Analytical Characteristics of Cleavable Isotope-Coded Affinity Tag-LC-Tandem Mass Spectrometry for Quantitative Proteomic Studies

Cecily P. Vaughn,* David K. Crockett,*
Megan S. Lim,*† and
Kojo S. J. Elenitoba-Johnson*†

From the ARUP Institute for Clinical and Experimental
Pathology,* Salt Lake City; and the Department of Pathology,†
University of Utah Health Sciences Center, Salt Lake City, Utah

Quantitative proteomic studies using cleavable isotope-coded affinity tags (cICAT) in concert with tandem mass spectrometry (MS/MS) permit unbiased comparisons between biologically distinct samples. We sought to determine the analytical characteristics of cICAT-based studies by examining the cumulative results of multiple, separate cICAT-based experiments involving human lymphoma-derived cells. We found that the number of identified proteins increased with larger numbers of fractions analyzed. The majority of proteins were identified by single peptides. Only 24 to 41% of the peptides contained cysteine residues, but 85% of the cysteine-containing peptides yielded quantification data. Approximately 28% of all identified proteins yielded quantification data, with 57% of these being differentially expressed by at least 1.5-fold. The quantification ratios of peptides for proteins with multiple quantified peptides were concordant in trend in 87% of instances. cICAT-labeled peptides identified proteins in all subcellular compartments without significant bias. Analysis of the flow-through fraction did not increase the number of peptides identified per protein. Our studies indicate that cICAT-LC-MS/MS yields quantifications primarily based on single peptides, and analysis of flow-through peptides does not contribute significantly to the results. Nevertheless, identifications based on single cICAT-labeled peptides with tryptic ends provide sufficiently reliable protein identifications and quantification information in cICAT-LC-MS/MS-based proteomic studies. (*J Mol Diagn* 2006, 8:513–520; DOI: 10.2353/jmoldx.2006.060036)

quantitative proteomic analysis.¹ Despite recent improvements in the technology, two-dimensional gel electrophoresis is somewhat limited in its capacity to detect or discriminate low abundance proteins, co-migrating proteins, and proteins with very low or high isoelectric points or molecular weight.² A more recently developed method, stable isotope labeling by amino acids in cell culture, involves the incorporation of isotopically distinct amino acids into growing cells before analysis by MS/MS.^{3,4} However, because this method requires culturing of cells, it is not amenable to use for nonmetabolically active specimens such as clinical biopsy samples. By comparison, labeling with cleavable isotope-coded affinity tags (cICAT)⁵ is a global approach that is not biased against low abundance proteins, or proteins with extreme isoelectric points or molecular weights, and can be applied to samples derived from a variety of sources including clinical specimens.

The cICAT strategy involves the reduction and subsequent labeling of cysteine residues with isotopically distinct moieties in two protein samples meant for comparison.^{5–7} The cICAT moiety consists of biotin, a cleavable linker, an isotope-coded tag, and a thiol-reactive group.⁸ After labeling, the two differentially labeled protein samples are combined and proteolytically digested to produce peptide fragments appropriate for MS/MS analysis. Avidin affinity chromatography is used to enrich the sample for labeled peptides. A flow-through fraction, which predominantly contains nonlabeled peptides, is first collected from the column followed by elution of the cICAT-labeled peptides. The cICAT-labeled fraction and, optionally, the flow-through fraction are then separated by liquid chromatography and analyzed by MS/MS. MS analysis provides relative quantification of peptides, and MS/MS provides peptide sequence identification that is subsequently matched to a protein in translated databases.

To evaluate the performance characteristics of cICAT-based experiments in proteomes derived from human samples, we analyzed the results from six separate cICAT experiments, including examination of the number

Several strategies have been used for quantitative assessment of differential expression of proteins between different physiological and/or disease states. Two-dimensional gel electrophoresis followed by tandem mass spectrometry (MS/MS) is an established approach for

Accepted for publication June 15, 2006.

Address reprint requests to Kojo S.J. Elenitoba-Johnson, M.D., Department of Pathology, University of Utah Health Science Center, 50 North Medical Dr., Salt Lake City, UT 84123. E-mail: kojo.elenitobaj@path.utah.edu.

