragmentation was not observed for modified Pro and Cys either as single amino acid derivatives or as N-terminal amino acids of peptides. In the case of Cys the peroxycarbamate may not form or



Scheme 3. Preparation of a deuterated peroxycarbonate reagent.



Scheme 4. Proposed mechanism for dissociation of peroxycarbamate.

may not be stable due to the reactive thiol group nearby. The observation that methionine does not present this problem supports the hypothesis of the negative effect of the free thiol group. In most cases in 'real' proteomics analyses, Cys residues will be alkylated (e.g., with iodoacetamide) to produce *S*-carboxamido derivatives. These, similarly to methionine, will likely be better behaved than a free Cys. In the case of modified Pro, the N-terminus is part of a five-membered ring structure, which would prevent side-chain loss by the mechanism shown in Scheme 4. The presence of Cys or Pro anywhere else in the peptide sequence, however, did not affect the formation of the N-terminal peroxycarbamates or N-terminal side-chain cleavage. Modified Gly did not appear to yield N-terminal loss of the side chain hydrogen by CID, but instead, a fragment consistent with the *m*/*z* of the aminyl radical (Scheme 4) was observed. It is unclear why modified Asn does not exhibit side-chain loss; however it is evident that this is not a problem when Asn is at the N-terminus of a peptide.

3.2. Peptide modifications

In order to selectively modify single peptides at the N-terminus, pH and reagent for the modification reaction were altered from the conditions used previously for modification of lysine side chains.[16] Reagent **1b** is less reactive towards nucleophilic attack by the amine nitrogen than **1a**, therefore modification occurs at a slower rate with this reagent. The reaction can be accelerated by increasing the temperature to 37 °C, conditions under which reagent **1a** undergoes decomposition. Lowering the pH below the pK_a of the terminal α-amino group (approximate pK_a = 8) increases the selectivity for reaction at the N-terminus with respect to the ε-amino group of lysine (approximate pK_a = 10.7).[18] Results shown in Fig. 2 show that at pH 5.5 modification with reagent **1b** occurs mostly at the N-terminus. When modifying the myoglobin tryptic digest, reagent **1c** appeared to provide better conversion to peroxycarbamates than either **1a** or **1b**. Reagent **1c** is more reactive than **1a** and **1b**, which would suggest that this reagent would not be as selective for the N-terminus as reagent **1b**; however, both **1a** and **1b** required longer reaction times under conditions necessary for selective N-terminus modification (over 4 h). The modification reactions described here are pseudo 1st order with rate = kobs[peptide] due to the large excess reagent used.[27] The lower concentration of peptides in the digest modification reaction (μM) compared to the peptide concentrations in the modification reactions shown in Table 2 (mM) would slow the reaction rate so that a more reactive reagent is needed.

3.3. Applications of N-terminal amino acid side-chain fragmentation

Results shown in Fig. 3 illustrate the use of N-terminal side-chain loss as a quick diagnostic peptide identification method without the need for a complete sequencing experiment. These results lead the way to the use of this technique in peptide mass mapping studies, where N-terminus identification can be used to eliminate false positive matches in database searching.[28] Because MS/MS spectra of unmodified peptides lack information about the N-terminus (b_1 and y_{n-1} ions),[29] identification of N-terminal residues as described here would aid in obtaining complete sequences in de novo analyses. In addition, results in Table 3 demonstrate that gas phase cleavage of the N-terminal side chain by CID labels the N-terminus so that a subsequent CID process (MS^3) leads to a mass shift in the *b*-ions when compared to the *b*-ions from the intact peptide. This mass shift corresponds to the *m*/*z* of the N-terminal side chain, and serves to elucidate the *b*-ion series in de novo sequencing.

The analysis of a myoglobin tryptic digest demonstrates that this N-terminus identification method is feasible in a proteomics setting. The SALSA algorithm finds charged and neutral losses, a particular product ion, or a sequence motif and it does not require a peptide sequence input. This analysis identifies MS/MS spectra of possible peptides of interest, which can then be confirmed by visual inspection.

