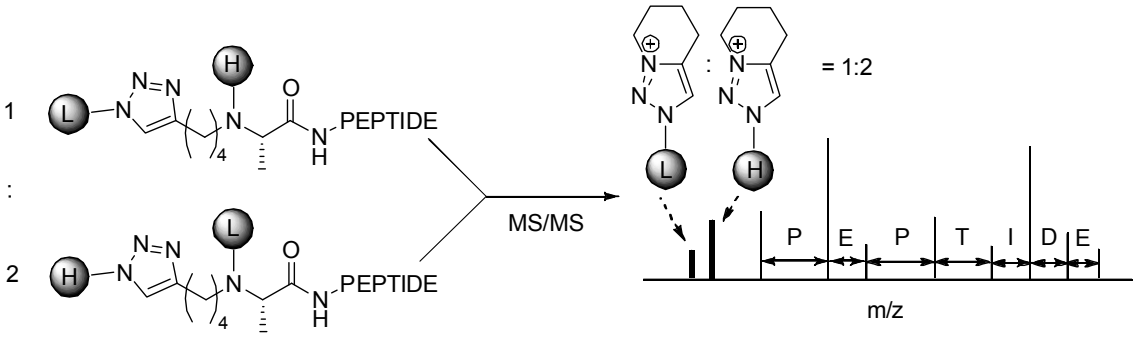
4. Designer Reagents for Mass Spectrometry-Based Proteomics: Click Chemistry Facilitates Synthesis of Amine-reactive Multiplexed Isobaric Tags for Protein Quantification

Abstract Quantitative mass spectrometry (MS)-based proteomics has been employed to investigate various biological processes by measuring the relative and absolute expression levels of proteins in cells. These studies are enabled by tandem mass spectrometry (MS/MS)-based isobaric tags, made popular by their capability for multiple sample comparison in a single MS experiment. Here, we develop novel isobaric tags for protein quantification, referred to as Caltech Isobaric Tags (CITs), which offer several advantages over other isobaric tags (*e.g.*, iTRAQ and TMT). These include 1) the formation of reporter ions based on a newly discovered low-energy fragmentation pathway, a nucleophilic displacement reaction by the 1,2,3-triazole ring, 2) an unlimited number of isobaric combinations of CIT reagents in principle, 3) an easily tunable reporter ion mass to access clear windows of m/z values not overlapping with peptide MS/MS fragments, and to avoid the low mass–cut off problems inherent in ion trap mass spectrometers, and 4) synthetic methodology that permits preparation of CITs with minimal expense and effort. These advantages are demonstrated by preparing duplex CIT reagents whose reporter ions appear at *m/z* 164 and 169. CIT reagents are applied to label a model tryptic peptide, protein mixture digests (bovine serum albumin, α and β caseins, ovalbumin, and lysozyme; enolase, aldolase, hemoglobin, creatine kinase, and alcohol dehydrogenase), and Cul1 protein complexes affinity-purified from HEK 293 cells with various ratios. The resulting CIT-labeled peptides are analyzed by either pulsed Q-dissociation (PQD) or higher energy collision dissociation (HCD) in LTQ-Orbitrap mass spectrometers. Heavy to light ratios of the CIT reporter ions provide excellent

quantitative results, indicating relative responses over a two-orders-of-magnitude dynamic range, in agreement with initial mixing ratios.

Graphical Abstract



4.1.Introduction

Recent achievements in mass spectrometry (MS)-based proteomics have provided essential methodologies for a deeper understanding of protein expressions in cells.[37,189] MS-based proteomics allows high-throughput identification and quantification of proteins of interest. Currently, state-of-the-art liquid chromatography (LC)-MS instruments can analyze the whole-cell yeast lysate within a day, identifying ~5000 proteins.[190,191] Quantitative approaches in MS-based proteomics are aiming to investigate the relative and absolute expression levels of proteins in cells.[16,17] By employing those approaches, various biological processes can be monitored by tracking changes in protein expression.[192]

The simplest method for quantitative mass spectrometric measurement is the label-free analysis. After successive runs of samples of interest under the same instrument conditions, protein abundances are determined by either integrating ion chromatograms or spectral counting of MS signals.[193] Yet, current label-free quantification approaches require highly consistent analyses, which are mostly hampered by fluctuations in ionization efficiencies and difficulties in subsequent data processing.[193,194]

In another label-free approach, selected reaction monitoring (SRM), and its extension plural multiple reaction monitoring (MRM),[195] examines the transitions (*i.e.*, one or more targeted fragment ions from the precursor ions) by scanning specific mass regions using triple quadruple mass spectrometers. This permits highly sensitive identification and concomitant quantification of peptides.[196-198] SRM, however, requires pre-knowledge of fragmentation behaviors of analytes, for which the transitions should be determined through tedious assays prior to actual SRM analyses. In addition, the high cost for the preparation of required synthetic peptides may limit its

wide application in shotgun proteomics involving complex mixtures. SRM also suffers from problems associated with fluctuation in ionization efficiency.

To address problems associated with label-free protein quantification, stable isotopes are incorporated into protein or peptide samples to be used as internal (or mutual) standards. A conceptual breakthrough was achieved by Aebersold and coworkers by introduction of isotope-coded tags that can be selectively labeled to peptide digests.[199] In this approach, cysteine-containing peptides from different sources are tagged by light or heavy isotope-coded affinity tag (ICAT) reagents, and enriched from the complex mixture utilizing an attached biotin affinity tag. Because these tagged peptides share the same physico-chemical properties, they are not differentiated by ionization and chromatography steps. Therefore, a simple comparison of MS signal intensities between light and heavy isotope-coded peptides directly yields the relative protein expression levels.

Another quantification approach takes advantage of *in vivo* incorporation of isotope labels. In SILAC (Stable Isotope Labeling with Amino acids in Cell culture),[200] two cell populations are grown with identical culture media except for stable isotope-labeled amino acids (*e.g.*, ^{13}C and/or ^{15}N labeled lysine and/or arginine). Resulting heavy isotope-coded cell populations behave equally as their light isotope-coded controls. After applying a perturbation to one of the light or heavy cell populations, the two samples are combined for MS analysis. Direct comparison of the peptide signals from the light- and heavy-labeled cell populations yields the relative protein expression levels. Unlike ICAT, which quantifies partial proteome by labeling cysteine-containing peptides, SILAC is capable of quantifying the global proteome, because all tryptic peptides are labeled with isotope-coded lysine and/or arginine.

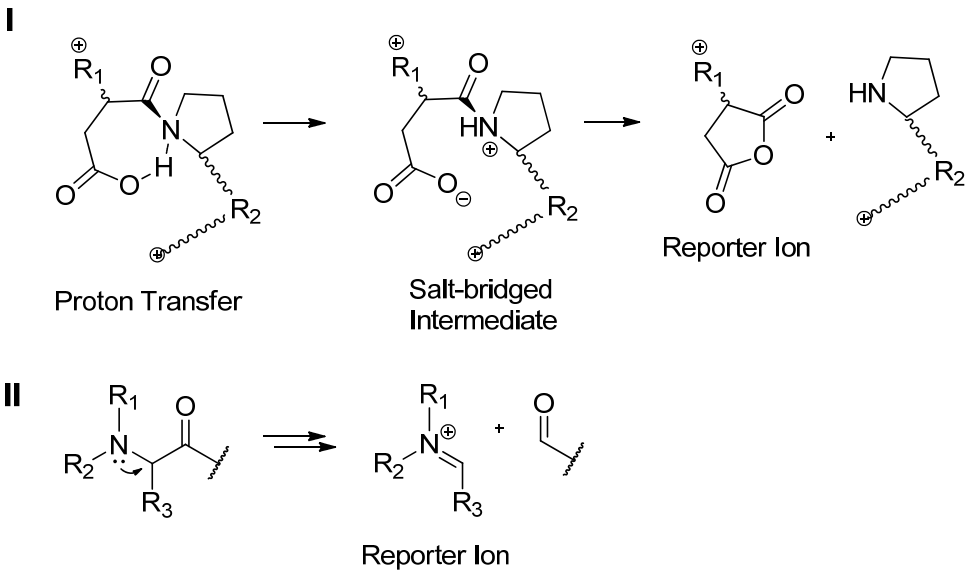
The LC-MS ion signals from isotope-coded peptides in both ICAT and SILAC methodologies are divided into two signals for each labeled peptide, causing an increase in the complexity of MS

scans and a reduction in the sensitivity of subsequent sequence analysis by MS/MS. As an alternative chemical-labeling method, isobaric tags such as Tandem Mass Tag (TMT)[18,201] and isobaric Tags for Relative and Absolute Quantification (iTRAQ)[19,202] were developed. An isobaric tag is composed of three parts: the reporter ion group, the mass balance group and the reactive group.[203] The combined mass of the reporter ion group and the mass balance group is isobaric despite each part having different masses. When the identical peptides from different biological experiments are labeled by isobaric tags, they possess the same masses, appearing as a single peak. Mass differentiated reporter ions are produced and detected in MS/MS scans, in which their relative intensities reflect the initial amounts of each peptide from the original sources. All backbone fragment ions are also isobaric, allowing simultaneous peptide sequencing and quantification. One advantage of isobaric tags over ICAT- and SILAC-based methodologies, especially when quantifying more than two system states, is that the combined signals from multiple biological samples reduce the complexity of the MS scans and increase the sensitivity for the following MS/MS analyses. Preparing each part of the tag with various isotopomers in principle allows for the facile construction of multiplexed reagents capable of quantifying multiple samples in a single MS analysis. The primary amines that are usually the target functional groups for isobaric tag labeling exist in virtually all peptides (*i.e.*, the N-terminal amine and ε-amine of lysine side-chain), enabling researchers to investigate the whole proteome. The isobaric tags have been employed for quantification of tissue samples and even isotopically non-transferable subjects such as human.[204]

In spite of the improvements achieved by isobaric tags, their applications have been limited by the cost of commercially available reagents.[205] In addition, the number of the current multiplexed isobaric tags (*e.g.*, iTRAQ) is limited to a maximum of eight[202] due to their inherent designs and difficulties in the synthesis of various isotope-coded functional groups. The low mass

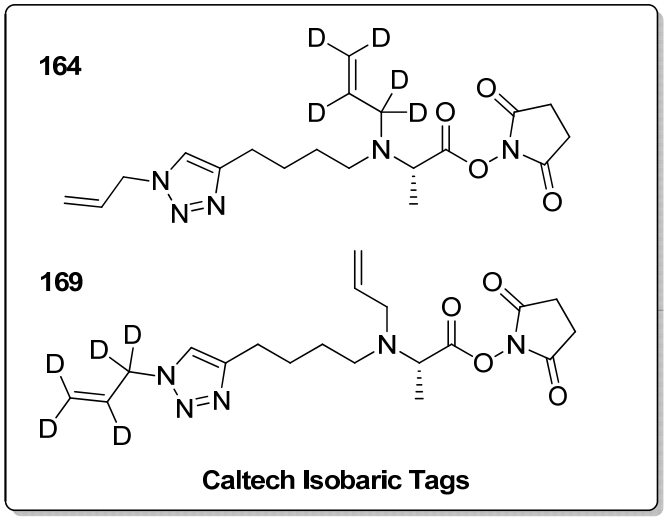
cut–off in resonance type ion trap mass spectrometers, one of the most popular proteomics platforms, also hinders the simultaneous monitoring of reporter ions and peptide sequence ions.

