Mass Spectrometric Analysis of Protein Mixtures at Low Levels Using Cleavable $^{13}$C-Isotope-coded Affinity Tag and Multidimensional Chromatography*

Kirk C. Hansen‡§, Gerold Schmitt-Ulms¶, Robert J. Chalkley‡, Jan Hirsch||, Michael A. Baldwin‡¶, and A. L. Burlingame‡¶**

In order to identify and compare the protein content of very low quantity samples of high complexity, a protocol has been established that combines the differential profiling strength of a new cleavable $^{13}$C isotope-coded affinity tag (cICAT) reagent with the high sequence coverage provided by multidimensional liquid chromatography and two modes of tandem mass spectrometry. Major objectives during protocol optimization were to minimize sample losses and establish a robust procedure that employs volatile buffer systems that are highly compatible with mass spectrometry. Cleavable ICAT-labeled tryptic peptides were separated from nonlabeled peptides by avidin affinity chromatography. Subsequently, peptide samples were analyzed by nanoflow liquid chromatography electrospray ionization tandem mass spectrometry and liquid chromatography matrix-assisted laser desorption/ionization tandem mass spectrometry. The use of two ionization/instrumental configurations led to complementary peptide identifications that increased the confidence of protein assignments. Examples that illustrate the power of this strategy are taken from two different projects: i) immunoaffinity purified complexes containing the prion protein from the murine brain, and ii) human tracheal epithelium gland secretions. In these studies, a large number of novel proteins were identified using stringent match criteria, in addition to many that had been identified in previous experiments. In the latter case, the ICAT method produced significant new information on changes that occur in protein expression levels in a patient suffering from cystic fibrosis. Molecular & Cellular Proteomics 2:299–314, 2003.

In recent years, the emphasis within the proteomics field has moved from the identification of isolated proteins (1, 2) to the challenge of characterizing complex mixtures (3). Frequently, the goal is to monitor changes in mixture composition and/or relative abundance under differing physiologically relevant conditions, an experimental approach commonly referred to as differential profiling (4, 5). While mass spectrometry has become the method of choice for the identification of proteins, several alternative approaches are being employed for the separation of complex mixtures that precedes their mass spectrometric analysis. Traditionally, the most frequently applied protein separation strategy has been based on two-dimensional polyacrylamide gel electrophoresis (6, 7), a technique that has the ability to separate up to 10,000 components (8). While the independent manual digestion of multiple gel spots is time-consuming, robotic digestion can improve this limitation (9). In the last few years, enzymatic digestion of unseparated protein mixtures followed by separation of peptides by multidimensional liquid chromatography (LC)$^1$ has been used as an alternative. Washburn et al. (10) assigned 5,540 peptides to 1,484 proteins in Saccharomyces cerevisiae using this approach. Among other benefits, this LC-based approach lends itself to automated peptide separation together with the acquisition and analysis of mass spectrometric data. A potential disadvantage lies in the fact that linkage of peptides belonging to any given protein is obscured (11).

Strategies for characterizing changes in complex mixtures have been developed using both of these technologies. For two-dimensional gels, samples may be run on separate gels, stained, and protein abundances compared with the use of imaging software (12). However, in practice, protein pattern comparisons can be difficult to achieve due to poor reproducibility of protein separations on two-dimensional gels. De-

---

$^1$ The abbreviations used are: LC, liquid chromatography; ICAT, isotope-coded affinity tag; MS, mass spectrometry; MS/MS, tandem mass spectrometry; ESI, electrospray ionization; MALDI, matrix-assisted laser desorption/ionization; ACN, acetonitrile; PrP, prion protein; SCX, strong cation exchange; CID, collision-induced dissociation; TOF, time-of-flight; cICAT, cleavable ICAT; BSA, bovine serum albumin; HPLC, high-performance LC; TFA, trifluoroacetic acid; S/N, signal-to-noise ratio; Qq, quadrupole selection, quadrupole collision cell; H/L, heavy-to-light ratio; HSA, human serum albumin.
focusing an individual spot may not be straightforward, as a result of which a large amount of manual work may be required to complement the software interpretation in obtaining a reliable analysis (13). An approach that has alleviated many of the problems due to gel-to-gel electrophoretic variability is differential gel electrophoresis (14, 15).

Mass spectrometry (MS) is not a quantitative technique per se as ion yields are highly dependent on the chemical and physical nature of the sample. However, isotopic labeling combined with MS has been extensively used for many years to produce accurate quantitation of small molecules (16, 17), and, more recently, this has been extended to peptides (18) and proteins (19–23). The development of isotope-coded affinity tag (ICAT) reagents allows for quantitation through isotopic labeling and simultaneously achieves a reduction in sample complexity (24). These reagents consist of three functional parts: i) an iodoacetamide group that reacts with the free sulfhydryl group of a reduced cysteine side chain, ii) a biotin moiety to aid isolation of modified peptides by avidin affinity chromatography, and iii) a linker group that contains either heavy or light isotopic variants. For the first generation ICAT reagent, this linker region contained either eight deuterium (heavy reagent: d8) or eight hydrogen atoms (light reagent: d0) and therefore conferred a difference in nominal mass of 8 Da between heavy and light reagents. In a typical side-by-side experiment, one sample would be labeled with light reagent and the other with heavy reagent. After attachment of ICAT labels, samples are combined and the cysteine-containing components are affinity purified by means of the biotin tag. After mass spectrometric data acquisition, the resulting mass spectra would be searched for pairs of isotope envelopes differing in mass by 8 Da, and relative quantities of proteins could be determined by comparison of the integrated peak areas of the two corresponding isotope profiles. The collision-induced dissociation (CID) of peptides of interest by tandem MS (MS/MS) would give rise to a sequence-specific fragmentation pattern, from which the identity of the parent protein could be derived by either data base search algorithms or de novo CID spectral interpretation.

However, despite the clear merits of this approach, several shortcomings of this first generation reagent were identified: i) the d8- and d0-modified peptides did not coelute by reverse-phase chromatography, making quantitation less accurate (25); ii) the tag itself was quite bulky, and consequently fragmentation of modified peptides produced many fragments in the CID spectrum related to the tag rather than the peptide (26); (iii) the substantial mass addition resulting from the attachment of the tag could shift the masses of larger peptides outside the optimum range for detection by standard MS instruments; and finally iv) the choice of 8 Da mass difference for the heavy ICAT reagent produced potential isobaric ambiguity between peptides containing two ICAT-labeled cysteine residues (ΔM = 16.100 Da) and the common oxidation of methionine residues (ΔM = 15.995 Da).

Hence, a second generation of such reagents has been developed. The first of these used ICAT reagents immobilized on beads and incorporated a photocleavable linker (27). Capture of the cysteine-containing peptides was followed by photocleavage-based elution of labeled peptides. Although this proved to be more sensitive than the first generation ICAT reagent, it retained the use of deuterium as the isotopic label and consequently still suffered from the chromatographic separation of light- and heavy-modified species.

Here we report the use of a commercial second generation ICAT reagent that contains an acid-cleavable linker group connecting the biotin moiety with the sulfhydryl reactive isotope tag. In this instance, ICAT labeling and biotin-based peptide affinity isolation is followed by acid cleavage, resulting in removal of the biotin moiety. The benefits of this step are the addition of a much smaller chemical moiety to the cysteine residue and improvement in the quality of CID fragmentation spectra obtained from modified peptides, especially larger species. Also, rather than using deuterium as the heavy isotope, this reagent employs nine 13C atoms as the isotopic label for the heavy reagent. Therefore, the heavy- and light-modified peptides coelute by reverse-phase chromatography, making quantitation simpler to achieve and the results more reliable.

In most ICAT studies reported thus far (28–31), sample availability has not been a limiting factor. Total amount of protein used in these studies ranged from 4.4 mg to 200 μg. Here we have sought to apply the technology to low-microgram sample quantities, consistent with the nature of many protein samples of interest in biomedical research. To reduce sample losses during the ICAT protocol, volatile buffers that could easily be removed by vacuum centrifugation were utilized wherever possible. Whereas in conventional ICAT analyses all noncysteine-containing peptides are typically discarded, here these peptides have been retained for separation and identification by multidimensional LC and MS/MS. Our approach serves to combine the differential profiling strength provided by the ICAT strategy with the high sequence coverage afforded by multidimensional LC.