4. Conclusion

The work presented here offers an approach to improve de novo sequencing that relies on the gas-phase chemistry of chemically modified peptides. We emphasize that relatively inexpensive instrumentation such as the ESI-triple quadrupole mass spectrometer used here can be sufficient to carry out N-terminus identification analyses. The ability to distinguish two isobaric peptides in a mixture by the N-termini without a com-

plete sequence has the potential to be applied in peptide mass mapping analyses. Studies on the N-terminal amino acid side-chain fragmentation induced by MALDI are currently underway and will be presented in the near future.

5. Experimental

5.1. Materials

Amino acids were purchased from Aldrich Chemical Co, Inc. (Milwaukee, MA) as either methyl, ethyl or *t*-butyl ester derivatives in their hydrochloride salt forms. Peptides were purchased from American Peptide Company (Sunnyvale, CA). Solvents were of HPLC grade, obtained from Fisher Scientific (Pittsburgh, PA). All other reagents were purchased from Aldrich Chemical Co, Inc. (Milwaukee, MA). Acetate buffers (10 mM) of various pH values were prepared from ammonium acetate titrated to pH 5.2, 5.5, and 5.8 with acetic acid. Tris buffers were 0.1 M solutions titrated to pH 7.5 and pH 8, respectively. Ammonium bicarbonate buffer was a 0.1 M solution at pH 8.6. Buffer systems used to modify peptides and amino acids consisted of a 1:1 mixture of the appropriate buffer and acetonitrile. The myoglobin tryptic digest standard mixture (500 pmol) was purchased from Alltech-Life Sciences (Deerfield, IL).

Mass spectrometry analyses of modified single amino acid derivatives and peptides were performed on a ThermoFinnigan (San Jose, CA) TSQ 7000 triple-quadrupole instrument, equipped with an Electrospray Ionization (ESI) source and a Waters (Milford, MA) Alliance HPLC system. The capillary temperature was kept at 200 °C with a voltage of 20 V. The electrospray needle voltage was 4.5 kV, and the tube-lens voltage was maintained between 70 and 100 V. The sheath and auxiliary gases (N_2) were adjusted to maximize signal. The offset voltage in CID experiments varied between 20 and 35 eV. Samples were introduced into the ESI source as either lithium solutions or acidic solutions at a rate of 10–20 μl/min in direct liquid infusion experiments. For LC–MS analyses of peptides, a Supelco (Bellefonte, PA) Discovery C18 column was used (25 cm × 4.6 mm, 5 μm). The solvent conditions consisted of a 20 min gradient from 5% to 50% solvent B (95% acetonitrile, 5% water, and 0.05% TFA) into solvent A (95% water, 5% acetonitrile, and 0.05% TFA). LC–MS/MS analyses of peptide derivatives were performed on a ThermoFinnigan (San Jose, CA) LCQ Deca XP ion trap instrument, equipped with an ESI source, and a HP Agilent (Palo Alto, CA) HPLC system. Instrument parameters were tuned using angiotensin II peptide (5 pmol) to maximize signal. LC–MS/MS analyses of the myoglobin digests were performed using a Grace Vydac (Hesperia, CA) C18 column (10 cm × 1 mm × 5 μm). Solvent conditions consisted of a 30 min gradient from 100% solvent A (5% acetonitrile, 95% water, and 0.5% acetic acid) to 100% solvent B (5% water, 95% acetonitrile, and 0.5% acetic acid).

5.2. Modification reagents

Compounds **1a**–**1c** (Scheme 1) were synthesized according to previous work in our laboratory[16] starting from 4-nitrophenylchloroformate, 3-nitrophenylchloroformate, and 2-chloro-4-nitrophenylchloroformate, respectively. The two chloroformates preceding **1b** and **1c** were prepared from their respective phenols by treatment with one equivalent of triethylamine followed by two equivalents of phosgene in methylene chloride at 0 °C under nitrogen. All compounds were characterized by ^1H and ^{13}C NMR (see supporting information).