Table 1. Project Summaries

ICAT experiment	Cell line(s)	Samples (test versus control)	Protein source	Amount of protein labeled (per condition)	Number of fractions analyzed	Number of replicates analyzed
1	SUDHL-1	IL-2-stimulated versus nonstimulated	Released proteins	80 μ g	1	2
2	SUDHL-4	PMA-stimulated versus nonstimulated	Released proteins	40 μ g	1	3
3	SUDHL-4	IGF-stimulated versus nonstimulated	Whole cell lysate	200 μ g	16	n/a
4	SUDHL-4	Rituximab-treated versus untreated	Whole cell lysate	400 μ g	37	n/a
5	SUDHL-1	Geldanamycin-treated versus untreated	Whole cell lysate	400 μ g	33	n/a
6	MAC1/MAC2A	Aggressive CD30 ⁺ T-cell lymphoma versus indolent	Whole cell lysate	400 μ g	44	n/a

of proteins identified relative to the number of fractions analyzed, the relative amount of +1, +2, and +3 ions identified in a cICAT experiment, and the percentage of proteins that are identified with multiple unique peptides. Additionally, we analyzed the quantification characteristics of cICAT experiments, including the percentage of identified peptides containing cysteine residues and the percentage of identified peptides with quantification data. We also examined the percentage of proteins with differential expression, the agreement between quantification values in proteins identified with multiple quantified peptides, and the subcellular location and molecular function of quantified proteins. Additionally, we examined the contribution of analysis of the flow-through sample to protein identification in cICAT-based proteomic experiments of complex protein mixtures derived from human lymphoma samples. Finally, we estimated the false-positive identification rates for the entire data sets of our cICAT experiments using composite decoy database approaches^{9–11} and confirmed differential protein expression of two selected proteins by Western blot analysis.

Materials and Methods

Protein Samples

The six experimental conditions are summarized in Table 1. All of the experiments used human lymphoma-derived cell lines. Three of the experiments compared stimulated cells to nonstimulated cells, two of the experiments compared drug-treated cells to untreated cells, and the final experiment compared a CD30⁺ cutaneous T-cell lymphoma-derived cell line to its indolent counterpart.

For experiments 1 and 2, the cells for each of the conditions were grown in serum-free media. This media was collected, and released proteins from the cells were precipitated using trichloroacetic acid. For experiments 3 to 6, cells for each of the conditions were collected, and lysed using RIPA buffer (20 mmol/L Tris, pH 7.5, 100 mmol/L NaCl, 1 mmol/L ethylenediaminetetraacetic acid, 1% Nonidet P-40, 0.5% sodium deoxycholate, and 0.1% sodium dodecyl sulfate) with 0.1% protease inhibitor cocktail (Sigma, St. Louis, MO). Protein concentrations were determined using the Coomassie protein assay (Pierce, Rockford, IL).

Cleavable Isotope-Coded Affinity Tag Labeling

Equal amounts of protein from the control sample and from the test sample (Table 1) were labeled using the cleavable ICAT reagent kit for protein labeling (Applied Biosystems, Foster City, CA) (Figure 1). All reagents were supplied with the kit and used according to the manufacturer's instructions. In brief, the control and test protein mixtures were denatured and mixed with reducing reagent for 10 minutes in a boiling water bath to prepare the cysteine residues for labeling. The light and heavy cICAT reagents were each mixed with 20 μ l of acetonitrile, added to the appropriate sample, and incubated for 2 hours at 37°C in the dark. The control sample was labeled with the light cICAT reagent, and the test sample was labeled with the heavy cICAT reagent. This step conjugated the appropriate reagent to the cysteine residues of the sample proteins. The isotope-coded affinity tag has the following components: biotin, cleavable linker, isotope-coded tag, and reactive group. The isotope-coded tag portion contains nine carbon atoms of either ¹²C (the light reagent) or ¹³C (the heavy reagent). This confers a mass difference of 9 d between identical peptides with a single-labeled cysteine residue from the control and test samples.