The design of new reagents is hindered by the fact that very few low-energy fragmentation pathways suitable for the production of the reporter ions are known in peptide tandem MS. One of the widely known fragmentation pathways, the preferential cleavage of the aspartic acid–proline peptide bond via a salt-bridged intermediate,[10,11] was applied in the first TMT report (**Scheme 4.1**, I).[18] In iTRAQ, the formation of the reporter ions proceeds through the facile *N*-methylpiperazine-acetyl bond−mediated cleavage process (**Scheme 4.1**, II). Most isobaric tags that have been proposed to provide cheaper synthetic routes are still based on similar tertiary amine-branched methylene-amide bonds.[206-208]



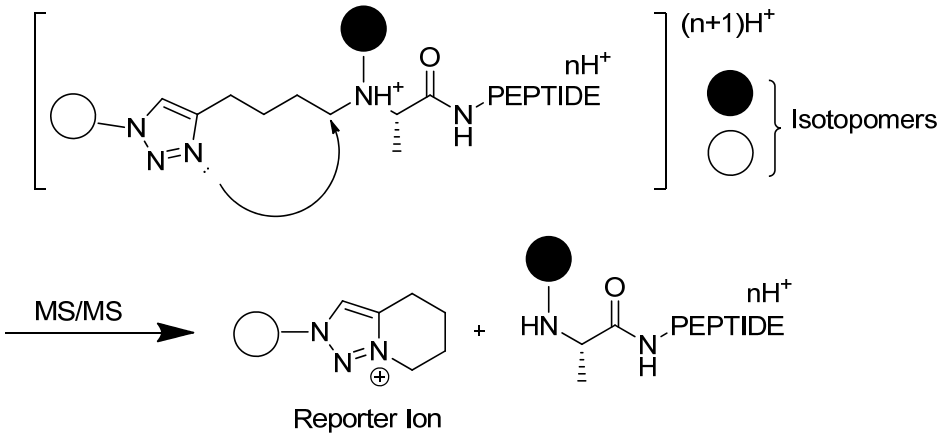
Scheme 4.1

Here, we report a novel isobaric reagent (**Scheme 4.2**), referred to as Caltech Isobaric Tag (CIT), which affords a significant advance in MS-based quantification methodology.



Scheme 4.2

CIT has many advantages over current isobaric labeling reagents. First, CIT is easy to produce and the possible number of multiplex isobaric tags, in theory, is unlimited. Furthermore, energetics for the reporter ion and peptide sequence ion formations are balanced, guaranteeing simultaneous quantification and sequencing of target peptides. CIT is inspired by the observation of a highly selective gas-phase fragmentation triggered by a nucleophilic substitution of N3 nitrogen in the 1,2,3-triazole ring (**Scheme 4.3**). The design of a duplex CIT reagent is described, and the applicability of the reagent is validated in model systems using various mass spectrometers.



Scheme 4.3

4.2.Experimental Sections

4.2.1. Materials

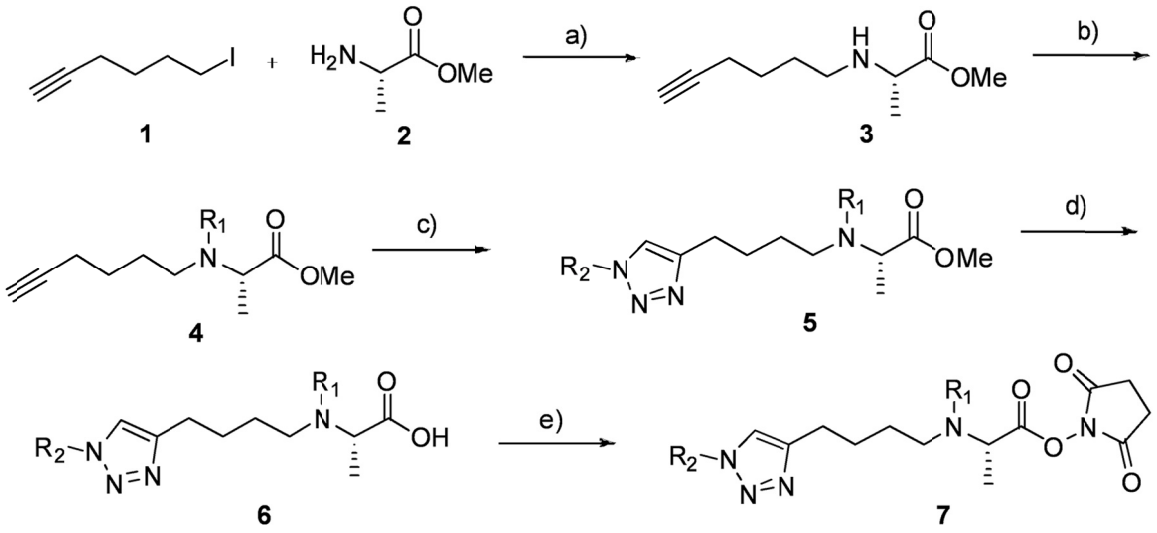
4.2.1.1.Chemicals

Allyl bromide-d_5 was purchased from C/D/N isotopes Inc. (Quebec, Canada). MagneHis Ni-Particles and sequencing grade trypsin was purchased from Promega (Madison, WI). The model peptide, VIP (residue 1-12), HSDAVFTDNYTR, was acquired from Anaspec (Fremont, CA). High Capacity Neutravidin Agarose Resin, n-dodecyl-β-D-maltoside, and SuperSignal West Dura Extended Duration chemiluminescent substrate were from Thermo Scientific (Rockford, IL). Lysyl endopeptidase (LysC) was from Wako Chemicals USA (Richmond, VA). Cell culture reagents, Flip-In T-REx 293 cells, plasmids and monoclonal antibodies for Cul1 and Cand1 were from Invitrogen (Carlsbad, CA). Plasmid DNA containing the human Cul1 sequence was purchased from Open Biosystems (Huntsville, AL). MLN4924 was a generous gift from Millennium: The Takeda Oncology Company (Cambridge, MA). All other general chemicals for buffers were purchased from Fisher Scientific (Hampton, NH), VWR International (West Chester, PA) and Sigma-Aldrich (St. Louis, MO) and used as received without further purification.

4.2.2. Synthesis of CITs

Synthetic schemes are summarized in **Figure 4.1**.

Figure 4.1 Synthesis of CIT reagents



Synthesis of CIT reagents. a) THF, K_2CO_3, TEAI, reflux, 18 h, 56%. b) THF, K_2CO_3, TEABr, reflux, 18 h, R_1=Allyl-d_0-bromide, 56%, R_1=Ally-d_5-bromide, 67%. c) 0.4 eq Na ascorbate, 0.1 eq CuSO4, 0.01 eq TBTA, DMSO/H_2O, RT, 4 h, R_2=Allyl-d_5-azide, 72% (heavy tag), R_2=Allyl-d_0-azide, 69% (light tag). d) 2M KOH, THF, RT, overnight, *quantitative* (heavy tag), 97% (light tag). e) TFA-NHS, DMF, overnight, 24% (heavy tag), 23% (light tag). THF = tetrahydrofuran, TEAI = tetraethylammonium iodide, TEABr = tetraethylammonium bromide, TBTA = tris[(1-*t*-butyl-1*H*-1,2,3-triazol-4-yl)methyl]amine, DMSO = dimethyl sulfoxide, TFA-NHS = trifluoroacetic *N*-hydroxysuccinimide ester, DMF = *N,N'*-dimethyl formamide.

4.2.2.1. 6-Iodo-hex-1-yne

In a flame-baked, one neck 250 mL round bottom flask, 40 mmol of 6-chloro-1-hexyne (~5mL) was dissolved in 25 mL of acetone with 80 mmol of sodium iodide (~12 g) and refluxed for 2 days with thin layer chromatography (TLC) check. After filteration, the mixture was properly diluted by diethylether. The organic layer was washed by water, $Na_2S_2O_4$ and brine and dried over $MgSO_4$. The resulting solution was concentrated by rotavap with caution (the product is slightly volatile). The desired product, 6-iodo-1-hexyne (5.58 g, 26.4 mmol) was acquired as a brownish oil. Yield: 66%

4.2.2.2. N-(5-hexynyl) L-alanine methyl ester

In a flame-baked, two neck 100 mL round bottom flask, 20 mmol of L-alanine methyl ester hydrochloride (2.8 g), 40 mmol of K_2CO_3 (5.53 g), and 20 mmol of tetrabutylammonium iodide (TBAI, 7.39 g) were charged under the stream of dry N_2 gas. 35 mL of tetrahydrofuran (THF) was slowly added and stirred for 15 min at room temperature. 2.6 mL of 6-iodo-hex-1-yne was added dropwise while the mixture was stirred. The reaction mixture was refluxed at ~70°C for 15-18 h with TLC check. After consumption of the starting material, the mixture was cooled to room temperature, diluted by diethylether, and filtered. The filtrate was further diluted by diethylether and filtered again. The solvent was removed by rotavap and the crude product was purified by flash chromatography on silica gel (1:2 = Hexane/EtOAc, 1% triethylamine) to give *N*-(5-hexynyl) L-alanine methyl ester (2.044 g, 11.2 mmol) as a yellow oil. Yield: 56%. R_f = 0.27 (1:1 = Hexane/EtOAc); ESI-MS $[M+H]^+$ = *m/z* 184.1; ^{1}H NMR (300 MHz, $CDCl_3$): δ 3.70 (s, 3H), 3.32 (q, *J* = 7.0 Hz, 1H), 2.53 (m, 2H), 2.18 (m, 2H), 1.92 (t, *J* = 2.6 Hz, 1H), 1.56 (m, 5H),

1.27 (d, J = 7.0 Hz, 3H); ^{13}C NMR (75 MHz, $CDCl_3$): δ 176.29, 84.21, 68.41, 56.61, 51.74, 47.41, 29.20, 26.01, 19.11, 18.23.

4.2.2.3. N,N-(5-hexynyl)(allyl-d_0) L-alanine methyl ester and N,N-(5-hexynyl)(allyl-d_5) L-alanine methyl ester

In a flame-baked, two neck 50 mL round bottom flask, 10 mL THF was charged under the stream of dry N_2 gas. 6 mmol of K_2CO_3 (0.83 g), and 6 mmol of tetrabutylammonium bromide (TBAB, 1.93 g) were slowly added and stirred for 15 min in room temperature. 3 mmol of *N*-(5-hexynyl) L-alanine methyl ester (0.55 g), and 4.5 mmol of allyl bromide-d_0 (0.544g, 0.38 mL) was slowly added dropwise using the syringe while the mixture was stirred. The reaction mixture was heated at ~55 °C and reacted for 8 h with TLC check. After consumption of the starting material, the mixture was cooled to room temperature, diluted by diethylether, filtered and repeated filteration twice to remove the remaining TBAB completely. The solvent was removed by rotavap and the crude product was purified by flash chromatography on silica gel (20:1 = Hexane/EtOAc) to give *N,N*-(5-hexynyl)(allyl-d_0) L-alanine methyl ester (0.378 g, 1.69 mmol) as a transparent oil. Yield: 56%. R_f = 0.4 (5:1 = Hexane/EtOAc); ESI-MS $[M+H]^+$ = *m/z* 224.2; ^{1}H NMR (300 MHz, $CDCl_3$): δ 5.79 (m, 1H), 5.12 (m, 2H), 3.67 (s, 3H), 3.54 (q, J = 7.3 Hz, 1H), 3.18 (m, 2H), 2.53 (m, 2H), 2.18 (m, 2H), 1.93 (t, J = 2.7 Hz, 1H), 1.52 (m, 4H), 1.24 (d, J = 7.3 Hz, 3H); ^{13}C NMR (75 MHz, $CDCl_3$): δ 174.45, 136.86, 116.78, 84.47, 68.24, 57.79, 54.20, 51.21, 49.85, 27.40, 26.07, 18.26, 14.88.

