

# **Implementation of Multidimensional Protein Identification Technology and its Application to the Characterization of Protein Complexes in Bakers Yeast**

Thesis by

Johannes Graumann

In Partial Fulfillment of the Requirements  
for the Degree of Doctor of Philosophy



California Institute of Technology  
Pasadena, California  
2006

(Defended May 24, 2006)



“... when one is writing a letter, he should think that the recipient will make it into a hanging scroll.”

Yamamoto Tsunetomo, “Hagakure”

# Acknowledgments

Many steps led to this work and every single one was helped along by people I would like to thank.

I thank my parents for shaping my curiosity and supporting my academic endeavors. Special credits for directing my inquisitiveness towards scientific questions goes to my grandfather Rudolf Budenz. Only the readiness of my wife Birgit to join me in California created the environment in which this work could eventually come about. I cannot thank her enough for her patience in dealing with a biochemists sense of time ( $t_{\text{biochem}} \approx 2t_{\text{standard}}$ ) and managing our life and son Leo—especially in the last months.

I also hold deep gratitude for the long line of teachers who laid the foundation to this thesis. Klaus Müller and Horst Christmann (both Gymnasium am Krebsberg, Neunkirchen/Saar, Germany) inspired me to take up biology in the first place, Rolf Knippers, Ulrich Krawinkel, and Peter Kroneck at the University of Konstanz (Germany) made me stick with it and Jon Beckwith (Harvard Medical School, Boston, MA), Jim Bardwell, and Ursula Jakob (both University of Michigan, Ann Arbor, MI) showed me that a future in academic research might be feasible. Given the technical nature of the work described here, I would like to thank especially Peter Kroneck (University of Konstanz, Germany) for acquainting me with instrumentation-driven biochemistry.

I thankfully acknowledge the contribution of the entire Division of Biology at the California Institute of Technology to the incredible scientific environment that formed the backdrop to much of this work. The members of my thesis committee—Pamela Björkman, Bill Dunphy, Bruce Hay, and Mel Simon—contributed to this

thesis their advice, criticism, and time. Barbara Wold provided invaluable insight into the politics of scientific interaction in general and adviser/advisee quarrels specifically. Many friends also contributed to the Caltech experience—too many to list them all, but a few shall be singled out. Dan Gold’s restaurant scouting was instrumental to stylish acquisition of graduate school weight and his critical reading of manuscripts was indispensable. Erik Griffin contributed discussion and inspiration, as did David Walker. Chris Hart, Diane Trout and Dave Mathog were essential to the fixing of the occasional going awry of the author’s computational autodidacticism. Dirk Walther, Kerstin Preuschoff, Daniela Dieterich and Marco Landwehr provided a little bit of home away from home.

Without John Yates and his laboratory at The Scripps Research Institute (La Jolla, CA) the work described here would simply not have been possible. I would like to thank John for generously hosting me for nine months and sharing the know-how of his lab. Special thanks go to the former Yates’ laboratory members Eberhard Dürr (Merck, Philadelphia, PA), Hayes McDonald (ORNL, Oak Ridge, TN), Mike MacCoss (University of Washington, Seattle, WA) and Dave Tabb (Vanderbilt University, Nashville, TN). Their mentorship regarding mass spectrometry and data analysis yielded the foundation to this thesis.

I thank all former and present members of the Deshaies’ laboratory at Caltech for advice, criticism and composure with which they have escorted this thesis and its author. I would like to especially thank Rusty Lipford (Amgen, Thousand Oaks, CA) for many an insight into the English language, Rati Verma for her bottomless treasure trove of biochemical advice, Geoff Smith for his unwavering reliability, Leslie Dunipace for countless packed capillary columns and Gabi Alexandru and Thibault Mayor, my fellow Euro-room constituents, for being there in times good

and not so good, scientifically and otherwise.

Last, but not least, I would like to express my gratitude toward my adviser and mentor Ray Deshaies. I am grateful for the challenge he presented me with in the shape of the project described here and the confidence he put in me while letting me go off into scientific territory entirely new to him as well. I would like to thank him for his patience while putting me back onto the right path numerous times and the constant challenge of his scientific wit. I highly value the countless political lessons taught by him—consciously or not—and the example he sets by his ambition. Thank you for being available when needed and thank you also for many insightful discussions on science and academic life in general.

# Abstract

The analysis of complex polypeptide mixtures poses a central and ubiquitous problem to biochemistry, molecular and cellular biology. Historically the problem has been approached by means of gel electrophoretic separation, coupled to immunology or Edman degradation (Edman 1949) based identification of separated components. These approaches as well as those based on liquid chromatography are hampered by a central issue: the wide spectrum of polypeptide characteristics that renders their separation difficult. A recent strategy termed *multidimensional protein identification technology* (MudPIT) tackles this problem by capillary chromatographic separation of not the complete polypeptides, but rather peptides yielded by them through proteolytic digest and analyzing them in-line using ion trap mass spectrometry (Link et al. 1999; Washburn et al. 2001; and Wolters et al. 2001).

This work describes the implementation of MudPIT outside of the analytical chemistry environment of its inception. Robustness and generalizability of the technique are tested by analysis of polypeptide complexes copurified with 25 selected gene products from *Saccharomyces cerevisiae* (Graumann et al. 2004). The pilot study reveals MudPIT to be mature enough for use outside of specialized environments and, by yielding with Rtt102p a novel component of the Swi/Snf and RSC chromatin remodelling complexes, to have potential for delivering new insights even into extensively studied systems.

Subsequent application of MudPIT to the characterization of components of the ubiquitin-proteasome system (Verma et al. 2004; and Mayor et al. 2005) and mitochondrial fission (Griffin et al. 2005) in *S. cerevisiae* further emphasize its potential to contribute to biochemical research.

# Table of Contents

Acknowledgments	iv
Abstract	vii
Table of Contents	viii
Figures	xii
1 Introduction	1
1.1 The Problem of a Complex Protein Mixture	1
1.2 Multidimensional Protein Identification Technology	7
1.3 Mass Spectrometric Quantification of Polypeptides	11
1.4 MudPIT at Caltech	14
1.5 References	19
2 Applicability of TAP–MudPIT to Pathway Proteomics in Yeast	28
2.1 Summary	28
2.2 Introduction	29
2.3 Experimental Procedures	32
2.4 Results	39
2.5 Discussion	53
2.6 Acknowledgements	59
2.7 Note Added in Proof	59
2.8 References	59

Part Appendices	66
A Multiubiquitin Chain Receptors Define a Layer of Substrate Selectivity in the Ubiquitin–Proteasome System	67
A.1 Abstract	67
A.2 Introduction	68
A.3 Results and Discussion	72
A.4 Experimental Procedures	94
A.5 References	95
B MS1, MS2, and SQT—Three Unified, Compact, and Easily Parsed File Formats for the Storage of Shotgun Proteomic Spectra and Identifications	103
B.1 Abstract	103
B.2 Introduction	104
B.3 Format Descriptions	106
B.4 Results and Discussion	114
B.5 References	118
C Analysis of Polyubiquitin Conjugates Reveals that the Rpn10 Substrate Receptor Contributes to the Turnover of Multiple Proteasome Targets	121
C.1 Abstract	121
C.2 Introduction	122
C.3 Experimental Procedures	125
C.4 Results	129

C.5	Discussion	140
C.6	Acknowledgments	146
C.7	References	146
D	The WD40 Protein Caf4p is a Component of the Mitochondrial Fission Machinery and Recruits Dnm1p to Mitochondria	153
D.1	Abstract	153
D.2	Introduction	154
D.3	Results	156
D.4	Discussion	175
D.5	Materials and Methods	178
D.6	Acknowledgments	187
D.7	References	187
	About this Document	192
	References	193

# Figures

2.1	Schematic Representation of the HPM Tag.	40
2.2	SDS–Polyacrylamide Gel Analysis of Glc7p–HPM, Mcd1p–HPM, Pds1p–HPM and Gcn5p–HPM Affinity Purifications.	41
2.3	Reproducibility of Results Between Independent Gcn5p–HPM TAP–MudPIT Experiments.	42
2.4	Comparison with Ho et al. (2002).	51
2.5	An <i>rtt102Δ</i> Strain Partially Recapitulates the Phenotype of Mutants Lacking Swi/Snf Complex.	53
A.1	Structural and Functional Characterization of 26S Proteasomes Isolated from <i>rpn10Δ</i> and <i>rad23Δ</i> Mutants by Affinity Chromatography.	73
A.2	The Degradation and Deubiquitination Defects of <i>rpn10Δ</i> and <i>rad23Δ</i> 26S Proteasomes Can be Rescued by Recombinant Proteins.	76
A.3	Complementation of <i>rad23Δ</i> Proteasomes Requires Both the Ub Binding UBA Domains and the Proteasome Binding UbL Domain of Rad23.	81
A.4	26S Proteasomes from <i>rpn10Δ</i> Are Defective in Binding UbMbpSic1.	82
A.5	Rpn10 UIM Domain and Rad23 Serve Redundant Roles in Sic1 Turnover <i>In Vivo</i> .	85
A.6	UPS Substrates Have Differential Requirements for Multiubiquitin Chain Receptors <i>In Vivo</i> .	87
A.7	Hypothetical Model for Physiological Targeting Pathways that Deliver Ubiquitinated Substrates to the 26S Proteasome.	91

B.1	MS1 File Format Description.	108
B.2	MS2 File Format Description.	111
B.3	SQT File Format Description.	112
C.1	Flow Diagram for the Two-step Purification of Polyubiquitin Conjugates.	131
C.2	Two-step Affinity Purification Specifically Enriches for Polyubiquitylated Proteins.	132
C.3	Proteins Identified by LC/LC-MS/MS After Two-step Purification of Ubiquitin Conjugates.	134
C.4	Protein Representation and Reproducibility Overview.	136
C.5	Putative Ubiquitylated Proteins Identified in <i>rpn10Δ</i> But Not Wild-type Cells.	138
C.6	Analysis of Candidate Substrates of the Rpn10-dependent Targeting Pathway.	139
D.1	Construction of M <sub>9</sub> TH- <i>FIS1</i> and Caf4p/Mdv1p Alignment.	157
D.2	Caf4p and Mdv1p Coimmunoprecipitation Experiments.	160
D.3	<i>CAF4</i> Regulates Mitochondrial Morphology.	163
D.4	<i>CAF4</i> Mediates Residual Fission in <i>mdv1Δ</i> Cells.	166
D.5	Caf4p Overexpression Blocks Mitochondrial Fission.	168
D.6	Suppression of the Glycerol Growth Defect of <i>fzo1Δ</i> Cells.	170
D.7	Mitochondrial Localization of Caf4p and Mdv1p Requires Fis1p.	172
D.8	Fis1p Mediates Dnm1p-GFP Localization Through Either Mdv1p or Caf4p.	174

# 1 Introduction

The work presented in this thesis describes the implementation of a set of techniques termed “Multidimensional Protein Identification Technology” or MudPIT (Link et al. 1999; Washburn et al. 2001; and Wolters et al. 2001), that enables the analysis of complex protein mixtures. This chapter provides an introduction to the significance of the analysis of complex protein mixtures in molecular biology and biochemistry, as well as describing MudPIT in detail.

## 1.1 The Problem of a Complex Protein Mixture

Polypeptides dominate the spectrum of biological functions as both mediators and catalysts. Although knowledge of biological processes mediated by nucleic acids has expanded dramatically as a result of whole genome sequencing projects (Storz 2002), polypeptides provide the greater variety of building blocks—20 amino acids vs. four nucleotides—and as a result the larger spectrum of possible conformations and chemistries. The array of possible posttranscriptional modifications of nucleic acids (Gott and Emeson 2000) is met by an equally extensive variety of posttranslational modifications in polypeptides (Creighton 1993) and does not shift the balance.

A significant part, if not the majority, of protein-mediated biological reactions is dependent not on a single functional polypeptide, but rather a group of polypeptides working together in a concerted manner, often forming subunits of one protein complex, one “molecular machine” (Alberts 1998). Gavin et al. (2006), for example, estimate that *S. cerevisiae* contains 800 core “protein complexes”—condition-

independent protein complexes, whose composition is modulated in a condition-dependent manner by “attachment proteins.” Conservatively assuming two polypeptide chains per “core complex” and disregarding all transient interactions with “attachment proteins,” this amounts to 27 % of all systematically named open reading frames in yeast being assembled into complexes. This number rises to 41 % for an average of three polypeptide chains per complex.<sup>1</sup>

These two points—the domination of biological processes by proteinaceous agents and the prevalence of these polypeptides in heterogeneous complexes—present a challenge: separation of a complex protein mixture and the identification of its components, even if one is interested in a single biological process rather than questions of global changes in a cellular or organellar protein complement.