After labeling, the two samples for each experiment were combined and digested with trypsin overnight at 37°C. After the tryptic digest, the combined sample was adjusted to pH 3 and run through a cation exchange column. A single fraction was collected for experiments 1 and 2 because of the relatively low complexity of these samples. For experiments 3 to 6, multiple fractions were collected from the cation exchange column (Table 1) using elution with a salt gradient (0 to 100% solution B for 2 hours; solution A = 10 mmol/L KH₂PO₄, 25% acetonitrile, pH 3; solution B = 10 mmol/L KH₂PO₄, 350 mmol/L KCl, 25% acetonitrile, pH 3). The eluted peptide samples were neutralized to pH 7 and loaded onto an avidin affinity column, and the flow-through off the column was collected. This fraction contained unlabeled peptides from the samples and was retained for analysis in experiments 2, 4, and 6. The column was washed, and the cICAT-labeled peptides were eluted. During the final step the biotin portion of the tag was removed using the acidic cleaving reagents supplied with the cICAT-labeling kit.

Figure 1. Schematic of an ICAT-labeling experimental protocol comparing control and test samples. Protein from each sample is collected, denatured, and reduced. Equal quantities of each sample are then labeled with the ICAT reagents. Typically, the control sample is labeled with the light reagent, and the test sample is labeled with the heavy reagent. The samples are then combined and digested with trypsin. Digested samples are fractionated using a cation exchange column. Each fraction is run through an avidin affinity column to separate ICAT-labeled peptides from unlabeled peptides. The flow-through sample, containing unlabeled peptides, is collected from the avidin affinity column first, followed by elution of the ICAT-labeled peptides. The flow-through sample may be retained for analysis by LC-MS/MS. The ICAT-labeled sample is subjected to LC-MS/MS analysis and quantification.

Tandem Mass Spectrometry

Digested samples were analyzed using the LCQ Deca XP ion trap mass spectrometer (Thermo Electron Corp., San Jose, CA). Ten to 15 μ L of each collected fraction were injected by the autosampler (Surveyor, Thermo Electron) into a reverse-phase column (75- μ m ID fused silica packed in-house with 10 cm of 5- μ m C18 particles). For experiments 1 and 2, multiple replicates of the samples were analyzed (Table 1). The samples were eluted through the column using an acetonitrile gradient (0 to

60% solution B for 80 to 240 minutes; solution A = 5% acetonitrile, 0.4% acetic acid, and 0.005% heptafluorobutyric acid; solution B = 95% acetonitrile, 0.4% acetic acid, and 0.005% heptafluorobutyric acid) followed by electrospray ionization. Mass spectrometry scans were performed from 400 to 2000 m/z followed by data-dependent MS/MS scans of the three most abundant peptides in each MS scan. Dynamic exclusion was set to a repeat count of 2 with an exclusion duration of 3 minutes.

Analysis of Flow-Through Samples

Flow-through samples from experiments 2, 4, and 6 were also analyzed by LC-MS/MS. We sought to determine whether the analysis of these samples, containing unlabeled peptides, would provide additional peptides to help confirm the identity of proteins identified by cICAT-labeled peptides. The flow-through samples were analyzed by LC-MS/MS as described above.

Protein Database Searching

The acquired MS/MS spectra for each sample were searched using the SEQUEST algorithm (BioWorks 3.1 SR1, Thermo Electron) against the UniProt database (5.26.05 download; 125,244 entries). Searches were performed with trypsin specified as the enzyme with an allowance for up to two missed cleavage sites. Searches from multiple fractions or replicates within an experiment were combined in BioWorks using a MultiConsensus report to generate a comprehensive list of peptides and proteins identified in a particular experiment.

Acceptance levels for positive peptide identification were determined using cross-correlation scores (X_{corr}) and delta correlation scores (ΔC_n). These scores aid in the determination of true positives, with higher scores increasing confidence in correct identifications. The minimum acceptable X_{corr} for identified peptides was 1.5 for +1 peptides, 2.5 for +2 peptides, and 3.5 for +3 peptides, with a $\Delta C_n \geq 0.100$. Analysis of the MS scans generated by the cICAT-labeled peptides was performed in BioWorks using the XPRESS software tool for quantification.¹² The specified mass difference was 9.0 d with mass tolerance of ± 0.7 amu.