0.457g of *N,N*-(5-hexynyl)(allyl-d_5) L-alanine methyl ester (2.0 mmol) was obtained by the same procedure described above using 3.55 mmol of allyl bromide-d_5 (0.448 g). Yield: 67%. R_f = 0.4 (5:1 = Hexane/EtOAc); ESI-MS $[M+H]^+$ = *m/z* 229.3; ^{1}H NMR (500 MHz, $CDCl_3$): δ 3.68 (s, 3H), 3.54 (q, J = 7.08 Hz, 1H), 2.53 (m, 2H), 2.18 (m, 2H), 1.93 (t, J = 2.68 Hz, 1H), 1.52 (m,

4H), 1.25 (d, J = 7.08 Hz, 3H); ^{13}C NMR (125 MHz, $CDCl_3$): δ 174.47, 136.16 (t, J = 23.48 Hz), 116.25 (quintet, J = 23.47 Hz), 84.47, 68.27, 57.76, 53.28 (quintet, J = 18.41 Hz), 51.22, 49.79, 27.39, 26.07, 18.26, 14.90.

4.2.2.4. Allyl-d_0 azide and allyl-d_5 azide

0.5 M NaN_3 in DMSO was prepared as described in the literature by stirring the mixture at room temperature overnight.[209] 1.1 eq of 0.5 M NaN_3 solution was mixed with the appropriate amount of allyl bromide (d_0/d_5) and stirred overnight. TLC was monitored for the complete consumption of the starting material and unwanted dimerization of allyl azides was not observed in this condition. The resulting mixtures were used for the next steps without further purification or analysis.

4.2.2.5. N,N-(4-(1-allyl-d_5-1H-1,2,3-triazol-4-yl)butyl)(allyl-d_0) L-alanine methyl ester and N,N-(4-(1-allyl-d_0-1H-1,2,3-triazol-4-yl)butyl)(allyl-d_5) L-alanine methyl ester

The literature procedure was followed with some modifications.[210] To *in situ* prepared allyl-d_5 azide solution (~1.2 eq), 1.7 mmol of *N,N*-(5-hexynyl)(allyl-d_0) L-alanine methyl ester (0.378 g), 0.17 mmol of $CuSO_4 \cdot 5H_2O$ (42.5 mg, 0.1 eq), 0.68 mmol of sodium ascorbate (134.7 mg, 0.4 eq), 0.017 mmol of tris[(1-*t*-butyl-1*H*-1,2,3-triazol-4-yl)methyl]amine (TBTA) (7.3 mg, 0.01 eq), and additional 2 mL of DMSO were added and stirred for 2 h at room temperature. 2.4 mL of water was then added and stirred for additional 2 h with monitoring TLC. After the complete consumption of the starting material, 4 mL of 1M NH_4OH was added to remove residual CuN_3 and $(Cu)_2N_3$. The mixture was diluted by additional water and ethyl acetate. The aqueous layer turned to be blue by the coordination of ammonia to copper ions. The organic layer was separated, and further extracted by ethyl acetate three times. The combined organic layer was then washed

by brine, dried over $MgSO_4$ and concentrated by rotavap. The crude product was purified by flash chromatography on silica gel (1:1 = Hexane/EtOAc, 1% triethylamine) to give *N,N*-(4-(1-allyl-d_5-1*H*-1,2,3-triazol-4-yl)butyl)(allyl-d_0) L-alanine methyl ester (0.379 g, 1.22 mmol) as a transparent oil. Yield: 72%. R_f = 0.3 (1:1 = Hexane/EtOAc); ESI-MS $[M+H]^+$ = *m/z* 312.3, CID of $[M+H]^+$ produced *m/z* 169.1 fragment; ^{1}H NMR (500 MHz, $CDCl_3$): δ 7.25 (s, 1H), 5.76 (m, 1H), 5.14 (dd, *J* = 17.1, 1.22 Hz, 1H), 5.04 (d, *J* = 10 Hz, 1H), 3.64 (s, 3H), 3.51 (q, *J* = 7.08 Hz, 1H), 3.21 (dd, *J* = 14.6, 5.6 Hz, 1H), 3.09 (dd, *J* = 14.4, 6.9 Hz, 1H), 2.68 (t, *J* = 7.6 Hz, 2H), 2.52 (m, 2H), 1.54 (m, 4H), 1.21 (d, *J* = 7.08 Hz, 3H); ^{13}C NMR (125 MHz, $CDCl_3$): δ 174.46, 148.43, 136.81, 130.94 (t, *J* = 24.9 Hz), 120.40, 119.21 (quintet, *J* = 24.9 Hz), 116.82, 57.85, 54.25, 51.85 (quintet, *J* = 23.9 Hz), 51.24, 50.20, 27.97, 27.06, 25.54, 14.83.

1.59 mmol of *N,N*-(5-hexynyl)(allyl-d_5) L-alanine methyl ester (0.363 g) was used for the same reaction described above to give *N,N*-(4-(1-allyl-d_0-1*H*-1,2,3-triazol-4-yl)butyl)(allyl-d_5) L-alanine methyl ester (0.341 g, 1.10 mmol) as a transparent oil. Yield: 69%. R_f = 0.3 (1:1 = Hexane/EtOAc); ESI-MS $[M+H]^+$ = *m/z* 312.3, CID of $[M+H]^+$ produced *m/z* 164.1 fragment; ^{1}H NMR (500 MHz, $CDCl_3$): δ 7.25 (s, 1H), 5.98 (m, 1H), 5.30 (dd, *J* = 10, 0.9 Hz, 1H), 5.25 (d, *J* = 17.1 Hz, 1H), 4.91 (d, *J* = 6.1 Hz, 2H), 3.65 (s, 3H), 3.51 (q, *J* = 7.1 Hz, 1H), 2.69 (t, *J* = 7.6 Hz, 2H), 2.52 (m, 2H), 1.55 (m, 4H), 1.22 (d, *J* = 7.1 Hz, 3H); ^{13}C NMR (125 MHz, $CDCl_3$): δ 174.49, 148.45, 136.12 (t, *J* = 23 Hz), 116.29 (quintet, *J* = 22.6 Hz), 120.43, 119.78, 57.84, 53.33 (quintet, *J* = 19.3 Hz), 52.55, 51.24, 50.15, 27.98, 27.07, 25.55, 14.85.

4.2.2.6. N,N-(4-(1-allyl-d_5-1H-1,2,3-triazol-4-yl)butyl)(allyl-d_0) L-alanine and N,N-(4-(1-allyl-d_0-1H-1,2,3-triazol-4-yl)butyl)(allyl-d_5) L-alanine

1.22 mmol of *N,N*-(4-(1-allyl-d_5-1*H*-1,2,3-triazol-4-yl)butyl)(allyl-d_0) L-alanine methyl ester (0.375 g) was charged to a 10mL one neck flask with 2 mL of THF and 2 mL of 2M KOH and

stirred at room temperature for 10 h. The reaction was monitored by TLC and ESI-MS until the starting material was completely consumed. Upon completion of the reaction, THF was removed by rotavap and the aqueous layer was neutralized by ~2 mL of 2M HCl. Water was then completely removed by rotavap and the residue was reconstituted by acetonitrile (ACN). Insoluble KCl salt was filtered and ACN was removed by rotavap. The free acid of the alanine derivative, *N,N*-(4-(1-allyl-d_5-1*H*-1,2,3-triazol-4-yl)butyl)(allyl-d_0) L-alanine was recovered as a pale yellow gleasy oil. Yield: *quantitative*. ESI-MS $[M+H]^+$ = *m/z* 298.1; ^{1}H NMR (500 MHz, DMSO-d_6): δ 10.62 (br, 1H), 7.90 (d, *J* = 2.2 Hz, 1H), 6.05 (m, 1H), 5.53 (d, *J* = 17.1 Hz, 1H), 5.45 (d, *J* = 10.5 Hz, 1H), 4.18 (q, *J* = 7.1 Hz, 1H), 3.83 (m, 2H), 3.15 (br, 2H), 2.64 (t, *J* = 7.6 Hz, 2H), 1.69 (m, 4H), 1.52 (d, *J* = 6.9 Hz, 3H).

N,N-(4-(1-allyl-d_0-1*H*-1,2,3-triazol-4-yl)butyl)(allyl-d_5) L-alanine methyl ester (0.247 g, 0.79 mmol) was used for hydrolysis by the same procedure described above and 0.230 g of *N,N*-(4-(1-allyl-d_0-1*H*-1,2,3-triazol-4-yl)butyl)(allyl-d_5) L-alanine (0.773 mmol) was obtained as a pale yellow gleasy oil. Yield: 97%. ESI-MS $[M+H]^+$ = *m/z* 298.1; ^{1}H NMR (500 MHz, DMSO-d_6): δ 10.54 (br, 1H), 7.89 (d, *J* = 4.2 Hz, 1H), 6.03 (m, 1H), 5.25 (m, 1H), 5.16 (m, 1H), 4.98 (m, 2H), 4.18 (q, *J* = 7.1 Hz, 1H), 3.15 (br, 2H), 2.64 (t, *J* = 7.3 Hz, 2H), 1.69 (m, 4H), 1.52 (d, *J* = 7.1 Hz, 3H).

4.2.2.7.N,N-(4-(1-allyl-d_5-1H-1,2,3-triazol-4-yl)butyl)(allyl-d_0) L-alanine N-hydroxysuccinimide ester and N,N-(4-(1-allyl-d_0-1H-1,2,3-triazol-4-yl)butyl)(allyl-d_5) L-alanine N-hydroxysuccinimide ester

In a flame-baked 50 mL one neck flask, 2.75g of *N*-hydroxysuccinimide was added to 14 mL of trifluoroacetic anhydride at room temperature under the stream of dry N_2 gas and stirred for 4 h. The solvent was removed by rotavap and further eliminated by highvac overnight. The white

crystal product, trifluoroacetic *N*-hydroxysuccinimide ester (TFA-NHS) was obtained, stored in the dry desiccator, and used just before activation of free acids.

In a flame-baked 50 mL one neck flask, 87 mg of *N,N*-(4-(1-allyl-d_5-1*H*-1,2,3-triazol-4-yl)butyl)(allyl-d_0) L-alanine (0.29 mmol) and 75 mg of TFA-NHS were added to 1 mL of dry DMF, and stirred overnight at room temperature. After the complete consumption of the starting material by monitoring TLC, the reaction mixture was separated by flash chromatography on silica gel (1:1 = Hexane/EtOAc) and yielded 28 mg of *N,N*-(4-(1-allyl-d_5-1H-1,2,3-triazol-4-yl)butyl)(allyl-d_0) L-alanine *N*-hydroxysuccinimide ester (~0.7 mmol) as a yellow oil. (Note: the poor recovery yields for the final products are observed due to the retained products in silica gel through the coordination of the highly *N*-substituted residues such as the 1,2,3-triazole and the tertiary amine. Other purification methods such as crystallization would improve the overall yield.) Yield: 24%. ESI-MS (100% ACN) $[M+H]^+$ = *m/z* 395.1. The stock solution of the heavy tag (*m/z* 169 reporter ion) was prepared without further analysis by adding 20 μL dry DMSO to 1 mg of the NHS-ester product into each vials, and stored in −80 °C. Each vial contains 1mg of the reagent and used for each labeling experiment appropriately.

