

In Vitro Modification of Human Centromere Protein CENP-C Fragments by Small Ubiquitin-like Modifier (SUMO) Protein

DEFINITIVE IDENTIFICATION OF THE MODIFICATION SITES BY TANDEM MASS SPECTROMETRY ANALYSIS OF THE ISOPEPTIDES*[§]

Received for publication, May 20, 2004, and in revised form, July 20, 2004
Published, JBC Papers in Press, July 22, 2004, DOI 10.1074/jbc.M405637200

Tung-Liang Chung^{‡§¶}, He-Hsuan Hsiao[¶], Yuh-Ying Yeh[¶], Hui-Ling Shia^{**}, Yi-Ling Chen^{**},
Po-Huang Liang[§], Andrew H.-J. Wang[§], Kay-Hooi Khoo^{§¶‡‡}, and Steven Shoen-Lung Li^{‡§§}

From the [‡]Department of Biotechnology, College of Life Sciences, Kaohsiung Medical University, Kaohsiung 807, the [¶]Core Facilities for Proteomics Research at Academia Sinica, Taipei 11529, the ^{**}Institute of Biomedical Sciences, National Sun Yat-Sen University, Kaohsiung 804, and the [§]Institute of Biological Chemistry, Academia Sinica, Taipei 11529, Taiwan

Protein sumoylation by small ubiquitin-like modifier (SUMO) proteins is an important post-translational regulatory modification. A role in the control of chromosome dynamics was first suggested when SUMO was identified as high-copy suppressor of the centromere protein CENP-C mutants. CENP-C itself contains a consensus sumoylation sequence motif that partially overlaps with its DNA binding and centromere localization domain. To ascertain whether CENP-C can be sumoylated, tandem mass spectrometry (MS) based strategy was developed for high sensitivity identification and sequencing of sumoylated isopeptides present among in-gel-digested tryptic peptides of SDS-PAGE fractionated target proteins. Without a predisposition to searching for the expected isopeptides based on calculated molecular mass and relying instead on the characteristic MS/MS fragmentation pattern to identify sumoylation, we demonstrate that several other lysine residues located not within the perfect consensus sumoylation motif ψ KXE/D, where ψ represents a large hydrophobic amino acid, and X represents any amino acid, can be sumoylated with a reconstituted *in vitro* system containing only the SUMO proteins, E1-activating enzyme and E2-conjugating enzyme (Ubc9). In all cases, target sites that can be sumoylated by SUMO-2 were shown to be equally susceptible to SUMO-1 attachments which include specific sites on SUMO-2 itself, Ubc9, and the recombinant CENP-C fragments. Two non-consensus sites on one of the CENP-C fragments were found to be sumoylated in addition to the predicted site on the other fragment. The developed methodologies should facilitate future studies in delineating the dynamics and substrate specificities of SUMO-1/2/3 modifications and the respective roles of E3 ligases in the process.

Protein modification by covalent attachment of SUMO¹ (small ubiquitin-like modifier) proteins is emerging as an important regulatory mechanism in a diverse range of cellular processes (1, 2). Through a three-step enzymatic pathway analogous to ubiquitylation, sumoylation is initiated by ATP-dependent formation of a thioester bond between the C-terminal glycine of a SUMO protein and a catalytic cysteine of the SUMO-specific E1-activating enzyme, known as SAE1/SAE2 and Uba2p/Aos1p for human and yeast, respectively. The activated SUMO protein is then trans-esterified from the E1 enzyme to the catalytic cysteine of an E2-conjugating enzyme, Ubc9, which in turn catalyzes the formation of an isopeptide bond between the C-terminal glycine of SUMO and the ϵ -amino group of lysine residue in the substrate proteins. Ubc9 was reported to recognize and bind directly to a consensus sumoylation motif ψ KXE/D, where ψ represents a large hydrophobic amino acid L, I, V, or F, and X represents any amino acid (3–5). Although sumoylation can be accomplished *in vitro* without an E3-like enzyme, several E3 ligases, *e.g.* PIAS, RanBP2, and PC2, have been identified which may aid Ubc9 in substrate selection and ligation efficiency *in vivo* (2, 6).

Unlike yeast and other invertebrates, which contain only a single SUMO gene, vertebrates carry three. Human SUMO-1 protein exhibits 44% sequence identity with human SUMO-2 and SUMO-3 proteins, while SUMO-2 and SUMO-3 proteins share 86% sequence identity. All SUMO proteins from yeast to human share the conserved ubiquitin domain and the C-terminal diglycine cleavage/attachment site (3, 7). The most prominent difference between the SUMO proteins and ubiquitin is the presence of a highly variable N-terminal extension in all SUMO proteins, which is rich in charged amino acids, glycines, and prolines. This extension varies from 16 to 23 amino acids and is reasonably well conserved within, but not between, different SUMO-1/2/3 proteins. At present, human SUMO-1 and SUMO-2/3 are thought to be functionally non-overlapping and respond differently to stress signaling (2). Although all three SUMO isoforms utilize the same E1-activating and E2-conjugating enzymes, the molecular basis for their substrates preference and the additional recruitment of E3 ligases, especially in relation to the selective use of sumoylation sites conforming and not conforming to the minimal consensus motif, remain unclear. SUMO-2/3, like yeast SUMO (Smt3), but not

* This work was supported by National Science Council Grants NSC92-3112-B-110-001 and NSC93-3112-B-110-001 (to S. S.-L. L.) and NSC92-3112-B-001-006-Y and NSC93-3112-B-001-010-Y to the Core Facilities and in part by Academia Sinica (to A. H.-J. W. and K.-H. K.). The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked "advertisement" in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

[§] The on-line version of this paper (available at <http://www.jbc.org>) contains supplemental Figs. S1 and S2.

[¶] These two authors contributed equally to this work.

^{‡‡} To whom correspondence may be addressed. Tel.: 886-2-2785-5696; Fax: 886-2-2788-2043; E-mail: kkhoo@gate.sinica.edu.tw (for K.-H. K.).

^{§§} To whom correspondence may be addressed. Tel.: 886-7-313-5162; Fax: 886-7-312-5339; E-mail: lissl@kmu.edu.tw (for S. S.-L. L.).

¹ The abbreviations used are: SUMO, small ubiquitin-like modifier; aa, amino acid; CID, collision-induced dissociation; ESI, electrospray ionization; MALDI, matrix-assisted laser desorption ionization; MS, mass spectrometry; MS/MS, tandem MS; nanoESI, nanospray; nanoLC, nanoflow liquid chromatography.

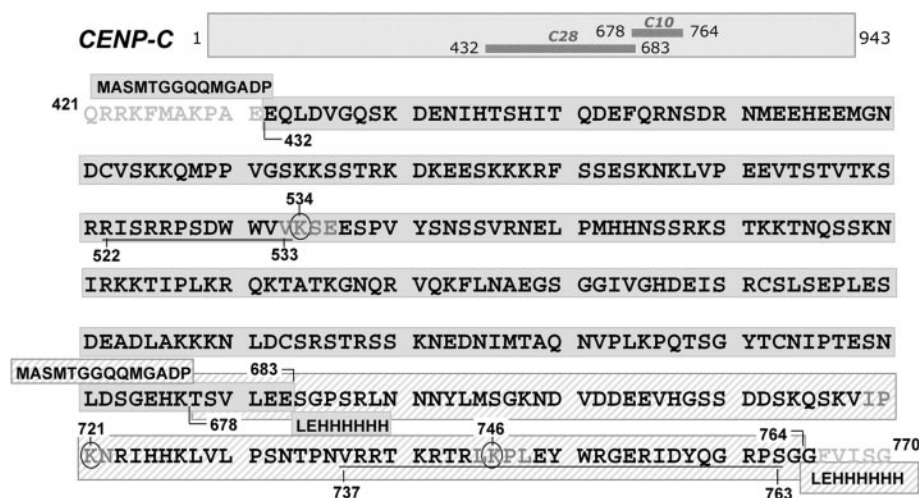


FIG. 1. **Partial amino acid sequences of human centromere protein CENP-C.** Both C28 (aa 432–683) and C10 (aa 678–764) fragments used in this study were constructed with the leader sequence MASMTGGQMGADP and His-tag sequence LEHHHHHH as indicated, which were subsequently cleaved off prior to *in vitro* sumoylation. Two distinct regions have been identified for centromere targeting, *i.e.* a central region around aa 522–534 (*underlined*) (18, 20) and another region at the C terminus (aa 690–943) (17, 20), which includes the Mif2 homology domain (aa 737–763, *underlined*) (14, 16). The implicated DNA-binding domains of CENP-C within the sequence shown include aa 330–551 (19), aa 433–520 (19), and aa 426–537 and 638–934 (21). The sumoylation sites identified in this study are Lys⁵³⁴ located within the consensus site VKSE (aa 533–536), and the non-consensus sites at Lys⁷²¹ and Lys⁷⁴⁶, as *circled*.

SUMO-1, contain a consensus ψ KXE sumoylation site in their N-terminal extension. Accordingly, oligomerization of SUMO-2/3 or Smt3 chains can be demonstrated *in vivo* and effected with the reconstituted *in vitro* sumoylation system in the absence of E3 (8, 9). In contrast, formation of SUMO-1 chains with as yet unknown linkage apparently can only be demonstrated in the presence of an E3 ligase activity (10).

Sumoylation is a dynamic, reversible process, and often only little or no modified protein can be detected under physiological conditions (1). It should be noted that direct biochemical evidence for the implicated isopeptide or the sumoylation site has been generally lacking for most reported sumoylation cases due to its low abundance. To date, only a handful of isopeptides have been physically defined by mass spectrometry (MS) analysis (8, 9, 11, 12). The isopeptide formation and hence the sumoylation site were directly inferred from the molecular mass detected but in general not further verified by MS/MS sequencing analysis. This approach is problematic in view of the increasing probability in finding a non-consensus sumoylation site nor can it distinguish the exact sumoylated lysine residue in the event that the implicated endoprotease-digested peptide carries more than one lysine residues. To better define the sumoylation site specificity and to delineate the respective roles of SUMO-1 *versus* SUMO-2/3, a high sensitivity MS/MS sequencing strategy for unambiguous identification of the isopeptide and sumoylation site is critically needed. By adopting an *in vitro* sumoylation system with defined components, we have thus undertaken to first develop the enabling analytical techniques while gearing toward addressing the role of sumoylation in regulating chromosome dynamics.

