Quantitative analysis of complex protein mixtures using isotope-coded affinity tags

Steven P. Gygi[1#], Beate Rist[1#], Scott A. Gerber[2], Frantisek Turecek[2], Michael H. Gelb[2,3], and Ruedi Aebersold[1*]

Departments of [1]Molecular Biotechnology, [2]Chemistry and [3]Biochemistry, University of Washington, Box 357730, Seattle WA 98195-7730.
*[#]These two authors contributed equally to this work. *Corresponding author (e-mail: ruedi@u.washington.edu).*

Received 13 July 1999; accepted 2 August 1999

We describe an approach for the accurate quantification and concurrent sequence identification of the individual proteins within complex mixtures. The method is based on a class of new chemical reagents termed isotope-coded affinity tags (ICATs) and tandem mass spectrometry. Using this strategy, we compared protein expression in the yeast *Saccharomyces cerevisiae,* using either ethanol or galactose as a carbon source. The measured differences in protein expression correlated with known yeast metabolic function under glucose-repressed conditions. The method is redundant if multiple cysteinyl residues are present, and the relative quantification is highly accurate because it is based on stable isotope dilution techniques. The ICAT approach should provide a widely applicable means to compare quantitatively global protein expression in cells and tissues.

Keywords: gene expression, functional genomics, proteomics, protein profiling, mass spectrometry

With the completion of an increasing number of genomic sequences, attention is currently focused on how the data contained in sequence databases might be interpreted in terms of the structure, function, and control of biological systems. Approaches for global profiling of gene expression at the mRNA level as a function of the cellular state have been developed[1–3] and are widely used to identify clusters of genes for which the expression is idiotypic for a specific state. These methods, though exquisitely sensitive, do not indicate changes in protein expression. Quantitative proteome analysis, the global analysis of protein expression, is a complementary method to study steady-state gene expression and perturbation-induced changes. In comparison to gene expression analysis at the mRNA level, proteome analysis provides more accurate information about biological systems and pathways because the measurement directly focuses on the actual biological effector molecules.

Most approaches to quantitative protein analysis are accomplished by combining protein separation, most commonly by high-resolution two-dimensional polyacrylamide gel electrophoresis (2D-PAGE), with mass spectrometry (MS)-based or tandem mass spectrometry (MS/MS)-based sequence identification of selected, separated protein species[4–8]. This method is sequential, labor intensive, and difficult to automate. In addition, it selects against specific classes of proteins, such as membrane proteins, very large and small proteins, and extremely acidic or basic proteins. However, the technique's most significant flaw lies in its bias toward highly abundant proteins, as lower abundance regulatory proteins (e.g., transcription factors, protein kinases) are rarely detected when total cell lysates are analyzed[4–8].

The development of methods and instrumentation for automated, data-dependent electrospray ionization (ESI) MS/MS, in conjunction with microcapillary liquid chromatography (μLC) and database searching, has significantly increased the sensitivity and speed for the identification of gel-separated proteins. Moreover, μLC-MS/MS has also been used successfully for the large-scale identification of proteins directly from mixtures without gel electrophoretic separation[9,10]. These analyses, though fast and easily automated, are not quantitative and are also incompatible with the analysis of low-abundance proteins. Thus, there is a great need for a general and quantitative technology for proteome analysis.

We describe an approach to quantitative proteome analysis based on a class of reagents termed isotope-coded affinity tags (ICAT), which consist of three functional elements: a specific chemical reactivity, an isotopically coded linker, and an affinity tag.

Results and discussion

Principles of the isotope-coded affinity tag strategy. An ICAT reagent with specificity toward sulfhydryl groups, an eightfold deuterated linker, and a biotin affinity tag is shown in Figure 1. The ICAT method includes the following sequential steps (Fig. 2): (1) The side chains of cysteinyl residues in a reduced protein sample representing one cell state are derivatized with the isotopically light form of the ICAT reagent. The equivalent groups in a sample representing a second cell state are derivatized with the isotopically heavy reagent. (2) The two samples are combined and enzymatically cleaved to generate peptide fragments, some of which are tagged. (3) The tagged (cysteine-containing) peptides are isolated by avidin affinity chromatography. (4) Finally, the isolated peptides are separated and analyzed by μLC-MS/MS. In this last step, both the quantity and sequence identity of the proteins from which the tagged peptides originated are determined by automated multistage MS. This is achieved by the operation of the mass spectrometer in a dual mode in which it alternates in successive scans between measuring the relative quantities of peptides eluting from the capillary column and recording the sequence information of selected peptides. Peptides are quantified by measuring in the MS mode the relative signal intensities for pairs of peptide ions of identical sequence that are tagged with the isotopically light or heavy forms of the reagent, respectively, and that therefore differ in mass by the mass differential encoded within the ICAT

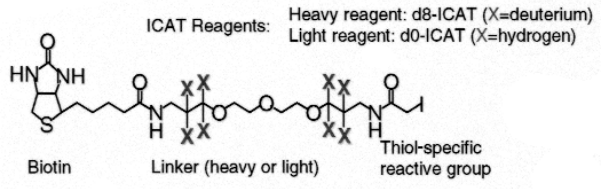


Figure 1. Structure of the ICAT reagent. The reagent consists of three elements: an affinity tag (biotin), which is used to isolate ICAT-labeled peptides; a linker that can incorporate stable isotopes; and a reactive group with specificity toward thiol groups (cysteines). The reagent exists in two forms, heavy (contains eight deuteriums) and light (contains no deuteriums).