The same procedure was used for NHS ester activation of 52 mg of *N,N*-(4-(1-allyl-d_0-1*H*-1,2,3-triazol-4-yl)butyl)(allyl-d_5) L-alanine (0.175 mmol) and yielded 16 mg of the NHS-ester product. Yield: 23%. ESI-MS (100% ACN) $[M+H]^+$ = *m/z* 395.1. The stock solution vials of the light tag (*m/z* 164 reporter ion) were prepared as described above and stored in −80 °C.

4.2.3. Synthesis of iTRAQ-113 Reagent

4.2.3.1. Methyl 2-(4-methylpiperazin-1-yl) acetate

In a flame-baked, two neck 100 mL round bottom flask, 5 mmol of 1-methyl piperazine (0.5 g), and 6 mmol of K_2CO_3 (0.83 g) were charged under the stream of dry N_2 gas. Fifteen mL of

tetrahydrofuran (THF) was slowly added and stirred for 15 min at room temperature. Zero point fivesix mL of methyl boromoacetate was added dropwise while the mixture was stirred. The mixture was reacted at room temperature for 15-18 h with TLC check. After consumption of the starting material, the mixture was cooled to room temperature, diluted by diethylether, and filtered. The solvent was removed by rotavap and the crude product was purified by flash chromatography on silica gel (5:1 = $CHCl_3$/MeOH) to methyl 2-(4-methylpiperazin-1-yl)aceate (0.111 g, 0.65 mmol) as a pale yellow oil. Yield: 13%. R_f = 0.36 (5:1 = $CHCl_3$/MeOH); ESI-MS $[M+H]^+$ = *m/z* 173.1, CID of $[M+H]^+$ produced *m/z* 113.0 fragment; ^1H NMR (300 MHz, $CDCl_3$): δ 3.68 (s, 3H), 3.18 (s, 2H), 2.52 (m, broad, 8H), 2.26 (s, 3H); ^{13}C NMR (75 MHz, $CDCl_3$): δ 170.70, 59.36, 54.71, 52.94, 51.70, 45.91.

4.2.3.2. 2-(4-methylpiperazin-1-yl)acetic acid

51.3 mg of methyl 2-(4-methylpiperazin-1-yl)acetate was charged to a 10mL one neck flask with 2 mL of THF and 2 mL of 2M KOH and stirred at room temperature for 10 h. The reaction was monitored by TLC and ESI-MS until the starting material was completely consumed. Upon completion of the reaction, THF was removed by rotavap and the aqueous layer was neutralized by ~2 mL of 2M HCl. Water was then completely removed by rotavap and the residue was reconstituted by acetonitrile (ACN). Insoluble KCl salt was filtered and ACN was removed by rotavap. The free acid of the product was recovered as a transparent gleasy oil. Yield: *quantitative*. ESI-MS $[M+H]^+$ = *m/z* 159.1.

4.2.3.3. N-hydroxylsuccinimide 2-(4-methylpiperazin-1-yl)acetate ester (iTRAQ-113)

In a flame-baked 15 mL glass vial, ~0.65 mmol of 2-(4-methylpiperazin-1-yl)acetic acid and 33.0 mg of TFA-NHS were added to 0.5 mL of dry DMF and 101 mg of *N*,*N*-diisopropylethylamine, and stirred overnight at room temperature. After the complete consumption of the starting material by monitoring TLC, the reaction mixture was used for peptide labeling without further purification. ESI-MS (100% ACN) $[M+H]^+$ = *m/z* 256.1. The stock solution vials of the iTRAQ-113 reagent (*m/z* 113 reporter ion) were prepared as described above and stored in −80 °C.

4.2.4. Protein Mixture Digestion

Two sets of equal amounts (by weight) of protein mixtures, 1) proteins, bovine serum albumin, ovalbumin, α and β caseins, lysozyme and 2) enolase, aldolase, hemoglobin, creatine kinase and alcohol dehydrogenase were digested by trypsin (50:1, w/w) in 50 mM ammonium bicarbonate buffer for 15 h at 37 °C. The reaction was quenched by 5% formic acid, and the resulting peptide products were desalted with C_{18} spin columns (Satorious Stedim Biotech, Aubagne Cedex, France).

4.2.5. Affinity Purification and Digestion of Cul1 and Its Associated Proteins

The applicability of CIT to the quantification of protein complexes in the cell was evaluated with Cul1. Cul1 is a ubiquitin ligase that attaches a ubiquitin chain on target substrates for proteasome-catalyzed degradation.[211] Cul1 is a prototype of the cullin ligase family, and constitutes modular ligase complexes with many confirmed binding partners. The purification of Cul1 and its binding partners from the cells were carried out as described previously with minor

modifications.[212] Briefly, to facilitate the purification of Cul1, a HEK 293-derived stable cell line capable of expressing tagged Cul1 upon tetracycline treatment was constructed using the T-REx™ (Tetracycline-regulated Expression) system (Invitrogen, Carlsbad, CA). The tandem tag we adopted consists of a hexa-histidine sequence and a biotinylation signal sequence.[213] Biotinylation is catalyzed by endogenous biotin ligases, which are present in all eukaryotic cells.[214] A specific lysine residue in the biotinylation signal sequence functions as an acceptor site for biotin *in vivo*.[215]

Tagged Cul1 was induced with 1.0 μg/mL tetracycline for 4 h in experiments aimed to quantify Cul1 complexes with an initial 1:2 mixing ratio for the light and heavy CIT labeling. For the quantification of differentially expressed Cul1 in the cell, 0.5 or 2.0 μg/mL tetracycline were added to the growth medium for 1 h or 4 h, respectively. Twenty four hours after induction, cells were lysed for 30 min at 4 °C with the lysis buffer (0.050 M HEPES, pH 7.5, 0.0050 M Mg(OAc)2, 0.070 M KOAc, 10% glycerol, and 0.4% IGEPAL CA630). The lysate was centrifuged at 16,600 g at 4 °C for 20 min and the supernatant was used for purification. Tandem purification of tagged Cul1 and associated proteins was carried out using MagneHis Ni-particles from Promega and Streptavidin-coupled Dynabeads from Invitrogen. Purified proteins were subjected to sequential on-bead protease digestions, first with Lys-C (35 ng/mg lysate) for 4 h at 37 °C in 8 M urea, and then with trypsin (30 ng/mg lysate) for 12 h at 37 °C in 2 M urea. The resulting tryptic peptides were desalted with C_{18} spin columns.

4.2.6. CIT Labeling

The model peptide, VIP (residue 1-12, HSDAVFTDNYTR; 50 μg), was dissolved in 50 μL of 100 mM tetraethylammonium bicarbonate (pH 8.5) and 100 μL of ACN (66.7% organic phase) and labeled with 5 μL of 5 μg/μL DMSO stock solution of either light or heavy tag, by incubating

for 2 h at room temperature. The reaction was quenched by adding 50 μL of 100 mM hydroxylamine and incubated for 7 h at room temperature. The mixture was acidified by adding 4 μL of formic acid and completely dried by speedvac. The residue was reconstituted by 100 μL of 0.1% formic acid, desalted by the C_{18} desalting tip, and eluted to 100 μL of 0.1% formic acid, 50% ACN and 50% water. The CIT labeled VIP peptide eluent was properly diluted (~x20), and analyzed by various mass spectrometers.

An aliquot of 1 μg of the protein digest was labeled by light or heavy CIT reagents under the same solvent system used for the model peptide, and mixed with the 1:1 ratio. The combined samples were desalted, and injected to a nanoLC-LTQ-Orbitrap mass spectrometer (Thermo Scientific, San Jose, CA) at Caltech. Approximately 3 μg of Cul1 digests were labeled by light or heavy CIT reagents with the 1:2 ratio under the same solvent system used for the model peptide. After conjugation, light and heavy CIT-labeled Cul1 digests were combined, desalted and injected to a nanoLC-LTQ-Orbitrap mass spectrometer at Caltech. For HCD/CID experiments, differentially expressed Cul1 digest samples were labeled by light or heavy CIT reagents. The resulting peptides were subject to a nanoLC-LTQ-Orbitrap XL mass spectrometer (Thermo Scientific, San Jose, CA) at UCLA for HCD/CID analyses.

4.2.7. Instruments

Matrix-assisted laser desorption/ionization time-of-flight (MALDI TOF) spectra were acquired using Voyager DE-Pro mass spectrometer (Applied Biosystems, Foster City, CA) in the reflectron mode with a 20 kV acceleration voltage, a 150 ns delay extraction time and a 75% grid voltage. A 0.5 μL sample of the derivatized peptide solution was mixed with 0.5 μL of 10 mg/mL CHCA matrix solution in 0.1% TFA, 50% ACN, and 50% H_2O and the mixed spots were dried and introduced to the mass spectrometer for analyses. For all spectra, 100 shots were averaged.

Beam-type CID experiments were performed using Micromass Q-TOF ultra-2 (Waters, UK) in positive ion mode using the Z-spray ion source with a 2.65 kV spraying voltage, a 15 V cone voltage, a 6 V extractor voltage and 22 V (for 3+ ions) or 52 V (for 2+ ions) collision voltages. For beam-type CID spectra, ~100 scans were averaged and used for data analysis.

PQD experiments via direct infusion for model peptide studies were performed by ion trap scans in an LTQ-FTICR mass spectrometer equipped with the Nanomate (Advion BioSciences Inc., Ithaca, NY, USA) nanospray unit. The spraying voltage was 1.4 kV and the gas pressure was 0.3 psi. Critical parameters of the mass spectrometer include capillary voltage 49 V, capillary temperature 200 °C and tube lens voltage 180 V. Other ion optic parameters were optimized by the auto-tune function in the LTQ tune program for maximizing the signal intensity. The parameters for PQD experiments are the precursor isolation window 2.0 *m/z*, collision energy 29%, isolation Q 0.70, and activation time 0.1 ms. For PQD spectra, 100 scans were recorded.