The centromere is a highly ordered structure of the eukaryotic chromosomes which provides the site for microtubule (spindle) attachment during cell division (13). The first implication of sumoylation in centromere functions came from the original isolation of the yeast SUMO (Smt3) as a high copy suppressor of a mutation in the Mif2 gene, the yeast homologue of the gene for mammalian centromere protein CENP-C (14). CENP-C is an essential component of the kinetochore inner plate, which contributes to the formation of functional centromeres for correct chromosome segregation. Temperature-sensitive CENP-C mutants in vertebrate cells could likewise be suppressed by overexpression of SUMO-1 (15). Despite several studies, which

demonstrated the localization of SUMO at or adjacent to the kinetochore and how sumoylation may affect centromeric chromosome cohesion (reviewed in Refs. 1 and 2), the mechanistic details remain largely unknown nor is it clear whether critical centromeric proteins like CENP-C are themselves sumoylated. Molecular dissection of CENP-C revealed that the protein is composed of several functional domains (16–21) including a specific region of 12 amino acids at aa 522–533 that could confer DNA binding as well as centromere targeting functions (Fig. 1). Notably, its partial overlapping with a sumoylation consensus sequence of VKSE at aa 533–536 raises an intriguing possibility that regulated sumoylation of CENP-C may affect its functions.

Using the methodologies developed, which were first validated against SUMO-2/SUMO-2 and SUMO-2/Ubc9 conjugation, we report here that a C28 fragment of CENP-C encoding aa 432–683 (Fig. 1), which includes the consensus sumoylation site, DNA binding and centromere localization regions, and two nuclear localization signals, could be *in vitro* sumoylated by both SUMO-1 and SUMO-2 at the expected site in the absence of any E3 ligases. In addition, another C10 fragment of CENP-C encoding aa 678–764, which comprises a Mif2 homology block at aa 737–759, could also be sumoylated *in vitro* by both SUMO-1 and SUMO-2 at two distinct sites, neither of which corresponds to the consensus sequence motif. We further demonstrate that SUMO-1 could be attached to SUMO-2 and the characteristic MS/MS fragmentation pattern associated with SUMO-2 isopeptides could be utilized as an effective fingerprint in identifying sumoylation site at high sensitivity.

EXPERIMENTAL PROCEDURES

cDNA Cloning, Protein Expression, and Purification—The C28 cDNA encoding aa 432–683 of human CENP-C was amplified by reverse transcriptase-PCR using poly(A)-containing RNAs from HeLa cells, as well as primers of 5'-CGGATCCAGAACACCTTGATGTGGGAC-3' and 5'-GTGCTCCAGTTCCTCTAAAAGT-3', and then cloned into the BamHI-XhoI sites of bacterial expression vector pET-21b (His-tag at the C terminus). The C10 cDNA encoding aa 678–764 of human CENP-C was also amplified using primers of 5'-CGGATCCACTTCA-GTTTATAGAGGAA-3' and 5'-GTGCTCCAGTTCCTGATGGCCTTCCTT-GATA-3' and then cloned into the BamHI-XhoI sites of pET-21b. The human cDNA encoding the active form (exposing diglycine motif) of SUMO-1GG was obtained by PCR from the FLAG-SUMO-1 cDNA described previously (22), and the cDNA encoding the active form of

SUMO-2GG was obtained by PCR from the hSMT3 cDNA reported previously (23). Both SUMO-1GG and SUMO-2GG cDNAs were subsequently cloned into pET-32 Xa (both His-tag and S-tag at the N terminus). The human Ubc9 cDNA was amplified by reverse transcriptase-PCR using poly(A)-containing RNAs from HeLa cells and then cloned into pET-21b and pET-32 Xa. The *Escherichia coli* BL21 strain was used as host cells for protein expression. His-tag proteins were purified using nickel-nitrilotriacetic acid-agarose column (Qiagen). The purified proteins from pET-32 Xa were cleaved with Factor Xa to remove both His-tag and S-tag.

In Vitro Sumoylation Assay—The *in vitro* sumoylation system employed in this investigation contains E1 enzyme (159 ng of SAE1/SAE2, LAE Biotechnology, Taichung, Taiwan), E2 enzyme (2 μ g Ubc9), SUMO-1GG or SUMO-2GG protein (5 μ g), and C28-His or C10-His protein substrate (5 μ g) in sumoylation buffer (2 μ l of 500 mM Tris, pH 7.5, 1 μ l of 100 mM MgCl₂, 1 μ l of 100 mM ATP, 0.5 μ l of 40 mM dithiothreitol, an ATP-regenerating system, including 0.5 μ l of 400 mM creatine phosphate, 0.5 μ l of 140 units/ml creatine kinase, 0.5 μ l of 24 units/ml inorganic pyrophosphatase). The sumoylation reactions were incubated at 37 °C for 3 h. After termination with SDS buffer containing 1 mM dithiothreitol, reaction products were separated on 12% SDS-PAGE, and the gels were stained with Coomassie Blue.

Identification of Sumoylation Sites by Tandem Mass Spectrometry Analyses—Following SDS-PAGE, Coomassie Blue-stained protein bands were excised from gels, reduced and alkylated with iodoacetamide, and in-gel-digested with sequencing grade, modified trypsin (Promega, Madison, WI). Peptides were extracted and subjected first to nanoLC-nanoESI-MS/MS analysis for protein identification as described previously (24). Subsequently, MALDI-MS detection and MS/MS sequencing of isopeptides in reflectron mode were performed on an Applied Biosystems 4700 Proteomics Analyzer mass spectrometer (Applied Biosystems, Framingham, MA) equipped with an Nd:YAG laser (355 nm wavelength, <500-ps pulse, and 200 Hz repetition rate in both MS and MS/MS modes). 1000 and 10,000 shots were accumulated in positive ion mode MS and MS/MS modes, respectively. The tryptic digested peptide samples were dissolved in 50% acetonitrile with 0.1% formic acids and premixed with a 5 mg/ml matrix solution of α -cyano-4-hydroxycinnamic acid in 70% acetonitrile with 0.1% formic acid for spotting onto target plate. For collision-induced dissociation (CID) MS/MS operation, the indicated collision cell pressure was increased from 3.0×10^{-8} torr (no collision gas) to 5.0×10^{-7} torr, with the potential difference between the source acceleration voltage and the collision cell set at 1 kV. The resolution of timed ion selector for precursor ion was set at 100, which would allow in a mass window of about 50 Da for precursors at m/z 5000. Both MS and MS/MS data were acquired using the instrument default calibration. At a resolution above 10,000 in MS mode, accurate mass measurement (<50 ppm) of the monoisotopic isopeptide signals is possible when further adjusted against an internal reference peak at m/z 5557.8 derived from autolytic cleavage of the modified trypsin.

RESULTS

Sumoylation of SUMO-2 and Ubc9—The *in vitro* sumoylation system employed in this investigation contained E1 (SAE1/SAE2)-activating enzyme, E2 (Ubc9)-conjugating enzyme, the active SUMO proteins (SUMO-1GG or SUMO-2GG), with and without the centromere proteins (C28-His or C10-His), and an ATP-generating system, including creatine kinase. To better define the sumoylation reaction products as visualized by SDS-PAGE analysis, reactions of the active SUMO proteins with the E1/E2 enzymes were first investigated prior to the addition of the substrate proteins. Fig. 2 shows that in the absence of E1 enzyme, no reaction product could be detected (lanes 2 and 4), but one (lane 1) and two (lane 3) additional protein bands were clearly visible when E1 enzyme was added to the SUMO-1GG/E2 and SUMO-2GG/E2 reaction mixtures, respectively. By LC-ESI-MS/MS analysis of the respective in-gel tryptic digests, bands 2 and 3 were found to contain peptides derived from SUMO-1 and SUMO-2, respectively, in addition to those from Ubc9 (data not shown). Band 1 on the other hand only afforded peptides matching to SUMO-2. Since the tagged SUMO proteins and Ubc9 alone ran as protein bands close to 20 kDa (data not shown), the results indicated that the SUMO proteins can be stably conjugated to

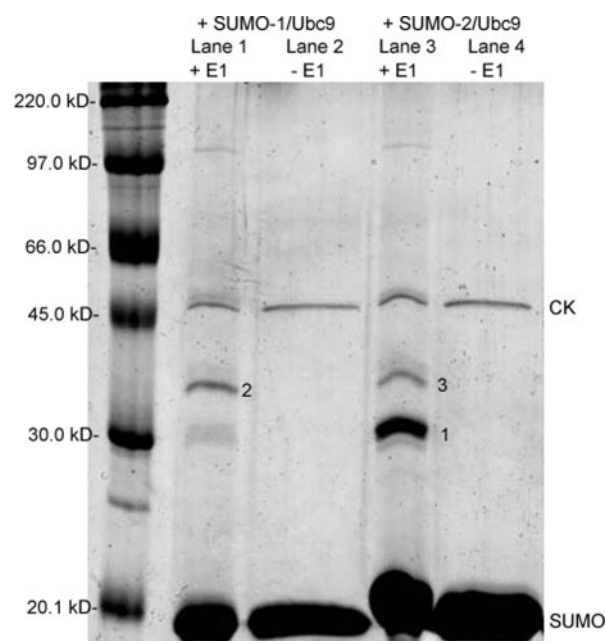


Fig. 2. SDS-PAGE analysis of sumoylation reaction products. The *in vitro* sumoylation reactions contained the E2 (Ubc9) and SUMO-1GG/2GG with and without the E1 enzyme, as well as creatine kinase (CK). The gel was stained with Coomassie Blue for visualization of the protein bands.

Ubc9 and ran as bands 2 and 3 and that only SUMO-2, but not SUMO-1, can itself be sumoylated to give a higher molecular weight product (band 1).

MS/MS Characteristics of SUMO-2-conjugated Isopeptides—The monoisotopic mass of the SUMO-2-conjugated SUMO-2 isopeptide, in which the Gly⁹³ of C-terminal tryptic peptide of aa 62–93 from one SUMO-2 was conjugated to the sumoylation site at Lys¹¹ within the tryptic peptide of aa 8–21 from another SUMO-2, was calculated to be 5159.34. By nanoESI-MS analysis on a quadrupole/time-of-flight instrument of sufficiently high mass accuracy and resolution, a pair of well resolved, quadruply charged isotopic signal clusters could be detected with the respective monoisotopic molecular ions $[M+4H]^{4+}$ occurring at m/z 1290.848 and 1294.847 (Fig. 3A, inset). The former corresponds exactly (within 10 ppm) to the expected isopeptide, while the latter at 16 mass units higher most likely was derived from the same peptide but containing an oxidized Met⁷⁸ residue. The relative ratio of these two peaks varied from batch to batch, implying different degree of induced oxidation during sample preparation. To confirm the isopeptide linkage assignment predicted by previous MALDI-TOF MS study (8), the isopeptides detected here as singly charged $[M+H]^+$ molecular ions on a MALDI-TOF/TOF instrument were selected as precursors for MALDI CID-MS/MS analysis (Fig. 3). Two well known fragmentation characteristics facilitated the spectral assignment. First, peptide fragmentation by MALDI-MS/MS is often dominated by facile cleavages at C terminus of an Asp residue (25). For tryptic peptides where the positive charge tends to be localized to C-terminal Lys or Arg residue, the MS/MS spectra will be dominated by a series of y ions, the most abundant of which are derived from cleavages at Asp. In the case of the SUMO-2 conjugated SUMO-2 isopeptide carrying an oxidized Met (m/z 5180.6, Fig. 3A), the major ions at m/z 624.5 (y_5), 2456.5 (y_{22}), 2785.7 (y_{25}), 3029.7 (y_{27}), 4060.1 (y_{36}), and 4915.3 (y_{44}) effectively and unambiguously confirmed the peptide sequence. Other y ions, y_5 – y_{10} , followed by the apparent absence of y_{11} – y_{13} and the reappearance of the y ion series at y_{14} clearly define the SUMO-2 sumoylation site at Lys¹¹.