### 1.1.1 Separation of complex protein mixtures

Separation of protein mixtures is commonly handled by one of two technically divergent approaches: gel electrophoresis or liquid chromatography. Gel electrophoresis separates proteins by a combination of their electrostatic and size properties, whether native or conferred by agents such as sodium dodecyl sulfate as introduced by Laemmli (1970). While immensely popular, the technique in both its one- as well as two-dimensional form (see, e. g., Klose 1975; and O’Farrell 1975), has inherent disadvantages:

---

<sup>1</sup> Based on 5872 nondubious and nonpseudogene open reading frames present in the *Saccharomyces* Genome Database (Cherry et al. 1998) as of 03/31/2006.

1. Both one- and two-dimensional gel electrophoresis have severely reduced resolving power for polypeptides of extremely small or large sizes.
2. Similarly, the isoelectric-focusing-based first dimension of conventional two-dimensional gel electrophoresis biases against polypeptides with very high or low isoelectric points.
3. Gel electrophoresis is notoriously unsuited for the separation of polypeptides with extreme hydrophobicity, such as membrane proteins.
4. Although more mechanized approaches have been made (see, e. g., Gavin et al. 2002), selection of separated polypeptide chains is commonly done visually, opening the technique to bias against weakly staining or diffusely migrating polypeptides.
5. Gel electrophoresis delivers the separated polypeptide chains embedded in a gel matrix, which implies the potential for low extraction efficiency.
6. The conventional workflow (see, e. g., Shevchenko et al. 2002) of gel electrophoretic mixture separation, gel block excision and in-gel digest results in the case of complex mixtures in massive sample parallelization, requiring a significant degree of automation. This problem is partially remedied by slicing groups of polypeptide bands rather than individual bands and subsequent chromatographic separation of the electrophoretically prefractionated mixture (Gavin et al. 2002).

Traditional protein mixture separation by chromatographic methods—the mixture is carried through a column of chromatography matrix by a liquid phase and separated by differential interaction with the matrix—implies a similar set of problems, for example:

1. Depending on the polypeptide property by which the matrix separates in a given

experiment (e. g., hydrophobicity, hydrophilicity, charge, size), there are gel electrophoresis analogous problems in separating polypeptides in the extremes of the property spectrum. Extremely hydrophobic proteins are, for example, difficult to separate by reverse phase liquid chromatography, as are extremely hydrophilic ones.

2. Conventional chromatography workflows involving fraction collection potentially yield, just as in the case of gel electrophoretic separation, massive sample parallelization and the necessity of automation.
3. Varying with the liquid phase throughput through the column, chromatographic methods have the inherent problem of volume expansion, necessitating additional procedures as precipitation or lyophilization, implying the risk of sample loss by, e. g., surface coating.

The preceding lists concentrate on the drawbacks of the two most popular means to separate complex polypeptide mixtures. Evidently, the two approaches also have distinct advantages. Gel electrophoresis for example is uniquely suited for separation of posttranslationally modified polypeptide forms (for an example see Larsen et al. 2001), while liquid chromatographic methods are very well suited for subsequent biochemical manipulations as functional assays or crystallographic analysis of the separated polypeptides. Approach-specific problems aside, the methodologies essentially struggle with the same issue: the wide spectrum of biochemical/biophysical characteristics associated with polypeptides in the complex mixtures to be separated. Section 1.2 describes in detail how MudPIT tries to remedy this problem.

### 1.1.2 Identification of the components of a complex protein mixture

The second step in characterization of a complex polypeptide mixture is the identification of separated components. The methodologies to achieve this fall into two categories: before and after the application of mass spectrometry to the problem. Pre-mass-spectrometry methods for the analysis of an unknown proteinaceous agent include Edman degradation (Edman 1949) as well as raising antibodies against a purified polypeptide, which is then identified by screening through of a phage expression library and sequencing. The first mass spectrometric approach to join the canon of techniques applied to polypeptide analysis was peptide mass fingerprinting (Henzel et al. 1993; James et al. 1993; Mann et al. 1993; Pappin et al. 1993; and Yates et al. 1993). This technique is based on the proteolytic digest of the polypeptide to be analyzed with a site-specific protease and the subsequent mass spectrometric analysis of the resulting peptide mixture. The measured peptide masses are matched with the *in silico* digest of a protein database, yielding the protein with the closest hypothetical spectrum as the identification candidate. Peptide mass fingerprinting shares a major drawback with pre-mass-spectrometric methods: they require polypeptide mixture components to be highly purified, which poses a significant challenge when dealing with highly complex mixtures.

While peptide mass fingerprinting already took advantage of some of the following innovations (e. g., electrospray ionization), the application of the complete set was necessary in order for mass spectrometry to emerge as the dominating technique with respect to polypeptide mixture analysis:

1. The development of postsource decay (PSD, Spengler et al. 1992) and collision induced dissociation (CID, Hunt et al. 1986) changed the field dramatically: the

techniques allow the direct sequencing of the amino acid composition of peptides, which are not necessarily present in highly purified form but can be isolated from an injected peptide mixture by mass filtration in the mass spectrometer.

2. “Soft” ionization techniques such as matrix assisted laser desorption ionization (MALDI, Karas and Hillenkamp 1988) and electrospray ionization (ESI, Fenn et al. 1989) enable the analysis of chemically fragile biomolecules such as polypeptides without significant decomposition.
3. The introduction of a new class of mass analyzers to the characterization of biological samples proved to be crucial to the success of mass spectrometry: quadrupole–ion trap mass spectrometers (ITMS, Jonscher and Yates 1997) not only enable rapid rounds of selection of a single ion from an injected mixture of peptides, but also multiple stages of collision induced dissociation—and therefore sequencing of multiple fragmentation ions. Constant improvements of ITMS systems focus mainly on scan speed—crucial for example to the sampling depth in a chromatographic sample eluted via ESI directly into the mass spectrometer—and better ion statistics (Blackler et al. 2006) and mass accuracy, resulting in higher sequence confidence (Olsen et al. 2005). A small but important feature of these instruments designed for high–throughput analyses is the so called “dynamic exclusion,” a mechanism preventing the refragmentation/sequencing of ions in an injected mixture by imposing a temporary exclusion of mass over charge values already attended to.
4. The last innovation to pave the way for mass spectrometry in the analysis of polypeptide mixtures was the creation of software which automatically matches experimental peptide fragmentation spectra to hypothetical spectra derived from organism specific protein sequence databases. Eng et al. (1994) pioneered this

approach with their program **Sequest**, but a number of competing programs as, e. g., **Mascot** (Perkins et al. 1999) and **X! Tandem** (Craig and Beavis 2004) have followed suit.

MudPIT incorporates a number of these innovations to tackle the problem of the analysis of a complex polypeptide mixture, which is described in detail in section 1.2.

## 1.2 Multidimensional Protein Identification Technology

Multidimensional protein identification technology (MudPIT) was introduced by Link et al. (1999) as “Direct Analysis of Large Protein Complexes” (DALPC). Generalization of the concept lead to the coining of the term MudPIT (Washburn et al. 2001; and Wolters et al. 2001). The workflow established by the authors combines multidimensional capillary chromatography of complex polypeptide mixtures digested in solution with in-line electrospray-ionization ion-trap tandem mass-spectrometry and automated matching of the acquired fragmentation spectra to translated genomic sequence via **Sequest** (Eng et al. 1994). The strategy addresses many of the challenges to the analysis of complex polypeptide mixtures laid out in section 1.1.

MudPIT strives to separate proteins that have been digested into peptides rather than the intact polypeptides. This approach—also termed *bottom up* (Wysocki et al. 2005) or *shotgun proteomics* (Wolters et al. 2001)—levels the biochemical/biophysical properties and therefore reduces the problems polypeptides pose to separation techniques with their wide property spectrums.

Reliance on capillary chromatography with low liquid phase flow rates<sup>2</sup> remedies the issue of volume expansion connected to conventional high pressure liquid chromatography (HPLC), while directly interfacing the separation setup to the mass spectrometer via electrospray ionization. The latter prevents the need for sample collection, thereby rendering further automation unnecessary, reducing manual intervention and preempting sample loss by surface coating. The use of a two-dimensional chromatography column significantly improves the resolution of the setup by utilizing two independent biophysical characteristics of the peptides to be separated: charge by the strong cation exchanger (SCX) phase and hydrophobicity by the reverse phase. It extends prior work (e.g., Lundell and Markides 1992; and Takahashi et al. 1985) and transfers the principles long utilized in twodimensional gel electrophoresis (O’Farrell 1975) to liquid chromatography. McDonald et al. (2002) further enhanced the approach by adding a third phase—a second reverse phase chromatography matrix—to the capillary column, allowing sample desalting in-line to the mass spectrometer, further reducing handling requirements and capturing a class of hydrophilic peptides missed when using the twophasic column layout.

The utilization of iontrap mass spectrometers—for reasons of patent protection of key scan features predominantly ThermoElectron’s line of Deca, DecaXP, and LTQ mass spectrometers (historically successive in this order)—enables the analysis of ions eluting into the mass spectrometer with increasing speed and sensitivity (for the LTQ see Blackler et al. 2006). Together with the mechanism of dynamic exclusion discussed above, this renders possible increasingly comprehensive analysis

---

<sup>2</sup> 50  $\mu$ l/min in the original publication, in further works reduced to the order of 100 nl/min.

of peptides of ever lower abundance eluting into the spectrometer.

All the advantages of MudPIT aside, the technique also meets with valid criticisms. Through the projects described in this work, the hand-crafted, single-use capillary chromatography columns used emerged incontestably as the weakest link in the chain of procedures constituting MudPIT. Packing the 100  $\mu\text{m}$  inner diameter columns on customized pressure vessels is tedious work, often requiring multiple attempts. After successful packing, some columns clog during sample loading<sup>3</sup> or produce suboptimal electrospray due to inadequate tip shape. Custom capillary columns are commercially available (e.g., New Objective, Woburn, MA), but the high price together with the triphasic nature of the columns, which interferes with effective cleaning and leave the column a single-use item, did prohibit their use for this work. Although there are promising microfluidic approaches emerging (e.g., Xie et al. 2005), mass production of multiphase capillary columns seems far in the future. The single-use characteristic of the capillary chromatography columns implies (together with the stochastic nature of peak sampling by the ion trap mass spectrometer) relatively low reproducibility when analyzing the same sample on different columns. Multidimensional chromatography is also possible with traditional HPLC columns, but their reliability, reproducibility, reusability and ready commercial availability comes with a significant hit to analysis sensitivity<sup>4</sup>—a fact very much undesirable when analyzing highly complex mixtures with low polypeptide

---

<sup>3</sup> Promoted by the high urea concentrations regularly present.

<sup>4</sup> According to Abian et al. (1999), the maximum peak concentration of the sample eluate  $C_{\text{max}}$  increases by a factor of 100 when reducing the column diameter from the commonly used 1 mm to 100  $\mu\text{m}$ .

abundance (as, e. g., polyubiquitin conjugates; see appendix C).

Another critical point is the use of in-line electrospray ionization itself. While reducing manual intervention and analysis time by directly linking the chromatography setup with the mass analyzer, this methodology also requires extreme spectrum acquisition speeds to be able to sample deeply into the injected ion mixtures. It also possesses the inherent drawback of producing multiply charged ions (which complicates subsequent spectrum matching) and—in conjunction with in-line chromatography—restricts the time window for analyzing a chromatographically separated peptide peak to its actual elution from the column.

The speed of spectrum acquisition in an ion trap mass spectrometer traditionally comes at the expense of mass accuracy, but this criticism is slowly disappearing due to the combination of ion traps with high mass accuracy mass analyzers as Fourier-transform mass spectrometers (e. g., ThermoElectron's LTQ-FTMS, see Olsen et al. 2004) and orbi-traps (e. g., ThermoElectron's LTQ-Orbitrap, see Olsen et al. 2005).