The samples were analyzed by a nanoflow HPLC, Proxeon easy-nLC-System (Proxeon Biosystems) coupled on-line via a nanoelectrospray ion source (Proxeon Biosystems) to a LTQ-Orbitrap mass spectrometer at Caltech. Samples were loaded onto a C_{18}-reversed phase column (15 cm long, 75 μm inner diameter, packed in-house with ReproSil-Pur C_{18}-AQ 3 μm resin (Dr. Maisch)) in buffer A (5% ACN, 0.2% formic acid) with a flow rate of 500 nl/min for 24 min and eluted with a linear gradient from 0% to 36% buffer B (80% ACN 0.2% formic acid) over 110 minutes, followed by 10 minutes at 100% buffer B, at a flow rate of 350 nl/min. The column was re-equilibrated with buffer A. Mass spectra were acquired in the positive ion mode applying data-dependent acquisition with automatic switching between survey scan and tandem mass spectrum acquisition. Samples were analyzed with a top 5 method; acquiring one Orbitrap survey scan in the mass range of *m/z* 400-1600 followed by MS/MS of the five most intense ions in the LTQ in the mass range of *m/z* 100-1600. The target value in the LTQ-Orbitrap was 500,000 for survey

scan at a resolution of 60,000 at *m/z* 400. Fragmentation in the LTQ was performed by Pulsed Q-Dissociation (PQD) with a target value of 5,000 ions. Selected sequenced ions were dynamically excluded for 30 s. General mass spectrometric conditions were: spray voltage, 2.4 kV; no sheath and auxiliary gas flow; ion transfer tube temperature, 200 °C; normalized collision energy (29%) using wide band activation mode for MS/MS. An activation of q = 0.55 and delay time of 0.4 ms were applied in MS/MS acquisitions.[216]

For HCD/CID analyses, CIT-labeled peptides were loaded and washed on a 100 μm×35 mm CVC microtech (Fontana, CA). Peptide trapping was performed using a New Objective Halo C_{18} 75μm × 100 mm, 90 Å, 2.7 μm (Woburn, MA) column by flushing a mobile phase of 0.1% formic acid in water (A). Peptides were subsequently eluted from the column at 300 μL/min using an Eksigent nanoLC 2D pump (Dublin, CA) with a 110 min gradient (0.1% formic acid in water (buffer A) and ACN containing 0.1% formic acid (buffer B); 0-30% phase B over 90min, 30-80% phase B over 20 min). The HPLC system was coupled to an LTQ-Orbitrap XL mass spectrometer and the source conditions were as follows: capillary temperature, 180 C; capillary voltage, 49 V; ESI spray voltage, 1.8 kV. The automatic gain control target was fixed at 5×10^5 ions for MS and 5×10^4 for MS/MS scans. The instrument was operated in data-dependent acquisition mode, with MS survey scan (*m/z* 400-1400*)* performed in the Orbitrap using a resolution set at 60,000. CID and HCD activations were performed on the 3 most abundant ions over 5000 counts (charge state +1 rejected) using normalized collision energies of 30 and 40, respectively, and detected using the linear ion trap. Ions with masses within 10 ppm of previously fragmented ions were excluded for 120s.

4.2.8. Data Processing

The raw files from LTQ-Orbitrap mass spectrometers were converted to mascot generic format (MGF) files using ReAdW4Mascot (version 20090305a, available from the National Institute of Standards and Technology at http://peptide.nist.gov/software/ReAdW4Mascot2_20090305a.zip). PQD of the CIT-labeled model peptide acquired for testing of the linear dependency in reporter ion formation was then analyzed using in-house software and best-fit lines were calculated using linear regression. Error bars are displayed for the middle 95% reported ion ratios. For the five protein mixture, a database was constructed containing the five protein sequences as well as a small contaminant protein database. For the Cul1 pull-down search, a target sequence database was constructed from the human IPI database (version 3.54) and a small containment protein database. A decoy database was constructed from the target following the protocol as described elsewhere.[217] The decoy database was then appended to the target and used to estimate the false discovery rate of the database search. The database search was performed using mascot (version 2.2.06, Matrix Science, http://www.matrixscience.com). The database search parameters were as follows: 0.5 Da fragment ion mass tolerance, 10 ppm precursor ion mass tolerance, trypsin enzyme specificity, up to two missed cleavages, fixed carbamidomethyl (57.02146 Da) modification of cysteine, variable modifications of oxidation (15.99491 Da) of methionine, carbamylation (43.005814 Da) of the N-terminal, and quantitation enabled. The mascot quantitation parameters were as follows: fixed the N-terminal modification of 279.210745 Da with reporter ions at *m/z* 164.1188 and 169.1502, respectively. Reported proteins have at least one unique peptide sequence and two peptide ratios. Reported peptide ratios are included for those peptides whose score is above the homology level and outlier peptide ratios are discarded using the Mascot auto outlier detection. Reported mascot protein quantitation ratios are the median of the top scoring peptide reporter ion ratios.

4.2.9. Density Functional Calculation

The formation of the reporter ion is simulated by the *N*-protonated *N,N*-dimethyl-4-(1-methyl-1*H*-1,2,3-triazol-4-yl)butan-1-amine. Initial coarse geometries were constructed by the MC/MM conformer search using Macromol 8.0 (Schrödinger Inc., Portland, OR, USA) as implemented in Maestro 8.0 (Schrödinger Inc., Portland, OR, USA). The OPLS 2005 was used for the force field model. Within 5 kcal/mol energy, all low energy conformers were initially recorded. Low-energy conformers were selected for further structure optimization by density functional theory (DFT). Each conformer was subject to a geometry optimization using Jaguar 7.5 (Schrödinger Inc., Portland, OR, USA) at the B3LYP/6-311++G(d,p) level. Thermochemical parameters of optimized conformers were estimated by vibrational frequency calculation at 1 atm and 298.15 K at the same level of theory. The transition state structures were searched using the QST method by interpolating initial guesses for reactants, products and transition states. All calculations were performed using computational resources kindly provided by the Material and Process simulation center at the Beckman Institute, Caltech.

4.3.Results and Discussion

4.3.1. Rationale of CIT Design

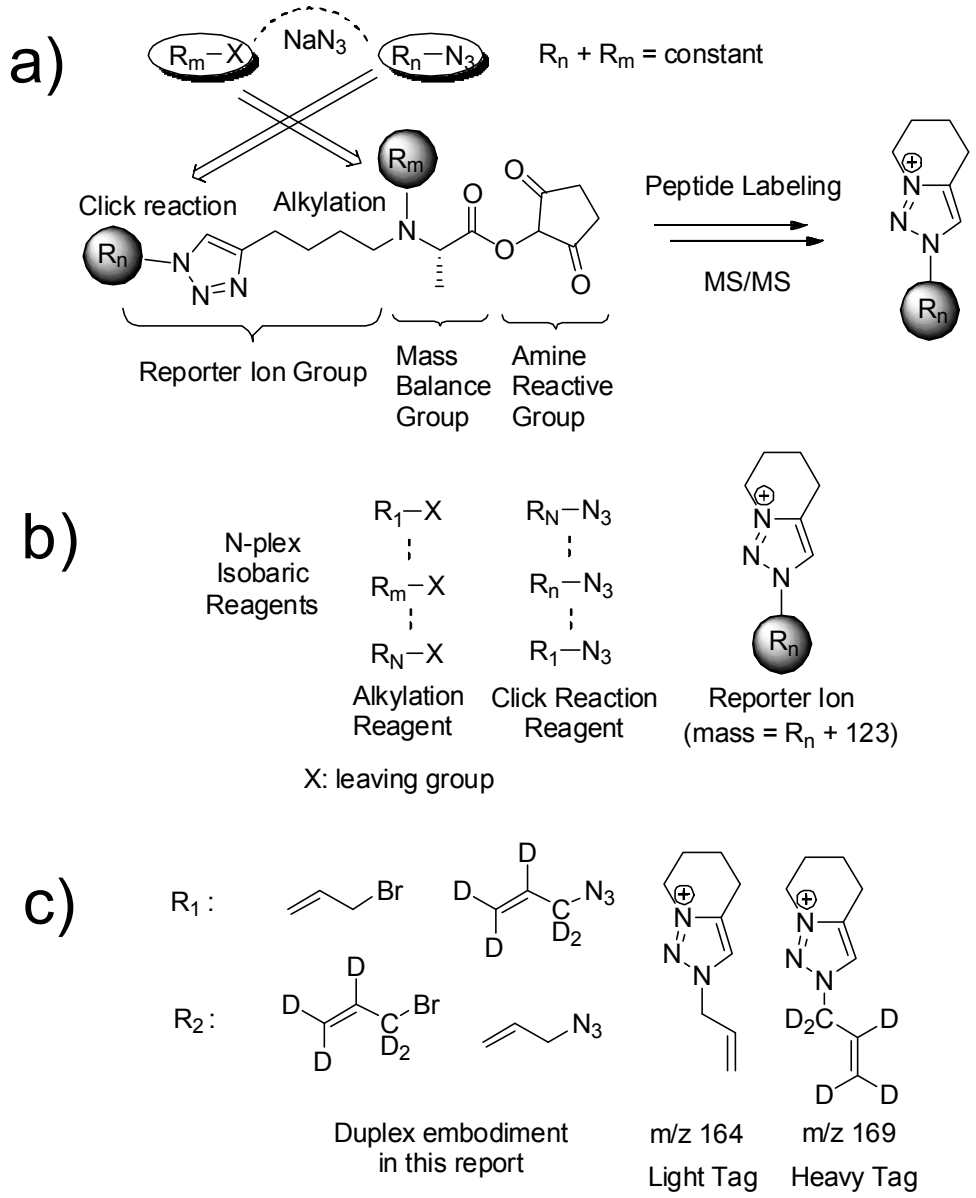
At the inception of this study, a key goal was to find an appropriate gas-phase fragmentation pathway for the formation of the reporter ions. At that time, we were interested in the application of bio-orthogonal azide-alkyne "click" cycloaddition reactions[218-220] to MS-based proteomics studies. Observation of a highly selective gas-phase fragmentation triggered by a nucleophilic substitution of N3 nitrogen in the 1,2,3-triazole ring competitive with the formation of b- and y-

type ions in CID of covalently labeled peptides inspired us to create novel isobaric tags (Scheme 3).

Figure 4.2 depicts the structure of CIT and the construction of the theoretical N-plex reagents. CIT is composed of three parts: the reporter ion group, the mass balance group and the amine reactive group found in other commercially available isobaric tags.

The major improvement of CIT that distinguishes it from other isobaric tags is the *modularization* of the isotope-coded residues, both for the reporter ion group and the mass balance group. Any groups (R) that do not contain other reactive or interfering functionalities can be implemented into the current CIT design by inserting a good leaving group such as bromine, iodine or tosylate. Via a simple S_N2 reaction in the mild conditions employing DMF/NaN_3, the isotope-coded reporter ion group can be easily prepared from an activated R group (**Figure 4.2**). Each isobaric pair of R_m−X (X: leaving group) and R_n−N_3 forms a building block for an isobaric tag with a certain reporter ion mass (R_n + 123 Da). By preparing a set of the N different isotope-coded R−X, it is possible to construct the N-plex isobaric reagents. This modularity of CIT significantly reduces the effort and cost of synthesis. This feature is made possible by the newly discovered low energy gas phase fragmentation pathway depicted in **Scheme 4.3**, which occurs regardless of the attached R groups. In addition, the mass of the reporter ion is *tunable*; this property enables us to bypass the mass cut-off problem in ion trap mass spectrometers and target open windows of m/z values normally found in peptide tandem MS (*e.g.*, sequence ions, immonium ions or internal fragments).

Figure 4.2 Design and structure of CIT



a) The components of N-plex CIT reagents: the reporter ion group, the mass balance group and the amine reactive group. b) Each reporter ion group and mass balance group can be prepared from a series of isotope-coded iodinated R_n groups. CIT labeled peptides are fragmented by various ion activation methods (*e.g.*, PQD, beam-type CID, and High energy Collisional Dissociation (HCD)), yielding the reporter ions whose masses are R_n + 123 Da. c) the duplex embodiment of the CIT reagents in this report by using allyl bromide-d_0 and d_5. Note that the reporter ion is formed regardless of the structure of the attached R_n or R_m groups.