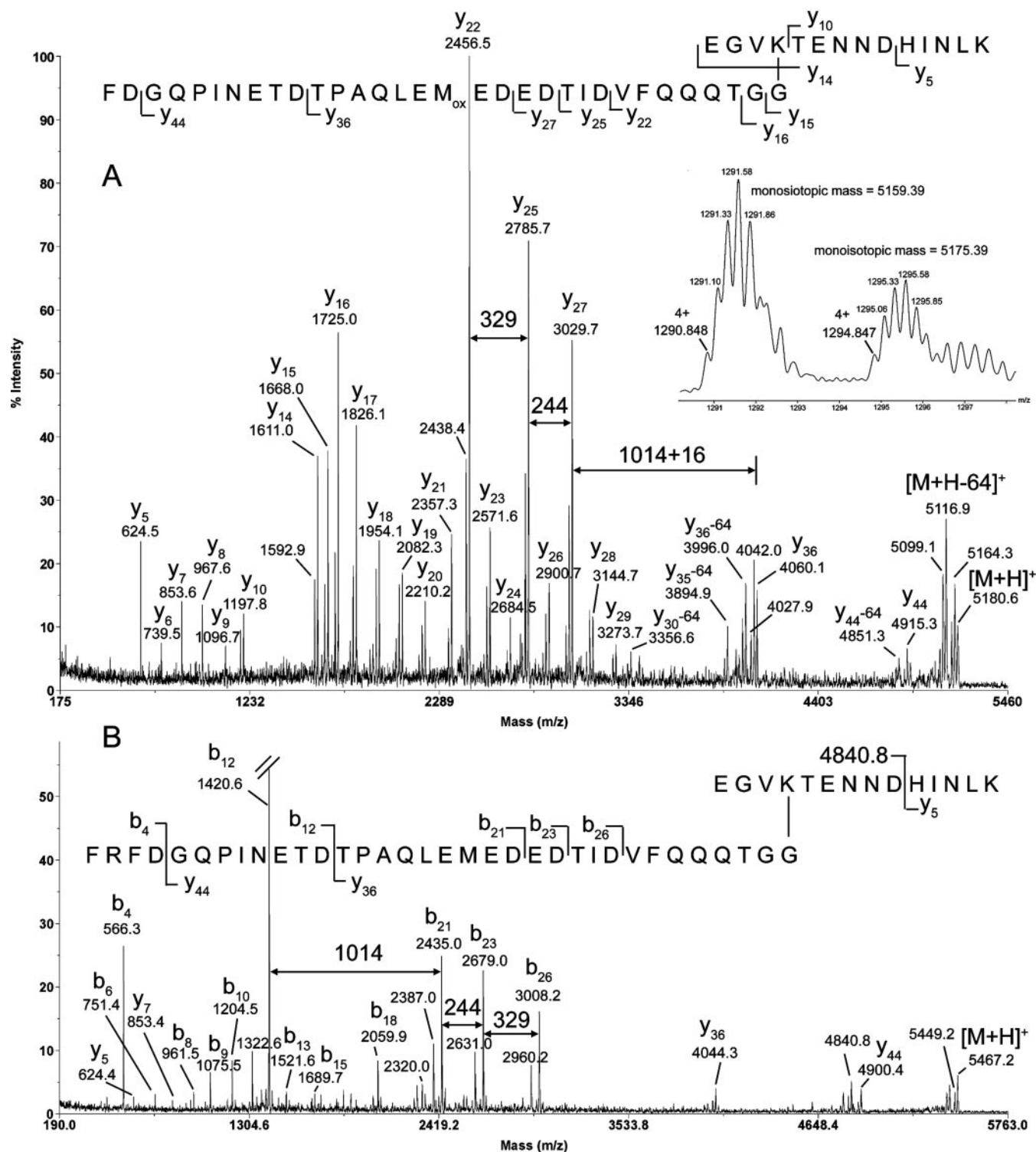


FIG. 3. MALDI-MS/MS spectra of the SUMO-2-conjugated SUMO-2 isopeptides carrying an oxidized Met⁷⁸ residue (A) or a trypsin missed cleavage site at Arg⁶¹ (B). In A, the majority of the fragment ions could be assigned to the y ion series as annotated, which were commonly accompanied by neutral loss of H₂O (18 mass units) giving rise to pairs of signals. Ions that carry the oxidized Met also afforded neutral loss of HSOCH₃ (64 mass units). The quadruply charged molecular ion signals corresponding to the isopeptides containing a non-oxidized and oxidized Met as detected by nanoESI-MS analysis are shown as the inset in A. The spectrum in B is normalized to the intensity of the b₁₂ ion signal, but only the portion below 50% intensity is shown. Localization of an Arg residue near the N terminus contributed to the predominance of the b ion series except for a couple of y ions as annotated. Neutral loss of H₂O (18 mass units) is also evident for the b₂₁, b₂₃, and b₂₆ ions. The labeled *m/z* values for this and other mass spectra figures mostly correspond to monoisotopic accurate masses as originally annotated by the instrument except for signals at above *m/z* 3500, especially those of relatively weak intensity, which may not afford full isotopic resolution and may thus be labeled according to the detected peak top, usually at several mass units higher than the theoretical monoisotopic masses.

Second, consistent with the presence of an oxidized Met residue, the major fragment ions y₃₆ and y₄₄ and the molecular ion that contains the oxidized Met were accompanied by facile neutral loss of HSOCH₃ (64 mass units) (26, 27), giving rise to

ions at *m/z* 3996.0, 4951.3, and 5116.9, respectively. These neutral losses were not observed in the MS/MS spectra of isopeptides with non-oxidized Met, e.g. that of the monoisotopic [M+H]⁺ molecular ion signal at *m/z* 5164.3 (data not shown).

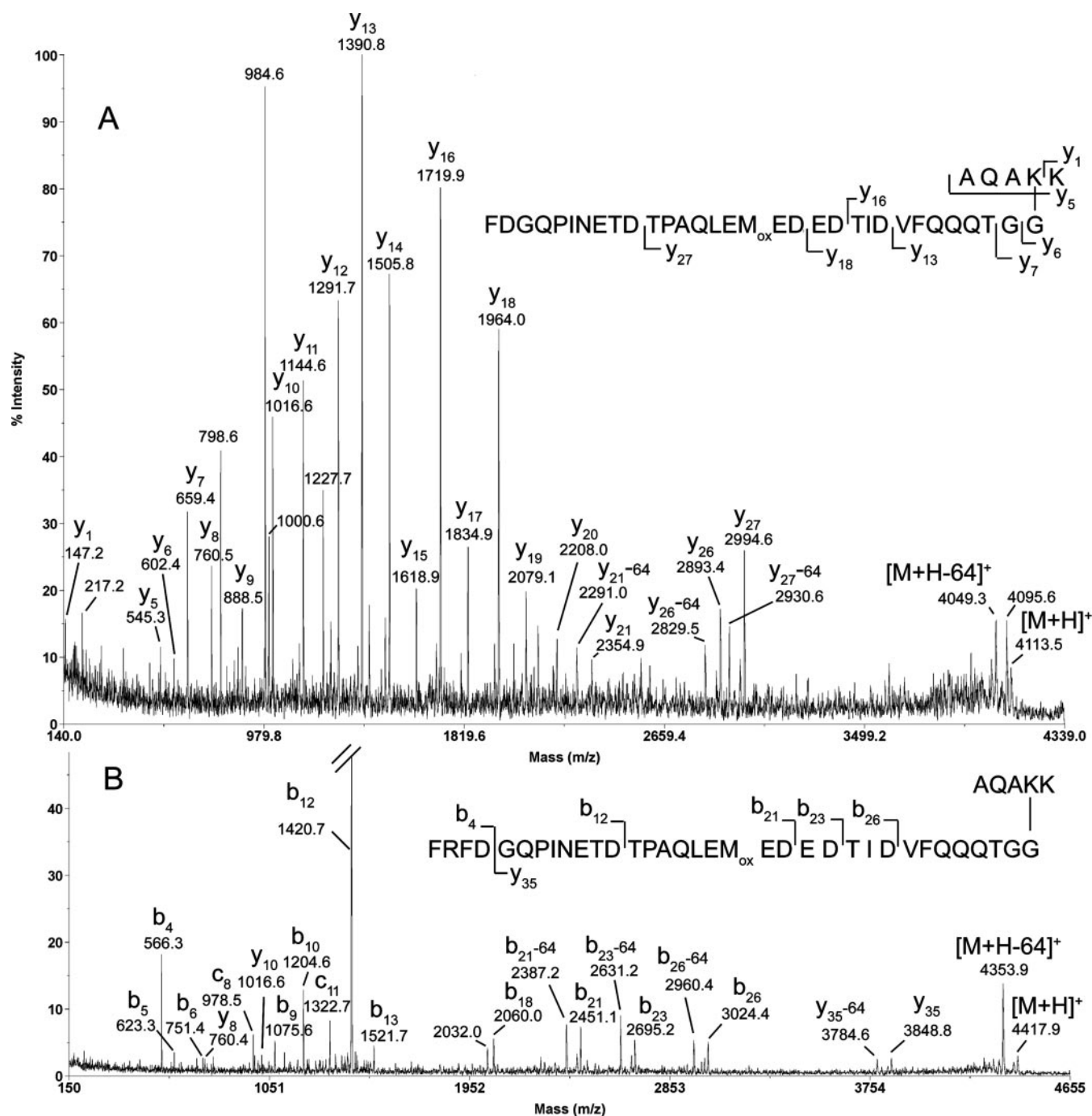


FIG. 4. MALDI MS/MS spectra of the SUMO-2-conjugated Ubc9 isopeptide (A) and that with additional N-terminal FR due to trypsin-missed cleavage at Arg⁶¹ (B). The spectrum in B is normalized to the intensity of the b₁₂ ion signal, but only the portion below 50% intensity is shown. The detected y ion series in A and the b ion series in B are analogous to those of SUMO-2-conjugated SUMO-2 isopeptides (Fig. 2). Oxidized Met was indicated by facile loss of 64 mass units. The characteristic mass intervals of 1030/244/329 for SUMO-2 isopeptide are defined by y₂₇, y₁₈, y₁₆, and y₁₃ in A and b₁₂, b₂₁, b₂₃, and b₂₆ in B.