Shotgun shotgun proteomic data pose significant analysis challenges (Steen and Mann 2004). Improved precursor scan mass accuracy, as delivered by instruments similar to the ones described in the previous paragraph, remedies a part of that problem, but what remains—especially when dealing with higher eukaryotes—is the problem of polypeptide isoform multiplicity due to differential splicing, alternative promoter usage and other mechanisms, as well as often extensive groups of homologous polypeptides, which make pinpointing a polypeptide from a collection of sequenced peptides very difficult (Nesvizhskii and Aebersold 2005; and Godovac-Zimmermann et al. 2005). While this problem is triggering the field to revisit *top down* proteomics with its significant separation challenges (see above), the work pre-

sented here concentrates on the model organism *Saccharomyces cerevisiae*, which carries introns in only  $\approx 4\%$  of its open reading frames (Spingola et al. 1999) leaving a *bottom up* approach as MudPIT in a favorable light.

After evaluation of these arguments, MudPIT emerges as a viable candidate for complex polypeptide analysis—especially for *Saccharomyces cerevisiae* as a model organism—for the foreseeable future.

A more trivial data analysis problem also arises in conjunction with *shotgun* approaches: the sheer scale of spectra to be searched and their computational handling. The original **Sequest** Eng et al. (1994) read in input files containing information for a single spectrum and produced an output file for every single one of those. Given the tens of thousands of spectra a single MudPIT experiment produces, this strategy taxed even industry grade UNIX file systems to their limits. For **Sequest** the problem was fixed with unified input and output formats (Eng et al. 1994; Sadygov et al. 2002; and McDonald et al. 2004), providing all spectral information from one MudPIT step in one single file and the results inferred from it in another. J. G. was involved in the setup of this infrastructure, which is covered in detail in appendix B (p. 103). All spectrum matching programs in use today apply similar approaches.

### 1.3 Mass Spectrometric Quantification of Polypeptides

Knowing the constituents of a complex mixture of polypeptides represents valuable information in itself, but a large group of biological problems require the identification of differences in polypeptide representation between different biological states, such as wildtype versus mutant or untreated versus pharmacologically manipulated. MudPIT alone delivers excellent data on the composition of a polypeptide mixture,

but the issues of reproducibility discussed above render comparison of independent MudPIT analyses for different biological states difficult (see appendix C), which leaves MudPIT largely incompatible with so-called label-free approaches (Old et al. 2005) to the problem of polypeptide quantification that rely either on measurements and comparison of ion intensities (Bondarenko et al. 2002; Chelius and Bondarenko 2002; and Wang et al. 2003) or spectral counting (Liu et al. 2004).

The major class of solutions to the quantification problem that remains available is isotope or mass tag labeling (Old et al. 2005). The different approaches that can be combined in this category follow one theme: isotopically marking the states to be compared differentially and comparing the abundance of different forms of the same peptide mass spectrometrically—in the same analysis, using the same column, which implies compatibility with MudPIT despite its low reproducibility.

The first subclass in this collection includes the approaches termed ICAT (isotope coded affinity tags; Gygi et al. 1999), its successor cICAT (cleavable ICAT; Hansen et al. 2003; and Yu et al. 2004) and iTRAQ (isobaric tags for relative and absolute quantification; Ross et al. 2004).

(c)ICAT works by mass-differential chemical derivatization of peptides on cysteine residues. The restriction to cysteine-containing peptides along with differential reverse-phase elution behavior of heavy and light forms (Goshe and Smith 2003; Leitner and Lindner 2004; and Wu et al. 2006) are major criticisms facing the techniques. iTRAQ overcomes the residue specificity problem by targeting amines, so that all peptide N-termini, along with lysine side chains, are potential tag receptors. It also enables the direct comparison of up to four samples in the same experiment—a feat no other technique described here accomplishes. Since the tagged peptides have the same mass independent from which of the up to four tested conditions

they arise, the peptide mixture complexity is not increased (all other isotope labeling strategies described here raise it by a factor of two), which relieves the scan burden of the mass spectrometer (Wolff et al. 2006). However, iTRAQ requires high mass resolution fragmentation spectra, since quantification is achieved from small, low  $m/z$ -difference daughter ions of the fragmented linkers the peptides are derivatized with (114, 115, 116, 117 kDa), which in turn takes a toll in the achievable sequencing speed. (c)ICAT and iTRAQ share a central disadvantage: samples to be compared using these techniques have to be prepared in parallel, independently derivatized and then mixed, which obviously opens the door to asymmetric processing errors.

This caveat is not present with the second subclass of isotope tag labeling techniques: metabolic labeling. These approaches are based on the utilization of isotopically different polypeptide precursors in one of the biological samples to be compared. Polypeptides from the tagged and untagged samples are as a result available *in vivo* in mass spectrometric distinguishable populations and analytes are prepared from mixed samples rather than in parallel.

Metabolic labeling is generally available in two flavors: SILAC-like (stable isotope labeling by amino acids in cell culture; Ong et al. 2002) approaches based on the incorporation of selected, isotopically labeled amino acids and approaches providing solely heavy nitrogen ( $^{15}\text{N}$ ) in the form of ammonium acetate to the organism under study (Oda et al. 1999; and MacCoss et al. 2003). SILAC approaches elegantly combine applicability to difficult model systems such as culture cells (see acronym) with easy predictability of sister ion mass: when using arginine and lysine as the isotopically tagged amino acids, each peptide produced by trypsin—which hydrolyzes polypeptides specifically c-terminally of those two amino acids— will

carry the additional mass conferred by its c-terminal residue. Isotopically modified amino acids are, however, very expensive and imply the problem of being rerouted by the organisms metabolism, potentially resulting in the isotopic labeling of unintended amino acids, posing problems to accurate peptide sequencing. Approaches based on minimal diets solely providing heavy nitrogen ( $^{15}\text{N}$ , mostly in the form of ammonium acetate) are in comparison significantly more affordable and, since all nitrogen atoms indiscriminately represent the heavy or light form, do not suffer from metabolic rerouting. This makes them applicable to all systems able to grow on minimal media (a minimal diet). Wu et al. (2004) even managed to raise a rat (*R. norvegicus*) on a diet including  $^{15}\text{N}$ -grown algae as the only source of nitrogen.

Given the model organism this study centers on—*S. cerevisiae*—and the considerations above, metabolic labeling by  $^{15}\text{N}$  on minimal media was implemented in the course of the work, additionally profiting from seamless integration of the appropriate quantification software **Relex** (MacCoss et al. 2003) in the data analysis infrastructure consisting from **Sequest** (Eng et al. 1994; Sadygov et al. 2002; and McDonald et al. 2004) and **DTASelect/Contrast** (Tabb et al. 2002).

## 1.4 MudPIT at Caltech

Motivated by the advantages of MudPIT laid out in the preceding sections, we set out to implement the technique in a biochemistry laboratory at Caltech. MudPIT was—and may still be—considered experimental technology and had not spread far beyond the labs of John R. Yates III at The Scripps Research Institute (TSRI) and Torrey Mesa Research Institute (TMRI) in La Jolla, California, where the Yates groups were honing the technique (Washburn et al. 2001; and Wolters et al. 2001)

after having moved there from its birthplace at the University of Washington in Seattle (Link et al. 1999). The main challenge in doing so was to implement the setup without the analytical chemistry environment that had bred it: our laboratory had, as is to be expected for the majority of biochemistry/molecular biology laboratories, no expertise in mass spectrometry and very little in information technology and HPLC separation, yet all three fields are required for running a MudPIT facility. To overcome this obstacle J. G. spent 9 months in the Yates lab at TSRI, intensely immersed in all aspects of the labs operation: sample preparation, mass spectrometric and data analysis, as well as hardware maintenance.

Back at Caltech, we proceeded to emulate the Yates lab setup in small scale. The setup initially consisted of

- A P-2000 LASER needle puller by Sutter Instrument Co. (Novato, CA). This instrument is used to outfit the fused silica capillaries from which capillary chromatography columns are constructed (inner diameter: 100  $\mu\text{m}$ ) with a  $\approx 5 \mu\text{m}$  diameter tip required for electrospray ionization and chromatography matrix retention.
- Two capillary chromatography column packing stations. These stainless steel pressure vessels, which were produced in-house according to drawings provided by the Yates lab, utilize helium pressures of up to  $7 \times 10^6 \text{ N/m}^2$  to pack chromatography matrices into tipped capillary chromatography columns and after equilibration load sample onto the finished columns.
- ThermoElectron's DecaXP<sup>+</sup> ion trap mass spectrometer for spectrometric analysis of sample peptides eluted from MudPIT columns.
- An HP-1100 HPLC pump and solvent degasser combination with four solvent

- channels by Agilent (Palo Alto, CA). The extremely low flow rates used by Mud-PIT (100 nl/min and less) in combination with the required solvent gradients between low and high organic solvents with preceding salt bumps cannot be delivered by standard HPLC pumps. The HP-1100 system is therefore used at 100  $\mu$ l/min and interfaces to the mass spectrometer via
- A custom capillary column electrospray ionization source, which splits the column flow (100 nl/min) from the pump-delivered solvent flow (100  $\mu$ l/min). The source also provides a liquid phase/voltage junction, applying the 2.4 kV electrospray ionization voltage to the waste arm of the split flow, which prevents gas bubbles resulting from electrochemistry on the electrode from entering the capillary chromatography column. The original design for this ionization source was also provided by the Yates lab and the source manufactured in-house.
  - A Linux cluster for data analysis. The cluster consists from twenty 1.8 GHz RS-1200 computation nodes, provided by Verari Systems (formerly RackSaver, San Diego, CA). Mass spectrometric data is transferred to a central file server and undergoes charge state analysis as well as data quality filtration by `2to3` (Sadygov et al. 2002). `Sequest` search jobs using unified input and output formats (Eng et al. 1994; Sadygov et al. 2002; and McDonald et al. 2004) can then be queued on the cluster, using `GridEngine` (<http://gridengine.sunsource.net/>). The queuing mechanism as well as user account dependence are significant enhancements in comparison with the original Yates lab setup, where all members logged on as one user to execute `Sequest`, verbal agreements were necessary concerning how many jobs to run at one time and manual checking of running processes with low-level system commands provided the only handle on available slots. In order to coerce the experimental `Sequest` binary provided by

the Yates lab through a collaboration (Graumann et al. 2004) into conforming to the requirements of this system, GNU `screen` (<http://www.gnu.org/software/screen/>) has to be used to mimic a terminal interactively open to the binary—on the original setup remote users had to leave a terminal open on the computer they were accessing the cluster from. After `Sequest` analysis on the cluster, data filtration and annotation is performed by `DTASelect` (Tabb et al. 2002) on the fileserver and the results are immediately available for browsing through a `html` interface from the outside.

The MudPIT setup was later enhanced by the acquisition of ThermoElectron’s next generation ion trap mass spectrometer: the LTQ linear ion trap instrument (Blackler et al. 2006). This instrument provides much higher scan rates as well as higher sensitivity and better signal-to-noise ratios due to a bigger ion capacity of the trap. It interfaces to a Surveyor four solvent channel HPLC pump (ThermoElectron, Waltham, MA) and a MicroAS autosampler (ThermoElectron, Waltham, MA). In this setup the HPLC pump only provides the low to high organic solvent gradients specific to the reverse phase parts of a MudPIT column, while the salt bumps necessary to elute subsets of peptides from the SCX phase are provided by injection of defined volumes and concentrations through the autosampler—potentially delivering much sharper salt peaks. Beyond that, quantitative mass spectrometric polypeptide analysis via metabolic incorporation of  $^{15}\text{N}$  and the program `ReLEx` (MacCoss et al. 2003) have been included as well.

To test the MudPIT setup established, we proceeded to analyze a diverse collection of affinity purified polypeptide complexes using baits mainly involved in cell cycle progression and transcription in the yeast *Saccharomyces cerevisiae* (Grau-

mann et al. 2004). This pilot study, which represents the core of this thesis, is documented in detail in chapter 2 (p. 28). The ORFs (open reading frames) to be studied were chromosomally tagged with a tandem affinity purification tag (TAP tag) analogous to the pioneering construct by Rigaut et al. (1999) and purified under native conditions from whole cell extract. Twentytwo out of 26 attempted chromosomal taggings succeeded. The study revealed 102 previously known and 279 potential new physical interactions to the set of tagged gene products. It includes among other things the characterization of a new subunit of the intensely studied Swi/Snf (Fry and Peterson 2001) and RSC (Sanders et al. 2002; Damelin et al. 2002; and Cairns et al. 1998) chromatin remodelling complexes. MudPIT proved mature enough for migration into less specialized environments and presented immediately new insights into systems extensively studied with more traditional techniques.