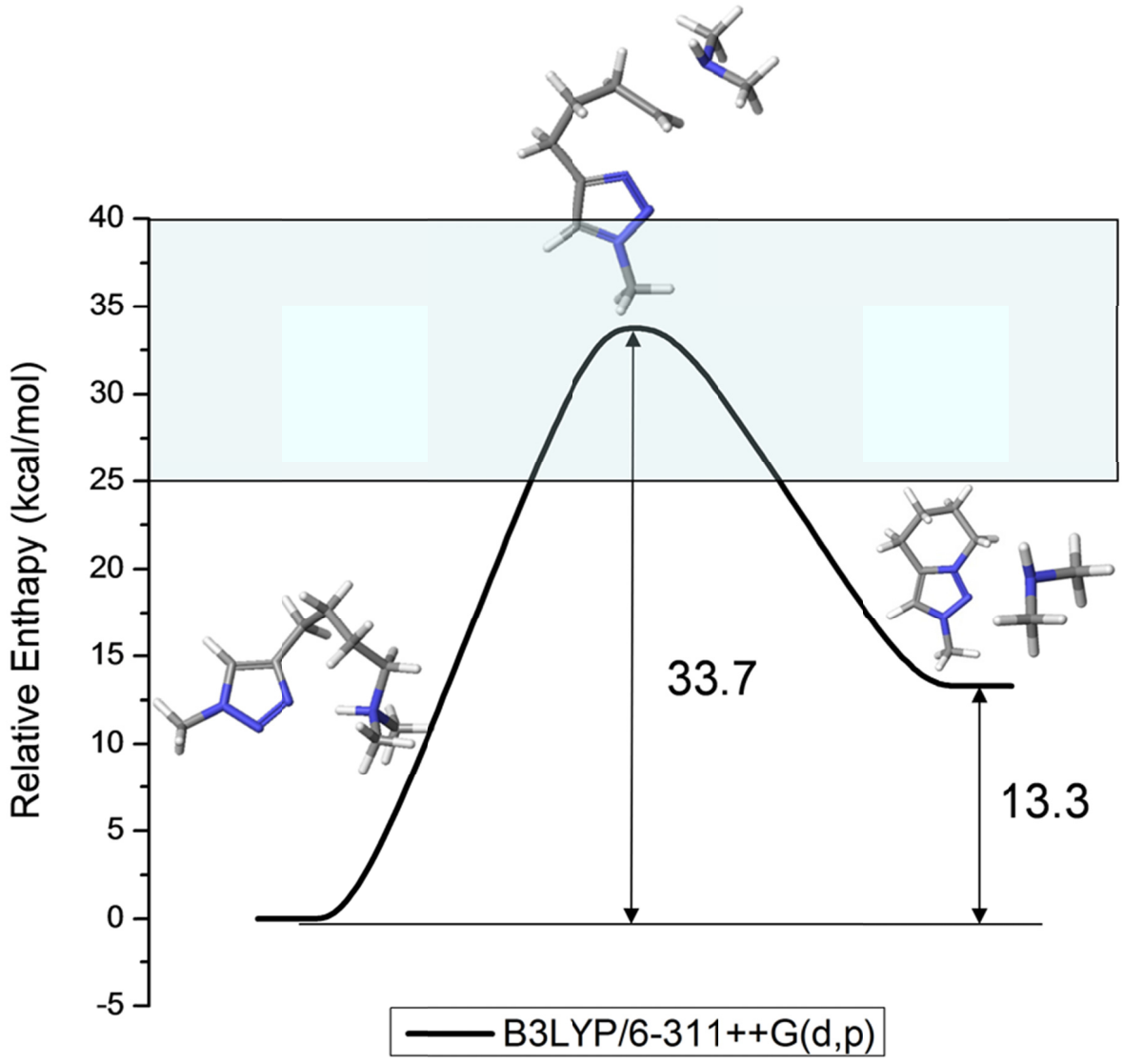
These new discoveries are applied to the creation of a prototype CIT using allyl bromide-d_0 and d_5 as the isotope-coded starting materials. Duplex CIT reagents were synthesized, with heavy and light isotopes having the reporter ions at *m/z* 164 and 169, respectively. The size of the overall modification by this duplex CIT reagent is 279 Da, which is not much larger than most of the commercially available isobaric tags (iTRAQ 4-plex, 144 Da; TMT 6-plex, 224 Da; iTRAQ 8-plex, 304 Da). Only effective collision between peptide amine and *N*-hydroxysuccinimide (NHS) of CIT induces the actual coupling reaction. Therefore, if the size of a certain isobaric tag is relatively large, the kinetics of conjugation can be adversely affected due to decreased frequency of the effective collision. Because the size of CIT is comparable to other isobaric tags, no significant slowdown in CIT labeling is expected.

We adopt the NHS group for facile amine-reactive coupling to peptides as in other commercially available isobaric tags. NHS has been popular in bioconjugation due to the compatibility with most biological buffer solutions. Most importantly, its target functional groups (N-termini of peptides and the ε-amine of lysine) are ubiquitous among tryptic peptides.

The reporter ion fragmentation pathway of CIT is proposed to occur as depicted in **Scheme 4.3**. In multiply protonated CIT-labeled peptides, the tertiary amine in the CIT reagent would be protonated due to its higher proton affinity than most backbone amides and amino acid side-chains. A nucleophilic attack of N3 of the 1,2,3-triazole ring to the C_α of the protonated *N*,*N*-alkylated alanine residue in the CIT reagent releases a stable quaternary ammonium reporter ion, forming a six-membered ring.

The energetics of reporter ion formation is investigated by density functional theory calculations (**Figure 4.3**). If this process is significantly favored compared to backbone fragmentation, less sequence information would be acquired by having fewer and weaker intensity b- and y-type ions in the MS/MS spectrum. It is desirable that activation parameters

Figure 4.3 Energetics of reporter ion formation



The model system, *N,N*-dimethyl-4-(1-methyl-1*H*-1,2,3-triazol-4-yl)butan-1-amine was chosen for calculation. Geometry optimization, thermochemical calculation were performed using B3LYP/6-311++G(d,p) level of theory. The shaded area indicates the range of enthalpies of activation for amide cleavage to form b- and y-type ions via collisional activation.

associated with reporter ion formation are balanced with those of backbone fragmentation. This ensures that accurate protein quantification is achieved while not reducing sequencing efficacy using MS/MS. In our calculation model, the formation of the reporter ion is simulated by the *N*-protonated *N,N*-dimethyl-4-(1-methyl-1*H*-1,2,3-triazol-4-yl)butan-1-amine. At the B3LYP/6-311++G(d,p) level of theory, the reaction barrier and enthalpy at 1 atm and 298.15 K are determined as 33.7 and 13.3 kcal/mol, respectively (**Figure 4.3**). The usual reaction barrier for the amide bond cleavage is ranged from 25 to 40 kcal/mol.[161] Therefore, it is expected that most of the backbone cleavages occur as efficiently as the formation of the reporter ion.

4.3.2. MS/MS of CIT-Labeled Peptides

The light or heavy duplex CIT reagents were labeled to the model tryptic peptide, HSDAVFTDNYTR. The masses of CIT-labeled peptides are 279 Da larger than the original peptides as expected (**Figure 4.4**). The labeling yields of light and heavy CIT reagents are both ~99% estimated by the peak height comparison between unmodified and CIT-labeled peaks in the MALDI TOF MS spectra (**Figure 4.4**). The exact masses of light- and heavy-labeled peptides are identical, appearing as one peak in all mass spectrometric analyses.

Two collisional activation methods, beam-type CID by qTOF and pulsed-Q dissociation (PQD) by a linear ion trap in the LTQ-Orbitrap mass spectrometers, were utilized to fragment CIT-labeled peptides. Beam-type CID of the 2:3 mixture of the light and heavy CIT-labeled peptides in qTOF generates abundant reporter ions at *m/z* 164.1 and 169.1 as well as sequence ions, confirming N-terminal labeling (**Figure 4.5**). CID of the triply protonated precursor ion yields abundant backbone fragment ions along with reporter ions, but only few backbone fragments are observed for the doubly protonated precursor ion. These results are presumably caused by sequestering of mobile protons at the CIT and arginine residues, increasing the reaction barrier

for backbone cleavage. The heavy to light ratio (H/L) of reporter ions is determined to be 1.50 for the triply protonated peptides and 1.26 for the doubly protonated peptides (**Figure 4.5**).

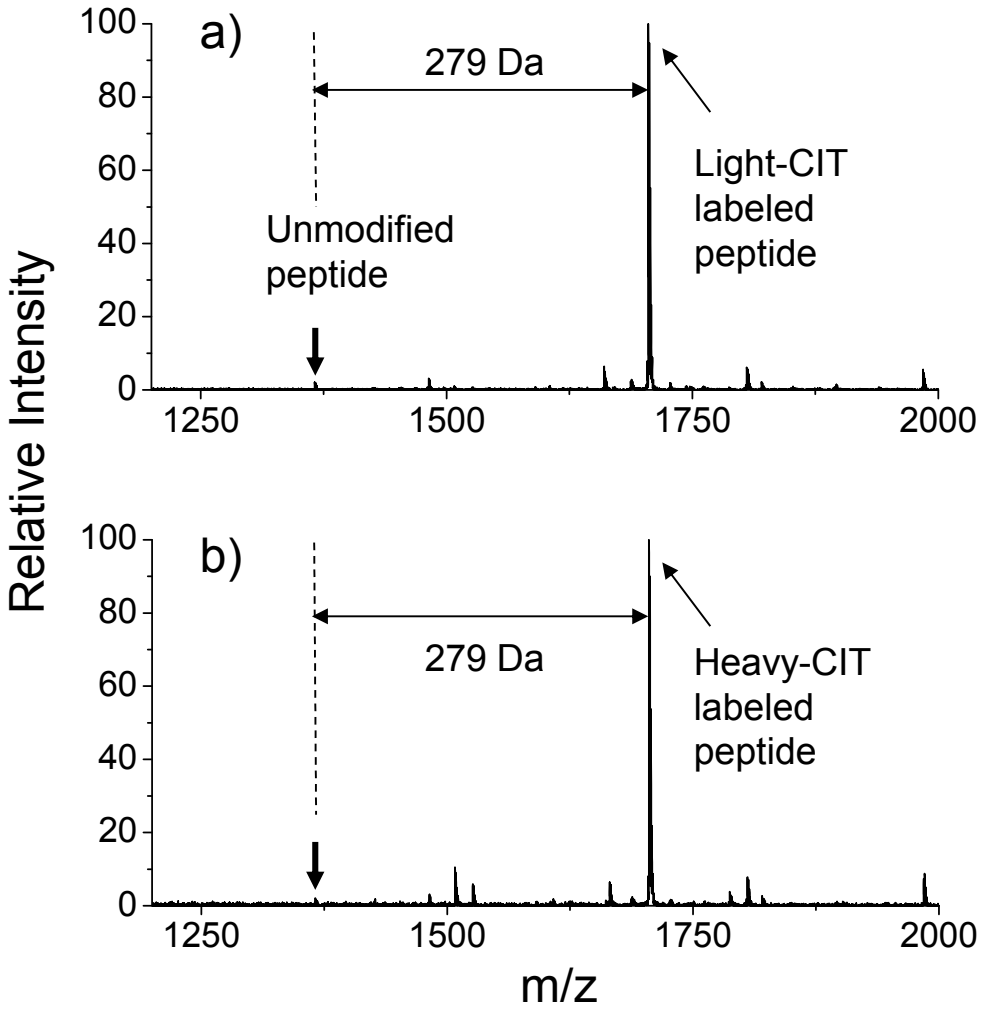
PQD of the 1:1 mixture of light and heavy CIT-labeled peptides in LTQ-Orbitrap produces abundant reporter ions and sequence ions (**Figures 4.6a-b**). More backbone cleavages are observed in PQD of the doubly protonated precursor ions than in beam-type CID. This feature of PQD is attractive for the bottom-up MS-based proteomics, considering dominant doubly protonated peptide ions generated by electrospray ionization (ESI). The H/L ratio of the CIT reporter ion is determined to be ~0.8 in both +2 and +3 charge states (**Figures 4.6a-b**), which may include the initial experimental mixing error.[221]

For the comparison of the energetics for the formation of the reporter ions, the iTRAQ-113 reagent is labeled to the same model peptide and the resulting peptides are subject to PQD (**Figures 4.6c-d**). The overall sequence coverage of the iTRAQ-labeled peptide by PQD is very similar to that of the CIT-labeled peptides. Yet, the relative reporter ion intensities in the PQD spectra of the doubly and triply protonated iTRAQ-labeled peptide ions are lower compared to those of the CIT-labeled peptides (**Figures 4.6c-d**). This result indicates that the process for reporter ion formation of CIT is slight favored over that of iTRAQ, enabling more reliable quantification using abundant CIT reporter ions.