To exemplify the power of this technology, we present data from two projects of biological significance that place high demands on sensitivity and analysis: i) the characterization of proteins binding to the murine prion protein, and ii) the characterization of proteins from human tracheal epithelium gland secretions. Data were collected using two different MS platforms, a MDS-Sciex QSTAR with electrospray ionization (ESI) for on-line nanoflow-high-pressure LC (HPLC) analysis and an Applied Biosystems 4700 Proteomics Analyzer utilizing matrix-assisted laser desorption/ionization (MALDI) with off-line analysis of previously separated nanoflow-HPLC fractions. The former is a quadrupole selection, quadrupole collision cell, orthogonal acceleration time-of-flight instrument (QqTOF), whereas the latter is an axial TOF/TOF instrument. We present a comparison of results on the basis of total number

300 Molecular & Cellular Proteomics 2.5
of peptides detected, total number of proteins identified, and proteins detected and quantitated by cleavable ICAT (cICAT) versus those identified in the flow through of the avidin chromatography. We also consider the accuracy of mass measurement in MS and MS/MS modes, sensitivity in both modes, sample throughput, and ease of use.

EXPERIMENTAL PROCEDURES

**Materials**—Original ICAT reagents (d0/d8) and cICAT reagents (13C0/13C9) and kits were obtained from Applied Biosystems (Framingham, MA). BSA and other control and calibration peptides and proteins were obtained from Sigma (St Louis, MO). Siliconized 0.65-mI tubes from PGC Scientifics (Frederick, MD) were washed with methanol and water prior to use. The polysulfoethyl A strong cation exchange (SCX) columns (50 x 1 mm and 10 x 2.1 mm) were obtained from Poly LC through Western Analytical Products (Murietta, CA). An experimental 20-μl avidin cartridge column was supplied by Applied Biosystems. Reverse-phase packing material was obtained from Phenomenex (Torrance, CA), and fused-silica capillary tubing was purchased from Dionex (Sunnyvale, CA). The matrix solution containing 10% methanol, followed by 1000 μl of 100 mM NH4HCO3, pH 8.0, was obtained from Applied Biosystems. Reverse-phase packing material was obtained from Phenomenex (Torrance, CA), and fused-silica capillary tubing was purchased from Dionex (Sunnyvale, CA). The matrix solution containing 10% methanol, followed by 1000 μl of 100 mM NH4HCO3, pH 8.0, was obtained from Applied Biosystems.

**ICAT Labeling**—Bovine serum albumin (BSA) was reduced, alkylated, and digested according to the standard protocol supplied with the original d0/d8 ICAT kit. The complex protein samples (5–15 μg) were labeled with the cleavable 13C-ICAT reagent using a modified protocol. Briefly, protein samples were denatured in 6 M urea/20 mM NH4HCO3, pH 8.2. Reduction with 1 mM trichloroethylphosphine was allowed to proceed for 20 min at 70 °C. ICAT reagents were dissolved in 20 mM NH4HCO3, pH 8.2, 10% acetonitrile (ACN), and labeling was carried out for 2 h at room temperature. Samples were combined and diluted 4-fold to reduce the concentration of urea to below 1.5 M. Tryptic digestion was initiated by the addition of 1% (w/w) of side-chain modified, tosylphenylalanine chloromethyl ketone-treated porcine trypsin and was allowed to proceed at 37 °C for 4 h.

**Cation Exchange Chromatography**—SCX chromatography was used to remove neutral species from the tryptic peptides and to achieve peptide fractionation of the digest mixture. Tryptic digest samples were adjusted to 25% ACN and acidified (pH 3.0) by the addition of formic acid. HPLC was carried out using a Beckman Gold system equipped with an analytical μ-flow upgrade, with Rheodyne injection port and a 35-μl dead volume ultraviolet cell. Separation was achieved using multiple sample injections onto a 2.1 x 10 mm polysulfoethyl A column with a 240-μl injection loop. Solvent A consisted of 25% ACN, 0.05% formic acid, and solvent B consisted of solvent A with 400 mM NH4HCO3. A typical separation employed 0% B from 0–15 min to allow for sample loading and removal of nonpeptide species, followed by a gradient of 0–50% B from 15–22 min, 50–100% from 22–23 min, and finally the column was washed with a solution of 1 M KCl in solvent A. Fractions were collected in 0.65-ml siliconized tubes.

**Avidin Affinity Purification**—The SCX-eluted fractions were neutralized by the addition of 2 volumes of 100 mM NH4HCO3, pH 9.5, using NH4OH (30%) as necessary to bring the pH up to 8.0. The above-mentioned HPLC was also utilized for the avidin affinity chromatography. The 20-μl avidin cartridge was primed with 1 ml 0.4% trifluoroacetic acid (TFA) in 30% ACN followed by 1 ml of 100 mM NH4HCO3 at pH 8.0. Samples were loaded using multiple injections on a 240-μl injection loop. The column was washed with 500 μl of 50 mM NH4HCO3, pH 8.0, followed by 500 μl of the same solution containing 10% methanol, followed by 1000 μl of HPLC-grade water. The SCX fractions were passed through the avidin column one at a time using a flow rate of 100 μl/min, and the flow through was collected for LC-MS/MS analysis. Labeled peptides were eluted with 200–400 μl of 0.4% TFA in 30% ACN as determined by the absorbance at 218 nm.

**Nano-LC-MALDI-TOF/TOF Mass Spectrometry Analysis**—The SCX fractions and avidin-eluted samples were subjected to nanoflow HPLC using the Ultimate LC system (Dionex) at a flow rate of 300 nL/min. Separation of peptides was achieved by a gradient of increasing ACN in water (2–34%) over 100 min using 0.1% w/v TFA as the ion-pairing agent on a 75-μm ID self-packed column. HPLC eluent was spotted directly onto the MALDI target plate using a Probot spotting robot (Dionex), supplemented with a sheath flow of 500
Mass Spectrometric Analysis of Low Level Protein Samples Using Cleavable $^{13}$C-ICAT

**Protocol Optimization**—To ascertain the extent of derivatization, quantitation efficiency, and accuracy and for overall method optimization when analyzing complex samples in the low-microgram range, several ICAT experiments were performed on peptide and protein standards. The efficiency of ICAT labeling and cleavage was first tested on the peptide laminin (CDPGYIGSR) using a light-to-heavy ratio of 2:3. Reverse-phase HPLC was used to remove hydrophobic contaminants after labeling by using a step elution of 0–80% ACN, conditions that allowed both unlabeled and labeled peptides to be retained. One-half of the resulting peptide mixture was spotted for direct MALDI analysis, and the other half was cleaved without the avidin affinity purification step then dried down and spotted as before. MS spectra obtained with the MALDI-TOF/TOF instrument are illustrated in Fig. 1. The upper panel shows laminin before derivatization at $m/z$ MH$^+ 967.43$ (1 pmol on target). After reaction with the cICAT reagent, the light- and heavy-labeled peptides were observed at $m/z$ 1875.02 and 1884.03 (middle panel, 500 fmol). Upon cleavage, the resulting labeled peptides were detected at $m/z$ 1194.56 and 1203.55, as shown in the lower panel (500 fmol). Although a small peak was observed corresponding to cleaved cICAT-laminin prior to the intentional cleavage reac-

---

**RESULTS**

**Protein Quantitation**—Protein quantitation using ICAT pairs was performed by an initial analysis using two different software systems from Applied Biosystems: GPS Explorer in the case of TOF/TOF data and ProICAT in the case of Qq-TOF data. This was followed by manual confirmation of ICAT-labeled ions using these software programs and manual analysis for those that were not identified in an automated fashion. For manual quantitation, monoisotopic peak intensities were used initially followed by isotope envelope area for proteins that were of significant interest.

---

**Fig. 2. Flow chart of the protocol used for quantitative analysis of low-level protein mixtures using the cICAT reagent.**

nl/min matrix solution (1:1 dilution of $\alpha$-cyano-4-hydroxycinnamic acid with 70% methanol/0.4% TFA) spotting one fraction per minute. The Probot plumbing was replaced with capillary tubing using polyetheretherketone (PEEK) sleeves to reduce void matrix volume. Using this mixture, it was necessary to ensure that the elution capillary protruded no further than 2 mm from the matrix sheath needle, thereby preventing crystallization of the matrix on the tip.