Subcellular Localization and Functional Categorization of Proteins

Proteins with quantification data were grouped into subcellular locations and functional categories using the GoMiner software tool.¹³ In addition, a non-cICAT-labeled protein sample taken from a human lymphoma-derived cell line (SUDHL-4) was analyzed by LC-MS/MS as described above. The resulting MS/MS spectra were searched as described above and the protein identities were also subjected to analysis by the GoMiner software tool. The average percentages of proteins in each category for the quantified proteins in the cICAT experiments

Table 2. Peptide and Protein Identifications

ICAT experiment	Number of peptide identifications*	Number of protein identifications
1	26	22
2	86	82
3	130	82
4	215	203
5	289	280
6	765	679

*Peptide identifications met the following criteria: minimum cross correlation score (X_{corr}) of 1.5 for +1 ions, 2.5 for +2 ions, and 3.5 for +3 ions, and a minimum Δ cross correlation score of 0.1.

were compared to the percentages observed for the proteins identified from the non-cICAT-labeled sample.

Estimation of Error Rates

A minimum of 20% of the files for each data set was subjected to estimation of error using the composite target decoy reversed database approach as previously described.^{9,10} We also used a composite target decoy scrambled database approach in which the decoy database was obtained by random shuffling of the translated amino acid sequences in the database.¹¹ These searches were used to estimate false-positive rates for each data set.

Western Blot Analysis

The untreated control and Rituximab-treated samples from experiment 4 were subjected to Western blot analysis to validate the differential protein expression determined by cICAT-LC-MS/MS. The protein samples (40 μ g per lane) were resolved on a NuPage 4 to 12% Bis-Tris gel (Invitrogen, Carlsbad, CA) and transferred to a polyvinylidene difluoride membrane (Invitrogen). The membrane was probed with antibodies against α -tubulin (1:500 dilution; Santa Cruz Biotechnology, Santa Cruz, CA) and PLK-1 (1:200 dilution; Zymed Laboratories, San Francisco, CA). In addition, the membrane was probed with actin (1:10,000 dilution; Calbiochem, San Diego, CA) to serve as a protein loading control. Blots were visualized with the enhanced chemiluminescence Western blotting detection reagents from Amersham Biosciences (Buckinghamshire, UK).

Results

Peptide and Protein Identifications

The number of peptides and proteins identified that met our criteria for X_{corr} and ΔC_n scores in each of the six experiments is shown in Table 2. In general, the number of identifications increased with increased numbers of fractions analyzed per experiment, as shown in Figure 2. The average percentages of +1, +2, and +3 ions identified were 54.7, 20.4, and 24.9%, respectively.

Figure 2. The number of peptide and protein identifications versus the number of fractions analyzed. Multiple fractions were collected for four of the ICAT experiments. The number of peptides and proteins identified in these experiments is shown relative to the number of fractions analyzed. As expected, increasing the number of fractions, and therefore reducing the relative complexity of each fraction analyzed, generally results in identification of more peptides and proteins.

Multiple Unique Peptides

The percentage of proteins identified with multiple unique peptides ranged from 1.2 to 28%. For three of the six experiments, less than 4% of the proteins were identified with more than one unique peptide, as shown in Table 3.

Quantification Data

Between 24.4 and 40.8% of the peptides identified in each of the experiments contained cysteine residues. However, 69 to 100% of the cysteine-containing peptides in each experiment yielded corresponding quantification data (Figure 3). We observed that 15.9 to 36.4% (mean, 27.0%) of the total number of proteins identified in each of the experiments was detected with relative quantification data. More significantly, 6.1 to 21.1% (mean, 14.3%) of the proteins were differentially expressed by greater than 1.5-fold (Figure 4).