The technique has consequently been applied to a variety of problems linked with moderately complex polypeptide mixtures as delivered by affinity purified protein complexes. Appendices A (p. 67) and D (p. 153) present two such examples in detail.

We have also extended the use of MudPIT—in the spirit of Washburn et al. (2001), who analyzed whole cell lysates from *S. cerevisiae* by MudPIT—to much more complex mixtures with low abundant components. Appendix C (p. 121) presents an example for this with the analysis of affinity purified multiubiquitin conjugates from whole cell lysate.

## 1.5 References

- Abian, J., Oosterkamp, A. J. and Gelpi, E. (1999). Comparison of conventional, narrow-bore and capillary liquid chromatography/mass spectrometry for electrospray ionization mass spectrometry: Practical considerations. *J Mass Spectrom*, 34(4):244–54.
- Alberts, B. (1998). The cell as a collection of protein machines: preparing the next generation of molecular biologists. *Cell*, 92(3):291–4.
- Blackler, A. R., Klammer, A. A., MacCoss, M. J. and Wu, C. C. (2006). Quantitative comparison of proteomic data quality between a 2D and 3D quadrupole ion trap. *Anal Chem*, 78(4):1337–44.
- Bondarenko, P. V., Chelius, D. and Shaler, T. A. (2002). Identification and relative quantitation of protein mixtures by enzymatic digestion followed by capillary reversed-phase liquid chromatography–tandem mass spectrometry. *Anal Chem*, 74(18):4741–9.
- Cairns, B. R., Erdjument-Bromage, H., Tempst, P., Winston, F. and Kornberg, R. D. (1998). Two actin-related proteins are shared functional components of the chromatin-remodeling complexes RSC and SWI/SNF. *Mol Cell*, 2(5):639–51.
- Chelius, D. and Bondarenko, P. V. (2002). Quantitative profiling of proteins in complex mixtures using liquid chromatography and mass spectrometry. *J Proteome Res*, 1(4):317–23.
- Cherry, J. M., Adler, C., Ball, C., Chervitz, S. A. and Dwight, S. S. et al. (1998). SGD: *Saccharomyces* genome database. *Nucleic Acids Res*, 26(1):73–9.

- Craig, R. and Beavis, R. C. (2004). TANDEM: Matching proteins with tandem mass spectra. *Bioinformatics*, 20(9):1466–7.
- Creighton, T. E. (1993). *Proteins, Structures and Molecular Properties*. W. H. Freeman and Company, New York, 2nd edition.
- Damelin, M., Simon, I., Moy, T. I., Wilson, B. and Komili, S. et al. (2002). The genome-wide localization of Rsc9, a component of the RSC chromatin-remodeling complex, changes in response to stress. *Mol Cell*, 9(3):563–73.
- Edman, P. (1949). A method for the determination of the amino acid sequence in peptides. *Arch Biochem*, 22(3):475–476.
- Eng, J. K., McCormack, A. L. and Yates III, J. R. (1994). An approach to correlate tandem mass-spectral data of peptides with amino-acid sequences in a protein database. *J Am Soc Mass Spectr*, 5(11):976–989.
- Fenn, J. B., Mann, M., Meng, C. K., Wong, S. F. and Whitehouse, C. M. (1989). Electrospray ionization for mass spectrometry of large biomolecules. *Science*, 246(4926):64–71.
- Fry, C. J. and Peterson, C. L. (2001). Chromatin remodeling enzymes: Who’s on first?. *Curr Biol*, 11(5):R185–97.
- Gavin, A. C., Aloy, P., Grandi, P., Krause, R. and Boesche, M. et al. (2006). Proteome survey reveals modularity of the yeast cell machinery. *Nature*, 440(7084):631–636.
- Gavin, A. C., Bosche, M., Krause, R., Grandi, P. and Marzioch, M. et al. (2002). Functional organization of the yeast proteome by systematic analysis of protein complexes. *Nature*, 415(6868):141–7.
- Godovac-Zimmermann, J., Kleiner, O., Brown, L. R. and Drukier, A. K. (2005). Perspectives in spicing up proteomics with splicing. *Proteomics*, 5(3):699–709.

- Goshe, M. B. and Smith, R. D. (2003). Stable isotope-coded proteomic mass spectrometry. *Curr Opin Biotechnol*, 14(1):101–9.
- Gott, J. M. and Emeson, R. B. (2000). Functions and mechanisms of RNA editing. *Annu Rev Genet*, 34:499–531.
- Graumann, J., Dunipace, L. A., Seol, J. H., McDonald, W. H. and Yates III, J. R. et al. (2004). Applicability of tandem affinity purification MudPIT to pathway proteomics in yeast. *Mol Cell Proteomics*, 3(3):226–37.
- Griffin, E. E., Graumann, J. and Chan, D. C. (2005). The WD40 protein Caf4p is a component of the mitochondrial fission machinery and recruits Dnm1p to mitochondria. *J Cell Biol*, 170(2):237–48.
- Gygi, S. P., Rist, B., Gerber, S. A., Turecek, F. and Gelb, M. H. et al. (1999). Quantitative analysis of complex protein mixtures using isotope-coded affinity tags. *Nat Biotechnol*, 17(10):994–9.
- Hansen, K. C., Schmitt-Ulms, G., Chalkley, R. J., Hirsch, J. and Baldwin, M. A. et al. (2003). Mass spectrometric analysis of protein mixtures at low levels using cleavable  $^{13}\text{C}$ -isotope-coded affinity tag and multidimensional chromatography. *Mol Cell Proteomics*, 2(5):299–314.
- Henzel, W. J., Billeci, T. M., Stults, J. T., Wong, S. C. and Grimley, C. et al. (1993). Identifying proteins from two-dimensional gels by molecular mass searching of peptide fragments in protein sequence databases. *Proc Natl Acad Sci USA*, 90(11):5011–5.
- Hunt, D. F., Yates III, J. R., Shabanowitz, J., Winston, S. and Hauer, C. R. (1986). Protein sequencing by tandem mass spectrometry. *Proc Natl Acad Sci USA*, 83(17):6233–7.

- James, P., Quadroni, M., Carafoli, E. and Gonnet, G. (1993). Protein identification by mass profile fingerprinting. *Biochem Biophys Res Commun*, 195(1):58–64.
- Jonscher, K. R. and Yates III, J. R. (1997). The quadrupole ion trap mass spectrometer—a small solution to a big challenge. *Anal Biochem*, 244(1):1–15.
- Karas, M. and Hillenkamp, F. (1988). Laser desorption ionization of proteins with molecular masses exceeding 10,000 daltons. *Anal Chem*, 60(20):2299–301.
- Klose, J. (1975). Protein mapping by combined isoelectric focusing and electrophoresis of mouse tissues. A novel approach to testing for induced point mutations in mammals. *Humangenetik*, 26(3):231–43.
- Laemmli, U. K. (1970). Cleavage of structural proteins during the assembly of the head of bacteriophage T4. *Nature*, 227(5259):680–5.
- Larsen, M. R., Larsen, P. M., Fey, S. J. and Roepstorff, P. (2001). Characterization of differently processed forms of enolase 2 from *Saccharomyces cerevisiae* by two-dimensional gel electrophoresis and mass spectrometry. *Electrophoresis*, 22(3):566–75.
- Leitner, A. and Lindner, W. (2004). Current chemical tagging strategies for proteome analysis by mass spectrometry. *J Chromatogr B Analyt Technol Biomed Life Sci*, 813(1–2):1–26.
- Link, A. J., Eng, J., Schieltz, D. M., Carmack, E. and Mize, G. J. et al. (1999). Direct analysis of protein complexes using mass spectrometry. *Nat Biotechnol*, 17(7):676–82.
- Liu, H., Sadygov, R. G. and Yates III, J. R. (2004). A model for random sampling and estimation of relative protein abundance in shotgun proteomics. *Anal Chem*, 76(14):4193–201.

- Lundell, N. and Markides, K. (1992). Two-dimensional liquid chromatography of peptides: An optimization strategy. *Chromatographia*, 34(5–8):369–375.
- MacCoss, M. J., Wu, C. C., Liu, H., Sadygov, R. and Yates III, J. R. (2003). A correlation algorithm for the automated quantitative analysis of shotgun proteomics data. *Anal Chem*, 75(24):6912–21.
- Mann, M., Hojrup, P. and Roepstorff, P. (1993). Use of mass spectrometric molecular weight information to identify proteins in sequence databases. *Biol Mass Spectrom*, 22(6):338–45.
- Mayor, T., Lipford, J., Graumann, J., Smith, G. and Deshaies, R. (2005). Analysis of polyubiquitin conjugates reveals that the rpn10 substrate receptor contributes to the turnover of multiple proteasome targets. *Mol Cell Proteomics*, 4(6):741–51.
- McDonald, W. H., Ohi, R., Miyamoto, D. T., Mitchison, T. J. and Yates III, J. R. (2002). Comparison of three directly coupled hplc ms/ms strategies for identification of proteins from complex mixtures: Single-dimension lc-ms/ms, 2-phase mudpit, and 3-phase mudpit. *Int J Mass Spectrom*, 219(1):245–251.
- McDonald, W. H., Tabb, D. L., Sadygov, R. G., MacCoss, M. J. and Venable, J. et al. (2004). MS1, MS2, and SQT—three unified, compact, and easily parsed file formats for the storage of shotgun proteomic spectra and identifications. *Rapid Commun Mass Spectrom*, 18(18):2162–2168.
- Nesvizhskii, A. I. and Aebersold, R. (2005). Interpretation of shotgun proteomic data: The protein inference problem. *Mol Cell Proteomics*, 4(10):1419–40.
- O’Farrell, P. H. (1975). High resolution two-dimensional electrophoresis of proteins. *J Biol Chem*, 250(10):4007–21.

- Oda, Y., Huang, K., Cross, F. R., Cowburn, D. and Chait, B. T. (1999). Accurate quantitation of protein expression and site-specific phosphorylation. *Proc Natl Acad Sci USA*, 96(12):6591–6.
- Old, W. M., Meyer-Arendt, K., Aveline-Wolf, L., Pierce, K. G. and Mendoza, A. et al. (2005). Comparison of label-free methods for quantifying human proteins by shotgun proteomics. *Mol Cell Proteomics*, 4(10):1487–502.
- Olsen, J. V., de Godoy, L. M., Li, G., Macek, B. and Mortensen, P. et al. (2005). Parts per million mass accuracy on an Orbitrap mass spectrometer via lock mass injection into a C-trap. *Mol Cell Proteomics*, 4(12):2010–21.
- Olsen, J. V., Ong, S. E. and Mann, M. (2004). Trypsin cleaves exclusively C-terminal to arginine and lysine residues. *Mol Cell Proteomics*, 3(6):608–14.
- Ong, S. E., Blagoev, B., Kratchmarova, I., Kristensen, D. B. and Steen, H. et al. (2002). Stable isotope labeling by amino acids in cell culture, SILAC, as a simple and accurate approach to expression proteomics. *Mol Cell Proteomics*, 1(5):376–86.
- Pappin, D. J., Hojrup, P. and Bleasby, A. J. (1993). Rapid identification of proteins by peptide-mass fingerprinting. *Curr Biol*, 3(6):327–32.
- Perkins, D. N., Pappin, D. J., Creasy, D. M. and Cottrell, J. S. (1999). Probability-based protein identification by searching sequence databases using mass spectrometry data. *Electrophoresis*, 20(18):3551–67.
- Rigaut, G., Shevchenko, A., Rutz, B., Wilm, M. and Mann, M. et al. (1999). A generic protein purification method for protein complex characterization and proteome exploration. *Nat Biotechnol*, 17(10):1030–2.