MALDI-MS data were acquired in an automated mode using a 4700 Proteomics Analyzer (Applied Biosystems). This instrument employed a neodymium: yttrium aluminum garnet (Nd:YAG) frequency-tripled laser operating at a wavelength of 354 nm and a laser repetition rate of 200 Hz. Initially, a MALDI-MS spectrum was acquired from each spot (1000 shots/spectrum), then peaks with a signal-to-noise ratio (S:N) greater than 15 in each spectrum were automatically selected for CID (7500 shots/spectrum). A collision energy of 1 keV was used with air as the collision gas for CID accumulation. After acquisition, the data were subjected to automatic baseline correction, mathematically smoothed, and stored in an Oracle data base. Assuming that all ions were singly charged, peaklists from all MS/MS spectra were automatically extracted from the Oracle data base and submitted for batch analysis data base searching using an in-house copy of Protein Prospector (version 4.3, University of California San Francisco, San Francisco, California) with the new program, LCBatchTag, or an in-house copy of Mascot, version 1.8 (Matrix Science). The latter was managed using the Mascot Daemon running on the same computer. MS/MS mass values submitted to both search engines were limited using the following criteria: minimum S:N threshold 8–10, masses of 0–60 Da, and within 20 Da of the precursor were excluded, and a maximum of 60 peaks per spectrum were submitted.

Protein Prospector searches were performed by specifying the inclusion of high-energy fragment ions characteristic of the TOF/TOF instrument, whereas Mascot searches included only the low-energy fragment ions and internal ions. For externally calibrated spectra, the allowed mass tolerance specified between expected and observed masses for searches was $\pm$75 ppm for MS data, $\pm$200 for MS/MS parent ions, and $\pm$250 ppm for MS/MS fragment ions. In cases where internal calibrants were used, the analogous values were $\pm$25, $\pm$25, and $\pm$150 ppm. All samples were searched against the nonredundant National Center for Biotechnology Information data base (NCBiN.10.25.2002).

Nano-LC-ESI-Qq-TOF Mass Spectrometry Analysis—Tryptic peptides were subject to LC-MS/MS analysis on a QSTAR Pulsar mass spectrometer (MDS Sciex, Concord, Ontario, Canada) operating in positive ion mode. Chromatographic separation of peptides was performed as above except that formic acid was used as the ion pairing agent. The LC eluent was directed to a micro-ionspray source. Throughout the running of the LC gradient, MS and MS/MS data were recorded continuously based on a 6-s cycle time. Within each cycle, MS data were accumulated for 1 s, followed by two CID acquisitions of 2.5 s each on ions selected by preset selection parameters of the information-dependant acquisition method. In general, the ions selected for CID were the most abundant in the MS spectrum, except that singly charged ions were excluded and dynamic exclusion was employed to prevent repetitive selection of the same ions within a preset time. Collision energies were programmed to be adjusted automatically according to the charge state and mass value of the precursor ions. Peak lists for data base searching were created using a script from within the Analyst software. Searches were performed using the two search engines as above except that only the low-energy CID fragments characteristic of the ESI Qq-TOF instrument were considered. The allowed mass tolerance range between expected and observed masses for searches was $\pm$100 ppm for MS peaks and $\pm$0.1 Da for MS/MS fragment ions.

Protein Quantitation—Protein quantitation using ICAT pairs was performed by an initial analysis using two different software systems from Applied Biosystems: GPS Explorer in the case of TOF/TOF data and ProICAT in the case of Qq-TOF data. This was followed by manual confirmation of ICAT-labeled ions using these software programs and manual analysis for those that were not identified in an automated fashion. For manual quantitation, monoisotopic peak intensities were used initially followed by isotope envelope area for proteins that were of significant interest.
tion being carried out, this was likely attributable to in-source MS fragmentation. Thus, these results point to relatively clean and efficient labeling.

To further characterize the yield of the derivatization reaction, the efficiency of the chromatographic separation, and the fragmentation properties of the derivatized peptides, low levels of protein standards were analyzed using single proteins and protein mixtures. These consisted of BSA, apo-transferrin in an equimolar ratio (an aliquot of 830 fmol each), and a mixture of these two plus α-galactosidase (40 fmol), cytochrome c (1 fmol), lactoferrin, and carbonic anhydrase (4 fmol each). Samples were solubilized in 9M urea or 0.1% SDS solution, reduced, split based on the desired ratios to be tested, alkylated with either heavy or light cICAT reagent, and then digested with trypsin. Sample clean-up and peptide separation was achieved using three modes of chromatographic separation (Fig. 2).

The first two chromatographic steps were conducted off-line in order to i) increase sample capacity (i.e. columns with different loading capacities and requiring different flow rates could be used), ii) permit the use of larger amounts of ACN, thereby minimizing hydrophobic interactions during SCX chromatography, iii) give superior peptide separation, iv) remove detergents and solubilizing agents, and v) achieve higher overall robustness. SCX was used for the first dimension of chromatographic separation due to its high recovery, ability to remove detergents and reagents, orthogonal separation to reverse phase, and adequate resolving power. To optimize the sensitivity of peptide detection, it was desirable to have the highest concentration of each peptide elute during a given LC-MS/MS run, which requires maximizing the resolution of the preceding chromatographic steps. Furthermore, to maximize protein sequence coverage while at the same time preserving sample, a fairly fast salt gradient was used for elution of the peptides. In addition, the number of fractions collected was dependent on the concentration of peptides as

| TABLE I |
| Cysteine-containing tryptic peptides for transferrin in the range of 700-4000 Da |

Peptides that contain two or more cysteine residues are highlighted. X indicates identification of the given peptide by automated data base searching (no missed cleavages); bold X indicates peptide identified with tryptic missed cleavage; underlined X indicates peptide identified by both tryptic missed cleavage and fully cleaved forms.

| m/z (M) | Start | End | Peptide | ESI | MALDI |
|--------|-------|-----|---------|-----|-------|
| 818.4466 | 542 | 546 | (R) CLVEK (G) | x | x | x | 0.9685 | 0.9470 |
| 667.4147 | 38 | 42 | (K) CGSFR (D) | x | x | x | 1.0380 | 0.9294 |
| 895.4416 | 450 | 494 | (K) INHCR (F) | x | x | 1.0976 |
| 879.4246 | 510 | 515 | (K) DSSLCK (L) | x | 1.0755 |
| 913.4566 | 613 | 618 | (K) EACVHK (I) | x | x | 1.0439 |
| 1057.521 | 468 | 475 | (K) SCHTAVGR (T) | x | x | x | 0.9501 | 0.9466 |
| 1057.521 | 136 | 143 | (K) SCHTGLGR (S) | x | x | x | 1.0258 | 0.9595 |
| 1091.541 | 652 | 659 | (K) DTTGCLK (L) | x | x | 0.9744 | 0.9499 |
| 1167.503 | 62 | 69 | (K) ASYLDCIR (A) | x | x | x | 1.0007 | 1.1912 |
| 1365.449 | 363 | 371 | (K) WCALSHHER (L) | x | x | 0.9868 | 0.9718 |
| 1450.664 | 374 | 384 | (K) CDEWSVNSVGK (I) | x | x | 1.0146 |
| 1487.665 | 27 | 37 | (R) WOAVEHEATK (C) | x | x | x | 1.0376 | 1.0478 |
| 1524.737 | 577 | 587 | (K) DYELLCLGTR (K) | x | x | 0.9973 |
| 1585.826 | 47 | 60 | (K) SVIPSDPGSVACVK (K) | x | x | 1.0146 |
| 1648.658 | 421 | 433 | (K) CGLYVFLAENYK (S) | x | x | 1.0146 |
| 1709.817 | 240 | 251 | (R) DOYELCLDNR (K) | x | x | 0.8139 |
| 1747.764 | 495 | 508 | (R) DFEPEFEGCPGGSK (K) | x | x | x | 1.0309 | 0.9332 |
| 1758.881 | 598 | 600 | (R) KPVEEYANCVALAR (A) | x | x | x | 0.9585 | 0.9741 |
| 1869.965 | 259 | 279 | (K) DCHLAQVSPTVVAR (S) | x | x | x | 1.0020 | 0.9659 |
| 1871.989 | 694 | 696 | (K) CTTSSILAEACTFR (R) | x | x | 1.2440 |
| 2049.378 | 516 | 530 | (K) LOMGSGGLNCELPIK (E) | x | x | 1.0972 |
| 2065.379 | 395 | 399 | (K) IECSAEITDEDOAK (I) | x | x | 1.0972 |
| 2158.016 | 347 | 362 | (R) EGTCPKAPTECPKPKV (K) | x | x | 1.0724 |
| 2242.034 | 434 | 452 | (K) SDNCGDTPASTPGYFAAVVK (K) | x | x | 1.0724 |
| 2341.202 | 144 | 162 | (R) SAGWNPQIGLLCPDPEPR (K) | x | x | 0.9928 | 1.0145 |
| 2492.435 | 3 | 23 | (R) LVSLGIVLAVGADCLAVPK (I) | x | x | 1.0724 |
| 2855.396 | 622 | 642 | (R) QQQHLPINGSNVTDCSNGNCLFR (S) | x | x | 1.0934 |