It is important to note that, depending on the substrate proteins, the molecular weight of the SUMO-2-conjugated isopeptide will naturally be different. However, the tryptic fragment corresponding to the C-terminal peptide of SUMO-2 will remain the same. The four Asp residues, namely Asp⁷¹, Asp⁸⁰, Asp⁸², and Asp⁸⁵, will always induce four major fragment ions due to cleavage at their C termini with mass intervals corresponding to 1014 (TPAQLEMED) or 1030 (TPAQLEM_{ox}ED), 244 (ED), and 329 (TID) mass units. Detection of this characteristic MALDI-MS/MS pattern is thus a first indication that the parent ion signal is most probably a SUMO-2-sumoylated isopeptide. As reported by others (8), there is a

tendency of missed cleavage by trypsin at the Arg⁶¹ located immediately upstream of the 62–93 C-terminal peptide. As a consequence of the additional two residues Phe⁶⁰-Arg⁶¹ at the N terminus, strong fragment ion signals corresponding to the b ion series could be found to dominate the MS/MS spectrum in place of the y ions, a phenomenon best rationalized as charge localization at the N-terminal Arg in preference to the C-terminal Lys. Consequently, the mass intervals of 1014/244/329 are now defined by the b ions, namely b₁₂ (m/z 1420.6), b₂₁ (m/z 2435.0), b₂₃ (m/z 2679.0), and b₂₆ (m/z 3008.2) ions for all SUMO-2-conjugated isopeptides carrying the missed cleavage which, together with the prominent b₄ ion at m/z 566.3, consti-

tute the signature pattern (Fig. 3B). The complementary y ions could still be detected albeit at lower intensity, e.g. y_{44} and y_{36} at m/z 4900.4 and 4044.3, respectively, which can be further used as supporting evidence for confident sequence mapping. Interestingly, we noted that if the SUMO-2-sumoylated tryptic peptide from the substrate protein is itself terminating with an Arg residue instead of Lys, the presence of Arg at both the C and N termini of the isopeptide would lead to both y and b ion series as exemplified by the SUMO-2-sumoylated S-tag isopeptides (supplemental Fig. S1). In this case, the isopeptide, which did not carry the extra FR residues, afforded the expected y ion series dominated by y_{17} , y_{20} , and y_{22} (supplemental Fig. S1A). In contrast, the isopeptide with additional N-terminal FR residues due to missed cleavage afforded strong b_4 , b_{12} , b_{21} , and b_{23} ion signals, as well as y_{17} , y_{20} , y_{22} , y_{31} , and y_{39} ions (Fig. S1B), due to the Arg residue at the C terminus of the sumoylated tryptic peptide, ETAAAKFER.

Sumoylation of Ubc9 at the Non-consensus Site—Identification of peptides matching to both SUMO-2 and Ubc9 from band 3 (Fig. 2) implicated a covalent complex of the two proteins. Two putative isopeptides were detected by MALDI-TOF/TOF MS analysis at high mass range, with their monoisotopic $[M+H]^+$ molecular ion signals accurately mass-measured at m/z 4110.9 and 4414.0, respectively, corresponding to a FR residual mass difference. Both isopeptides were subjected to MS/MS sequence mapping and the resulting y and b ions detected (Fig. 4) confirm their identities, localizing the Lys sumoylation site to the C-terminal peptide of Ubc9, QAQAKKFAPS-COOH, similar to analogous position reported for yeast Ubc9 (AKQYSK-COOH) conjugation with Smt3p (9). The putative isopeptide carrying the catalytic Cys⁹³ of Ubc9 with thiol ester bond to SUMO-2 was not detected. This may reflect that such isopeptide is transient and unstable whereby the conjugated SUMO protein will be rapidly transferred to substrate proteins. In the absence of other substrate proteins in the *in vitro* system, Ubc9 may itself act as an acceptor, utilizing a non-consensus site near its C terminus. It is reasonable to assume that SUMO-1 can likewise sumoylate Ubc9, probably at the same site, to yield the protein band 2 in Fig. 2. Indeed, a candidate SUMO-1/Ubc9 isopeptide was barely detectable by MALDI-MS analysis of its tryptic digestion products, further MS/MS analysis of which yielded a series of y ions consistent with the expected isopeptide sequence (data not shown). However, since the isopeptide molecular ion signal was very weak and that the MALDI-TOF/TOF system employed could not allow high resolution precursor isolation, the fragment ions contribution from nearby precursors prevented an unambiguous assignment. The data nevertheless indicated that the same site on Ubc9 could be sumoylated *in vitro* by either SUMO-1 or SUMO-2 depending on availability of the SUMO proteins, a conclusion further echoed by our results on C28/C10 from CENP-C (see later under “Results”).

SUMO-2 and SUMO-1 Sumoylation of C28 at the Consensus Site—SDS-PAGE analysis revealed that the purified C28-His gave an intense major band at about 45 kDa (Fig. 5, lane 5), slightly below the creatine kinase band (46 kDa). Several additional major bands could be detected when the C28 sample was subjected to *in vitro* sumoylation with either SUMO-1 (lane 2) or SUMO-2 (lane 4). At first glance, several bands of 31 and 34 kDa (Fig. 5, bands 3–14) were commonly found irrespective of whether E1 was added to the reaction mixtures. LC-MS/MS analyses (Table I), however, revealed that in the absence of E1 when the sumoylation reaction could not proceed (lanes 1 and 3), only C28 could be identified from these protein bands (bands 3–5 and 9–11), which probably represent the various degraded forms of C28. When E1 was added, sumoyla-

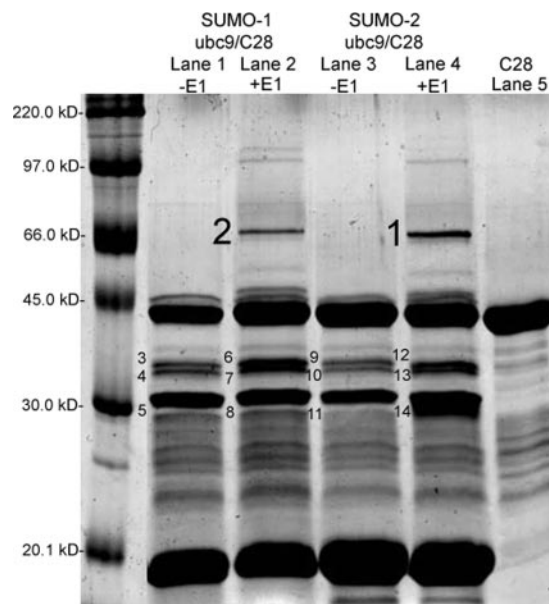


FIG. 5. SDS-PAGE analysis of *in vitro* sumoylated C28 products with either SUMO-1GG or SUMO-2GG. The sumoylation reactions contain C28-His, either SUMO-1GG or SUMO-2GG, E2 (Ubc9) enzyme with and without E1 enzyme. Bands 1 and 2 are the C28-sumoylated by SUMO-2GG and SUMO-1GG, respectively. Bands 3–14 correspond to partially degraded fragments of C28, as well as sumoylated SUMO-2 (band 14), sumoylated Ubc9 (bands 6, 7, 12, and 13). Results from LC-MS/MS identification of the protein bands are tabulated in Table I.

TABLE I
LC-MS/MS identification of protein bands from SDS-PAGE analysis of the *in vitro* sumoylated C28 (Fig. 5)

Band	Protein identified (number of matched peptides/mascot score) ^a			
	C28 (CENP-C)	SUMO-1	SUMO-2	Ubc9
1	5/243	—	2/119	—
2	2/172	4/35	—	—
3	8/305	—	—	—
4	8/308	—	—	—
5	9/491	—	—	—
6	8/340	4/132	—	6/201
7	8/348	2/56	—	8/246
8	10/466	—	—	—
9	7/338	—	—	—
10	10/412	—	—	—
11	8/469	—	—	—
12	9/395	—	3/93	6/173
13	8/336	—	3/104	7/261
14	10/444	—	3/124	—

^a Proteins were identified by standard LC-MS/MS analyses of the tryptic digestion products. MS/MS data were processed into peak lists and searched against Swiss-Prot database using the Mascot program. CENP-C was identified as top hit in all bands. Other non-relevant but significant hits attributable to contaminants from the reaction mixtures were not listed. “—” denotes not identified from the respective analysis.

tion could occur and resulted in additional identification of Ubc9 and the SUMO proteins in bands 6 and 7 and 12 and 13. These barely separated doublets may thus contain sumoylated Ubc9 in addition to C28. It is also clear that band 14 appeared to be broader than other bands (bands 5, 8, and 11) and was found to also contain SUMO-2 proteins. Subsequent MALDI MS/MS analysis led to detection of the implicated SUMO-2 conjugated SUMO-2 isopeptide (data not shown). These findings are therefore consistent with those described above (Fig. 2) when no C28 was added, and none of the 31/34-kDa bands afforded any candidate isopeptide suggestive of sumoylated C28 when screened by MALDI-MS and -MS/MS. Of interest