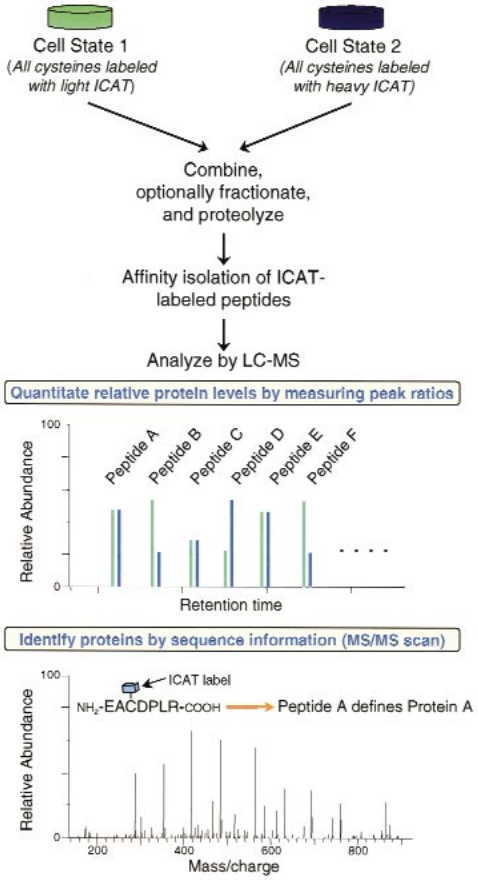


Figure 2. The ICAT strategy for quantifying differential protein expression. Two protein mixtures representing two different cell states have been treated with the isotopically light and heavy ICAT reagents, respectively; an ICAT reagent is covalently attached to each cysteinyl residue in every protein. Proteins from cell state 1 are shown in green, and proteins from cell state 2 are shown in blue. The protein mixtures are combined and proteolyzed to peptides, and ICAT-labeled peptides are isolated utilizing the biotin tag. These peptides are separated by microcapillary high-performance liquid chromatography. A pair of ICAT-labeled peptides are chemically identical and are easily visualized because they essentially coelute, and there is an 8 Da mass difference measured in a scanning mass spectrometer (four m/z units difference for a doubly charged ion). The ratios of the original amounts of proteins from the two cell states are strictly maintained in the peptide fragments. The relative quantification is determined by the ratio of the peptide pairs. Every other scan is devoted to fragmenting and then recording sequence information about an eluting peptide (tandem mass spectrum). The protein is identified by computer-searching the recorded sequence information against large protein databases.

reagent. Peptide sequence information is automatically generated by selecting peptide ions of a particular mass-to-charge (m/z) ratio for collision-induced dissociation (CID) in the mass spectrometer operating in the MS/MS mode[4,6,11]. Using sophisticated computer-searching algorithms[12–15], the resulting CID spectra are then automatically correlated with sequence databases to identify the protein from which the sequenced peptide originated. Combination of the results generated by MS and MS/MS analyses of ICAT-labeled peptides therefore determines the relative quantities as well as the sequence identities of the components of protein mixtures in a single automated operation.

The ICAT approach is based on two principles. First, a short sequence of contiguous amino acids from a protein (5–25 residues) contains sufficient information to identify that unique protein. Second, pairs of peptides tagged with the light and heavy ICAT reagents, respectively, are chemically identical and therefore serve as ideal mutual internal standards for accurate quantification. In MS,