Average 1.0007 1.0003
Standard Deviation 0.0619 0.0934
indicated by the ultraviolet absorbance. The average mAU range for eluting peptides was 0.002–0.04 with 3–8 fractions being collected during individual runs. Formic acid was used in the cation exchange buffers to control the pH, thereby improving compatibility with downstream mass spectrometry of nonlabeled peptides. For the same reason, cation exchange fractions were neutralized with ammonium bicarbonate buffer for avidin affinity chromatography.

To improve the quality of the quantitative results and to maximize the signal, the ICAT-labeled peptides were retained on the monomeric avidin column until all cation fractions had been passed through and were then eluted in a single fraction. The flow-through fractions and the avidin eluate resulted in \( n \) samples for mass spectrometric analysis. Each of these was split into two identical subfractions, which were subjected to 100 min reverse-phase nanocapillary LC-MS/MS analyses; one using the online ESI-Qq-TOF and the other using the off-line MALDI-TOF/TOF instrument.

One experiment consisted of preparing two samples from the stock six-protein mixture, each containing 5 \( \mu \)g of total protein to give a theoretical ratio of 1. Data from this experiment is described in Tables I and II. Here we compare the number of peptides predicted to contain cysteine with the number detected after the standard cICAT procedure. Labeling of the cysteine thiols and the avidin affinity chromatography both proved to be efficient because no cysteine-containing peptides were detected in the flow-through fractions analyzed. The avidin elution fraction contained primarily cICAT-labeled peptides and, in the case of BSA, one to four nonlabeled peptides. In experiments carried out on BSA alone, the hydrophobic peptide DAFLGSFLYEYSR was found in both labeled and nonlabeled fractions. It was observed that the abundance of this peptide in the cICAT fractions could be decreased by the use of stronger wash conditions. Despite the observation of a limited amount of nonspecific peptide binding, overall the affinity chromatographic separation was highly specific and efficient.

In the case of transferrin, 20 of the 27 predicted cysteine-containing peptides were identified using the cICAT reagent. This corresponded to 33 different peptides with unique heavy masses.

| Table II  |
| Cysteine-containing tryptic peptides for bovine serum albumin in the range of 700-4000 Da |

| mz (m/z) | Start | End | Peptide | ESI L | H | MALDI L | H | ESI | MALDI |
|---------|-------|-----|---------|-------|----|--------|----|-----|-------|
| 876.4613 | 223   | 228 | (R) CASIQK (F) | x     | x | 0.9802 |
| 928.529  | 198   | 204 | (K) GACLKLP (I) |       |    | 0.9911 |
| 1068.588 | 483   | 489 | (R) LCVLHEK (T) | x     | x | 0.7850 |
| 1179.514 | 581   | 587 | (K) CCAADDK (E) | x     | x | 1.3856 |
| 1204.579 | 123   | 130 | (R) NECFLSHK (D) | x     |    | 1.2400 |
| 1233.548 | 413   | 420 | (K) QNCQDFEK (L) | x     |    |         |
| 1242.615 | 310   | 318 | (K) SHCAAEYK (D) | x     | x | 0.9998 |
| 1277.62   | 588   | 597 | (K) EACFAYEAGPK (L) | x     |    |         |
| 1478.71   | 409   | 507 | (R) CTTUSSLVNR (F) | x     | x |         |
| 1506.705  | 400   | 468 | (R) CTCPEESER (M) | x     | x | 0.0080 |
| 1589.8    | 89    | 100 | (K) SLHTFWFDELCK (V) | x     | x | 0.0080 |
| 1613.748  | 286   | 297 | (K) YICDQDQITISSK (L) | x     | x | 0.9072 |
| 1631.814  | 300   | 309 | (K) ECDKPYLIEK (S) | x     | x | 0.5384 |
| 1724.759  | 317   | 399 | (K) DDPHACSTYVFKD (L) | x     | x | 0.9432 |
| 1745.874  | 139   | 151 | (K) KDDPINTLQDFEK (A) | x     | x | 0.9818 |
| 1803.801  | 76    | 88  | (R) TVCAEDHAGCEK (S) | x     | x | 0.4149 |
| 1816.735  | 106   | 117 | (R) ETYQDMADCEK (Q) | x     | x | 0.0548 |
| 1842.825  | 375   | 386 | (K) EYATLLEEOAK (D) | x     | x |         |
| 1894.941  | 469   | 482 | (R) MPMTEYLSLNR (L) | x     | x |         |
| 2051.027  | 506   | 523 | (R) RPOCSALTPEPTYVPK (A) | x     | x | 0.8565 |
| 2073.901  | 184   | 197 | (K) YNGVGFQCCQAEKD (G) | x     | x | 0.9469 |
| 2078.027  | 829   | 844 | (K) LPTFRADICTLQDFEK (Q) | x     | x | 0.9837 |
| 2255.068  | 267   | 280 | (K) ECCHGKLLCADER (A) | x     | x | 0.9920 |
| 2562.37   | 45    | 65  | (K) GLVLAFSQQYLOQCPFDEHKV (L) | x     | x | 0.5160 |

Average 0.7901 0.8783
Standard Deviation 0.2170 0.2870

Average (single cyste peptides only) 0.9284 1.0012
Standard Deviation 0.0750 0.1303

* Masses are calculated for cysteine residues modified with cICAT \(^{13}\)C.0.
or light label in the ESI analysis and 28 in the MALDI data. These numbers are derived from including peptides containing missed tryptic cleavage sites, but not counting peptides containing oxidized methionine residues when the nonoxidized version was also observed. Five peptides were uniquely identified by ESI and three by MALDI. Several peptides were observed containing a missed trypsin cleavage site, but in many of these cases the fully cleaved form of the same peptide was also identified. The occurrence of missed cleavages did not appear to affect the quantitation accuracy, as one would expect considering the ICAT labeling was performed prior to digestion.

The cICAT ratios were calculated using several methods: i) integration of the isotope envelope in a single MS spectrum or the averaged spectrum over the elution time of the peptide; ii) integration of the monoisotopic peak in a single MS spectrum; or iii) measurement of the monoisotopic peak intensity in a single MS spectrum. The maximum standard deviation for these measurements on individual peptides in ESI and MALDI spectra for peptides derived from transferrin was 0.108, with an average standard deviation of 0.047 (data not shown). Thus, all of these approaches gave reasonably accurate quantitation measurements, and there was no significant difference observed in quantitation values or reliability between the two mass spectrometer platforms. After adding a correction factor of 0.16 for ESI and 0.88 for MALDI to each ratio, the average heavy-to-light (H:L) ratio for all peptides derived from transferrin are 1.001 for ESI and 1.003 for MALDI with standard deviations of 0.062 and 0.093, respectively. The correction factor was calculated by taking the average of all ratios in the data set and subtracting from the theoretical ratio of 1. While the difference in the correction factors for these two platforms is unclear, their origin can be explained in part by the amount of $^{13}$C$_8$ reagent impurity in the heavy-labeled tag (see Fig. 9, m/z 582.83). This example is an indication of the accuracy and precision achievable with this method when several cICAT-labeled peptides are identified from the same protein and a protein is present with a known ratio so that an accurate correction factor can be applied.