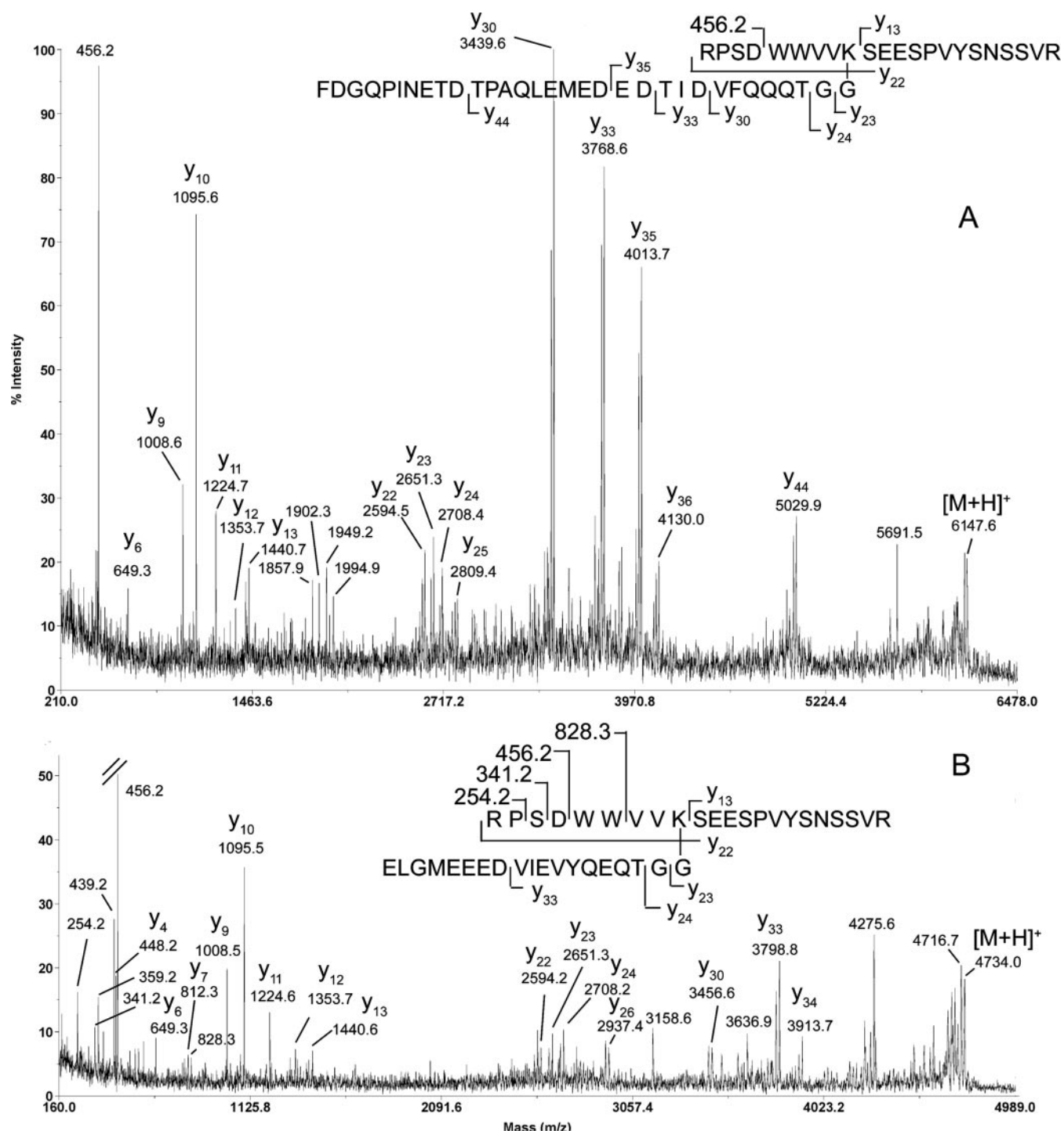


FIG. 6. MALDI MS/MS spectra of the SUMO-2-conjugated C28 isopeptide (A) and SUMO-1-conjugated C28 isopeptide (B). The spectrum in B is normalized to the intensity of the b_{1,2} ion signal, but only the portion below 50% intensity is shown. The detected y ion series in A are analogous to those of SUMO-2 conjugated SUMO-2 isopeptides (Fig. 2) with the characteristic mass intervals of 1014/244/329 for SUMO-2 isopeptide being defined by y₄₄, y₃₅, y₃₃, and y₃₀. The same series of y₆, y₉₋₁₃, and y₂₂₋₂₄ ions afforded by both SUMO-1 and SUMO-2 isopeptides unambiguously defined the same tryptic peptide from C28 of CENP-C, sumoylated at Lys⁵³⁴ of the consensus site.

then are the two bands at ~66 kDa that were also clearly absent from the reaction mixtures if E1 was omitted.

In addition to peptides from C28, SUMO-2 and SUMO-1 peptides were found in the tryptic digests from bands 1 and 2, respectively (Table I), which suggested that these two bands correspond to sumoylated C28. For band 1, MALDI-TOF/TOF MS analysis led to identification of a weak molecular ion signal, the more abundant precursor signal of which was detected at *m/z* 6146.8, corresponding exactly to the expected tryptic peptide fragment of C28, ⁵²⁶RPSDWWVVKSEESPVYSNSSVR⁵⁴⁷,

conjugated with the SUMO-2 C-terminal peptide. The base peak at *m/z* 456.2 in its MS/MS spectrum (Fig. 6A) could be assigned as the b₄ ion of the C28 tryptic peptide due to facile cleavage at Asp⁵²⁹ and charge retention on the N-terminal Arg⁵²⁶. Sequence mapping and localization of the sumoylation site to Lys⁵³⁴ of the ψ KXE consensus motif were afforded by a series of prominent y ions (Fig. 6A). Importantly, all the b ions related to the C28 tryptic peptides and the y ion series up to the C-terminal GG of SUMO-2 (y₂₃ and y₂₄) were reproduced in the MS/MS spectrum of a putative SUMO-1 sumoylated C28

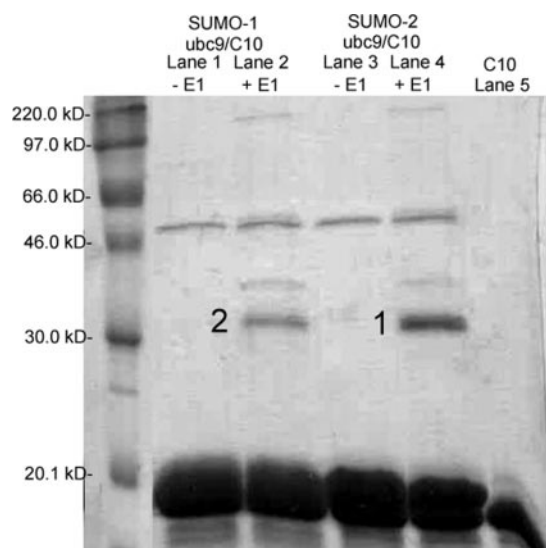


FIG. 7. SDS-PAGE analysis of *in vitro* sumoylated C10 products by either SUMO-1GG or SUMO-2GG. The sumoylation reactions contain C10-His, SUMO-1GG or SUMO-2GG, E2 (Ubc9) enzyme with and without E1 enzyme. Bands 1 and 2 are the C10 sumoylated by SUMO-2GG and SUMO-1GG, respectively.

isopeptide from band 2 (Fig. 6B). The accurate mass measurement of the monoisotopic $[M+H]^+$ molecular ion signal at m/z 4730.2 and other detectable y ions, especially the intense y_{33} ion, confirmed that the same C28 tryptic peptide (aa 526–547) could be sumoylated by SUMO-1 at the same Lys⁵³⁴ site, giving rise to the analyzed isopeptide. Thus, protein bands 1 and 2 (Fig. 5) were confirmed to be SUMO-2- and SUMO-1-sumoylated C28 of CENP-C, respectively.

SUMO-2 and SUMO-1 Sumoylation of C10 at Non-consensus Sites—In the case of C10-His, no consensus sumoylation site could be located, but SDS-PAGE analysis of the *in vitro* sumoylation products suggested that C10 was sumoylated by SUMO-2 and SUMO-1, yielding the protein bands 1 and 2, respectively (Fig. 7). MALDI-MS analysis failed to detect any molecular ion signals that may correspond to the predicted isopeptide comprising the weakly potential sumoylation site ⁶⁹⁷GKND⁷⁰⁰. To search for other putative sumoylated isopeptides, MS/MS analyses were performed on all reasonably strong signals at above m/z 4000. Four putative isopeptides with the characteristic MS/MS signature pattern established for SUMO-2 isopeptide were identified, two of which correspond to SUMO-2-conjugated SUMO-2 isopeptides with and without Met oxidation, affording a pair of monoisotopic $[M+H]^+$ signals at m/z 5160.3 and 5176.3, respectively. The other two signals, which gave monoisotopic $[M+H]^+$ at m/z 4292.0 and 4670.0, could be assigned as corresponding to SUMO-2-conjugated tryptic peptides of C10, ⁷¹⁸VIPK^{NR}⁷²³, and ⁷⁴⁵LK^{PLEYWR}⁷⁵², respectively, from MS/MS analyses (Fig. 8, A and B). The mass intervals of 1030/244/329 could be clearly attributed to the dominant y ions detected along with neutral loss of 64 mass units consistent with the presence of an oxidized Met.

As in the case of Ubc9 sumoylation, assuming that the same sites that were sumoylated by SUMO-2 might likewise be sumoylated by SUMO-1, the predicted SUMO-1/C10 isopeptides were indeed detected by MALDI-MS analysis of the tryptic digests from band 2 (Fig. 7) at m/z 2878.4 and 3256.6 (monoisotopic masses). MS/MS analyses further confirmed the sequence assignment as shown in Fig. 8, C and D. Thus, the two non-consensus sites, Lys⁷²¹ and Lys⁷⁴⁶, as in the C10 fragment of CENP-C, could be sumoylated *in vitro* by both SUMO-1 and SUMO-2. Under the experimental conditions em-

ployed, each C10 fragment molecule was apparently sumoylated at only one of the two sites and not both, since only one major band could be detected. Since C10 is an artificially created small polypeptide, it is not known whether the sumoylation detected may bear any *in vivo* significance, but our findings shed light on the utilization of consensus versus non-consensus sites in the substrate proteins. It is of interest to note that of all the seven Lys residues in C10, the two identified as sumoylation sites both correspond to Lys with adjacent Pro residue. In contrast, the weakly potential site ⁶⁹⁷GKND⁷⁰⁰ could not be detected as sumoylated, but the corresponding tryptic peptide terminating with an unmodified Lys⁶⁹⁸ could, thus arguing that the *in vitro* sumoylation reaction was not a random event targeting any Lys residue.

DISCUSSION

We have demonstrated in this work that tandem mass spectrometry is a powerful and effective analytical means to detect and unequivocally map the sumoylation sites. Based on MS/MS characteristics of SUMO-2-conjugated isopeptides, a concerted strategy could be formulated to rapidly screen for SUMO-2-conjugated isopeptides among the tryptic digestion products. The excised band on SDS-PAGE would first be tryptic digested followed by LC-MS/MS analysis for protein identification. The particular band of interest should be found to contain peptide matches from both the SUMO and substrate proteins. The use of trypsin not only preserves the useful localization of Arg/Lys to C termini for facile peptide MS/MS sequencing but should in principle also cleave all non-sumoylated Lys sites. By definition, the isopeptide should have a minimum molecular mass corresponding to the tryptic C-terminal peptide of SUMO-2 at 3567.5 or 3870.7 Da (monoisotopic mass) depending on whether a missed cleavage at Arg⁶¹ is included, taking also into consideration that the Met contained within may be oxidized to give additional 16-mass unit increment. Rapid second tier MALDI-MS mapping will be followed by acquisition of MS/MS spectra for all detectable signals at above m/z 3600 on a MALDI-TOF/TOF instrument, which allows for high sensitivity CID-MS/MS of precursor ions of high m/z values. If the desirable characteristic patterns were detected, possible related isopeptide signals are sought at ± 16 mass units for Met oxidation and ± 303 mass units for Phe-Arg residual mass increment. Detection of SUMO-1-conjugated isopeptide is more problematic as it lacks the critical Asp residues to induce the dominant fragmentation pattern. However, our work demonstrated that, at least for the *in vitro* system, sites that can be sumoylated by SUMO-2 will also likely to be sumoylated by SUMO-1. This may form the basis for identifying SUMO-1-conjugated non-consensus sites.