- Ross, P. L., Huang, Y. N., Marchese, J. N., Williamson, B. and Parker, K. et al. (2004). Multiplexed protein quantitation in *Saccharomyces cerevisiae* using amine-reactive isobaric tagging reagents. *Mol Cell Proteomics*, 3(12):1154–69.
- Sadygov, R. G., Eng, J., Durr, E., Saraf, A. and McDonald, H. et al. (2002). Code developments to improve the efficiency of automated MS/MS spectra interpretation. *J Proteome Res*, 1(3):211–5.
- Sanders, S. L., Jennings, J., Canutescu, A., Link, A. J. and Weil, P. A. (2002). Proteomics of the eukaryotic transcription machinery: identification of proteins associated with components of yeast TFIID by multidimensional mass spectrometry. *Mol Cell Biol*, 22(13):4723–38.
- Shevchenko, A., Chernushevic, I., Shevchenko, A., Wilm, M. and Mann, M. (2002). “De novo” sequencing of peptides recovered from in-gel digested proteins by nanoelectrospray tandem mass spectrometry. *Mol Biotechnol*, 20(1):107–18.
- Spengler, B., Kirsch, D., Kaufmann, R. and Jaeger, E. (1992). Peptide sequencing by matrix-assisted laser-desorption mass spectrometry. *Rapid Commun Mass Spectrom*, 6(2):105–8.
- Spingola, M., Grate, L., Haussler, D. and Ares Jr., M. (1999). Genome-wide bioinformatic and molecular analysis of introns in *saccharomyces cerevisiae*. *RNA*, 5(2):221–34.
- Steen, H. and Mann, M. (2004). The ABC’s (and XYZ’s) of peptide sequencing. *Nat Rev Mol Cell Biol*, 5(9):699–711.
- Storz, G. (2002). An expanding universe of noncoding RNAs. *Science*, 296(5571):1260–3.

- Tabb, D. L., McDonald, W. H. and Yates III, J. R. (2002). DTASelect and Contrast: Tools for assembling and comparing protein identifications from shotgun proteomics. *J Proteome Res*, 1(1):21–6.
- Takahashi, N., Ishioka, N., Takahashi, Y. and Putnam, F. W. (1985). Automated tandem high–performance liquid chromatographic system for separation of extremely complex peptide mixtures. *J Chromatogr*, 326:407–18.
- Verma, R., Oania, R., Graumann, J. and Deshaies, R. J. (2004). Multiubiquitin chain receptors define a layer of substrate selectivity in the ubiquitin–proteasome system. *Cell*, 118(1):99–110.
- Wang, W., Zhou, H., Lin, H., Roy, S. and Shaler, T. A. et al. (2003). Quantification of proteins and metabolites by mass spectrometry without isotopic labeling or spiked standards. *Anal Chem*, 75(18):4818–26.
- Washburn, M. P., Wolters, D. and Yates III, J. R. (2001). Large–scale analysis of the yeast proteome by multidimensional protein identification technology. *Nat Biotechnol*, 19(3):242–7.
- Wolff, S., Otto, A., Albrecht, D., Stahl Zeng, J. and Buttner, K. et al. (2006). Gel–free and gel–based proteomics in *Bacillus subtilis*: A comparative study. *Mol Cell Proteomics*.
- Wolters, D. A., Washburn, M. P. and Yates III, J. R. (2001). An automated multidimensional protein identification technology for shotgun proteomics. *Anal Chem*, 73(23):5683–90.
- Wu, C. C., MacCoss, M. J., Howell, K. E., Matthews, D. E. and Yates III, J. R. (2004). Metabolic labeling of mammalian organisms with stable isotopes for quantitative proteomic analysis. *Anal Chem*, 76(17):4951–9.

- Wu, W. W., Wang, G., Baek, S. J. and Shen, R. F. (2006). Comparative study of three proteomic quantitative methods, DIGE, cICAT, and iTRAQ, using 2D gel- or LC-MALDI TOF/TOF. *J Proteome Res*, 5(3):651-8.
- Wysocki, V. H., Resing, K. A., Zhang, Q. and Cheng, G. (2005). Mass spectrometry of peptides and proteins. *Methods*, 35(3):211-22.
- Xie, J., Miao, Y., Shih, J., Tai, Y. C. and Lee, T. D. (2005). Microfluidic platform for liquid chromatography-tandem mass spectrometry analyses of complex peptide mixtures. *Anal Chem*, 77(21):6947-53.
- Yates III, J. R., Speicher, S., Griffin, P. R. and Hunkapiller, T. (1993). Peptide mass maps: A highly informative approach to protein identification. *Anal Biochem*, 214(2):397-408.
- Yu, L. R., Conrads, T. P., Uo, T., Issaq, H. J. and Morrison, R. S. et al. (2004). Evaluation of the acid-cleavable isotope-coded affinity tag reagents: Application to camptothecin-treated cortical neurons. *J Proteome Res*, 3(3):469-77.

## 2 Applicability of TAP–MudPIT to Pathway Proteomics in Yeast

This chapter describes the exploration of the use of multidimensional protein identification technology for the analysis of moderately complex polypeptide mixtures as resulting from affinity purification of protein complexes in a nonspecialized academic laboratory setting. It was published as

Graumann, J., Dunipace, L. A., Seol, J. H., McDonald, W. H. and Yates III, J. R. et al. (2004). Applicability of tandem affinity purification MudPIT to pathway proteomics in yeast. *Mol Cell Proteomics*, 3(3):226–37.

The copyright for the presented material is held by the American Society for Biochemistry and Molecular Biology, who has authorized its use in this work. The supplementary material referred to is available at <http://www.mcponline.org/cgi/data/M300099-MCP200/DC1/1>.

### 2.1 Summary

A combined multidimensional chromatography–mass spectrometry approach known as “MudPIT” enables rapid identification of proteins that interact with a tagged bait while bypassing some of the problems associated with analysis of polypeptides excised from SDS–polyacrylamide gels. However, the reproducibility, success rate, and applicability of MudPIT to the rapid characterization of dozens of proteins have not been reported. We show here that MudPIT reproducibly identified bona fide partners for budding yeast Gcn5p. Additionally, we successfully applied MudPIT to rapidly screen through a collection of tagged polypeptides to identify new

protein interactions. Twenty-five proteins involved in transcription and progression through mitosis were modified with a new TAP tag. TAP-MudPIT analysis of 22 yeast strains that expressed these tagged proteins uncovered known or likely interacting partners for 21 of the baits, a figure that compares favorably with traditional approaches. The proteins identified here comprised 102 previously-known and 279 potential physical interactions. Even for the intensively studied Swi2p/Snf2p, the catalytic subunit of the Swi/Snf chromatin remodeling complex, our analysis uncovered a new interacting protein, Rtt102p. Reciprocal tagging and TAP-MudPIT analysis of Rtt102p revealed subunits of both the Swi/Snf and RSC complexes, identifying Rtt102p as a common interactor with, and possible integral component of, these chromatin remodeling machines. Our experience indicates it is feasible for an investigator working with a single ion trap instrument in a conventional molecular/cellular biology laboratory to carry out proteomic characterization of a pathway, organelle, or process (i. e. “pathway proteomics”) by systematic application of TAP-MudPIT .

## 2.2 Introduction

To understand the function of a protein, it is crucial to characterize its physical environment: what other proteins is it interacting with under various conditions? Traditionally, this question has been addressed by biochemical fractionation of cell extracts under mild conditions and subsequent identification of the members of a purified protein complex by immunoblotting or peptide sequencing.

Primed by the dawning of the postgenomic era, genome-wide yeast two-hybrid interaction screens (Ito et al. 2001; and Uetz et al. 2000) and protein chip based

methods (Zhu et al. 2001) have supplemented traditional purification and identification techniques, allowing broader insight into the interaction networks that constitute a functional cell. Both of these approaches require the creation and maintenance of libraries of tagged proteins and in the case of protein chips the daunting task of purifying and spotting them under conditions that preserve their activity. The potential for detecting nonphysiological protein–protein interactions and the necessity to piece together interaction networks from a catalog of resulting binary interactions further complicate these approaches.

Developed in parallel with two–hybrid and protein chip technologies, mass spectrometry of protein complexes purified through single or tandem affinity steps eliminates the need for complex–specific immunochemicals and enables analysis of very small amounts of sample on a proteome wide scale (Gavin et al. 2002; and Ho et al. 2002). This approach can be performed under more physiological conditions and substitutes whole complex analysis for the reconstruction of interaction networks from binary interaction data. However, the Gavin et al. (2002) and Ho et al. (2002) studies employed SDS–PAGE to separate affinity–purified protein mixtures prior to mass spectrometric analysis, thereby encountering the problems linked to this technique including: limitations of dynamic range of detection, considerable sample parallelization, variable elution efficiency of peptides from the polyacrylamide matrix, and potential selection against proteins with properties that impede analysis by SDS–PAGE (e. g., unusually high or lowmolecular weight, diffuse migration, comigration with contaminants, and poor binding to stain).

To circumvent these problems McCormack et al. (1997) demonstrated the possibility of analyzing digested protein complexes directly using single dimensional liquid chromatography. An improvement of this method—multidimensional protein

identification technology (MudPIT; Link et al. 1999)—extended its applicability to large protein complexes and is a bona fide alternative to gel-based protein separation. MudPIT relies on digestion in solution of the protein mixture to be analyzed, and separation of the resulting complex peptide mixture by multidimensional capillary chromatography connected in-line to an ion trap mass spectrometer. Owing to its unique advantages, MudPIT is an attractive alternative to traditional methods for the rapid identification of protein-protein interactions for stoichiometric and substoichiometric partners. MudPIT can also be applied to deconvolve complex sets of proteins related by a common property. For example, Peng et al. (2003) applied a multidimensional approach similar to MudPIT to identify hundreds of candidate ubiquitinated proteins in budding yeast cells.

Despite its considerable power, some potential limitations to MudPIT remain to be addressed. For example, it is unclear how reproducible such analyses are. This is of particular concern for analysis of samples that contain many proteins, like that reported by Peng et al. (2003). Second, since only individual analyses have been reported to date, it remains unclear what the likelihood of success is for any given MudPIT experiment. The success rate of individual experiments, in turn, is important for the question of whether it will be profitable to scale the MudPIT approach to the rapid analysis of multiple baits. Third, because the issues of reproducibility and scalability have not been addressed, it is not known if the parallel application of MudPIT to multiple proteins will enable filtering approaches to separate bona fide interactors from nonspecific contaminants. Finally, it remains unclear how feasible it will be to transfer cutting-edge proteomic technologies like MudPIT from specialized environments to a conventional cell biology laboratory.

In this study, we address these various issues. We show that the combination

of a bipartite affinity tag with MudPIT allows for the rapid analysis of protein complexes. Pilot experiments with Gcn5p confirmed the reproducibility of the technique. Application of MudPIT to a set of 22 expressed baits revealed a success rate comparable to conventional approaches, and confirmed the scalability of the approach. Comparison of proteins identified across all MudPIT analyses, comprising diverse baits from different subcellular compartments and pathways, also enabled a filtering strategy to cull nonspecific contaminants. Our experience indicates that multidimensional chromatography in combination with mass spectrometry technology can be readily transferred from a specialized analytical chemistry environment to a traditional molecular cell biology laboratory. Routine application of MudPIT may thus enable cell biologists to dissect dynamic changes in protein interactions in response to specific chemical or biological ligands, environmental perturbations, or mutations.

## 2.3 Experimental Procedures

### 2.3.1 Construction of a Bipartite Affinity Purification Tag

To construct pJS-HPM53H, a 940 bp fragment was PCR amplified from pJS-TM53H (RDB1344, Seol et al. 2001) with the primers HTM A and B (see supplementary material, table 1). This was used as a template to PCR amplify a HPM tag containing 670 bp fragment with the primers HPM C and D (see supplementary material, table 1), which replaced the XhoI-EcoRI restriction fragment of pJS-TM53H.

### 2.3.2 Strain Construction

The bipartite affinity purification tags were amplified by PCR from pJS-HPM53H (HPM tag) or pKW804 (modified TAP tag, Cheeseman et al. 2001) with primers conferring sequence homology to the 3' end of targeted open reading frames (see supplementary material, table 1), using Expand High Fidelity PCR System (Roche, Indianapolis, IN). The resulting PCR products were transformed into the *Saccharomyces cerevisiae* strain RJD 415 (W303 background, *MATa*, *can1-100*, *leu2-3,-112*, *his3-11,-15*, *trp1-1*, *ura3-1*, *ade2-1*, *pep4Δ::TRP1*, *bar1Δ::HISG*; see supplementary material, table 2) with a modified Lithium acetate method (Ito et al. 1983). Integration and expression of the tagged gene product were checked by anti-myc western blotting of whole cell lysate using 9E10 monoclonal antibodies (Evan et al. 1985). Strain RJD 2067, carrying a TAP tagged (Rigaut et al. 1999) *GCN5* allele was a gift from Erin O'Shea, UCSF.