The quantitation results for BSA with its 35 cysteine residues resulted in a greater range of H:L ratios for individual peptides and as a result greater standard deviation for the overall protein H:L ratio. The measured ratios for albumin where 0.7901 for ESI and 0.8381 for MALDI with standard deviations of 0.2170 and 0.2728, respectively. Upon closer inspection, it was found that the ratios derived from peptides with multiple cysteine residues gave much lower H:L values than the theoretical ratio of 1. In fact, if these peptides are not used for the quantitation, the H:L ratios become 0.928 and 0.919, respectively, with standard deviations of 0.075 and 0.031. A possible explanation for this observation is that this cysteine-rich protein was not fully reduced in the sample that was to be labeled with the $^{13}$C$_9$ reagent. This lowered the
amount of reduced and alkylated peptides in the heavy-labeled sample and therefore the H:L ratio; for peptides with multiple cysteines the effect was magnified to give much lower ratios. This effect seemed to be specific to the protein rather than multiple-cysteine-containing peptides in general, as multiple-cysteine-containing peptides were found in other proteins in our studies that gave expected ratios (i.e. transferrin peptide EGTCPEAPTDECKPVK, ratio 0.97).

A shortcoming of the original ICAT reagent, and indeed of any labeling method that relies on the \(^1\)H and \(^2\)H isotopes, is that peptides labeled with this reagent pair do not coelute by reverse-phase chromatography (25). BSA was labeled in a 1:1 ratio using the original ICAT, and cICAT reagents. Five micrograms of total protein was carried through each procedure, but to simulate the low levels of protein anticipated in "real" biological samples, only 2.5% of the resulting sample from each was analyzed by LC-MALDI-MS and 2.5% by LC-ESI-MS (corresponds to \(\sim 830 \text{ fmol BSA and transferrin each}\)).

The peptide SHC\(^*\)IAEVEK labeled with the D8 tag eluted earlier than the D0-tagged peptide (Fig. 3a), whereas this same peptide when labeled with the cICAT reagent showed coelution of both the light and heavy variants (Fig. 2b). Panels a, b, and c indicate that the observed ratio of peptides derivatized with light or heavy original ICAT reagent changed substantially as they eluted from the LC column, whereas d, e, and f showed this was not the case for cICAT. Thus, quantitation accuracy was found to be more reliable using the cICAT reagent. These results were representative of all peptides studied, regardless of their elution time in a given LC run.

The success of a differential profiling method that uses MS/MS analysis relies on the use of reagents that do not detract from the quality of the resulting CID spectra. We compared tryptic digests of protein standards labeled with iodoacetic acid, the original ICAT, and cICAT reagents to evaluate this aspect of the reagents. Fig. 4 compares the fragmentation of a tryptic peptide with the sequence SLHTLFGDELC\(^*\)K derived from BSA derivatized with each of these three reagents. The majority of the potential b and y ions were observed in all of the spectra. However, upon closer inspection of ions in the lower m/z range it is apparent that the cICAT reagent and iodoacetic acid-labeled peptides contain significantly more ions from fragmentation of the peptide moiety itself, whereas the spectra of the original ICAT-labeled peptides resulted in more fragment ions that were not peptide related. In other studies, these nonpeptide-related fragment ions have been assigned to the ICAT reagent (26) and have aided in the identification of ICAT pairs by precursor ion scanning (32).

The development of ESI and MALDI methods have been pivotal in achieving the current level of power for the mass spectrometric characterization of biological macromolecules (33). It has been observed by us and others that the types of peptides that are detected using the two modes of ionization do not overlap completely (32). Data were collected using ESI on a Qq-TOF mass spectrometer and by MALDI on a TOF/TOF instrument. On the latter instrument, samples were analyzed either without chromatographic separation or after fraction collection from reverse-phase HPLC directly onto MALDI targets using a spotting robot. Data generated on both instruments were submitted to data base searching, both individual and combined data sets.

Shown in Fig. 5 are the CID spectra of the labeled peptide VVEQMC\(^*\)VTQYQK, in the cICAT heavy and light forms, acquired with these two different instrument systems. This peptide was obtained in the course of the identification of proteins that interact with the prion protein (PrP), and in fact is a peptide derived from PrP. The fragmentation pattern is representative of some general features differentiating the two mass spectrometric measurements. One obvious difference is the distribution of fragment ions: using ESI we observed more y series ladders, whereas the MALDI spectrum showed slightly less preference for y series but an increase in the internal ions. In general, for MALDI-TOF/TOF analysis at the level of tens of femtomoles (or S:N >30–50 in the MS scan), large numbers of fragment ions are observed. Such CID spectra tend to give high scores when matched by data base searching, as shown by the example in Fig. 5, c and d, and...
therefore higher protein identification confidence levels. Conversely, for precursor ions of low signal-to-noise ratios, the general trend is for only a few fragment ions to be observed (Fig. 6a). For analyses carried out with the ESI Qq-TOF platform, the number of fragment ions was found to be less dependent on the precursor ion intensity, although as expected the y and b fragment ions of higher m/z tend to regress into the noise for the lowest intensity precursor ions.

For online LC-MS/MS, there is a trade-off between the number of ions selected for CID and the quality of the resulting spectra, which to some extent is overcome by using off-line LC-MALDI on the TOF/TOF mass spectrometer. However, if too many ions are selected for CID from a given spot, the peptides will be depleted before all spectra have been acquired. This generally occurs only when the individual peptides are in the low-femtomole range and more than 10–20 different species are selected for CID analysis.

Prion Protein-Containing Complexes—In one collaborative project, the combined multidimensional chromatography/ciCAT protocol introduced above was employed to identify proteins that interact with the cellular PrP in mice. Our goal was to identify non-PrP components within immunoaffinity-purified protein complexes that contain the PrP. In particular, it was envisioned that such an approach could yield information about the cellular microenvironment and function of PrP.
In order to distinguish specific PrP interactors from proteins that copurify nonspecifically, a negative control sample in which the PrP-specific antibody was omitted was processed in parallel. High sensitivity was particularly important in this study because only a limited amount of sample (i.e. low-microgram quantities) could be obtained.

From analysis of the mass spectrometric data, the identification of several proteins not previously implicated in binding/interacting with the PrP was established with high confidence. A number of proteins were common to both the negative control sample and the PrP-specific pull down. These included BSA that was employed to saturate unspecific binding sites of chromatography matrices, avidin, and keratins 1 and 9. Furthermore, in the PrP-specific pull-down sample, both PrP and the PrP-directed antibody employed for immunofinity purification gave rise to a number of high-quality CID spectra. Nevertheless, of the ~50 proteins identified with high confidence, ~20 were unique to the test sample and absent from the control. Among these were N-CAM1, a known interactor of the cellular PrP (34), and N-CAM2, a low-abundance paralogue of N-CAM1 that is predominantly expressed in the olfactory bulb (35). Two of the CID spectra that aided the identification of N-CAM1 are shown in Fig. 6, a and b. Fifty-five peptides were identified as belonging to this protein, with 25 being nonredundant. Ten of these were identified using each of the mass spectrometer platforms, nine were unique to ESI-collected data, whereas six were unique to MALDI. Fig. 6c shows the sequence coverage obtained.

It was anticipated that quantitation by cICAT might permit the identification of proteins specifically involved in the devel-

![CID spectra](image-url)
velopment of prion diseases. To this end, a "dominant negative" mouse strain was employed that expresses physiological levels of a mutated PrP on a wild-type PrP-ablated background. Previously, it had been shown that this point mutation renders mice resistant to infection with prions (36). The quantitative comparison of samples derived from wild-type and PrP-mutant mice revealed no significant differences in the abundances of the identified proteins specific to the PrP pull-down samples. Some examples of identified proteins and corresponding H:L ratios are PrP (1.14), PrP-specific Fab (1.02), N-CAM1 (1.01), and contactin 1 (0.98). However, some of the proteins identified in the negative control did show changes in abundance ratios, such as the glycolytic enzyme glyceraldehyde-3-phosphate dehydrogenase (1.85), propionyl CoA carboxylase alpha subunit (1.67), and Na/K ATPase beta subunit (1.43). Therefore, on this occasion the combination of multidimensional chromatography and cICAT offered no significant advantage over multidimensional chromatography alone in terms of identifying proteins involved in the development of prion diseases. Detailed results of these experiments will be described elsewhere.2

Tracheal Epithelium Gland Secretions—In a second representative project, we investigated the changes in the proteome of the human airway lining fluid in a patient with cystic fibrosis in order to explore mechanisms of disease in these patients. Tracheal tissue was obtained from the explanted lung of a cystic fibrosis patient following lung transplantation. Tracheal tissue from the lungs of a donor that were not selected for transplantation was used as a control. After preparation of the trachea and cleaning of the tracheal epithelium, gland secretions were pipetted directly from the glands and collected. From each specimen, 3.5 μl of fluid was obtained.