Using this strategy, we have been able to not only detect the sumoylated isopeptides from the protein bands on SDS-PAGE but also provide additional MS/MS sequence mapping evidence for the identified sumoylation site, especially if it does not contain a consensus sequence motif (Table II). To our knowledge, this is the first report on direct MS/MS sequencing of sumoylated isopeptides. Distinctive features were identified that can be used to rapidly screen for SUMO-2/3-conjugated isopeptides within a complex digestion mixtures, as in shotgun proteomics. In principle, computer scripts can be written to automate real-time data-dependent MS/MS acquisition based on the characteristic mass intervals. A 329/244/1014 increment implies y ions for SUMO-2/3 C-terminal tryptic peptide without missed cleavage, whereas the reverse order, *i.e.* 1014/329/244 increment, defines the b ion series and indicates the presence of additional Phe-Arg residues at the N terminus. The molecular weight of the tryptic peptide fragment from the substrate protein can subsequently be computed and matched against the

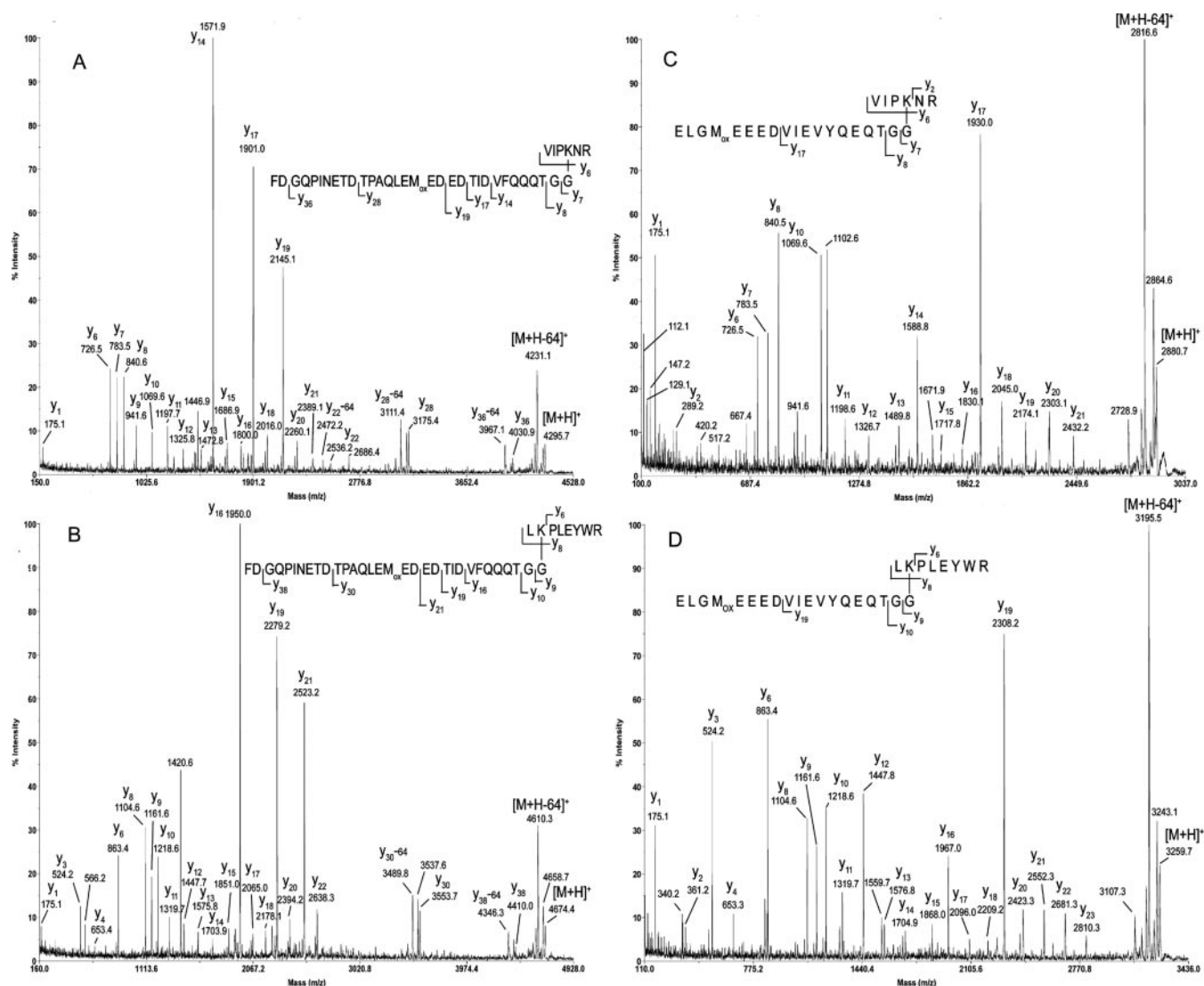


FIG. 8. MALDI-MS/MS spectra of the C10 isopeptides sumoylated by SUMO-2 at Lys⁷²¹ (A) and Lys⁷⁴⁶ (B) or SUMO-1 at Lys⁷²¹ (C) and Lys⁷⁴⁶ (D). The detected y ion series in A and B are analogous to those of SUMO-2-conjugated SUMO-2 isopeptides (Fig. 2) with the characteristic mass intervals of 1030/244/329 for SUMO-2 isopeptide being defined by y₂₈, y₁₉, y₁₇, and y₁₄ in A and y₃₀, y₂₁, y₁₉ and y₁₆ in B. The same series of y_{6–10} ions in A and C afforded by both SUMO-2 and SUMO-1 isopeptides unambiguously define the same tryptic peptide ⁷¹⁸VIPK^{NR}⁷²³ from C10 of CENP-C, sumoylated at Lys⁷²¹; whereas the same y₆, y₈, y₉, and y₁₀ ions in B and D define the same tryptic peptide ⁷⁴⁵LK^PLEYWR⁷⁵² from C10 of CENP-C, sumoylated by either SUMO-2 or SUMO-1 at Lys⁷⁴⁶.

TABLE II

Sumoylation sites identified in this study by MALDI-TOF/TOF MS/MS sequence mapping

Both SUMO-1- and SUMO-2-sumoylated isopeptides for all the implicated sites could be detected and verified by MS/MS analyses except for S-tag in which the candidate SUMO-1-sumoylated isopeptide has not been actively sought.

Protein	Site	Flanking sequence of the identified sumoylation site
SUMO-2	Lys ¹¹	H ₂ N-MADEKPKEGV K TENNDHINLK
Ubc9	Lys ¹⁵³	EYEKRVRAQA K KFAPS-COOH
S-tag	Lys ¹¹	GSGMKETA A A K FERQHMDSPD
C28	Lys ⁵³⁴	SRPSDWWV V KSEESPVSNS
C10	Lys ⁷²¹	DDSKQSKVIP K NRHHKLVL
	Lys ⁷⁴⁶	NVRRTKRTRL K PLEYWRGERI

known protein sequence, applying constraints for the presence of a non-terminal Lys, probable but not necessary within a ψ KXE motif. In cases when the MS/MS data are of sufficiently good quality, further confirmation of the sequence can be obtained through additional y ions detected.

This investigation therefore establishes the ground work for

future SUMO proteomics studies. Only when more sumoylation sites could be definitively confirmed can bioinformatics studies be meaningful to derive the actual consensus sumoylation site, taking into account its secondary and tertiary structures. Although the list of sumoylated sites continues to grow steadily, the vast majority were mapped only by mutagenesis of the putative acceptor lysines to arginines followed by analysis of loss of modifications and/or sumoylation-implicated functions. Such an experimental approach is, however, non-discriminative against lack of acceptor site or impairment of sumoylation enzyme-substrate interactions in the mutants. ψ KXE is apparently not sufficient nor absolutely necessary to ensure sumoylation (2, 3). In the case of RanGAP1, crystallographic analysis showed that the acceptor Lys is part of an accessible loop structure and reaches directly into the catalytic cleft of Ubc9 (5). It has also been noted that most sumoylated sequences contain Pro and/or Gly 2–5 aa upstream or downstream of the acceptor Lys, probably to help insuring the accessibility by introducing a turn. Alternatively, the sequence motif is located near the very N- or C-terminal ends, which could therefore be readily accessible (3). The two non-consensus

sumoylation sites on the C10 fragment of CENP-C identified in this study conform to the first scenario, namely with adjacent proline. The sumoylation site on Ubc9 appears to conform to the second scenario by virtue of its near C-terminal location. Yet, many other reported cases in the literature do not follow these simplistic rules and the issue of substrate specificity is far from resolved, especially in real physiological context, in the presence of other interacting and regulating factors including E3 ligases.

The *in vitro* system coupled with MS analysis could in fact serve as an assay system to screen for the identity and functional effect of putative E3 ligases, as well as isopeptidases, on the relative efficiency and specificity of sumoylation by SUMO-1 and SUMO-2/3 families. Interestingly, we could further detect isopeptide corresponding to SUMO-2 protein being sumoylated by SUMO-1 at consensus site Lys¹¹ in an *in vitro* sumoylation system where both SUMO-1 and SUMO-2 proteins were added (supplemental Fig. S2).² While polymerization of SUMO-2 has been reported (8, 9), it is unclear whether SUMO-1 may also be added on a growing chain of SUMO-2 to cap the reactions. This finding, coupled with our observation that most identified sites could be equally sumoylated by both SUMO-1 and SUMO-2, suggest a regulatory mechanism that may be more complicated than anticipated thus far.