A total of 31 proteins from the six cICAT experiments (of 1348 proteins identified, in total) had multiple peptides with quantification, per protein. The majority of these proteins (26 of 31) had quantification data from two peptides; the remaining five proteins had between three and five peptides with quantification data. The peptides were both unique and nonunique. The quantification values within protein were concordant in trend (either underexpression or overexpression) for 87% (27 of 31) of the proteins. The magnitude of the values for all but two of these proteins agreed within 20%. Specific, representa-

Table 3. Percentage of Proteins Identified with 1, 2, or 3 Unique Peptides

ICAT experiment	1 Unique peptide	2 Unique peptides	3 Unique peptides
1	90.9%	9.1%	0%
2	98.8%	1.2%	0%
3	72.0%	26.8%	1.2%
4	98.0%	2.0%	0%
5	97.9%	3.1%	0%
6	91.8%	8.0%	0.3%

Figure 3. Percentages of peptides with cysteine residues and with quantification data. The percentages of peptides containing cysteine residues (light gray bars) and the percentages of peptides with quantification data (dark gray bars) for each of six separate ICAT experiments is shown. Although only 24.4 to 40.8% of the peptides contained cysteine residues, the majority of the cysteine-containing peptides had quantification data.

tive examples of proteins with multiple unique peptides and multiple quantification values are shown in Table 4.

Subcellular Localization and Functional Categorization of Proteins

Proteins of diverse subcellular localization and molecular function were identified with quantification data in our cICAT experiments. Of the quantified proteins, an average of 27% were from the cytoplasm, 24% were from the nucleus, 34% were membrane-associated, 11% were extracellular, and 4% were of unknown cellular location. The distribution of subcellular localization of quantified proteins closely matched the distribution of proteins in the non-cICAT-labeled sample (Figure 5). The majority of proteins were involved in signaling, binding, and catalytic functions. The percentages of proteins identified in each functional category were comparable to the percentages observed for the non-cICAT-labeled sample (Figure 6).

Flow-Through Contribution

In three of the six cICAT experiments analyzed by LC-MS/MS, the flow-through sample (containing nonlabeled peptides) was also analyzed. In each of the three experiments, the analysis of the flow-through contributed to the

Figure 4. Percentages of proteins quantified and differentially expressed. The percentages of proteins with quantification data in each of six separate ICAT experiments are shown. Between 15.9 and 36.4% (mean, 27.0%) of the total number of proteins identified had quantification data (light gray bars). Proteins differentially expressed by 1.5-fold or greater (dark gray bars) comprised between 6.1 and 21.1% (mean, 14.3%) of the total identified proteins.

overall number of protein identifications for the experiment. However, the number of proteins identified both in the cICAT-labeled sample and in the flow-through sample was very low. The percentages of proteins identified in the cICAT-labeled sample for which additional unique peptides were identified in the flow-through sample were 0, 2.5, and 2.4% for experiments 2, 4, and 6, respectively. Only five peptides, in total, of 301 peptides identified from the flow-through analysis of the three experiments provided additional unique peptides for proteins with quantification values.

Estimated Error Rates

The estimated error rates for each data set (based on cICAT peptide identifications) using both the scrambled and reversed database-based approaches are shown in Table 5. Estimated error rates ranged from 8.1 to 17.3%. For each data set, the estimated error using the reversed database was slightly higher than the error estimated using the scrambled database.

Validation of Protein Expression

Western blot analysis was performed for select proteins from experiment 4 to confirm the presence of the proteins

Table 4. Comparison of Quantification Results from Multiple Peptide Values

Protein	Accession number	Peptide sequence	Ion charge	X _{corr} score	ΔC _n score	Fold-change
Acidic leucine-rich nuclear phosphoprotein 32	P39687	K.C*PNLTHLNLSGNK.I	2	3.72	0.54	1.02
		K.C*PNLTHLNLSGNK.I	1	1.93	0.27	1.22
		K.CPNLTHLNLSGNK.I	2	4.32	0.56	0.97
		R.VSGGLEVLAEKC*PNLTHLNLSGNK.I	2	2.52	0.31	1.23
Tubulin α1 chain	P68366	K.AYHEQLSVAEITNACFEPANQMVK.C	2	2.97	0.52	1.09
		R.SIQFVDWC*PTGFK.V	1	1.81	0.28	1.33
Midasin	Q9NU22	K.AGHWWVLDELNLASQSVLEGLNAC*FDHR.G	3	3.88	0.26	0.40
		R.HPDFRLFACMNPATDVGK.R	2	3.73	0.27	0.46
L-Lactatedehydrogenase C chain	P07864	K.EELFLSIPC*VLGRNGVSDVVKINLNSEEEALFK.1	3	4.87	0.34	2.86
		R.VHPVSTMVKGlyGIKEELFLSIPCVLGR.N	3	3.70	0.17	0.80