2 Schmitt-Ulms, G., Hansen, K. C., Liu, J., Cowdrey, C., Yang, J., DeArmond, S. J., Cohen, F. E., Prusiner, S. B., and Baldwin, M. A., manuscript in preparation.
An analysis was carried out using the protocols described above yielding close to 7000 CID spectra, from which a large number of proteins was identified. A summary of the proteins identified plotted versus the number of cysteine peptides they contain (in the mass range of 700–4000 Da) is presented in Fig. 7 and is reviewed in the “Discussion.” Several proteins appeared to be at a high concentration based on the number of peptides identified. These included human serum albumin (HSA), mucin, lactotransferrin, serotransferrin, immunoglobulin, and several types of keratins. In addition, several proteins were identified that are believed to be expressed at low levels, such as kinases, receptors, and other signaling proteins and peptides. As an example of sequence coverage among the strongly represented proteins, HSA gave almost 94% peptide sequence coverage of the protein. Within the ICAT fractions, coverage of the cysteine-containing peptides for HSA was 90% (20 of 22) with only the tryptic peptides CCK and ET-YGEMADCAK being missed.

By contrast with the PrP discussed above, the cICAT component of this study proved to be very informative. Of the top 100 protein matches, only 10 peptides were detected in the avidin eluant that were noncysteine containing. One such peptide was the HSA peptide HPYFYAPELLFFAK, which like the BSA peptide described earlier contains several aromatic residues. Eight proteins were identified purely on the basis of cICAT-labeled peptides, i.e. no noncysteine-containing peptides were identified for these proteins. One of these was the cystic fibrosis antigen calgranulin A, identified on the basis of the peptide LLETEC*POYIR13C9-cICAT. ESI-MS and CID spectra for this peptide are shown in Fig. 8. The cICAT ratio was used for the quantitation of this protein which, as predicted (37), was up-regulated in the cystic fibrosis sample. Other results were also consistent with previous studies that demonstrated increased defensin levels in inflammatory lung diseases, such as cystic fibrosis (38–40). Fig. 9 shows data for a tryptic peptide found in the defensin DPH-3, for which four cICAT pairs indicate a 3.66-fold up-regulation (standard deviation of 0.32, 9%). Besides their well-known role in antimicrobial defense, defensins also modulate inflammatory responses, stimulate the specific immune response, and contribute to tissue repair (38). These two examples based on the cICAT results clearly define proteins that are up-regulated in cystic fibrosis. Others were identified that will be presented elsewhere.3

Data Handling and Analysis—During the analysis of the large number of spectra that were acquired in this work, we employed three search engines: Mascot (Matrix Science, matrixscience.com), ProID (Applied Biosystems), and an in-house developmental version of Protein Prospector. For the ESI-Qq-TOF data, peak lists were extracted for Mascot searches using a purpose by designed mascot script supplied with the Analyst software. The peak list was filtered such that only peaks greater than 2% of the spectral base peak intensity were submitted for searching. Peak lists for Protein Prospector were filtered by removing all peaks below 3 counts in intensity.

For 4700 data, the script “peak to mascot” was used to create a peak list that was filtered on the basis of a minimum S:N threshold. In the case of Protein Prospector, we employed a new in-house program, PeakSpotter, to extract peak lists from TOF/TOF spectra that had been stored within the Oracle data base also filtering the peak list on the basis of S:N. For each data type, peak lists of all SCX fractions were combined into one text file for searching, and the cICAT fraction was searched both separately and as a combined list. For the identification of the major constituents of each sam-

3 Hirsch, J., Song, Y., Hansen, K. C., Thiagarajah, J. R., Matthay, M. A., Burlingame, A. L., and Verkman, A. S., manuscript in preparation.
the addition of 9 Da per cysteine, there is no ambiguity between peptides containing two cysteines and a peptide containing methionine sulfoxide, the common oxidation product of methionine. After cleavage of the linker, the resulting chemical tag is relatively small and does not itself yield major fragment ions under conditions used for high- or low-energy CID.

The key components required for optimizing the characterization of low-level biological samples are the quality of sample preparation, minimization of sample handling/losses, robustness of the protocol, mass spectrometric sensitivity, and the quality of the data processing and analysis. Sensitivity was important in this study because only limited amounts of sample were available. Other features critical for success of the technique were careful sample preparation to provide samples that were compatible with derivatization, efficient chromatographic separation, and the availability of high performance down-stream tandem mass spectrometric analysis. While the two biological projects outlined in this work were very different, both samples shared high complexity, low protein abundance, and were relatively difficult to extract from their natural biological sources.

It is well established that chromatographic separation of protein digests increases data density, i.e. the number of peptides that can be identified in complex samples by MS (32). In particular, peptides of low abundance or low ionization efficiency are more readily identified in separated digests, where suppression (41, 42) of ionization by other peptides is minimized. To further increase peptide coverage, datasets were collected from two different combinations of ionization strategies and machine architectures: on-line LC-ESI on a Qq-TOF instrument versus off-line collection of LC fractions for subsequent analysis using the TOF/TOF instrument (LC-MALDI). From these studies, we found that from 20 to 50% of all peptides detected were unique to the individual ionization strategies/instrument employed. The collection of such complementary datasets proved particularly beneficial whenever low-abundance proteins gave rise to low-confidence protein assignments based on a single ionization strategy/instrument. The overall amount of peptide needed for successful protein identification by MS/MS was found to be similar on both instruments (low numbers of femtomoles loaded or injected for an individual protein). In our hands, ESI tended to be more sensitive for the very lowest-level complex mixtures. Peptides of lower molecular mass were generally favored by LC-ESI, whereas LC-MALDI tended to identify fewer but larger peptides, thereby giving approximately equal percentage of protein coverage. It is also noteworthy that larger peptides generally give more definitive protein identifications, therefore LC-MALDI is a key element in the application of this methodology for all but the most sample-limited analysis.

Although the analysis of unseparated digests by MALDI MS/MS is ~10 times faster than ESI LC-MS/MS, off-line separation for MALDI as used in this work precedes mass analysis and throughput is approximately two to four times
slower than ESI LC-MS/MS experiments. This disadvantage of LC-MALDI may be outweighed by the fact that data can be collected from individual LC fractions for extended periods of time. On the other hand, the high sensitivity of the Qq-TOF instrument might allow for multiple injections of smaller fractions of material, combined with the use of exclusion lists created from previously identified proteins, thereby minimizing some of the restrictions imposed by the time-dependent nature of LC-ESI experiments. It should be noted that these comparisons are highly dependent on the sample and instrument configuration. For complex protein samples in the regime of hundreds of micrograms or more, the throughput may well be similar for either approach.

In our control experiments, quantitation accuracy for most peptides was within 10% deviation. However, ratios derived from some multiple-cysteine-containing peptides from albumin were off by as much as 60%. We have used as a threshold a greater than 30% ICAT ratio difference as a minimum change to report a change in protein abundance. In the PrP experiments, we expected to observe changes in protein levels for a few proteins at most. The experimentally introduced proteins identified in the experiment also served as a control for the cICAT ratios (Neutravidin 0.97, Fab or IgG-related 1.02). None of these internal standards gave cICAT ratios differing by greater than 10% from the theoretical ratio of 1, leading us to believe that the reduction and alkylation was complete for these proteins. Of the proteins that quantitation data was obtained that were specific to the PrP sample, none resulted in significant changes in expression levels.

In our initial analysis of the cystic fibrosis dataset, almost 7000 CID spectra submitted for data base searching yielded ~1500 protein identifications that were based on a Protein Prospector score of >10 and/or a Mascot score of >40. From our experience, such cut-off criteria lead to a high number of false positives. For identifications close to the cut-off, the false positive rate can be greater than 75%. Using more stringent criteria, such as not allowing protein IDs based on several low-scoring peptides, the final number of high-confidence identifications was 311. Based on manual interpretation of a random sampling of spectra, we believe that the rate of false positives within these protein identifications is less than 1 in 40.