To knock out *SNF2*, *ARP9* and *RTT102*, an *HIS3* carrying cassette was PCR amplified from pFA6a-His3MX6 (Longtine et al. 1998) and transformed into the strain RJD 415. The primers used (see supplementary material, table 1), allowed for complete replacement of the respective open reading frames by homologous recombination.

### 2.3.3 Preparation of Protein Complexes by Dual-step Affinity Purification

### 2.3.4 HPM-Tag

Yeast cells carrying a HPM-tagged gene were grown in 2.5 l YPD (1 % yeast extract, 2 % bacto-peptone, 2 % glucose) to  $OD_{600\text{ nm}} \approx 1.5$ . Cell extract was prepared by

glass beading in TNET (20 mM Tris·HCl pH 7.5, 150 mM NaCl, 0.1 mM EDTA, 0.2% Triton X-100), supplemented with 10  $\mu\text{g}/\text{ml}$  Aprotinin, 10  $\mu\text{g}/\text{ml}$  Leupeptin, 10  $\mu\text{g}/\text{ml}$  Chymostatin and 2  $\mu\text{g}/\text{ml}$  Pepstatin A. The extract was cleared by centrifugation at 100,000 g and 4°C for 30 min. Crude extract (300 mg total protein in 14 ml volume) was incubated with 200  $\mu\text{l}$  9E10  $\alpha$ -myc (Evan et al. 1985) coupled protein A sepharose beads (Sigma, St. Louis, MO) for 1.5 h at 4°C. The beads were washed three times in 50 bead volumes cold TNET, resuspended in 300  $\mu\text{l}$  TNET and adjusted to 1 mM DTT. Protein complexes were eluted for 25 min at room temperature by addition of 10 units of GST-tagged PreScission Protease (Amersham, Piscataway, NJ) and protease carryover was reduced by 10 min further incubation with 1/10 9E10 bead volumes of glutathione sepharose 4B beads (Amersham, Piscataway, NJ).

For the second affinity purification step 20  $\mu\text{l}$  of Ni-NTA agarose beads (Qiagen, Valencia, CA) were added to 200  $\mu\text{l}$  supernatant from the first step and the sample was rotated for 1 h at 4°C. The beads were washed three times with 25 bead volumes of cold TNET and twice with 25 bead volumes of cold TNE (20 mM Tris·HCl pH 7.5, 150 mM NaCl, 0.1 mM EDTA). Proteins were eluted by addition of 50  $\mu\text{l}$  100 mM EDTA and the resulting supernatant lyophilized.

### 2.3.5 TAP Tag

Purification of TAP-tagged Gcn5p was modified from Rigaut et al. (1999). Protein extractions for strain RJD 2067 (see supplementary material, table 2), carrying a TAP tagged *GCN5* allele was performed as described for HPM tagged strains, substituting IPP150 (10 mM Tris·HCl pH 8.0, 150 mM NaCl, 0.1% NP40) for TNET.

After protein extraction, 200  $\mu$ l of IgG sepharose (Amersham, Piscataway, NJ) was added to 300 mg total protein in a volume of 14 ml. This slurry was incubated at 4°C, rotating for 2 h. After incubation, the resin was washed 3 times with 50 bead volumes of IPP150, and once with 50 bead volumes of TEV protease cleavage buffer (10 mM Tris·HCl pH 8.0, 150 mM NaCl, 0.1 % NP40, 0.5 mM EDTA, 1 mM DTT). The IgG sepharose was resuspended in 300  $\mu$ l TEV protease cleavage buffer containing 100 U TEV protease (Invitrogen, Carlsbad, CA), and incubated at room temperature, rotating, for 45 min. The bead supernatant (280  $\mu$ l) was then retrieved and mixed with 840  $\mu$ l of calmodulin binding buffer (10 mM  $\beta$ -mercaptoethanol, 10 mM Tris·HCl pH 8.0, 150 mM NaCl, 1 mM Mg-acetate, 1 mM imidazole, 2 mM CaCl<sub>2</sub>, 0.1 % NP40), 0.84  $\mu$ l 1 M CaCl<sub>2</sub>, and 200  $\mu$ l calmodulin beads (Stratagene, La Jolla, CA). This mixture was incubated for 1 h at 4°C, with rotating. After incubation, the beads were washed 3 times with 5 bead volumes of calmodulin binding buffer and eluted 2 times with 250  $\mu$ l of calmodulin elution buffer (10 mM  $\beta$ -mercaptoethanol, 10 mM Tris·HCl pH 8.0, 150 mM NaCl, 1 mM Mg-acetate, 1 mM imidazole, 2 mM EGTA, 0.1 % NP40). The eluate was TCA precipitated, and the pellet was washed two times with ice cold acetone.

### 2.3.6 Modified TAP-Tag

The protocol for affinity purification of Gen5p tagged with the modified TAP tag was adapted from Cheeseman et al. (2001) and was identical to the TAP protocol up through the TEV protease treatment. After TEV protease digestion 50  $\mu$ l of S protein agarose (Novagen, Madison, WI) was added to 280  $\mu$ l of the supernatant and the slurry was incubated, rotating, at 4°C for 1.5 h. The beads were washed 3

times with 10 volumes of IPP150, once with IPP150 without NP40, and then with 50 mM Tris·HCl pH 8.5, 5 mM EGTA, 1 mM EDTA, 75 mM KCl. The protein was eluted in 50  $\mu$ l 100 mM Tris·HCl pH 8.5, 8 M urea for 30 min at room temperature.

### 2.3.7 Proteolytic Digest

Protein samples were proteolytically digested as follows: lyophilized protein mixtures were resolubilized in 40  $\mu$ l 8 M urea, 100 mM Tris·HCl pH 8.5 and reduced by incubation at a final concentration of 3 mM T-CEP (Pierce, Rockford, IL) for 20 min at room temperature. Reduced cysteines were subsequently alkylated by addition of iodoacetamide (10 mM final concentration) and incubation for 15 min at room temperature. Proteolysis was initiated with 0.1  $\mu$ g endoproteinase Lys-C (Roche, Indianapolis, IN) and allowed to proceed for 4 h at 37°C. The sample was then diluted fourfold by addition of 100 mM Tris·HCl pH 8.5 and adjusted to 1 mM CaCl<sub>2</sub>. Next, 0.5  $\mu$ g of sequencing grade trypsin (Roche, Indianapolis, IN) were added and the mixture incubated overnight at 37°C. The digest was quenched with the addition of formic acid to 5% and stored at -20°C.

### 2.3.8 Multidimensional Protein Identification Technology (MudPIT)

The peptide mixtures were separated utilizing a triphasic microcapillary column as described in McDonald et al. (2002). A fused silica capillary with an inner diameter of 100  $\mu$ m (PolyMicro Technology, Phoenix, AZ) and a 5  $\mu$ m diameter tip pulled with a P-2000 capillary puller (Sutter Instrument Company, Novato, CA) was packed with 6.5 cm 5  $\mu$ m Aqua C<sub>18</sub> reverse phase material (Phenomenex, Ventura,

CA), 3.5 cm 5  $\mu\text{m}$  Partisphere strong cation exchanger (Whatman, Clifton, NJ) and another 2.5 cm 5  $\mu\text{m}$  Aqua C<sub>18</sub> (in this order from the tip). The sample was pressure loaded onto the column.

In the event of irreversible column clogging, the 6.5 cm 5  $\mu\text{m}$  Aqua C<sub>18</sub> separation phase was replaced by an inline microfilter assembly (UpChurch Scientific, Oak Harbour, WA) and a 250  $\mu\text{m}$  ID fused silica collection capillary to reduce the overall back pressure. A 6.5 cm 5  $\mu\text{m}$  Aqua C<sub>18</sub> separation phase was spliced onto the setup after completion of loading. We noted that the presence of EDTA in the sample may increase the risk of clogging events.

The sample-loaded column was placed in line between a HP-1100 quaternary HPLC pump (Agilent, Palo Alto, CA) and a LCQ-DecaXP electrospray ion trap mass spectrometer (ThermoElectron, Palo Alto, CA). Sample separation was achieved with a six step chromatography program modified according to McDonald et al. (2002). Solutions used were 5% acetonitrile/0.1% formic acid (buffer A), 80% acetonitrile/0.1% formic acid (buffer B) and 500 mM ammonium acetate/5% acetonitrile/0.1% formic acid (buffer C). Step 1 consisted of an 80 min gradient to 40% buffer B, followed by a 10 min gradient to 100% buffer B and 10 min of 100% buffer B. Chromatography steps 2 to 5 followed the same pattern: 3 min of 100% buffer A followed by a 2 min buffer C pulse, a 10 min gradient to 15% buffer B and a 100 min gradient to 45% buffer B. The buffer C percentages used were 5%, 12.5%, 25% and 40%, respectively, for the steps. The terminal step consisted of 3 min 100% buffer A, 20 min 100% buffer C, a 10 min gradient to 15% buffer B and a 100 min gradient to 55% buffer B. The flow rate through the column was approximately 150 nl/min.

Eluting peptides were electrosprayed into the mass spectrometer with a distally

applied spray voltage of 2.4 kV. The column eluate was continuously analyzed during the whole six step chromatography program. One full range mass-scan (400–1400 m/z) was followed by three data dependent MS/MS spectra at 35 % collision energy in a continuous loop.

Both HPLC pump and mass spectrometer were controlled by the *Xcalibur* software (ThermoElectron, Palo Alto, CA).

### 2.3.9 Data Analysis

In a first step, MS/MS spectra recorded by *Xcalibur* were analyzed for their charge state and controlled for data quality by *2to3* (Sadygov et al. 2002). The data were then searched by *SEQUEST* (Eng et al. 1994) against the translated *Saccharomyces* Genome Database (SGD; Cherry et al. 1998; release time stamped 05/23/03) supplemented with common contaminants (e. g., Keratins) on a Linux cluster composed of twenty 1.8 GHz Athlon CPUs (Racksaver, San Diego, CA). *DTASelect* (Tabb et al. 2002) filtered the *SEQUEST* results according to the following parameters: minimum *XCORRs* of 1.8, 2.5 and 3.5 for singly, doubly and triply charged precursor ions, respectively, minimum  $\Delta C_n$  of 0.08, and a minimum requirement of two peptides per protein.

The resulting data was annotated and sorted with the Python script *RAYzer*. Annotation was added from SGD annotation tables (Cherry et al. 1998; table release time stamped 06/07/03) and interaction data curated by the MIPS Comprehensive Yeast Genome Database (MIPS CYGD; Mewes et al. 1997, 2002; release time stamped 04/29/03), the General Repository for Interaction Datasets (GRID; Breitkreutz et al. 2003; release 1.0) and the Yeast Protein Database (YPD; Garrels 1996;

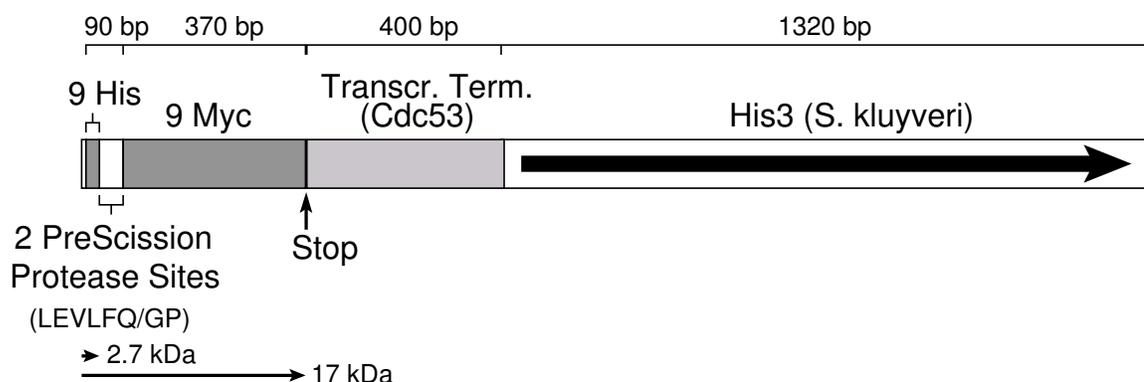
as of 06/09/03). Based on known interaction annotation and the frequency of appearance in a reference data set containing one representative experiment for every tagged open reading frame in this study ( $n = 22$ ), the data were then sorted into three tables: previously reported interactors retrieved in the experiment, potential new interacting proteins detected and likely contaminants (see supplementary online material). Proteins recovered in greater than 20% of the experiments in the reference data set were automatically considered contaminants (see section 2.5).