Figure 5. Classification of proteins by subcellular localization. The percentages and average percentages, respectively, of proteins for each cellular location are shown for proteins from the non-cICAT-labeled sample (light gray bars) and the quantified proteins from the cICAT experiments (dark gray bars). The percentages in each category are comparable between the unlabeled proteins and the quantified proteins.

and the relative expression levels between the control sample and the test sample. The cICAT ratios and the Western blot results for two of the proteins identified by LC-MS/MS are shown in Figure 7. In both instances, the trend in expression levels as determined by cICAT-LC-MS/MS was confirmed by Western blot analysis.

Discussion

Because cICAT in conjunction with MS/MS is increasingly used for quantitative proteomics, knowledge of the anticipated analytical results is important. Although a previous study has examined the analytical characteristics of cICAT studies of the *Escherichia coli* proteome,¹⁴ similar studies in complex proteomes of human cellular origin have not been performed. In this regard, we have examined multiple distinct cICAT-based experiments involving human-derived lymphoma cells. Our evaluation of the analytical characteristics of cICAT-based experiments

Figure 6. Classification of proteins by molecular function. The percentages and average percentages, respectively, of proteins for each molecular function are shown for proteins from the non-cICAT-labeled sample (light gray bars) and the quantified proteins from the cICAT experiments (dark gray bars). The percentages in each category are comparable between the unlabeled proteins and the quantified proteins.

Table 5. Estimated Error Rates Using Scrambled and Reversed Databases

ICAT experiment	Scrambled database	Reversed database
1	8.1%	9.7%
2	7.9%	9.1%
3	4.5%	6.1%
4	15.7%	17.3%
5	10.9%	11.5%
6	14.1%	14.6%

using proteomes of human origin provides information that will be helpful in the design of these types of experiments and in the determination of anticipated results from cICAT MS/MS-based quantitative proteomic experiments.

In the four experiments for which multiple fractions were collected from the cation-exchange column, increasing numbers of identifications correlated with the number of fractions analyzed. The remaining two experiments used less complex samples, were composed of released proteins rather than whole cell lysates, and yielded the fewest numbers of protein identifications, as expected. Fractionation of samples further simplifies the complex peptide mixture, thus allowing the identification of lower abundance proteins in addition to the identification of high abundance proteins.

The vast majority of the proteins identified in our cICAT experiments were based on single-peptide hits. Protein assignments based on a single peptide are at greater risk of being false-positive than those based on multiple peptides. In this regard, recent publication guidelines suggest two unique peptides per positive protein identification.¹⁵ Although this may be frequently applicable to the majority of MS-based experiments, our data show that this may not be feasible in cICAT-based MS experiments performed using complex mammalian proteomes because only a few of the proteins identified here were based on more than one unique peptide. The reduction in the number of identified peptides per protein is partially accounted for by the restriction of cICAT labeling to cysteine residues. After the labeling of cysteine residues, an avidin column selects for peptides containing labeled

Figure 7. Western blot validation of protein expression. The control and test protein samples from experiment 4 were resolved on a polyacrylamide gel and transferred to a polyvinylidene difluoride membrane. The membrane was probed with an anti-actin antibody as a loading control. The membrane was also probed with antibodies against two proteins identified by cICAT-LC-MS/MS, α-tubulin, and PLK-1. The expression levels shown by Western blot are concordant in trend with the cICAT ratio for each protein.

peptides. Although we did not find that the column exclusively retained cysteine-containing peptides (an average of two-thirds of the peptides did not contain a cysteine residue, as discussed below), this step in the sample preparation greatly reduces complexity and therefore may eliminate peptides that potentially would have contributed additional peptide identifications to proteins identified with cysteine-containing peptides.