Of the 311 proteins identified in the cystic fibrosis sample, 285 contain at least one cysteine residue. By contrast, 72 proteins were identified on the basis of cICAT-labeled peptides (Fig. 7b). There are many factors that contribute to this low overall efficiency of detection by ICAT. First, it should be noted that the average protein was predicted to give 41 peptides in the mass range 700–4000 Da, only 8 of which would contain cysteine. By contrast, the average number of peptides detected per protein was only 3, therefore many of the cysteine-containing peptides would be missed. Of course, many proteins have much fewer than 8 cysteines, and such proteins are less likely to be detected. Furthermore, such low-level analysis is likely to overlook certain peptides due to suppression effects. The presence of chemical or post-translational modifications and the occurrence of nonspecific cleavages will give peptides of unexpected mass that will not, in many cases, be correctly identified in the database search using the current strategies. Furthermore, the information-dependent acquisition methodology employed in the ESI Qq-TOF will overlook peptides that cannot be selected for CID as they coelute with others giving stronger signals or in some cases side-products of the cleavage reaction. In the case of LC-MALDI, suppression (even with LC separation) and limited sample amounts may restrict the number of peptides that can be analyzed in any one fraction.

The value of combining multidimensional chromatography and cICAT strategies is that we obtain very comprehensive analysis of the sample composition with the opportunity to derive accurate quantitation on a subset of proteins. Whether this extra layer of information is useful will depend upon the nature of any specific research goal. Thus, we obtained no new information on up- or down-regulation of the interaction partners of the PrP, whereas several proteins of interest in cystic fibrosis research were revealed by the cICAT ratios. As for the precision of the quantitation, changes in protein abundance of greater than 30% were detected with confidence, with 10% being typical. Thus, quantification was found to be
at least as accurate as gel analysis when using silver or Coomassie Brilliant Blue staining or mRNA profiling by expression array methods (43). While we have not performed a systematic analysis of the effective dynamic range for heavy- to-light quantitation, results from these experiments give a sense of the usable range. A number of protein expression differences of up to a factor of 5 were observed and showed deviations within the errors described above. Differences observed in the range of a factor of 10 were infrequent, and such assignments are, in most cases, ambiguous or unreliable as the peptide derived from the less-abundant sample would not be selected for CID in our analyses. If a clear identification was made for one of these “singlets,” it would not necessarily indicate that the protein was present in one sample and not the other. Fig. 8 illustrates the difficulty of achieving quantitative results when the intensity of one of the cICAT-labeled components is close to the signal-to-noise or chemical noise level of the MS scan. Precise measurement of the isotope envelope area of the light-labeled peptide in this example cannot be calculated accurately, although the ratio of the monoisotopic peak intensities gives an approximation of the ratio. In general, complex mixtures contain significant amounts of chemical noise in any given MS scan, which can be falsely interpreted as representing a potential cICAT partner. As a result, abundance differences greater than 5-fold between samples are likely to result in the more abundant ions being treated as singlets. This indicates the maximum realistic dynamic range of this technique when working with the small amounts of protein described in this work. However, for many biological questions a 5-fold or greater difference in protein abundance represents a significant change, and follow-up studies can provide more precise results. It should also be noted that a disadvantage of this or any other technique that acquires quantitative information at the peptide level rather than the protein level is the loss of information concerning post-translational modifications.

In summary, the technical methodology developed here allows for the comprehensive analysis of large protein complexes and substantially improves the sequence coverage obtained for low-level samples, thereby resulting in more protein identifications with higher confidence. This has been illustrated here by the successful application of this approach to samples of biological interest that would be difficult to analyze by other methods. The combination of SCX chromatography and cICAT labeling gives a broad and comprehensive picture of all proteins present in a complex sample, while simultaneously providing relative quantitative data on a significant fraction of the proteins identified. The combination of two different separation and analysis platforms also yields complementary information that greatly improves the confidence in the identifications of the less-abundant proteins, which incidentally may represent the species of greatest interest.

Acknowledgments—We thank David Maltby and Katalin Medzihradzsky for technical assistance. We also thank Lan Huang and Peter Baker for support and development of Protein Prospector software. The original ICAT reagents (d13/d16) and cICAT reagents (13C0/13C9) and kits were the generous gift of Applied Biosystems. We are grateful to Tony Hunt of Applied Biosystems for helpful information and advice. We also thank Michael A. Matthey, James A. Frank, Yuanlin Song, and Xiaohui Fang from the Cardiovascular Research Institute (University of California, San Francisco) for help with sample preparation.

REFERENCES

1. Clauser, K. R., Hall, S. C., Smith, D. M., Webb, J. W., Andrews, L. E., Tran, H. M., Epstein, L. B., and Burlingame, A. L. (1995) Rapid mass spectrometric peptide sequencing and mass matching for characterization of human melanoma proteins isolated by two-dimensional PAGE. Proc. Natl. Acad. Sci. U. S. A. 92, 5072–5076
2. Qiu, Y., Benet, L. Z., and Burlingame, A. L. (1998) Identification of the hepatic protein targets of reactive metabolites of acetaminophen in vivo in mice using two-dimensional gel electrophoresis and mass spectrometry. J. Biol. Chem. 273, 17940–17953
3. Huber, L. A. (2003) Opinion: Is proteomics heading in the wrong direction? Nat. Rev. Mol. Cell. Biol. 4, 74–80
4. Celis, J. E., and Gromov, P. (2003) Proteomics in translational cancer research: Toward an integrated approach. Cancer Cell 3, 9–15
5. Zabel, C., Chamrad, D. C., Priller, J., Woodman, B., Meyer, H. E., Bates, G. P., and Klose, J. (2002) Alterations in the mouse and human proteome caused by Huntington's disease. Mol. Cell. Proteomics 1, 366–375
6. Klose, J. (1975) Protein mapping by combined isolectric focusing and electrophoresis of mouse tissues. A novel approach to testing for induced point mutations in mammals. Humangenetik 26, 231–243
7. O'Farrell, P. H. (1975) High resolution two-dimensional electrophoresis of proteins. J. Biol. Chem. 250, 4007–4021
8. Gauss, C., Kalkum, M., Lowe, M., Lehrrach, H., and Klose, J. (1999) Analysis of the mouse proteome. I. Brain proteins: separation by two-dimensional electrophoresis and identification by mass spectrometry and genetic variation. Electrophoresis 20, 575–600
9. Huang, L., Baldwin, M. A., Maltby, D. A., Medzihradzsky, K. F., Baker, P. R., Allen, N., Rexach, M., Edmondson, R. D., Campbell, J., Juhasz, P., Martin, S. A., Vestal, M. L., and Burlingame, A. L. (2002) The identification of protein-protein interactions of the nuclear pore complex of Saccharomyces cerevisiae using high throughput matrix-assisted laser desorption ionization time-of-flight tandem mass spectrometry. Mol. Cell. Proteomics 1, 434–450
10. Washburn, M. P., Wolters, D., and Yates, J. R., 3rd. (2001) Large-scale analysis of the yeast proteome by multidimensional protein identification technology. Nat. Biotechnol. 19, 242–247
11. Huang, L., Jacob, R. J., Pegg, S. C., Baldwin, M. A., Wang, C. C., Burlingame, A. L., and Babbott, P. C. (2001) Functional assignment of the 20 S proteosome from Trypanosoma brucei using mass spectrometry and new bioinformatics approaches. J. Biol. Chem. 276, 28327–28339
12. Pun, T., Hochstrasser, D. F., Appel, R. D., Funk, M., Villars-Augsburger, V., and Pellegrini, C. (1998) Computerized classification of two-dimensional gel electrophoretograms by correspondence analysis and ascendant hierarchical clustering. Appl. Theor. Electrophor. 1, 3–9
13. Raman, B., Cheung, A., and Marten, M. R. (2002) Quantitative comparison and evaluation of two commercially available, two-dimensional electrophoresis image analysis software packages, Z3 and Melanie. Electrophoresis 23, 2194–2202
14. Unlu, M., Morgan, M. E., and Minden, J. S. (1997) Difference gel electrophoresis: a single gel method for detecting changes in protein extracts. Electrophoresis 18, 2071–2077

* This work was supported by Grants RR-01614, 14606, and 12961 from the National Institutes of Health.