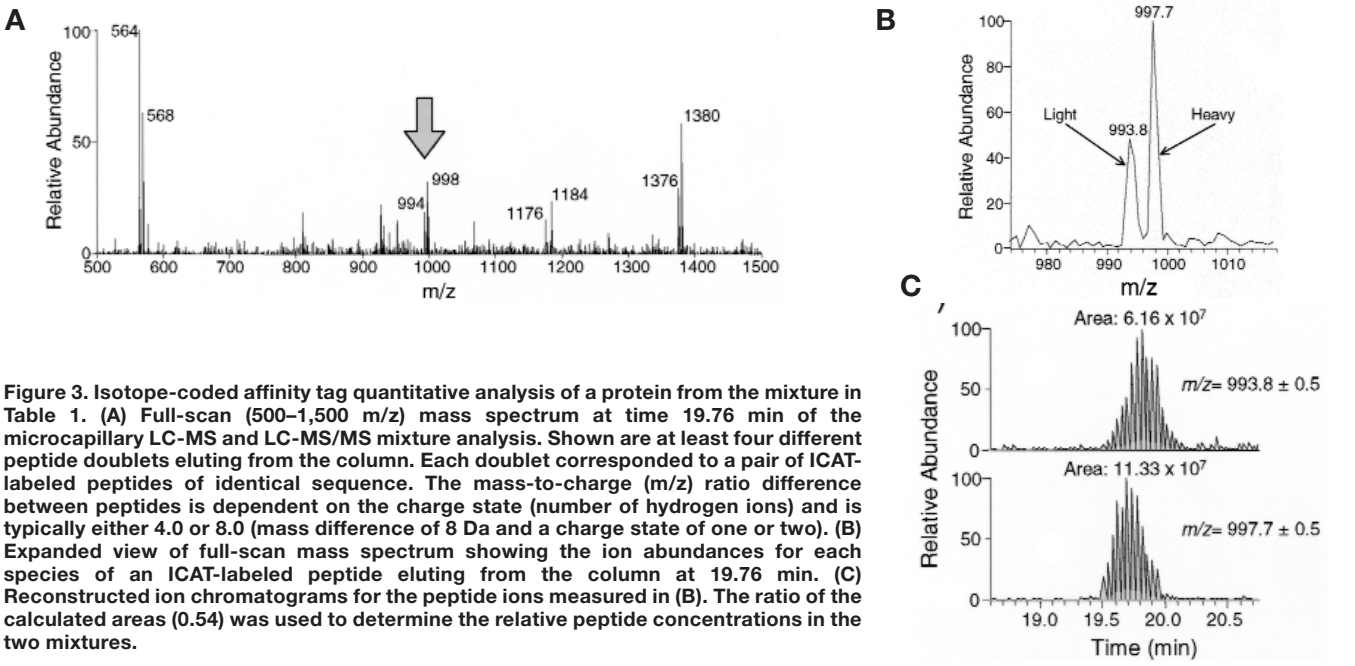
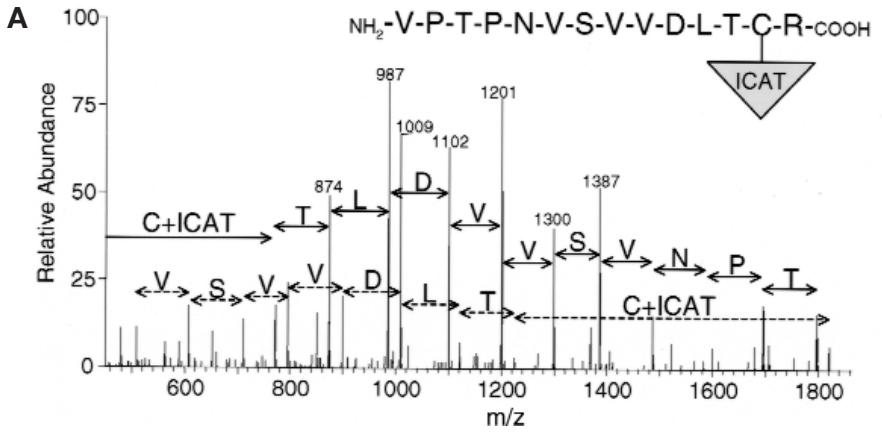


Figure 3. Isotope-coded affinity tag quantitative analysis of a protein from the mixture in Table 1. (A) Full-scan (500–1,500 m/z) mass spectrum at time 19.76 min of the microcapillary LC-MS and LC-MS/MS mixture analysis. Shown are at least four different peptide doublets eluting from the column. Each doublet corresponded to a pair of ICAT-labeled peptides of identical sequence. The mass-to-charge (m/z) ratio difference between peptides is dependent on the charge state (number of hydrogen ions) and is typically either 4.0 or 8.0 (mass difference of 8 Da and a charge state of one or two). (B) Expanded view of full-scan mass spectrum showing the ion abundances for each species of an ICAT-labeled peptide eluting from the column at 19.76 min. (C) Reconstructed ion chromatograms for the peptide ions measured in (B). The ratio of the calculated areas (0.54) was used to determine the relative peptide concentrations in the two mixtures.



B

```
s0319_hlnavcid.0364.0364.2.out
# amino acids = 93009033, # proteins = 290043, # matched peptides = 1973750
C:\LCQ\database\owl.v31.3, (C# +494.50)

 #   Rank/Sp  (M+H)+ C*10^4  Ions   Reference    Peptide
---  -------  ------ ------  ----  ----------  -------
 1.   1 /  1  1994.3 4.4675  17/26  G3P_RABIT  (R)VPTPNVSVVDLTC#R
 2.   2 /403  1995.1 2.7366  13/34  SLTRNGL     (E)LGKPVLTANQVTIWEGLR
 3.   3 /  3  1995.0 2.6591  16/36  FLP_LACCA   (N)IANPNVYTETLTAATVCTI
 4.   4 /209  1995.0 2.6335  14/36  A42912      (Y)LALLPSDAEGPHGQFVTDK
 5.   5 /381  1995.1 2.4634  13/38  H69373      (L)ALLVLVAPAMAAGNGEDLRN
```

Figure 4. The ICAT sequence identification of the peptide quantified in Figure 3 as derived from glyceraldehyde-3-phosphate dehydrogenase. Shown is the identification of the amino acid sequence from the peptide for which the abundance was measured in Figure 3C. (A) Tandem mass spectrum derived by collision-induced dissociation of the $(M + 2H)^{2+}$ precursor, m/z 998. Fragment ions in the spectrum represent mainly single-event preferential cleavage of peptide bonds resulting in the sequence information recorded from both the N and C termini of the peptide simultaneously. The one-letter code for encountered amino acids is shown. The ICAT modification remained fairly stable and attached to the cysteinyl residue. (B) Sequest output file for the above spectrum. The recorded sequence information was computer searched with the Sequest program[12] against the OWL nonredundant database containing 290,043 protein entries[23] with cysteinyl residues modified by the mass of either the light or the heavy ICAT reagent. The best-matching peptide was from glyceraldehyde-3-phosphate dehydrogenase and was a tryptic peptide with a modified cysteinyl residue.

the ratios between the intensities of the lower and upper mass components of these pairs of peaks provide an accurate measure of the relative abundance of the peptides (and hence the proteins) in the original cell pools because the MS intensity response to a given peptide is independent of the isotopic composition of the ICAT reagents[16]. The use of isotopically labeled internal standards is standard practice in quantitative MS and has been exploited to great advantage in, for example, the precise quantification of drugs and their metabolites in body fluids[16].