Relying on the SDS-PAGE to resolve potentially multi-sumoylated products, we could identify only a major band each for the sumoylated C28 and C10 fragments. In the case of C28, which contains a consensus ψ KXE sumoylation site, the only isopeptide identified was that expected, suggesting that C28 was mainly monosumoylated at the consensus site despite the presence of another Lys in C28 that is followed by Pro. In contrast, for the C10 fragment, which carries no apparent perfect consensus sumoylation site, the single major band detected was interpreted as corresponding to co-migrating C10 fragments monosumoylated at either of the two non-consensus sites identified. This conclusion is based primarily on positive identification of both isopeptides and the corresponding tryptic peptides spanning an unmodified Lys. Thus a weakly potential site such as GKND was found to be less preferred than Lys preceded or followed by Pro, which in turn indicates that a certain selectivity was maintained. It should be pointed out that our analytical strategy relies on tryptic cleavage, which would potentially cleave at all non-modified Lys, but not all resulting in-gel-digested tryptic peptides could be detected, especially those comprising only a few amino acids. It is therefore not possible to categorically rule out a sumoylation based on detecting the unmodified tryptic peptide. A major technical advance suggested by the current work is the successful application of MALDI-TOF/TOF MS/MS to identify instead the implicated isopeptides at the high mass end. When combined with size fractionation on SDS-PAGE, barring any abnormality in electrophoretic mobility, the approach should adequately establish the sumoylation pattern and the preferred site selection, as demonstrated here for the CENP-C fragments.

It remains to be established whether the identified CENP-C sumoylation sites are likewise sumoylated *in vivo*. Identification of SUMO-1 as a suppressor of CENP-C mutants provided a first

possible connection between CENP-C function and sumoylation (15). CENP-C and centromere were further associated, in a cell cycle-regulated manner, with nuclear dots ND10 proteins (PML, SP100, Daxx *etc.*) modified by SUMO-1 (28–30). Whether CENP-C itself is reversibly sumoylated by SUMO-1 or SUMO-2/3, or both, at a particular state of chromosome dynamics has been difficult to demonstrate via conventional approaches. The advent of MS and proteomics has been used successfully to define the yeast kinetochore protein complex and the centromere complex from HeLa interphase cells (31, 32). These analyses can now be further coupled with definitive identification of sumoylation sites among the principal components in conjunction with reconstituted *in vitro* sumoylation system to delineate a highly complicated and dynamic cellular process.

Acknowledgment—We thank Dr. A. Chen for helpful suggestions regarding the *in vitro* sumoylation system.

REFERENCES

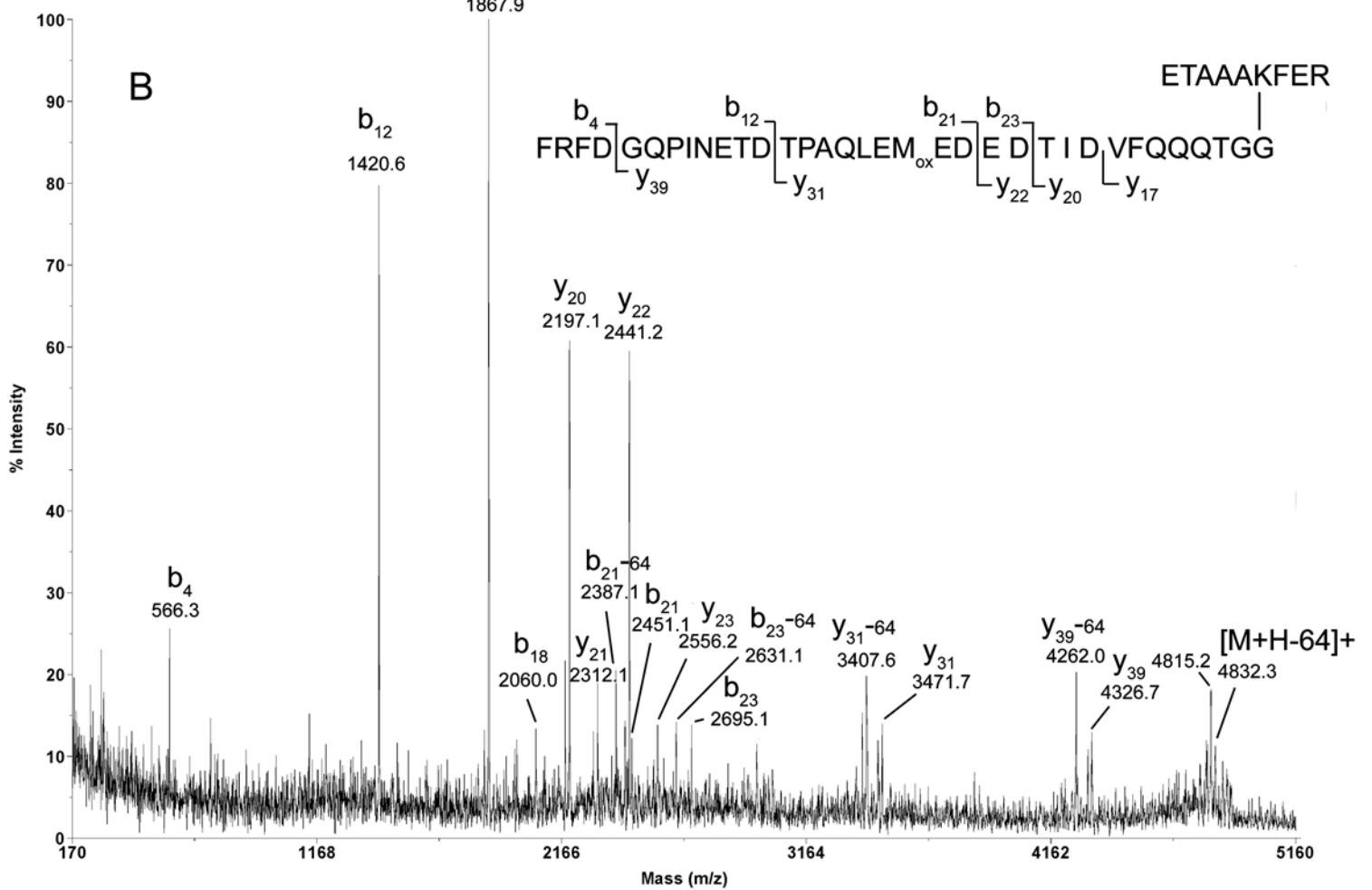
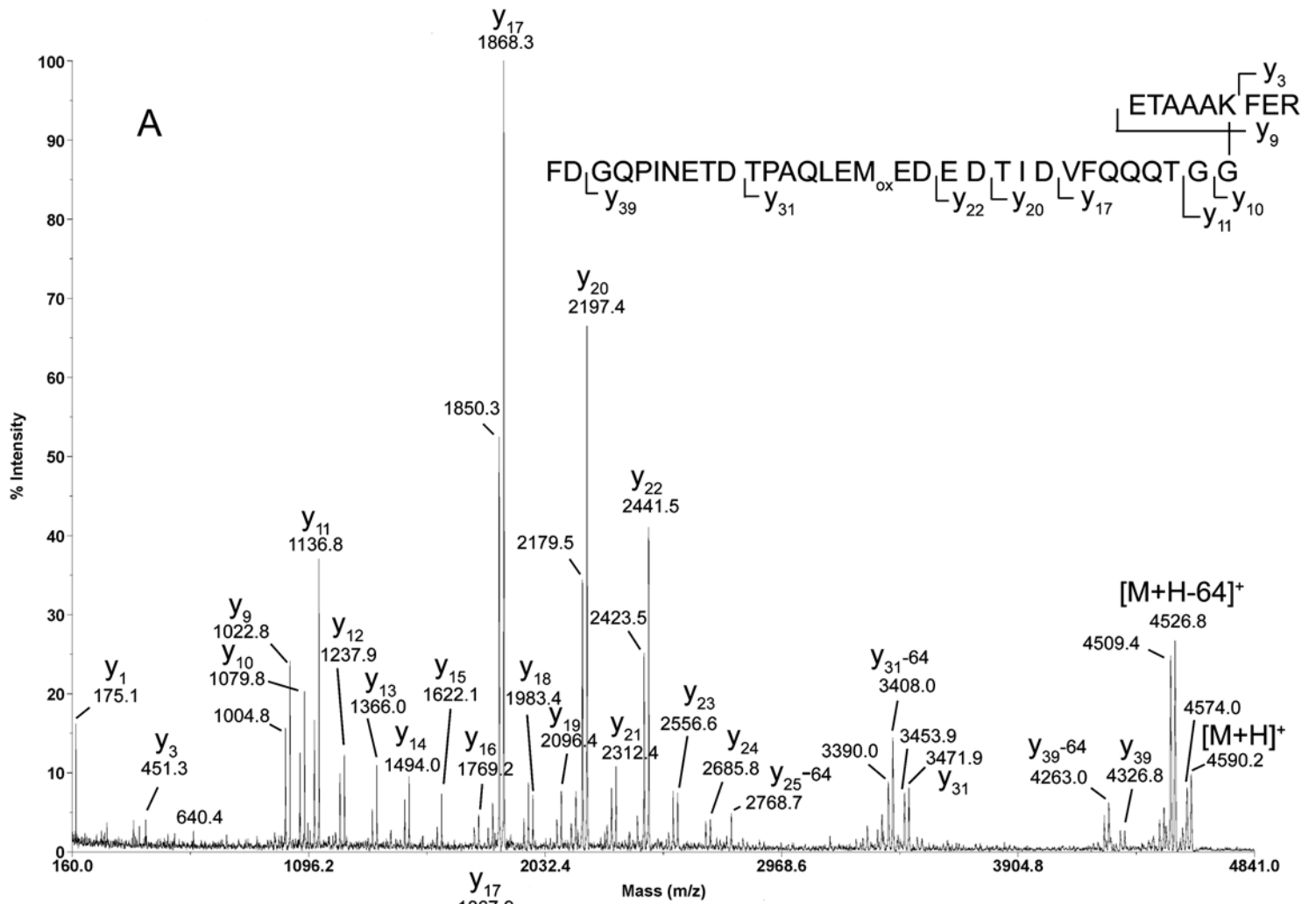
- Seeler, J. S., and Dejean, A. (2003) *Nat. Rev. Mol. Cell. Biol.* **4**, 690–699
- Johnson, E. S. (2004) *Annu. Rev. Biochem.* **73**, 355–382
- Melchior, F. (2000) *Annu. Rev. Cell Dev. Biol.* **16**, 591–626
- Sampson, D. A., Wang, M., and Matunis, M. J. (2001) *J. Biol. Chem.* **276**, 21664–21669
- Bernier-Villamor, V., Sampson, D. A., Matunis, M. J., and Lima, C. D. (2002) *Cell* **108**, 345–356
- Melchior, F., Schergaut, M., and Pichler, A. (2003) *Trends Biochem. Sci.* **28**, 612–618
- Mossessova, E., and Lima, C. D. (2000) *Mol. Cell* **5**, 865–876
- Tatham, M. H., Jaffray, E., Vaughan, O. A., Desterro, J. M., Botting, C. H., Naismith, J. H., and Hay, R. T. (2001) *J. Biol. Chem.* **276**, 35368–35374
- Bencsath, K. P., Podgorski, M. S., Pagala, V. R., Slaughter, C. A., and Schulman, B. A. (2002) *J. Biol. Chem.* **277**, 47938–47945
- Pichler, A., Gast, A., Seeler, J. S., Dejean, A., and Melchior, F. (2002) *Cell* **108**, 109–120
- Mahajan, R., Gerace, L., and Melchior, F. (1998) *J. Cell Biol.* **140**, 259–270
- Johnson, E. S., and Blobel, G. (1999) *J. Cell Biol.* **147**, 981–994
- Pluta, A. F., Mackay, A. M., Ainsztein, A. M., Goldberg, I. G., and Earnshaw, W. C. (1995) *Science* **270**, 1591–1594
- Meluh, P. B., and Koshland, D. (1995) *Mol. Biol. Cell* **6**, 793–807
- Fukagawa, T., Regnier, V., and Ikemura, T. (2001) *Nucleic Acids Res.* **29**, 3796–3803
- Brown, M. T. (1995) *Gene (Amst.)* **160**, 111–116
- Lanini, L., and McKeon, F. (1995) *Mol. Biol. Cell* **6**, 1049–1059
- Yang, C. H., Tomkiel, J., Saitoh, H., Johnson, D. H., and Earnshaw, W. C. (1996) *Mol. Cell. Biol.* **16**, 3576–3586
- Sugimoto, K., Kuriyama, K., Shibata, A., and Himeno, M. (1997) *Chromosome Res.* **5**, 132–141
- Song, K., Gronemeyer, B., Lu, W., Eugster, E., and Tomkiel, J. E. (2002) *Exp. Cell Res.* **275**, 81–91
- Trazzi, S., Bernardoni, R., Diolaiti, D., Politi, V., Earnshaw, W. C., Perini, G., and Della Valle, G. (2002) *J. Struct. Biol.* **140**, 39–48
- Su, H. L., and Li, S. S. L. (2002) *Gene (Amst.)* **296**, 65–73
- Mannen, H., Tseng, H. M., Cho, C. L., and Li, S. S. L. (1996) *Biochem. Biophys. Res. Commun.* **222**, 178–180
- Lee, C. L., Hsiao, H. H., Lin, C. W., Wu, S. P., Huang, S. Y., Wu, C. Y., Wang, A. H., and Khoo, K. H. (2003) *Proteomics* **3**, 2472–2486
- Wattenberg, A., Organ, A. J., Schneider, K., Tyldesley, R., Bordoli, R., and Bateman, R. H. (2002) *J. Am. Soc. Mass Spectrom.* **13**, 772–783
- Jiang, X., Smith, J. B., and Abraham, E. C. (1996) *J. Mass Spectrom.* **31**, 1309–1310
- Lagerwerf, F. M., van de Weert, M., Heerma, W., and Haverkamp, J. (1996) *Rapid Commun. Mass Spectrom.* **10**, 1905–1910
- Ishov, A. M., Sotnikov, A. G., Negorev, D., Vladimirova, O. V., Neff, N., Kamitani, T., Yeh, E. T., Strauss, J. F., 3rd, and Maul, G. G. (1999) *J. Cell Biol.* **147**, 221–234
- Everett, R. D., Earnshaw, W. C., Pluta, A. F., Sternsdorf, T., Ainsztein, A. M., Carmena, M., Ruchaud, S., Hsu, W. L., and Orr, A. (1999) *J. Cell Sci.* **112**, 3443–3454
- Pluta, A. F., Earnshaw, W. C., and Goldberg, I. G. (1998) *J. Cell Sci.* **111**, 2029–2041
- Westermann, S., Cheeseman, I. M., Anderson, S., Yates, J. R., III, Drubin, D. G., and Barnes, G. (2003) *J. Cell Biol.* **163**, 215–222
- Obuse, C., Yang, H., Nozaki, N., Goto, S., Okazaki, T., and Yoda, K. (2004) *Genes Cells* **9**, 105–120