§ To whom correspondence should be addressed. Tel.: 415-476-4895; Fax: 415-502-1655; E-mail: khansen@itsa.ucsf.edu.

** To whom correspondence should be addressed. E-mail: alb@itsa.ucsf.edu.
Mass Spectrometric Analysis of Low Level Protein Samples Using Cleavable $^{13}$C-ICAT

15. Gharbi, S., Gaffney, P., Yang, A., Zvelebil, M. J., Cramer, R., Waterfield, M. D., and Timms, J. F. (2002) Evaluation of two-dimensional differential gel electrophoresis for proteome expression analysis of a model breast cancer cell system. Mol. Cell. Proteomics 1, 91–98

16. Dube, G., Henrion, A., and Ohlendorf, R. (1999) Memory effects in combining isotope ratio monitoring with isotope dilution mass spectrometry. Rapid Commun. Mass Spectrom. 13, 1186–1188

17. Wolfe, R. R., Evans, J. E., Mullany, C. J., and Burke, J. F. (1980) Measurement of plasma free fatty acid turnover and oxidation using [1–13C]palmitic acid. Biomed. Mass Spectrom. 7, 168–171

18. Barr, J. R., Maggio, V. L., Patterson, D. G., Jr., Cooper, G. R., Henderson, L. O., Turner, W. E., Smith, S. J., Hannon, W. H., Needham, L. L., and Sampson, E. J. (1996) Isotope dilution–mass spectrometric quantification of specific proteins: Model application with apolipoprotein A-1. Clin. Chem. 42, 1676–1682

19. Stocklin, R., Vu, L., Vadas, L., Cerini, F., Kippen, A. D., Offord, R. E., and Rose, K. (1997) A stable isotope dilution assay for the in vivo determination of insulin levels in humans by mass spectrometry. Diabetes 46, 44–50

20. Goodlett, D. R., Keller, A., Watts, J. D., Newitt, R., Yi, E. C., Purvine, S., Eng, J. K., von Hailer, P., Aebersold, R., and Kolker, E. (2001) Differential stable isotope labeling of peptides for quantitation and de novo sequence derivation. Rapid Commun. Mass Spectrom. 15, 1214–1221

21. Yao, X., Freas, A., Ramirez, J., Demirev, P. A., and Fenselau, C. (2001) Proteolytic 18O labeling for comparative proteomics: model studies with two serotypes of adenovirus. Anal. Chem. 73, 2836–2842

22. Berger, S. J., Lee, S. W., Anderson, G. A., Pasa-Tosic, L., Tolic, N., Shen, Y., Zhao, R., and Smith, R. D. (2002) High-throughput global peptide proteomic analysis by combining stable isotope amino acid labeling and data-dependent multiplexed-MS/MS. Anal. Chem. 74, 4994–5000

23. Barnidge, D. R., Dratz, E. A., Martin, T., Bonilla, L. E., Moran, L. B., and Lindall, A. (2003) Absolute quantification of the G protein-coupled receptor rhodopsin by LC/MS/MS using proteolysis product peptides and synthetic peptide standards. Anal. Chem. 75, 445–451

24. Gygi, S., Han, D., Gingeras, A., Sonenberg, N., and Aebersold, R. (1999) Protein analysis by mass spectrometry and sequence database searching: tools for cancer in the post-genomic era. Electrophoresis 20, 310–319

25. Zhang, R., Sioma, C. S., Wang, S., and Regnier, F. E. (2001) Fractionation of isotopically labeled peptides in quantitative proteomics. Anal. Chem. 73, 5142–5149

26. Borisov, O. V., Goshe, M. B., Conrads, T. P., Rakov, V. S., Veenstra, T. D., and Smith, R. D. (2002) Low-energy collision-induced dissociation fragmentation analysis of cysteiny1-modified peptides. Anal. Chem. 74, 2284–2292

27. Zhou, H., Ranish, J. A., Watts, J. D., and Aebersold, R. (2002) Quantitative proteome analysis by solid-phase isotope tagging and mass spectrometry. Nat. Biotechnol. 20, 512–515

28. Gygi, S. P., Rist, B., Griffin, T. J., Eng, J., and Aebersold, R. (2002) Proteome analysis of low-abundance proteins using multidimensional chromatography and isotope-coded affinity tags. J. Proteome Res. 1, 47–54

29. Lee, H., Griffin, T. J., Gygi, S. P., Rist, B., and Aebersold, R. (2002) Development of a multiplexed microcapillary liquid chromatography system for high-throughput proteome analysis. Anal. Chem. 74, 4353–4360

30. Han, D. K., Eng, J., Zhou, H., and Aebersold, R. (2001) Quantitative profiling of differentiation-induced microsomal proteins using isotope-coded affinity tags and mass spectrometry. Nat. Biotechnol. 19, 946–951

31. Griffin, T. J., Han, D. K., Gygi, S. P., Rist, B., Lee, H., Aebersold, R., and Parker, K. C. (2001) Toward a high-throughput approach to quantitative proteomic analysis: expression-dependent protein identification by mass spectrometry. J. Am. Soc. Mass Spectrom. 12, 1238–1246

32. Baldwin, M. A., Medzhitov, K. F., Lock, C. M., Fisher, B., Settineri, T. A., and Burlingame, A. L. (2001) Matrix-assisted laser desorption/ionization coupled with quadrupole/orbital acceleration time-of-flight mass spectrometry for protein discovery, identification, and structural analysis. Anal. Chem. 73, 1707–1720

33. Burlingame, A. L., Boyd, R. K., and Gaskell, S. J. (1998) Mass spectrometry. Anal. Chem. 70, 647R–716R

34. Schmitt-Ulms, G., Legname, G., Baldwin, M. A., Ball, H. L., Bradon, N., Bosque, P. J., Crossin, K. L., Edelman, G. M., DeArmond, S. J., Cohen, F. E., and Prusiner, S. B. (2001) Binding of neural cell adhesion molecules (N-CAMs) to the cellular prion protein. J. Mol. Biol. 314, 1209–1225

35. Ronn, L. C., Hartz, B. P., and Bock, E. (1998) The neural cell adhesion molecule (NCAM) in development and plasticity of the nervous system. Exp. Gerontol. 33, 853–864

36. Penier, V., Kaneko, K., Safar, J., Vergara, J., Tremblay, P., DeArmond, S. J., Cohen, F. E., Prusiner, S. B., and Wallace, A. C. (2002) Dominant-negative inhibition of prion replication in transgenic mice. Proc. Natl. Acad. Sci. U. S. A. 99, 13079–13084

37. Passey, R. J., Xu, K., Hume, D. A., and Geczy, C. L. (1999) S100A8: emerging functions and regulation. J. Leukocyte Biol. 66, 549–556

38. Schutte, B. C., and McCray, P. B., Jr. (2002) [beta]-defensins in lung host defense. Annu. Rev. Physiol. 64, 709–748

39. Aarbiou, J., Ertmann, M., van Wetering, S., van Noort, P., Rook, D., Rabe, K. F., Litvinov, S. V., van Krieken, J. H., de Boer, W. I., and Hiemstra, P. S. (2002) Human neutrophil defensins induce lung epithelial cell proliferation in vitro. J. Leukocyte Biol. 72, 167–174

40. Aarbiou, J., Rabe, K. F., and Hiemstra, P. S. (2002) Role of defensins in inflammatory lung disease. Ann. Med. 34, 96–101

41. Cech, N. B., and Enke, C. G. (2000) Relating electrospray ionization response to nonpolar character of small peptides. Anal. Chem. 72, 2717–2723

42. Kratzer, R., Eckerskorn, C., Karas, M., and Lottspeich, F. (1998) Suppression effects in enzymatic peptide ladder sequencing using ultraviolet-matrix assisted laser desorption/ionization-mass spectrometry. Electrophoresis 19, 1910–1919

43. Griffin, T. J., Gygi, S. P., Ideker, T., Rist, B., Eng, J., Hood, L., and Aebersold, R. (2002) Complementary profiling of gene expression at the transcriptome and proteome levels in Saccharomyces cerevisiae. Mol. Cell. Proteomics 1, 323–333