We observed that 15.9 to 36.4% of the proteins identified in each experiment were detected with relative quantification information. As expected, all of these peptides were cysteine-labeled. An average of 14.3% of the proteins identified were differentially expressed by greater than 1.5-fold. Similar results were observed by Molloy and colleagues in a study examining the analytical characteristics of cICAT-based quantitative proteomics experiments in the *E. coli* proteome,¹⁴ which is substantially less complex than the human proteome. We also found that proteins of diverse subcellular localization and molecular function were represented among the quantified proteins in each data set and that cICAT-labeling of proteins did not change the relative distribution of proteins, by cellular location or function, as compared to a non-cICAT-labeled sample.

Ideally, the relative protein quantification should be based on multiple quantification values obtained from multiple peptides. Because only a few of the quantified proteins in our data sets had more than one quantification value, we analyzed the concordance of quantification ratios of those proteins in which multiple peptides with quantification values were identified. Based on our findings that the majority of quantification values within protein showed concordance in trends of differential expression, it may be acceptable to base quantification values obtained from cICAT experiments on a single value. We believe that the imposition of the additional constraint of a cysteine residue within a fully tryptic peptide increases the accuracy of identification and affords reliable assessment of differential expression based on data obtained from single unique peptides.

Because the majority of proteins identified in cICAT-LC-MS/MS experiments are based on single peptide hits, it is necessary to incorporate postanalytical approaches such as INTERACT-ProteinProphet^{16,17} or composite decoy database searches^{9–11} to assess the error rates of cICAT data as a measure of the quality of an entire data set. Because of ease of application across data sets from multiple experiments, we used the composite decoy approach using reversed and scrambled databases to assess error rates of identification in our cICAT studies. We found that the error estimated by scrambled database searching and reversed database searching was similar in each data set analyzed, with slightly higher error rates observed when the reversed database approach was used. Additionally, confirmation of protein identifications by orthogonal methods may be necessary in cICAT-based experiments because of the lack of multiple peptides per protein assignment. Methods such as immunofluorescence or Western blot analysis can be used to verify the presence of specific proteins. In this regard, our overall laboratory experience indicates concordance be-

tween the results of cICAT-MS/MS and Western blotting ~70% of the time.

Analysis of the flow-through samples for three of the experiments was undertaken to determine the value in collecting and analyzing these samples during the ICAT-labeling and subsequent MS analysis. We expected that analysis of the flow-through samples would identify additional peptides from the proteins identified in the cICAT portion of each experiment and thus increase confidence in those protein identifications that were based on single peptides. However, we did not find this to be the case, because less than 2% of the proteins identified by analysis of the cICAT samples (across the three experiments) were also identified in the flow-through samples. Based on these findings, we surmise that analysis of flow-through material does not contribute significantly to the results of cICAT-based experiments.

In summary, this study provides a comprehensive analysis of the characteristics of cICAT experiments for quantitative profiling of complex proteomes derived from human cellular specimens. We analyzed the results of six such experiments and found that identifications are primarily based on single peptide hits. Simplification of protein mixtures through the use of fractions resulted in higher numbers of both identifications and quantified proteins. Although most determinations of relative protein quantification between samples were based on single values, the majority of proteins with multiple values showed a general agreement between the multiple differential expression values. Finally, we found that analysis of collected flow-through material did not significantly contribute to the results of the experiments in identifying additional peptides from quantified proteins and can be omitted with no significant effect on the quality of data from cICAT-based studies.

Acknowledgments

We thank David Abbott, Kathryn Everton, Kalpana Reddy, and Jonathan Schumacher for providing samples for LC-MS/MS analysis.