Application of the ICAT method to a standard protein mixture. To illustrate the method, two mixtures consisting of the same six proteins at known but different concentrations were prepared and analyzed. The protein mixtures were labeled, combined, and treated as schematically illustrated in Figure 2. The isolated tagged peptides were quantified and sequenced in a single combined μLC-MS and μLC-MS/MS experiment on an ESI ion trap mass spectrometer. All six proteins were unambiguously identified and accurately quantified (Table 1). Multiple tagged peptides were encountered for each protein. The mean differences between the observed and expected quantities for the six proteins ranged between 2% and 12%.

The entire process is further illustrated for a single peptide pair in Figure 3. A single scan of the mass spectrometer operated in MS mode is shown in Figure 3A. Four pairs of peptide ions characterized by the ICAT-encoded mass differential are detected in this scan and indicated with their respective m/z values. The scan shown was acquired in 1.3 s. Over the course of the 1 h chromatographic elution gradient, more than 1,200 such scans were automatically recorded. Figure 3B shows the m/z range around the ion pair with m/z ratios of 993.8 and 977.7, respectively, and represents the integration of the entire eluting peaks (~10 s). Coelution and a detected mass differential of 4 U potentially identifies the ions as a pair of doubly charged ICAT-labeled peptides of identical sequence (mass difference of eight and a charge state of two). Figure 3C shows the reconstructed ion chromatograms for these two species. The relative quantities were determined by integrating the contour of the respective peaks. The

Table 1. Sequence identification and quantitation of the components of a protein mixture in a single analysis.

Gene name[a]	Peptide sequence identified[b]	Observed ratio (d0/d8)[c]	Mean ± SD	Expected ratio (d0/d8)[d]	% error
LCA_BOVIN	ALC#SEK	0.94	0.96 ± 0.06	1.00	4.2
	C#EVFR	1.03			
	FLDDLTDDIMC#VK	0.92			
OVAL_CHICK	ADHPFLFC#IK	1.88	1.92 ± 0.06	2.00	4.0
	YPILPEYLQC#VK	1.96			
BGAL_ECOLI	LTAAC#FDR	1.00	0.98 ± 0.07	1.00	2.0
	IGLNC#QLAQVAER	0.91			
	IIFDGVNSAFHLWC#NGR	1.04			
LACB_BOVIN	WENGEC#AQK	3.64	3.55 ± 0.13	4.00	11.3
	LSFNPTQLEEQC#HI	3.45			
G3P_RABIT	VPTPNVSVVDLTC#R	0.54	0.56 ± 0.02	0.50	12.0
	IVSNASC#TTNC#LAPLAK	0.57			
PHS2_RABIT	IC#GGWQMEEADDWLR	0.32	0.32 ± 0.03	0.33	3.1
	TC#AYTNHTVLPEALER	0.35			
	WLVLC#NPGLAEIIAER	0.30			

[a] Gene names are according to Swiss Prot nomenclature (www.expasy.ch).
[b] ICAT-labeled cysteinyl residue is denoted by # sign.
[c] Ratios were calculated for each peptide as shown in Fig. 3.
[d] Expected ratios were calculated from the known amounts of proteins present in each mixture.

ratio (light:heavy) was determined as 0.54 (Table 1). The peaks in the reconstructed ion chromatograms appear serrated because in every second scan the mass spectrometer switched between the MS and the MS/MS modes to generate the CID mass spectrum of a selected peptide ion for the purpose of identifying tagged proteins. Figure 4A shows the CID spectrum recorded from the peptide ion with m/z = 998 (marked with an arrow in Fig. 3A). Database searching with this CID spectrum identified the protein as glyceraldehyde-3-phosphate dehydrogenase (Fig. 4B).

Several features of the ICAT strategy are immediately apparent. First, at least two peptides were detected from each protein in the mixture. Therefore, quantification and protein identification can be redundant. Second, the identified peptides all contained at least one tagged cysteinyl residue. The presence of the relatively rare cysteinyl residue in a peptide adds an additional powerful constraint for database searching[17]. Third, tagging and selective enrichment of cysteine-containing peptides significantly reduced the complexity of the peptide mixture generated by the concurrent digestion of six proteins (293 potential tryptic peptides, 44 tryptic peptides containing at least one cysteinyl residue). Fourth, the peptide samples eluted from the avidin affinity column are directly compatible with analysis by μLC-MS/MS.

Application of the method to the quantitative analysis of protein expression in different cell states. The ICAT strategy was applied to study differences in steady-state protein expression in the yeast *Saccharomyces cerevisiae*, in two different states not subject to glucose repression (Table 2). Cells were harvested from yeast growing in log phase utilizing either 2% galactose or 2% ethanol as the carbon source. From each cell state, 100 μg of soluble yeast protein were labeled, combined, and analyzed as described in Figure 2, and 2% of the sample (the equivalent of cysteine-containing peptides from ~2 μg of protein from each cell state) was analyzed.

Glucose repression causes large numbers of proteins with metabolic functions significant to growth on other carbon sources to be minimally expressed[18,19]. Growth on galactose or ethanol with no glucose present results in the expression of glucose-repressed genes. Sequence information during the run was collected for more than 800 different peptides. Table 2 presents a selection of identified glucose-repressed genes encountered in the analysis (10 genes), along with 21 other genes for comparison (34 genes in total). Genes specific to either growth on galactose (*GAL1*, *GAL10*) or ethanol (*ADH2*, *ACH1*) were detected and quantified.