5.3. Modification reactions

Reactions with all 20 amino acids as methyl, ethyl, or *t*-butyl ester derivatives were carried out by combining 5 μl of amino acid solution (0.1 M in 50/50 acetonitrile and water), 20 μl of ammonium bicarbonate buffer system (50/50 acetonitrile/buffer), and 20 μl of 1a (0.1 M in anhydrous acetonitrile). Final concentrations for amino acids and reagent were 0.01 M, and 0.04 M, respectively. Reactions were incubated at room temperature overnight. The crude reaction mixtures were diluted to 1 ml with a 0.4 mM solution of LiCl in methanol for MS analysis.

The peptide YRVRFLAKENVTQDAEDNC (entry **21** in Table 2) was modified at pH range 5.2–8.6 by addition of 10 μl of buffer system (50/50 acetonitrile/buffer) to 10 μl of peptide solution (1 mg/ml in water), followed by 5 μl of **1a** or **1b** (0.1 M in acetonitrile). Final concentrations of peptide and reagent were 0.7 μM and 33 μM, respectively. The reaction mixture was incubated at 37 °C for 1 h. The crude mixtures were analyzed by reverse phase LC–MS to determine selectivity for N-terminus modification. A series of peptides bearing each of the 20 amino acids at the N-terminus (Table 2) were modified using the conditions which gave best selectivity for N-terminus modification (pH 5.5, buffer system, reagent **1b**). All peptide reaction mixtures were diluted to 1 ml with a 0.1% acetic acid solution and analyzed by MS/MS as direct liquid infusions.

5.4. Analysis of modified myoglobin tryptic digest

Modification of a myoglobin tryptic digest standard was achieved by combining 50 μl of digest solution (500 pmol in 200 μl of water), 50 μl of acetate buffer system (pH 5.5), and 25 μl of reagent **1c** (0.1 M in acetonitrile). The reaction mixture was incubated at room temperature for 2 h and analyzed by reverse phase LC–MS. MS/MS data were analyzed with the SALSA algorithm[17] to search for modified peptides and neutral losses of N-terminal amino acid side chains.

Acknowledgments

We thank the National Science Foundation, the Vanderbilt Center for Research in Toxicology, NIEHS P01 ES013125 and Vanderbilt University for financial support. We also thank Vanderbilt's Mass Spectrometry

Research Center, in particular Dr. David Hachey, Lisa Manier, and Dawn Overstreet, for advice in use of instrumentation. In addition, we thank the Proteomics Research Facility, in particular Dr. Amy Ham and Dr. David Freeman, for their advice with the myoglobin tryptic digest analysis.

Supplementary data

Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.bmc.2006.05.060.