References

1. Gorg A, Weiss W, Dunn MJ: Current two-dimensional electrophoresis technology for proteomics. *Proteomics* 2004, 4:3665–3685
2. Gygi SP, Corthals GL, Zhang Y, Rochon Y, Aebersold R: Evaluation of two-dimensional gel electrophoresis-based proteome analysis technology. *Proc Natl Acad Sci USA* 2000, 97:9390–9395
3. Ong SE, Blagoev B, Kratchmarova I, Kristensen DB, Steen H, Pandey A, Mann M: Stable isotope labeling by amino acids in cell culture, SILAC, as a simple and accurate approach to expression proteomics. *Mol Cell Proteomics* 2002, 1:376–386
4. Everley PA, Krijgsvelde J, Zetter BR, Gygi SP: Quantitative cancer proteomics: stable isotope labeling with amino acids in cell culture (SILAC) as a tool for prostate cancer research. *Mol Cell Proteomics* 2004, 3:729–735
5. Gygi SP, Rist B, Gerber SA, Turecek F, Gelb MH, Aebersold R: Quantitative analysis of complex protein mixtures using isotope-coded affinity tags. *Nat Biotechnol* 1999, 17:994–999
6. Hansen KC, Schmitt-Ulms G, Chalkley RJ, Hirsch J, Baldwin MA, Burlingame AL: Mass spectrometric analysis of protein mixtures at

- low levels using cleavable ¹³C-isotope-coded affinity tag and multi-dimensional chromatography. *Mol Cell Proteomics* 2003, 2:299–314
7. Li J, Steen H, Gygi SP: Protein profiling with cleavable isotope-coded affinity tag (ciCAT) reagents: the yeast salinity stress response. *Mol Cell Proteomics* 2003, 2:1198–1204
8. Yi EC, Li XJ, Cooke K, Lee H, Raught B, Page A, Aneliunas V, Hieter P, Goodlett DR, Aebersold R: Increased quantitative proteome coverage with (13)C/(12)C-based, acid-cleavable isotope-coded affinity tag reagent and modified data acquisition scheme. *Proteomics* 2005, 5:380–387
9. Moore RE, Young MK, Lee TD: Qscore: an algorithm for evaluating SEQUEST database search results. *J Am Soc Mass Spectrom* 2002, 13:378–386
10. Peng J, Elias JE, Thoreen CC, Licklider LJ, Gygi SP: Evaluation of multidimensional chromatography coupled with tandem mass spectrometry (LC/LC-MS/MS) for large-scale protein analysis: the yeast proteome. *J Proteome Res* 2003, 2:43–50
11. MacCoss MJ: Computational analysis of shotgun proteomics data. *Curr Opin Chem Biol* 2005, 9:88–94
12. Han DK, Eng J, Zhou H, Aebersold R: Quantitative profiling of differentiation-induced microsomal proteins using isotope-coded affinity tags and mass spectrometry. *Nat Biotechnol* 2001, 19:946–951
13. Zeeberg BR, Feng W, Wang G, Wang MD, Fojo AT, Sunshine M, Narasimhan S, Kane DW, Reinhold WC, Lababidi S, Bussey KJ, Riss J, Barrett JC, Weinstein JN: GoMiner: a resource for biological interpretation of genomic and proteomic data. *Genome Biol* 2003, 4:R28.1–R28.8
14. Molloy MP, Donohoe S, Brzezinski EE, Kilby GW, Stevenson TI, Baker JD, Goodlett DR, Gage DA: Large-scale evaluation of quantitative reproducibility and proteome coverage using acid cleavable isotope coded affinity tag mass spectrometry for proteomic profiling. *Proteomics* 2005, 5:1204–1208
15. Carr S, Aebersold R, Baldwin M, Burlingame A, Clauser K, Nesvizhskii A: The need for guidelines in publication of peptide and protein identification data: Working Group on Publication Guidelines for Peptide and Protein Identification Data. *Mol Cell Proteomics* 2004, 3:531–533
16. Keller A, Nesvizhskii AI, Kolker E, Aebersold R: Empirical statistical model to estimate the accuracy of peptide identifications made by MS/MS and database search. *Anal Chem* 2002, 74:5383–5392
17. Nesvizhskii AI, Keller A, Kolker E, Aebersold R: A statistical model for identifying proteins by tandem mass spectrometry. *Anal Chem* 2003, 75:4646–4658