## 2.4 Results

### 2.4.1 HPM-Tag

We constructed a bipartite affinity tag composed of nine histidines and nine myc-epitopes separated by two PreScission protease (Cordingley et al. 1990; and Walker et al. 1994) cleavage sites (HPM tag, figure 2.1; see section 2.3). Homologous recombination enables chromosomal integration of the PCR-amplified cassette in *Saccharomyces cerevisiae his3* strains at the 3' end of open reading frames targeted for affinity purification.

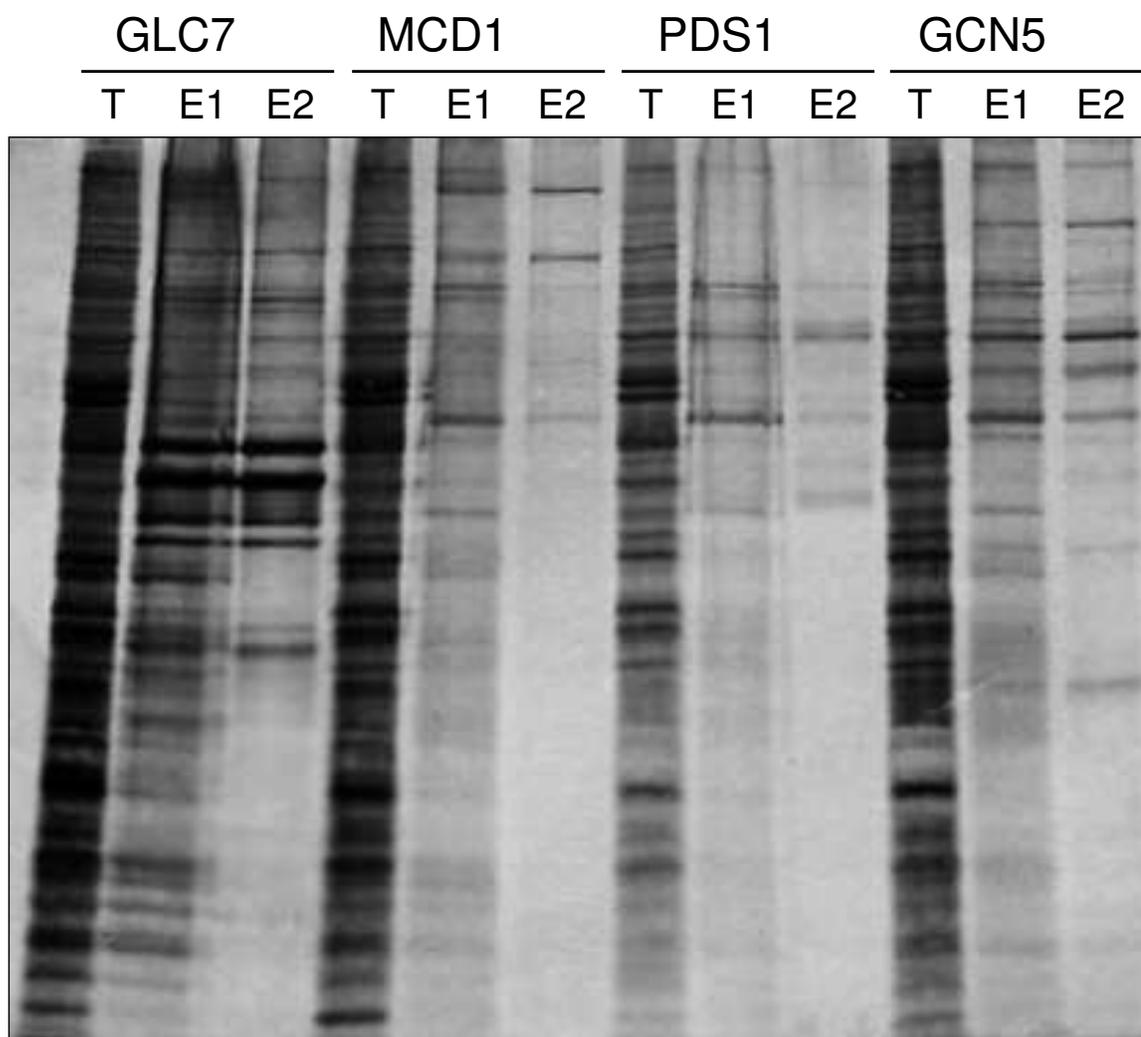
Using this cassette we tagged a test set of 25 gene products involved in transcription and progression through mitosis (see supplementary material, table 2) and established a variant of the “tandem affinity purification” (TAP) protocol (Rigaut et al. 1999) that employs affinity chromatography on a 9E10 monoclonal antibody resin followed by elution with PreScission Protease and adsorption to Ni-NTA resin (see section 2.3). For simplicity’s sake we refer to our protocol as “TAP,” even though



**Figure 2.1 Schematic Representation of the HPM Tag.** Nine histidines are separated from nine consecutive myc epitopes by two PreScission protease cleavage sites. The transcriptional terminator downstream of the Stop codon is from the *CDC53* locus. Chromosomal integration of the cassette can be selected for by restoring histidine prototrophy to *his3* mutant *S. cerevisiae* strains.

our tandem tag design requires a different purification protocol. Preliminary mass spectrometrical analyses showed that the eluates from the 9E10 resin still retained a high level of contaminating protein background (data not shown), and thus subsequent analyses were performed only on samples that were subjected to the complete TAP protocol. A representative SDS-PAGE analysis of the purification of four gene products is shown in figure 2.2.

The effectiveness and reproducibility of our overall approach was evaluated by analyzing the intensively studied histone acetyltransferase (HAT) Gcn5p (see fig. 2.3). Of the 23 previously reported interactors that were identified here, our experiments captured 15 (65%) in all three replicates and an additional 5 (22%) in two out of three attempts, including 18 known members of the SAGA/SLIK and ADA-HAT complexes (Sanders et al. 2002; Eberharter et al. 1999; Grant et al. 1998; and Pray-Grant et al. 2002). The majority of these validated partners ranked at the top of the list when the recovered proteins were sorted based on the size-normalized number of unique peptides sequenced per protein. These data indicate that TAP-MudPIT shows a high degree of reproducibility and robustness independent of fluctuations in



**Figure 2.2 SDS–Polyacrylamide Gel Analysis of Glc7p–HPM, Mcd1p–HPM, Pds1p–HPM and Gcn5p–HPM Affinity Purifications.** T: 2.5  $\mu$ g total cell extract protein. E1: 7% of material eluted by PreScission protease digest from  $\alpha$ -myc antibody beads. E2: 20% of EDTA eluate from the second affinity purification resin (Ni-NTA).

the sample quality of the individual experiment (see, e. g., varying peptide recovery for the bait in fig. 2.3).

Previous reports employed the original bipartite TAP tag and a modified TAP tag for tandem affinity purification (Rigaut et al. 1999; Gavin et al. 2002; and Cheeseman et al. 2001). A direct comparison of Gcn5p–TAP, Gcn5p–modified TAP,

**Figure 2.3 Reproducibility of Results Between Independent Gcn5p-HPM TAP-MudPIT Experiments.** Samples were prepared and analyzed as described in 2.3. Column “Known Interactor” indicates whether the gene product is a previously known Gcn5p interactor according to MIPS, GRID and YPD. Column “Gene Product” represents the name of the protein according to SGD. Red, yellow and plain background indicate recovery of the protein in three, two or one experiment out of three, respectively. Column “Frequency in Reference Set” lists the frequency with which the gene product was retrieved in the complete data set ( $n = 22$ ). Column “Length” represents the length of the ORF in amino acids according to SGD. Columns “Exp. 1–3” list the number of unique and total peptide hits assigned to the ORF for each of the three experiments. Gene products are listed in descending order starting with the highest average length-normalized number of unique peptide identifications. Data for highly homologous ORFs with identical length, identical peptide representation across experiments and identical frequency in the reference set have been merged. Ty-element related ORFs have been excluded from the analysis.

Known Interactor?	Gene Product	Frequency in Reference Set	Length (AA)	Exp. 1 Unique (All)	Exp. 2 Unique (All)	Exp. 3 Unique (All)
	Bait	Gon5p	4.55%	439	41 (41)	65 (65)
		YCR082Wp	4.55%	128	13 (13)	5 (5)
X		Ada2p	4.55%	434	22 (22)	15 (15)
X		Ngg1p	4.55%	702	30 (30)	29 (29)
X		Sgf29p	4.55%	259	19 (19)	4 (4)
		Rps22A/Bp	90.91%	130	6 (6)	3 (3)
		Rpl28p	68.18%	149	6 (6)	3 (3)
		Rpl2A/Bp	63.64%	254	7 (7)	5 (5)
X		Ahc1p	4.55%	566	16 (16)	7 (7)
X		Taf5p	4.55%	798	16 (16)	10 (10)
		Rps18A/Bp	68.18%	146	2 (2)	6 (6)
X		Hfi1p	4.55%	488	9 (9)	3 (3)
X		Taf9p	4.55%	157	2 (2)	2 (2)
X		Spt6p	4.55%	602	11 (11)	6 (6)
X		Spt20p	4.55%	604	8 (8)	6 (6)
X		Taf6p	4.55%	516	6 (6)	5 (5)
X		Taf10p	4.55%	206	—	4 (4)
		Rpl32p	63.64%	130	2 (2)	2 (2)
		Npl3p	95.45%	414	9 (9)	8 (8)
		Rps4A/Bp	59.09%	261	2 (2)	3 (3)
X		Taf12p	4.55%	539	7 (7)	6 (6)
X		Adh1p	86.36%	348	6 (8)	10 (14)
		Tef1/2p	90.91%	458	5 (5)	11 (11)
		Rps3p	54.55%	240	4 (4)	3 (3)
		Rps16A/Bp	27.27%	143	2 (2)	2 (2)
X		Sgf73p	4.55%	657	10 (10)	4 (4)
X		Spt3p	4.55%	337	4 (4)	—
		Rpp2Ap	18.18%	106	2 (2)	—
		Rps13p	22.73%	151	—	2 (2)
		Rps24A/Bp	22.73%	135	—	3 (3)
		Rpl19A/Bp	50.00%	189	—	6 (6)
		Rps15p	54.55%	142	—	3 (3)
X		Fba1p	68.18%	359	6 (6)	9 (9)
		Rps20p	31.82%	121	—	3 (3)
		Cts2p	81.82%	511	6 (6)	4 (4)
		Ugp1p	59.09%	499	8 (8)	12 (12)
		Rps6A/Bp	36.36%	261	3 (3)	2 (2)
		Rpl30p	22.73%	105	2 (2)	2 (2)
		Rpl14A/Bp	27.27%	138	2 (2)	—
X		Spt7p	4.55%	1332	9 (9)	10 (10)
		Sro9p	95.45%	466	4 (4)	6 (6)
		Rpl3p	40.91%	387	4 (4)	3 (3)
		YPL047Wp	4.55%	99	—	3 (3)
		Shm1p	63.64%	565	8 (8)	9 (9)
X		Ubp8p	4.55%	471	—	13 (13)
		Rpl10p	40.91%	221	—	6 (6)
		Sds22p	13.64%	338	—	9 (9)
		Cdc18p	50.00%	500	2 (2)	9 (10)
		Rpl38p	4.55%	78	2 (2)	—
		Rps8A/Bp	27.27%	261	2 (2)	2 (2)
		Rpl11A/Bp	9.09%	174	2 (2)	2 (2)
		Hta1/2p	4.55%	132	3 (3)	—
		Glc7p	13.64%	312	—	7 (7)
		Rpl25p	54.55%	142	—	2 (2)
		Rpl9Ap	95.45%	191	1 (10)	1 (4)
		Mis1p	90.91%	975	4 (4)	2 (2)
		Ssb1/2p	59.09%	613	2 (2)	10 (10)
		Psa1p	45.45%	361	3 (3)	—
		Rpp0p	45.45%	312	—	2 (2)
		Hyp2p	27.27%	157	3 (3)	0 (2)
		Rps25A/Bp	13.64%	108	—	2 (2)
		Rpp2Bp	9.09%	110	—	2 (2)
		Rps29Bp	18.18%	56	—	1 (2)
		Rpl31Ap	95.45%	113	1 (7)	0 (3)
		Rpl9Bp	90.91%	191	1 (10)	1 (4)
		Ura7p	13.64%	579	—	9 (10)
		Vma2p	22.73%	517	3 (3)	—
		Plk1p	63.64%	987	4 (4)	7 (7)
		Rps17A/Bp	22.73%	136	—	2 (2)
X		Tra1p	4.55%	3744	11 (11)	7 (7)
		Rps31p	31.82%	152	2 (4)	0 (4)
		Svp26p	0.00%	228	—	3 (3)
		Tdh1p	95.45%	613	2 (6)	6 (13)
		Ypl1p	13.64%	155	—	2 (2)
		Rps11A/Bp	22.73%	156	—	2 (2)
		Rib4p	13.64%	169	—	2 (2)
		Rps2p	9.09%	254	3 (3)	—
		Acs2p	40.91%	683	—	8 (8)
X		Eno2p	22.73%	437	3 (7)	2 (3)
		Tdh3p	100.00%	613	2 (8)	3 (19)
		Msn4p	4.55%	630	2 (2)	—
		Rpl17Ap	45.45%	184	0 (2)	1 (5)

Continued...