The quantitative nature of the ICAT method is apparent from the ability to obtain accurate measurements of small changes in relative protein levels. Evidence of this accuracy is evident on examination of peptide ratios for proteins for which multiple peptides were quantified. For example, the five peptides found from PCK1 had a mean ratio ±95% confidence intervals of 1.57 ± 0.15, and the percentage error was <10%. In addition, the observed changes fit the expected changes from the literature[18,19] and are in agreement with the changes in staining intensity for these same proteins examined by 2D-PAGE (data not shown).

The alcohol dehydrogenase family of isozymes in yeast facilitates growth on either hexose sugars (*ADH1*)

Table 2. Protein profiles from yeast growing on galactose or ethanol as a carbon source.

Gene name[a]	Peptide sequence identified[b]	Observed ratio[c] (Eth: Gal)[d]	Galactose-repressed[e]	Glucose-repressed[e]
ACH1	KHNC#LHEPHMLK	>100: 1	✓	
ADH1	YSGVC#HTDLHAW-HGDWPLPVK	0.57: 1		
	C#C#SDVFNQVVK	0.48: 1		
ADH2	YSGVC#HTDLHAW-HGDWPLPTK	>200: 1	✓	✓
	C#SSDVFNHVVK	>200: 1		
ALD4	TFEVINPSTEEEIC#-HIYEGR	>100: 1	✓	✓
BMH1	SEHQVELIC#SYR	0.95: 1		
CDC19	YRPNC#PIILVTR	0.49: 1		
	NC#TPKPTSTTETV-AASAVAAVFEQK	0.65: 1		
	AC#DDK	0.67: 1		
FBA1	SIAPAYGIPVVLH-SDHC#AK	0.60: 1		
	EQVGC#K	0.63: 1		
GAL1	LTGAGWGGC#TVH-LVPGGPNGNIEK	1: >200		✓
GAL10	HHIPFYEVDLC#DR	1: >200		✓
	DC#VTLK	1: >200		
GCY1	LWC#TQHHEPEVAL-DQSLK	0.34: 1		✓
GLK1	IC#SVNLHGDHTFS-MEQMK	0.65: 1		
GPD1	IC#SQLK	0.54: 1		✓
ICL1	GGTQC#SIMR	>100: 1		✓
IPP1	NC#FPHHGYIHNYGAFP-QTWEDPNVSHPETK	0.76: 1		
LPD1	VC#HAHPTLSEAFK	1.30: 1		✓
PEP4	KGWTGQYTLDC#NTR	2.60: 1		✓
PSA1	SVVLC#NSTIK	0.56: 1		
PGM2	C#TGGIILTASHNPG-GPENDMGIK	0.58: 1		✓
	LSIC#GEESFGTGS-NHVR	0.62: 1		✓
PCK1	C#PLK	1.59: 1		
	IPC#LADSHPK	1.47: 1		
	C#INLSAEKEPEI-FDAIK	1.52: 1		
	C#AYPIDYIPSAK	1.41: 1		
	IVEEPTSKDEIWWG-PVNKPC#SER	1.85: 1		
QCR6	ALVHHYEEC#AER	1.30: 1		✓
RPL1A[f]	SC#GVDAMSV-DDLKK	0.82: 1		
SAH1	HPEMLEDC#FGLSE-ETTTGVHHLYR	0.62: 1		
	EC#INIKPQVDR	0.74: 1		
SOD1	GFHIHEFGDATNGC#-VSAGPHFNPFK	0.46: 1		✓
TEF1	RGNVC#GDAK	0.81: 1		
	C#GGIDK	0.70: 1		
	FVPSKPMC#VEAFS-EYPPLGR	0.74: 1		
VMA2	IPIFSASGLPHNEI-AAQIC#R	0.70: 1		
YHB1	HYSLC#SASTK	0.69: 1		

[a] Gene names are according to the Yeast Proteome Database (YPD)[19]
[b] Sequence was identified by Sequest program[12].
[c] Protein expression ratios were calculated as described in Fig. 3.
[d] Carbon source for yeast growth was 2% ethanol (Eth) or 2% galactose (Gal).
[e] Gene is known to be galactose- or glucose-repressed[19].
[f] Eight other ribosomal proteins were detected at similar gene expression levels.
[#] Cysteinyl residue is ICAT-labeled.