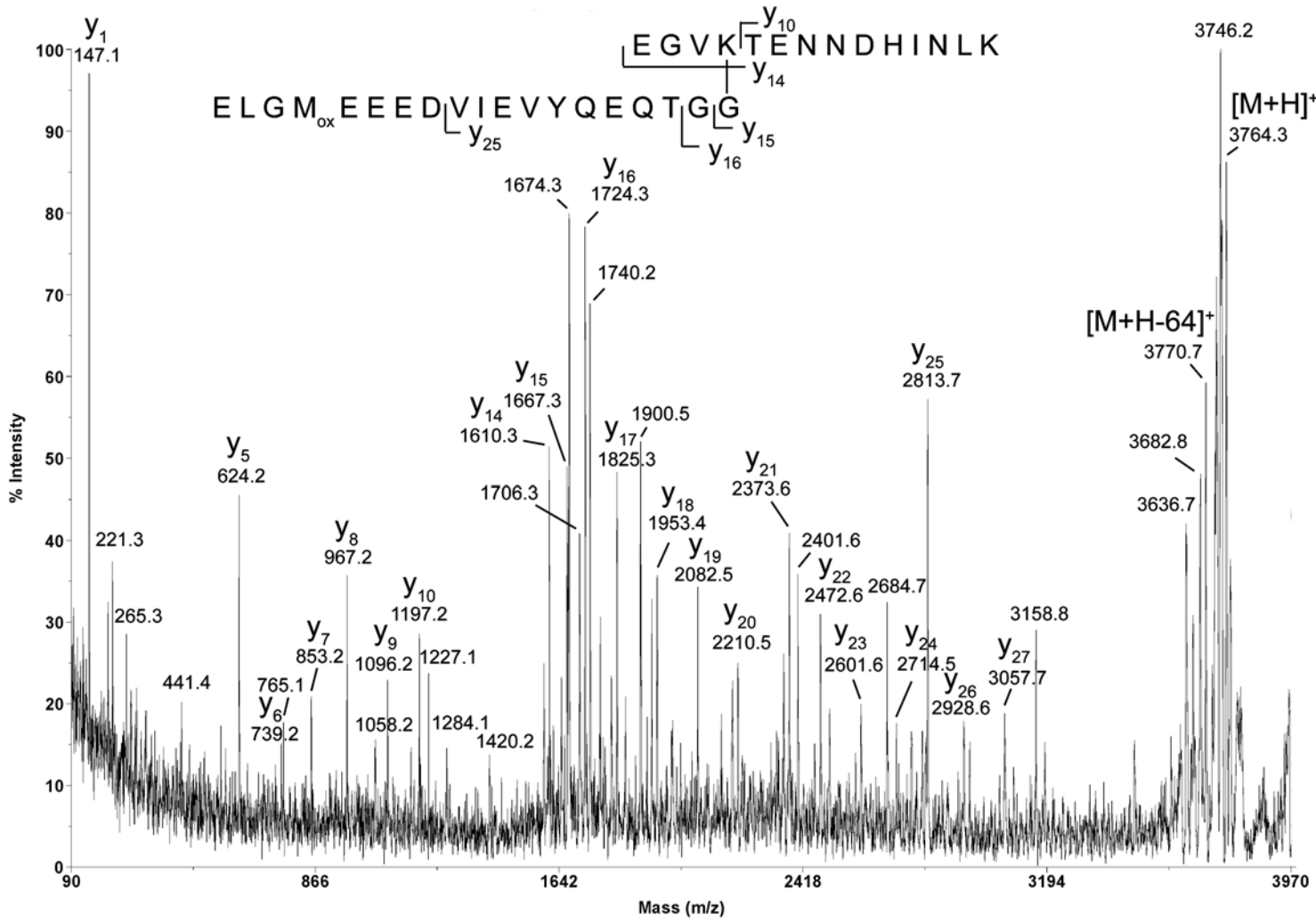
² T.-L. Chung, H.-H. Hsiao, Y.-Y. Yeh, A. H.-J. Wang, K.-H. Khoo, and S. S.-L. Li, unpublished data.

—

Fig. S1. MALDI MS/MS spectra of the SUMO-2 conjugated S-tag isopeptide (A) and that with additional N-terminal FR due to trypsin missed cleavage at Arg⁶¹ (B). The S-Tag (GSGMKETAAAKKFERQHMSDDLGTGGGSGIEGR) is commonly inserted at the N-termini of various proteins. The consensus sumoylation site present in S-Tag was found unexpectedly to be sumoylated by SUMO-2. This tag is therefore not suitable for insertion into substrate proteins or the SUMO proteins for sumoylation studies.

Fig. S2. MALDI MS/MS spectra of the SUMO-1 conjugated SUMO-2 isopeptide. A protein band containing both SUMO-1 and SUMO-2 peptides was identified by LC-MS/MS analysis of the *in vitro* sumoylation reaction products when both SUMO-1 and SUMO-2 were added. The predicted SUMO-2/SUMO-1 isopeptide with monoisotopic [M+H]⁺ molecular ion at *m/z* 3762.7 was then isolated for MALDI MS/MS analysis despite its low intensity in the MALDI mass spectrum. Not all fragment ions originated from the isopeptide as other neighboring precursors were also co-selected. The y series of fragment ions detected were however of sufficient quality to unambiguously confirm the identity of the implicated isopeptide, localizing the SUMO-1 conjugated site of SUMO-2 at the expected Lys¹¹.





***In Vitro* Modification of Human Centromere Protein CENP-C Fragments by Small Ubiquitin-like Modifier (SUMO) Protein: DEFINITIVE IDENTIFICATION OF THE MODIFICATION SITES BY TANDEM MASS SPECTROMETRY ANALYSIS OF THE ISOPEPTIDES**

Tung-Liang Chung, He-Hsuan Hsiao, Yuh-Ying Yeh, Hui-Ling Shia, Yi-Ling Chen, Po-Huang Liang, Andrew H.-J. Wang, Kay-Hooi Khoo and Steven Shoei-Lung Li

J. Biol. Chem. 2004, 279:39653-39662.

doi: 10.1074/jbc.M405637200 originally published online July 22, 2004

Access the most updated version of this article at doi: [10.1074/jbc.M405637200](https://doi.org/10.1074/jbc.M405637200)

Alerts:

- [When this article is cited](#)
- [When a correction for this article is posted](#)

[Click here](#) to choose from all of JBC's e-mail alerts

Supplemental material:

<http://www.jbc.org/content/suppl/2004/08/04/M405637200.DC1.html>

This article cites 32 references, 13 of which can be accessed free at <http://www.jbc.org/content/279/38/39653.full.html#ref-list-1>