To test the dynamic range of CIT, we monitored reporter ion formation upon mixing various ratios of light- and heavy-labeled peptides using PQD in an LTQ-Orbitrap mass spectrometer. The PQD spectra of doubly and triply charged CIT-labeled peptides were recorded in profile and centroid modes, and the intensities or areas of the reporter ions are used to plot the linear dependency on the initial mixing ratio (**Figure 4.7**, log2-log2 plot). All of the methods for data processing show a good correlation ($R^2 \sim 0.99$) between the initial mixing ratio of light and heavy CIT-labeled peptides. The overall linearity (slopes = ~1.0) and quality of fitting (R^2 = ~0.99)

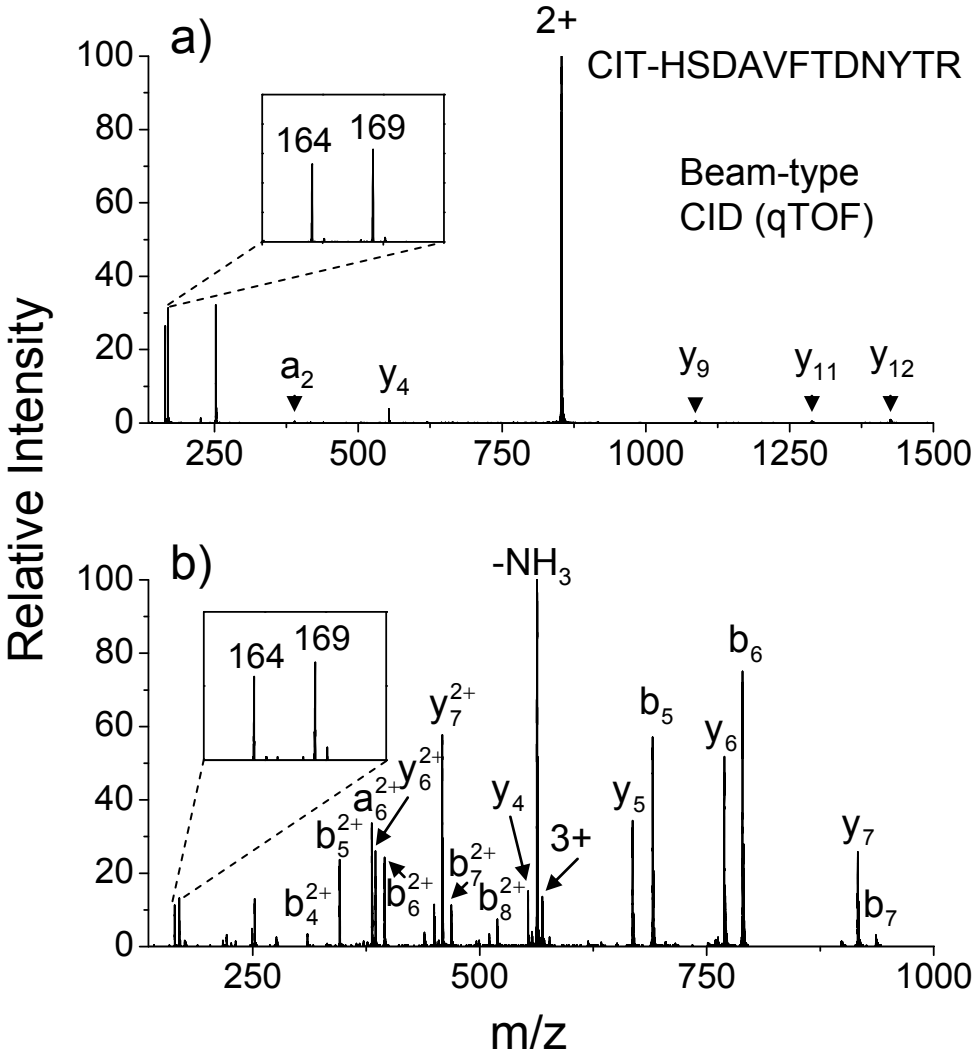
indicate a very good linear response of the CIT reagents from 1/9 to 9 H/L mixing ratios. This demonstrates roughly a two-orders-of-magnitude dynamic range performance of the CIT reagents in the relative quantification.

Figure 4.4 MALDI TOF MS spectra of CIT-labeled peptides



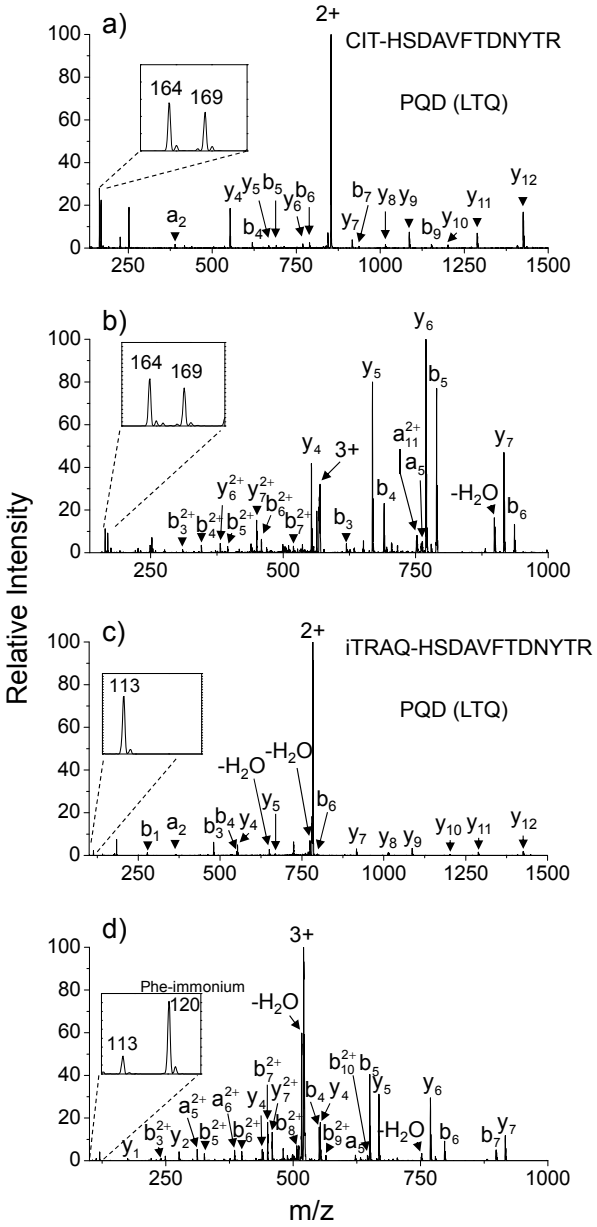
MALDI TOF MS spectra of a) Light and b) heavy CIT reagent labeling of the model peptide, VIP (1-12), HSDAVFTDNYTR. The labeling reaction was performed for 2 hr and quenched by 0.1 M hydroxylamine. The conversion yield is approximately ~99%. Quenching by 0.1 M hydroxylamine reverses unwanted byproducts which contain CIT reagent conjugation on tyrosine residues. Some of the impurities are observed but their contributions are appropriately considered for the calculation of H/L ratios.

Figure 4.5 Beam-type CID of CIT-labeled peptides



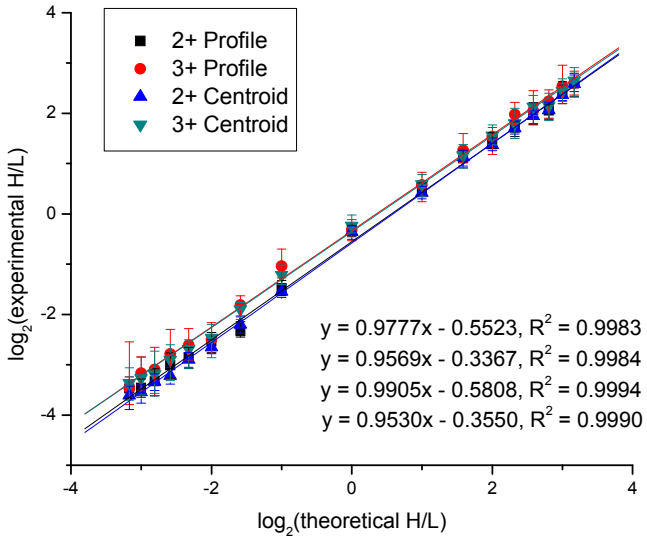
Beam-type CID of a) doubly and b) triply charged CIT-labeled model peptide ions, HSDAVFTDNYTR with 2:3 = light:heavy ratio. Both a) and b) show abundant reporter ions at *m/z* 164 and 169, respectively. Due to relatively short fragmentation time in qTOF, CID of the doubly charged peptide ions in a) shows poor sequence ion formation. CID of the triply charged peptide ions in b) contains rich sequence ion information. Experimentally observed reporter ion ratios (H/L) were 1.26 for 2+ and 1.499 for 3+ (theoretical: 1.5).

Figure 4.6 PQD of CIT-labeled peptides



PQD of doubly and triply protonated CIT- and iTRAQ-113-labeled model peptide ions, HSDAVFTDNYTR with 1:1 = light:heavy ratio. Both a) and b) show abundant reporter ions at *m/z* 164 and 169, respectively, whereas the intensities of the reporter ions at *m/z* 113 in c) and d) are relatively small, indicating the favored energetics of the fragmentation pathway used for the CIT reporter ions. Due to relatively longer fragmentation time in an ion trap compared to qTOF, PQD of the doubly peptide ions in a) and c) generates more sequence ions than that of beam type CID in **Figure 4.5a**. PQD of the triply protonated peptide ions in b) and d) also yields many sequence ions. Experimentally observed reporter ion ratios (H/L) were 0.80 for 2+ and 0.81 for 3+ (theoretical: 1.0). These deviations may result from initial experimental mixing errors.

Figure 4.7 Linearity test of CIT reporter ions



The linear fitting trend lines obtained by calculating the log2 of summations of a) the integration of all areas of isotopes in each reporter ions (*m/z* 164, 165, and 166 for the light tag; *m/z* 168, 169, 170, and 171 for the heavy tag), b) the integration of only *m/z* 164 and 169 peak areas, c) peak heights at *m/z* 164, 165, and 166 for the light tag, and at *m/z* 168, 169, 170, and 171 for the heavy tag, and d) peak heights only at *m/z* 164 and 169 for y-axis and the log2 of intended initial mixing ratios for x-axis. Relatively large (~0.4-0.5) y-axis intercepts in all figures are originated from systematic sources such as initial experimental mixing errors (See a note in the reference). Therefore, the overall linearity (slopes = ~1.0) and quality of fitting (R^2 = ~0.99) are not affected.