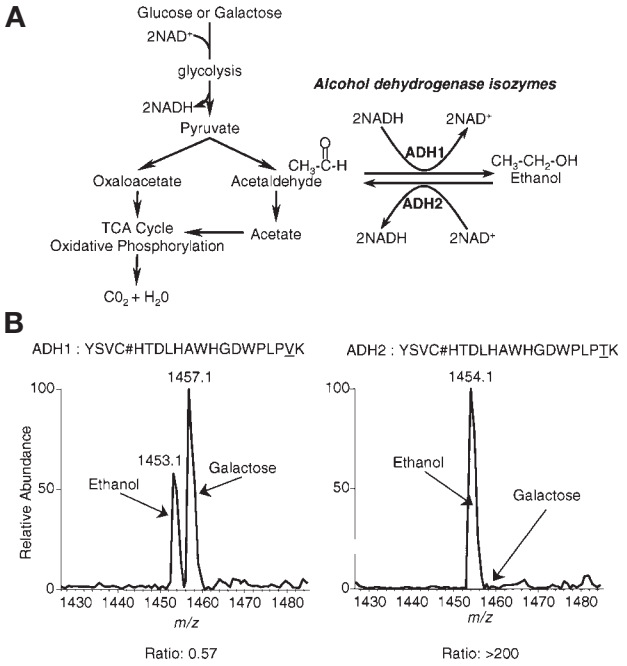


Figure 5. Isotope-coded affinity tag analysis of alcohol dehydrogenase isozymes in yeast. (A) ADH1 converts acetaldehyde to ethanol when cells are growing on hexose sugars. When other carbon sources are not available, ethanol can be utilized as the sole carbon source in yeast. This depends on ADH2 for conversion of ethanol back into acetaldehyde as the first step in the pathway where acetyl-CoA is used to fuel the tricarboxylic acid (TCA) cycle and the glyoxylate cycle for energy utilization in the cell[24]. (B) The peptides identified from ADH1 and ADH2 are 95% identical and differ only by a single amino acid substitution (valine for threonine). The substitution shifted the retention time of the ADH2 peptide pair by 2 min and shifted the mass by 2 Da, which allowed for unambiguous identification and analysis of gene expression levels.

or ethanol (*ADH2*). The gene *ADH2* encodes an enzyme that is both glucose and galactose repressed and permits a yeast cell to grow entirely on ethanol by converting it into acetaldehyde (Fig. 5A). In the presence of sugar, ADH1 performs the reverse reaction, converting acetaldehyde into ethanol. The regulation of these isozymes is key to carbon utilization in yeast[18]. Cross-hybridization sometimes compromises the ability to accurately measure differences in gene expression across families of isozymes using cDNA array techniques[1]. The ICAT strategy succeeded in measuring gene expression for each isozyme even though ADH1 and ADH2 share 93% amino acid (88% nucleotide) sequence homology. The detected ICAT-labeled peptides from each isozyme differed by a single amino acid residue (valine to threonine), which shifted the retention time by more than 2 min and the mass by 2 Da for the ADH2 peptides (Fig. 5B). ADH1 was expressed at approximately twofold higher levels when galactose was the carbon source, compared with ethanol. Ethanol induction of ADH2 expression resulted in more than a 200-fold increase compared with galactose induction.

These results illustrate the potential of the ICAT method for the quantitative analysis of protein mixtures and the identification of the protein components in a single automated operation. The ICAT method has the potential to become the basis for automated, quantitative, and global proteome analysis for a number of reasons. First, the ICAT method significantly reduces the complexity of the peptide mixture because biotinylated cysteine-containing peptides are selectively isolated. For example, a theoretical tryptic digest of the entire yeast proteome (6,113 proteins) produces 344,855 peptides, but only 30,619 of these peptides contain a cysteinyl residue. Thus, the complexity of the mixture is reduced, while protein quantification and identification are still achieved. Second, the chemical reaction in the ICAT alkylation can be performed in the presence of urea, sodium dodecyl sulfate (SDS), salts, and other chemicals that do not contain a reactive thiol group. Therefore, proteins are kept in solution with powerful stabilizing agents until they are enzymatically digested. Third, the sensitivity of the μLC-MS/MS system is critically dependent on the sample quality. In particular, commonly used protein-solubilizing agents are poorly compatible with MS. Avidin affinity purification of the tagged peptides completely eliminates contaminants incompatible with MS. Fourth, the quantification and identification of low-abundance proteins requires large amounts (milligrams) of starting protein lysate. Isotope-coded affinity tag analysis is compatible with any biochemical, immunological, or cell biological fractionation methods that reduce the mixture complexity and enrich for proteins of low abundance while quantification is maintained. It should be noted that accurate quantification is only maintained over the course of protein enrichment procedures if all manipulations preceding combination of the differentially labeled samples are strictly conserved. Fifth, the method can be redundant in both quantification and identification if multiple cysteines are detected. There is a dynamic range associated with the ability of the method to quantify differences in expression levels of ICAT-labeled peptides, which is dependent on both the intensity of the peaks corresponding to the peptide pair and the overall mixture complexity. In addition, the dynamic range will be different for each type of mass spectrometer used. The ion trap can collect impressive amounts of sequencing information in a data-dependent fashion but offers a more limited dynamic quantification range. The dynamic range of the ion trap (based on signal-to-noise ratios) varied depending on the signal intensity of the peptide pair and complexity of the mixture, but differences of up to 100-fold were generally detectable and even larger differences could be determined for more abundant peptides. Finally, the ICAT approach can be extended to include reactivity toward other functional groups. A small percentage of proteins (8% for *S. cerevisiae*) contain no cysteinyl residues and are therefore missed by using thiol-specific ICAT reagents. The ICAT reagents with different specificities also will make cysteine-free proteins susceptible to analysis by the ICAT method.

The ICAT strategy as a solution to quantitative proteome analysis. To date, large-scale quantitative protein expression analysis has been performed using radioactive[6] and stable isotope (^{15}N) metabolic labeling[20] followed by two-dimensional (2D) gel separation and quantitative analysis of separated proteins by scintillation counting or mass spectrometry. Any large-scale (proteome) analysis involving 2D gels has the unavoidable disadvantage of measuring only highly abundant proteins when total cell lysates are applied[4–8]. The general reason is that the quantities of protein that can be loaded onto an analytical 2D gel are limited to, at most, high microgram amounts. In contrast, any amount of starting material can theoretically be used with the ICAT strategy. Therefore, sufficient amounts of very low-abundance proteins can be prepared to make them detectable by mass spectrometry. Furthermore, the ICAT method is based on postisolation stable isotope labeling of proteins and is therefore not limited to cells and tissues compatible with metabolic labeling. In the postgenomic era, technologies are required that can quantitatively, globally, and automatically measure gene expression in cells and tissues. The ICAT approach provides a broadly applicable means for the quantitative cataloguing and comparison of protein expression in a variety of normal, developmental, and disease states, a task that is crucial in the emerging field of proteomics.