References and notes

1. Liebler, Daniel C. *Introduction to Proteomics: Tools for the New Biology*; Humana Press: Totowa, NJ, 2002, p. 198.
2. Reinders, J.; Lewandrowski, U.; Moebius, J.; Wagner, Y.; Sickmann, A. *Proteomics* **2004**, *4*, 3686–3703.
3. Taylor, J. A.; Johnson, R. S. *Anal. Chem.* **2001**, *73*, 2594–2604.
4. Ma, B.; Zhang, K.; Lajoie, G.; Doherty-Kirby, A.; Hendrie, C.; Liang, C.; Li, M. *Rapid Commun. Mass Spectrom.* **2003**, *17*, 2337–2342.
5. Frank, A.; Pevzner, P. *Anal. Chem.* **2005**, *77*, 964–973.
6. Roth, K. D. W.; Huang, Z. H.; Sadagopan, N.; Watson, J. T. *Mass Spectrom. Rev.* **1998**, *17*, 255–274.
7. Barofsky, D. F.; Chen, T. F. Peptide Sequencing Promoted by N-Terminal Derivatization. In *Mass Spectrometry and Hyphenated Techniques in Neuropeptide Research*; Silberring, J., Ekman, R., Eds.; John Wiley, 2002; pp 345–373.
8. Keough, T.; Youngquist, R. S.; Lacey, M. P. *Proc. Natl. Acad. Sci. U.S.A.* **1999**, *96*, 7131–7136.
9. Keough, T.; Youngquist, R. S.; Lacey, M. P. *Anal. Chem.* **2003**, *75*, 156A–165A.
10. Summerfield, S. G.; Steen, H.; O'Malley, M.; Gaskell, S. J. *Int. J. Mass Spectrom.* **1999**, *188*, 95–103.
11. Münchbach, M.; Quadroni, M.; Giovanni, M.; James, P. *Anal. Chem.* **2000**, *72*, 4047–4057.
12. Beardsley, R. L.; Reilly, J. P. *J. Am. Soc. Mass Spectrom.* **2004**, *14*, 158–167.
13. Goodlett, D. R.; Keller, A.; Watts, J. D.; Newitt, R.; Yi, E. C.; Purvine, S.; Eng, J. K.; Haller, P.; Aebersold, R.; Kolker, E. *Rapid Commun. Mass Spectrom.* **2001**, *15*, 1221–1224.
14. Quin, J.; Herring, C. J.; Zhang, X. *Rapid Commun. Mass Spectrom.* **1998**, *12*, 209–216.
15. Cannon, R.; Jarman, K. D. *Rapid Commun. Mass Spectrom.* **2003**, *17*, 1793–1801.
16. Masterson, D. S.; Yin, H.; Chacon, A.; Hachey, D. L.; Norris, J. L.; Porter, N. A. *J. Am. Chem. Soc.* **2003**, *126*, 720–721.
17. Hansen, B. T.; Jones, J. A.; Mason, D. E.; Liebler, D. C. *Anal. Chem.* **2001**, *73*, 1676–1683.
18. Stark, G. R. *Biochemistry* **1965**, *4*, 1030–1036.
19. Desbuquois, B. *Eur. J. Biochem.* **1975**, *60*, 335–347.
20. Bradbury, H. J.; Howell, J. R.; Johnson, R. N.; Warren, B. *Eur. J. Biochem.* **1978**, *84*, 503–511.
21. Wetzel, R.; Hlualani, R.; Stults, J. T.; Quan, C. *Bioconjugate Chem.* **1990**, *1*, 114–122.
22. Yates, J. R.; Eng, J. k.; McCormack, A. L. *Anal. Chem.* **1995**, *67*, 3202–3210.
23. Hansen, B. T.; Davey, S. W.; Ham, A. L.; Liebler, D. C. *J. Proteome Res.* **2005**, *4*, 358–368.
24. Porter, N. A.; Masterson, D. S.; Yin, H.; Caprioli, R. M.; Norris, J. L.; Chacon, A. Peptide and Protein Fragmentation by Lysine Residue Originated Reactions. WO2005037857, 2005.
25. Hodyss, R.; Cox, H. A.; Beauchamp, J. L. *J. Am. Chem. Soc.* **2005**, *127*, 12436–12437.
26. Milas, N. A.; Harris, A. S. *J. Am. Chem. Soc.* **1938**, *60*, 2434–2436.
27. Thomas, H. L.; Richardson, K. S. *Mechanism and Theory in Organic Chemistry*, 3rd ed.; Harper Collins Publishers: New York, 1987, p. 1090.
28. Brancia, F. L.; Butt, A.; Beynon, R. J.; Hubbard, S. J.; Gaskell, S. J.; Oliver, S. G. *Electrophoresis* **2001**, *22*, 552–559.
29. Reid, E. G.; Tichy, S. E.; Perez, J.; O'Hair, R. A. J.; Simpson, R. J.; Kenttamaa, H. I. *J. Am. Chem. Soc.* **2001**, *123*, 1184–1192.