4.3.3. Chromatographic Separation

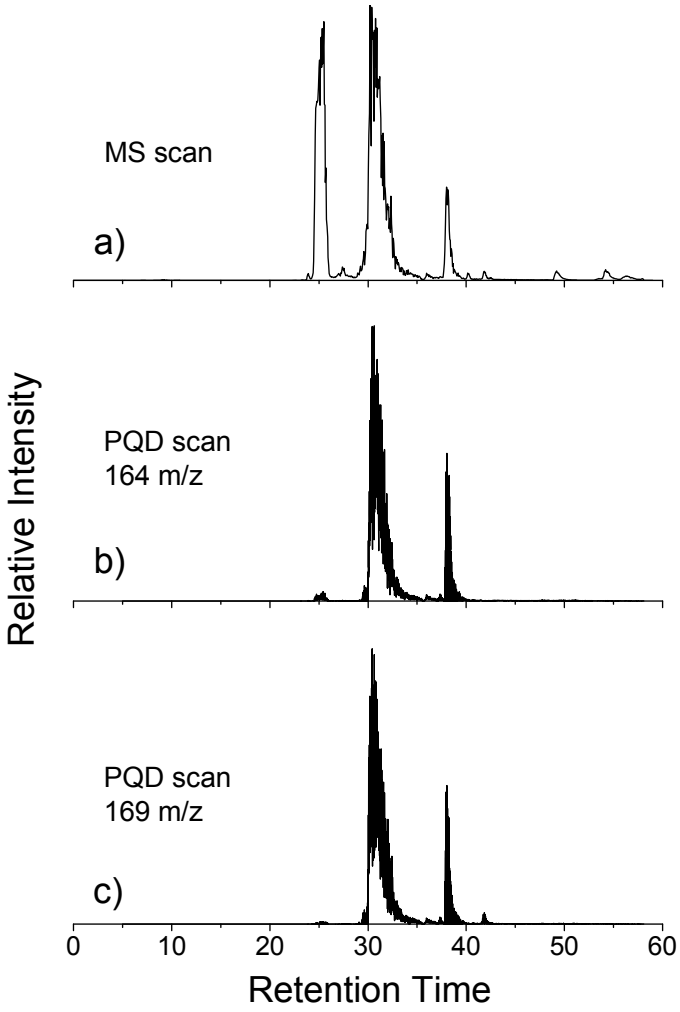
The tailing of deuterated peptides in LC elution profiles has been reported previously and may affect the accuracy of quantification.[206,222,223] To address this possibility, the retention times in nanoLC for both light and heavy peptide are measured. As seen in **Figure 4.8**, ion current diagrams for both reporter ions appear identical with no apparent tailing effect. This result indicates that both light and heavy CIT-labeled peptides have the same chromatographic properties, validating the suitability of the CIT reagents for protein quantification in the LC-MS platform.

4.3.4. Protein Labeling

The applicability of the CIT reagent is tested with model systems involving protein mixtures. Protein digests prepared from the mixtures of 1) bovine serum albumin, ovalbumin, α and β caseins and lysozyme and 2) enolase, aldolase, hemoglobin, creatine kinase and alcohol dehydrogenase were used for CIT labeling with the initial H/L mixing ratio of 1:1. The nanoLC-LTQ-Orbitrap analyses generally reproduce the initial mixing ratio, in which hundreds of peptides tagged by CIT reagents are quantified (**Table 4.1**). Note that geometric standard deviations of Mascot reported mixing ratios are relatively high. Judging from the same phenomenon observed independently in the previous report using PQD for the quantification of iTRAQ-labeled peptides,[224] we believe that the relatively high geometric standard deviation observed is, in large part, induced by poor performance of PQD.

Next, the CIT reagent is applied to quantify biologically relevant samples. Cul1 is a ubiquitin ligase that forms a large protein complex with dozens of known binding partners.[211] This protein complex was purified from the cell, and quantified using CIT after tryptic digest. To facilitate

Figure 4.8 Chromatographic identity of light and heavy CIT-labeled peptides



The nanoLC chromatograms of a) MS1 scans, b) *m/z* 164 reporter ion, and c) *m/z* 169 reporter ion observed in MS/MS scans generated by light and heavy CIT-labeled model peptides, HSDAVFTDNYTR. The base peaks in all chromatograms are related to CIT-labeled model peptides. Note that b) and c) are identical, indicating the same chromatographic property of light and heavy CIT-labeled peptides. The peak at 25 min in a) is a non-labeled model peptide. The peaks appearing around 37 min in b) and c) are from CIT-labeled peptide fragments, AVFTDNYTR.

purification of the Cul1 complex, we constructed a stable cell line that expresses tandem-tagged Cul1 upon tetracycline treatment.[212] Trypsin digests of Cul1 protein complexes affinity-purified from the HEK 293 cell line were split with the ratio of 2:1 for labeling with heavy or light CIT reagents. The labeled samples were combined, and the resulting mixture was analyzed by PQD in the LTQ-Orbitrap. **Table 4.2** lists identified proteins with the H/L ratios determined by Mascot. The calculated medians are close to 2 for all identified proteins, indicating that CIT is suitable for quantification of complex biological samples.

Lastly, the applicability of the CIT reagent for the quantification of relative protein expression levels *in vivo* is investigated. In this study, different amounts of Cul1 was expressed in stable HEK 293 cells by treating two populations of cells with either 0.5 μg/mL for 1 h or 2.0 μg/mL for 4 h. These two samples were subject to Western blot analysis. As shown in **Figure 4.9**, the level of Cul1 expression differs by ratios of 5.26 (Lane 2 / Lane1), and 5.20 (Lane 4 / Lane 3), respectively. For MS analyses, Cul1 was purified from the two differentially expressed samples and digested by Lys-C/trypsin. After CIT labeling, the resulting peptides were analyzed by HCD/CID in an LTQ-Orbitrap mass spectrometer. The median for the H/L ratio of 12 Cul1 tryptic peptides is 5.60, which agrees well with the ratio determined by Western blot experiments. These results demonstrate that the CIT-based quantification is an accurate, reliable methodology for the determination of protein abundance involving complex *in vivo* samples.

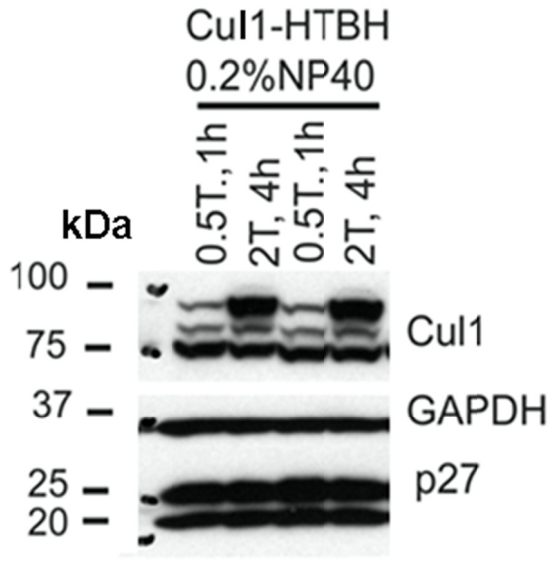
Table 4.1 Mascot Quantification Results of CIT-labeled Protein Mixtures

Mixture Set	Protein	H/L ratio	# of Peptide Hits	Geom. Std.
1	BSA	0.949	45	1.930
	Ovalbumin	0.872	44	1.791
	Lysozyme	1.049	4	1.262
	Alpha-S1-casein	0.887	27	2.086
	Beta-casein	1.063	7	2.232
2	Enolase 1, 2	0.951, 0.991	14, 10	1.651, 1.478
	Aldolase	0.929	18	1.818
	Hemoglobin alpha, beta units	0.790, 0.717	13, 8	1.757, 1.953
	Creatine Kinase	1.094	18	1.818
	Alcohol Dehydrogenase	1.131	2	2.431

Table 4.2 Mascot Quantification Results of Cul1 Complex

Protein	H/L ratio	Geom. Std.	# of Peptide Hits
CUL1	1.820	59	2.494
CAND1	1.779	35	N/A
COPS4	2.047	15	1.859
COPS5	1.635	6	1.749
COPS8	1.474	6	1.428
GPS1	1.642	6	1.123
DCUN1D1	2.834	2	1.079
COPS2	1.357	3	1.827
COPS6	2.277	2	1.156
FBXL18	1.734	3	1.031
SKP1	2.024	4	1.762
FBXO3	1.958	1	N/A
COPS7A	1.895	3	1.449
FBXO42	1.200	2	1.3344

Figure 4.9 Western blot analysis of cross-linked Cul1



Western blot analysis of differentially expressed Cul1 from HEK 293 cells. The amounts of GAPDH and p27 were analyzed as reference proteins. Differential induction was performed by adding 0.5 or 2.0 μg/mL tetracycline to the growth medium for 1 h or 4 h, respectively (Lane 1 and 2). Lane 3 and 4 are the replicates of Lane 1 and 2, respectively. The relative Cul1 expression level was 1:5.16 for Lane 1 and Lane 2, and 1:5.20 for Lane 3 and Lane 4. HCD/CID of CIT-labeled Cul1 digests yielded median H/L ratio of 5.60 using 12 peptides. The geometric standard deviation was 2.15.

4.4.Conclusion

A novel isobaric tag is developed for protein quantification, referred to as Caltech Isobaric Tag (CIT), with excellent demonstrated performance in a range of typical proteomics investigations employing model systems. The design of the CIT reagents is based on a novel gas-phase fragmentation pathway reported here for the first time. In this pathway, a nucleophilic attack of N3 of the 1,2,3-triazole ring releases a stable quaternary ammonium reporter ion with concomitant formation of a six-membered ring. The mass of the reporter ion can be easily tuned by varying azide groups in the preparation of the 1,2,3-triazole ring via Copper(I)-catalyzed azide-alkyne cycloaddition (CuAAC), better known as click reaction. The number of the possible isobaric tags is determined by the number of the isotope-tagged azide groups. These azides can be prepared from halogenated alkyl groups, which are also used for the alkylation of the linker amino acids, reducing both the cost of reagents and effort required for the synthesis of isobaric tags. This modular feature expands the possible number of combinations of CIT reagents. The properties of CIT reagents can be tuned by using larger isotope-coded halogenated alkyls that yield higher *m/z* reporter ions and these avoid the low mass–cut off problems normally associated with ion trap mass spectrometers. Mixtures of light and heavy CIT-labeled model peptides showed good linear correlations with a two-orders-of-magnitude dynamic rage. Observed ratios of the light and heavy CIT-labeled protein digests from the mixtures of 1) bovine serum albumin, ovalbumin, α and β caseins and lysozyme and 2) enolase, aldolase, hemoglobin, creatine kinase and alcohol dehydrogenase also exhibited good agreement with the initial mixing ratios. Lastly, we have demonstrated the applicability of CIT reagents in quantifying complex biological samples using affinity-purified Cul1 ubiquitin ligase complexes from HEK 293 cells.

4.5.Acknowledgement

This work was supported by the National Science Foundation through grant CHE-0416381 and the Beckman Institute at California Institute of Technology. Computational resources for density functional theory results were kindly provided by the Materials and Process Simulation Center at California Institute of Technology. C.H.S. acknowledges a fellowship from the Kwanjeong Educational Foundation. J.E.L. was supported by the Ruth L. Kirschstein NRSA fellowship from the NIH (CA138126).