Experimental protocol

Synthesis of ICAT reagents. The ICAT reagents used were synthesized with the following isotopically different substrates: 4,7,10-trioxa-1,13-tridecanediamine (A) (Aldrich, Milwaukee, WI) and 2,2',3,3',11,11',12,12'-octadeutero-4,7,10-trioxa-1,13-tridecanediamine (B)[21]. Synthesis of *N*-(13-amino-4,7,10-trioxatridecanyl) biotinamide (C) was as follows. To biotin-pentafluorophenyl ester (Pierce, Rockford, IL) in dry dimethylformamide containing excess *N,N*-diisopropylethylamine (Aldrich) were added five equivalents of (A) with stirring at room temperature for 3 h. Solvent was removed under reduced pressure and (C) was purified to homogeneity by reverse-phase HPLC. The heavy analog was prepared as per (C), but with five equivalents of (B). Synthesis of *N*-(13-iodoacetamido-4,7,10-trioxatridecanyl) biotinamide (D) was as follows. To (C) (or heavy analog) in dry dimethylformamide containing excess *N,N*-diisopropylethylamine was added two equivalents iodoacetic anhydride (Aldrich) with stirring at room temperature for 3 h. Solvent was removed under reduced pressure, and (D) was purified to homogeneity by reverse-phase HPLC and characterized by MS.

ICAT analysis of standard mixture. Two mixtures containing the same six proteins at different concentrations (shown as μg/ml in mixture 1, and μg/ml in mixture 2) were purchased from Sigma (St. Louis, MO): rabbit glyceraldehyde-3-phosphate dehydrogenase (40, 20); *Escherichia coli* β-galactosidase (50, 50); rabbit phosphorylase b (60, 20); chicken ovalbumin (30, 60); bovine β-lactoglobulin (10, 40); bovine α-lactalbumin (10, 10). Disulfide bonds in the denatured protein mixtures were reduced (50 mM Tris buffer pH 8.5, 6 M guanidine HCl, 5 mM tributyl phosphine) for 1 h at 37°C. Cysteinyl groups in each mixture were independently biotinylated with a fivefold molar excess of the appropriate ICAT reagent. Excess ICAT reagent was removed from the combined samples by gel filtration (Bio-Rad, Richmond, CA) in Tris buffer (50 mM, pH 8.5) with 0.1% SDS, and the protein fraction was digested with trypsin (Promega, Madison, WI) overnight at 37°C. The peptide solution was then passed over a prepared monomeric avidin column (Pierce). The column was washed with water, and biotinylated peptides were eluted with 0.3% formic acid (1 ml fractions). Peptide recovery across the entire procedure was estimated at approximately 70%.

An LCQ ion trap mass spectrometer (Finnigan MAT, San Jose, CA) was used with an in-house fabricated microelectrospray source[22] and an HP1100 solvent delivery system (Hewlett Packard, Palo Alto, CA). A 60 min binary gradient with 5–80% solvent B (acetonitrile and 0.005% heptafluorobutyric acid (HFBA)). Solvent A consisted of 0.4% acetic acid and 0.005% HFBA. A flow rate of 0.5 μl/min was used with a 100 μm × 12 cm fused silica capillary column in-house packed with Monitor spherical silica (Column Engineering, Ontario, CA). Functional chromatography has been achieved with this setup with peptide loads as high as 500 pmol. in H_2O. One microliter of the peptide mixture was pressure loaded onto the column. Eluting peptides were analyzed by μLC-MS and μLC-MS/MS techniques as described in the text and elsewhere[6,11]. The intensities of eluting peptide pairs were measured in the scanning mass spectrometer. There is a slight difference in the elution times of differentially tagged peptide pairs, with the heavy analog eluting 1–2 s before the light analog. For this reason, the entire peak area of each eluting peptide was reconstructed and used in the ratio calculation. To determine the amino acid sequence, the mass spectrometer operated in a data-dependent MS/MS mode (a full-scan mass spectrum is followed by a tandem mass spectrum), where the precursor ion is selected "on the fly" from the previous scan. An m/z ratio for an ion that had been selected for fragmentation was placed in a list and dynamically excluded for 1 min from further fragmentation. Proteins from the mixture were later identified automatically by the computer program Sequest[12], which searched tandem mass spectra against the OWL nonredundant sequence database[23].

Protein profiling in yeast utilizing either galactose or ethanol as a carbon source. The source of the yeast protein was YPH499 (*MATa ura3-52 lys2-801 ade2-101 leu2-1 his3-200 trp1-63*). Logarithmically growing cells (2–4 × 10^7/ml) utilizing either 2% galactose or 2% ethanol as a carbon source in YP media were harvested as described[6] and lyophilized. Protein samples were redissolved and the ICAT strategy proceeded as above using 100 μg of total protein from each cell state. The volume of sample eluted (in 0.3% formic acid) was reduced from 1,000 to 50 μl. One microliter of the peptide mixture was analyzed by μLC-MS/MS.

Acknowledgments

This work was supported in part by the NSF Science and Technology Center for Molecular Biotechnology; NIH grants RR11823, T32HG00035, and HD-02274; a grant from the Merck Genome Research Institute; University of Washington Royalty Research Fund grant; and a fellowship to B.R. from the Swiss National Science Foundation. We thank D.R. Goodlett and J. Watts for critical reading of the manuscript.

1. DeRisi, J.L., Iyer, V.R. & Brown, P.O. Exploring the metabolic and genetic control of gene expression on a genomic scale. *Science* **278**, 680–686 (1997).
2. Roth, F.P., Hughes, J.D., Estep, P.W. & Church, G.M. Finding DNA regulatory motifs within unaligned noncoding sequences clustered by whole-genome mRNA quantitation. *Nat. Biotechnol*. **16**, 939–945 (1998).
3. Velculescu, V.E. et al. Characterization of the yeast transcriptome. *Cell* **88**, 243–251 (1997).
4. Link, A.J., Hays, L.G., Carmack, E.B. & Yates, J.R.. Identifying the major proteome components of *Haemophilus influenzae* type-strain NCTC 8143. *Electrophoresis* **18**, 1314–1334 (1997).
5. Shevchenko, A. et al. Linking genome and proteome by mass spectrometry: large-scale identification of yeast proteins from two dimensional gels. *Proc. Natl. Acad. Sci. USA* **93**, 14440–14445 (1996).
6. Gygi, S.P., Rochon, Y., Franza, B.R. & Aebersold, R. Correlation between protein and mRNA abundance in yeast. *Mol. Cell. Biol.* **19**, 1720–1730 (1999).
7. Garrels, J.I. et al. Proteome studies of *Saccharomyces cerevisiae*: identification and characterization of abundant proteins. *Electrophoresis* **18**, 1347–1360 (1997).
8. Boucherie, H. et al. Two-dimensional gel protein database of *Saccharomyces cerevisiae*. *Electrophoresis* **17**, 1683–1699 (1996).
9. Link, J. et al. Direct analysis of large protein complexes using mass spectrometry. *Nat. Biotechnol.* **17**, 676–682 (1999).
10. Opiteck, G.J., Lewis, K.C., Jorgenson, J.W. & Anderegg, R.J. Comprehensive online LC/LC/MS of proteins. *Anal. Chem.* **69**, 1518–1524 (1997).
11. Gygi, S.P., Han, D.K.M., Gingras, A.C., Sonenberg, N. & Aebersold, R. Protein analysis by mass spectrometry and sequence database searching: tools for cancer research in the post-genomic era. *Electrophoresis* **20**, 310–319 (1999).
12. Eng, J., McCormack, A.L. & Yates, J.R. An approach to correlate tandem mass spectral data of peptides with amino acid sequences in a protein database. *J. Am. Soc. Mass Spectrom.* **5**, 976-989 (1994).
13. Mann, M. & Wilm, M. Error-tolerant identification of peptides in sequence databases by peptide sequence tags. *Anal. Chem.* **66**, 4390–4399 (1994).
14. Qin, J. et al. A strategy for rapid, high-confidence protein identification. *Anal. Chem*. **69**, 3995–4001 (1997).
15. Clauser, K.R. et al. Rapid mass spectrometric peptide sequencing and mass matching for characterization of human melanoma proteins isolated by two-dimensional PAGE. *Proc. Natl. Acad. Sci. USA* **92**, 5072–5076 (1995).
16. De Leenheer, A.P. & Thienpont, L.M. Application of isotope dilution-mass spectrometry in clinical chemistry, pharmacokinetics, and toxicology. *Mass Spectrom. Rev.* **11**, 249–307 (1992).
17. Sechi, S. & Chait, B.T. Modification of cysteine residues by alkylation. A tool in peptide mapping and protein identification. *Anal. Chem.* **70**, 5150–5158 (1998).
18. Ronne, H. Glucose repression in fungi. *Trends Genet.* **11**, 12–17 (1995).
19. Hodges, P.E., McKee, A.H., Davis, B.P., Payne, W.E. & Garrels, J.I. The Yeast Proteome Database (YPD): a model for the organization and presentation of genome-wide functional data. *Nucleic Acids Res.* **27**, 69–73 (1999).
20. Oda, Y., Huang, K., Cross, F.R., Cowburn, D. & Chait, B.T. Accurate quantitation of protein expression and site-specific phosphorylation. *Proc. Natl. Acad. Sci. USA* **96**, 6591–6596 (1999).
21. Gerber, S.A., Scott, C.R., Turecek, F. & Gelb, M.H. Analysis of rates of multiple enzymes in cell lysates by electrospray ionization mass spectrometry. *J. Am. Chem. Soc.* **121**, 1102–1103 (1999).
22. Figeys, D. et al. Electrophoresis combined with novel mass spectrometry techniques: powerful tools for the analysis of proteins and proteomes. *Electrophoresis* **19**, 1811–1818 (1998).
23. Bleasby, A.J., Akrigg, D. & Attwood, T.K. OWL—a non-redundant composite protein sequence database. *Nucleic Acids Res.* **22**, 3574–3577 (1994).
24. Johnston, M. & Carlson, M. in *The molecular and cellular biology of the yeast Saccharomyces*. (eds Jones, E.W., Pringle, J.R. & Broach, J.R.) 193–281 (Cold Spring Harbor Press, New York City; 1992).