Known Interactor?	Gene Product	Frequency in Reference Set	Length (AA)	Exp. 1 Unique (All)	Exp. 2 Unique (All)	Exp. 3 Unique (All)
		Rpl18A/Bp	31.82%	186	—	—
		Adh3p	50.00%	375	1 (2)	3 (4)
		Ilv1p	45.45%	576	—	3 (3)
		Rpl16Bp	27.27%	198	—	—
		Rpl16Ap	13.64%	199	—	—
		Rpl15Ap	36.36%	204	0 (2)	1 (2)
X		Pgk1p	0.00%	416	4 (4)	—
		Kcs1p	31.82%	1050	—	—
		Gua1p	9.09%	525	—	5 (5)
		Cdc33p	13.64%	213	—	2 (2)
		Pnc1p	18.18%	216	—	2 (2)
		Rpl1A/Bp	27.27%	217	—	—
		Rpl31Bp	90.91%	113	0 (6)	0 (3)
		YOR283Wp	4.55%	230	—	2 (2)
		Yef3p	54.55%	1044	—	9 (13)
		Sod2p	4.55%	233	—	—
		Shm2p	13.64%	469	—	4 (4)
		Trp5p	22.73%	707	—	6 (6)
		Tfp1p	36.36%	1071	2 (2)	7 (7)
		Prb1p	31.82%	635	2 (2)	3 (3)
		Rps1Ap	40.91%	255	2 (3)	0 (2)
		Rps1Bp	45.45%	255	1 (2)	0 (2)
		Ssa2p	100.00%	639	2 (12)	1 (8)
		Rpl8Ap	50.00%	256	0 (3)	0 (2)
		YJR023Cp	0.00%	133	—	1 (1)
		Ura2p	63.64%	2214	2 (2)	2 (2)
		Tub2p	13.64%	457	—	—
		Sec23p	18.18%	768	—	5 (5)
		Rpl21Ap	36.36%	160	—	0 (3)
		Trp3p	13.64%	484	—	—
		Sik1p	4.55%	504	3 (3)	—
		Adh2p	72.73%	348	2 (4)	0 (4)
		Hos3p	22.73%	697	4 (4)	—
		Rpl6Bp	27.27%	176	1 (3)	0 (2)
		Rpl6Ap	31.82%	176	1 (3)	0 (2)
		Act1p	18.18%	375	—	2 (2)
		Aro2p	9.09%	376	—	2 (2)
X		Pfk2p	36.36%	959	2 (2)	3 (3)
		Rpl13Bp	36.36%	199	0 (2)	0 (3)
		Yhb1p	18.18%	399	—	—
		Eno1p	18.18%	437	1 (5)	1 (2)
		Hsm3p	0.00%	480	—	2 (2)
X		Clu1p	45.45%	1277	5 (5)	—
		Rrb1p	13.64%	511	—	2 (2)
		Asn2p	22.73%	572	—	2 (5)
		Ssa1p	100.00%	642	2 (12)	0 (7)
		Ttc1p	4.55%	649	—	—
		Ppz2p	13.64%	710	—	—
		Gla1p	9.09%	717	—	—
		Pbp1p	9.09%	722	2 (2)	—
		Rpl4Ap	68.18%	362	0 (5)	1 (8)
		Ysh1p	4.55%	779	—	—
		Imd1p	22.73%	403	0 (2)	0 (2)
		Eft1/2p	18.18%	842	—	—
		Sec24p	4.55%	926	—	—
		Imd3p	27.27%	523	0 (4)	0 (2)
		Asn1p	9.09%	572	—	—
		Tdh2p	100.00%	613	0 (6)	1 (17)
		YPL137Cp	9.09%	1276	—	—
		YHL035Cp	0.00%	1592	2 (2)	—
		Glt1p	13.64%	2145	—	—
		Sth1p	9.09%	1359	1 (2)	—
		Ssa4p	72.73%	642	0 (4)	0 (2)
		Imd2p	22.73%	523	0 (4)	0 (2)
		Rpl8Bp	50.00%	256	0 (3)	0 (2)
		Rps14Bp	22.73%	138	0 (2)	—
		Rps14Ap	22.73%	137	0 (2)	—
		Rpl21Bp	36.36%	160	—	0 (3)
		Anb1p	0.00%	157	—	0 (2)
		Rpl20Ap	45.45%	180	0 (3)	0 (3)
		Rpl17Bp	40.91%	184	0 (2)	0 (4)
		Imd4p	9.09%	524	0 (3)	—
		Rpl13Ap	36.36%	199	0 (2)	0 (3)
		Rpl20Bp	45.45%	174	0 (3)	0 (3)
		Rpl15Bp	36.36%	204	0 (2)	—
		Ade3p	9.09%	946	—	—
		Ssa3p	81.82%	649	0 (4)	0 (3)
		Rpl4Bp	68.18%	362	0 (5)	0 (7)
		Hef3p	18.18%	1044	—	—
		Rpl40A/Bp	18.18%	128	0 (2)	0 (5)
		Ubi4p	22.73%	381	0 (2)	0 (4)

and Gcn5p–HPM revealed that the set of previously known interactors identified with the different tags are well within the margins of variability between independent experiments performed with the HPM tag (table 2.1).

Remarkably, our comparative analysis of Gcn5p purifications yielded strong candidates for six new Gcn5p interactors. YCR082W, a nonessential gene product (Winzeler et al. 1999; and Giaever et al. 2002) with unknown function, was found in all five Gcn5p purifications but was not recovered with any of the other baits that we analyzed. YCR082W exhibits a two–hybrid interaction with Ahc1p (Uetz et al. 2000; and Ito et al. 2001), which together with Gcn5p is a member of the ADA histone acetyltransferase complex (Eberharter et al. 1999). Another candidate is Msn4p, a nonessential (Estruch and Carlson 1993; and Winzeler et al. 1999) major transcriptional regulator of stress responses (Treger et al. 1998). Msn4p was recovered in four of the five Gcn5p pull down experiments but was not recovered with any of the other baits. This finding is interesting in the light of evidence that promoters activated by Msn4p and its partner Msn2p show increased histone H4 acetylation (Deckert and Struhl 2001). Other potential interaction partners include YPL047W (present in two of the HPM purifications and the TAP purification), histones Hta1p/Hta2p and Imd4p (in TAP, modified TAP and one HPM pulldown). Other gene products recovered in more than two of the experiments are mostly ribosomal proteins that are likely contaminants. Finally, the interaction observed between Gcn5p and Swi1p in the TAP tag experiment was previously proposed only on the basis of their synthetically lethal genetic interaction (Pollard and Peterson 1997).

**Table 2.1 Comparison of TAP–MudPIT Analyses Using Different Bipartite Affinity Tags to Gcn5p.** Samples were prepared and analyzed as described in section 2.3. Column “Gene Product” represents the name of the gene product recovered and known to interact with Gcn5p according to GRID, MIPS and YPD. “Exp. 1–3” represent three independent affinity purifications of Gcn5p–HPM. “TAP tag” and “Mod. TAP tag” represent tandem affinity purification–MudPIT experiments performed with strains in which the *GCN5* locus was tagged with either the TAP (Rigaut et al. 1999) or modified TAP tag from Cheeseman et al. (2001). The numbers of unique peptides from each ORF that were sequenced are shown (with the total number of sequenced peptides in parentheses). The last column lists the frequency with which the gene product is found in the entire data set ( $n = 22$ ). For example, a gene product found in association with a single bait has a frequency of 4.55 % (1/22). The GRID, MIPS, and YPD interaction databases contain 83 additional gene products classified as interacting with Gcn5p, but not recovered in our analyses.

Gene Product	HPM tag			TAP tag	Mod. TAP tag	Frequ. in Ref. Set
	Exp. 1	Exp. 2	Exp. 3			
Gcn5p	41 (41)	25 (25)	65 (65)	19 (19)	21 (21)	4.55 %
Ada2p	22 (22)	15 (15)	36 (36)	22 (22)	36 (36)	4.55 %
Adh1p	6 (8)	10 (14)	2 (4)	—	—	86.36 %
Ahc1p	16 (16)	7 (7)	22 (22)	7 (7)	31 (31)	4.55 %
Clu1p	5 (5)	—	—	—	—	45.45 %
Eno2p	3 (7)	2 (3)	—	—	—	22.73 %
Fba1p	6 (6)	9 (9)	—	—	—	68.18 %
Hfi1p	9 (9)	3 (3)	23 (23)	20 (20)	23 (23)	4.55 %
Ngg1p	30 (30)	29 (29)	53 (53)	43 (43)	68 (68)	4.55 %
Pfk2p	2 (2)	3 (3)	—	—	—	36.36 %
Pgk1p	4 (4)	—	—	—	—	0.00 %
Rpg1p	—	—	—	—	5 (5)	0.00 %
Sgf29p	19 (19)	4 (4)	17 (17)	21 (21)	32 (32)	4.55 %
Sgf73p	10 (10)	4 (4)	18 (18)	25 (25)	29 (29)	4.55 %
Spt20p	8 (8)	6 (6)	26 (26)	26 (26)	29 (29)	4.55 %
Spt3p	4 (4)	—	12 (12)	12 (12)	8 (8)	4.55 %
Spt7p	9 (9)	10 (10)	28 (28)	49 (49)	52 (53)	4.55 %
Spt8p	11 (11)	6 (6)	23 (23)	18 (18)	20 (20)	4.55 %
Swi1p	—	—	—	3 (3)	—	9.09 %
Taf10p	—	4 (4)	9 (9)	7 (7)	11 (11)	4.55 %
Taf12p	7 (7)	6 (6)	16 (16)	28 (28)	23 (23)	4.55 %
Taf5p	16 (16)	10 (10)	35 (35)	46 (46)	37 (37)	4.55 %
Taf6p	6 (6)	5 (5)	23 (23)	24 (24)	26 (26)	4.55 %
Taf9p	2 (2)	2 (2)	7 (7)	7 (7)	15 (15)	4.55 %
Tra1p	11 (11)	7 (7)	35 (35)	82 (82)	99 (99)	4.55 %
Ubp8p	—	—	13 (13)	17 (17)	18 (18)	4.55 %
Yap1p	—	—	—	6 (6)	12 (12)	0.00 %

**Table 2.2a Potential New Interactors for a Test Set of HPM Tagged Proteins.** Samples were prepared and analyzed as described in 2.3. Column “Known interactors—Total” lists the number of physical/genetic interactions reported for the bait in the combined GRID/MIPS/YPD databases. “Known interactors—Recovered” represents the number of known physical/genetic interactors experimentally retrieved in this study. Partners marked “\*” are reported to interact physically as well as genetically. Column “Potential new interactors” contains all gene products identified by TAP–MudPIT, which are not listed as known interactors and are recovered in association with less than 20% of the baits analyzed ( $n = 22$ ).

Bait	Known interactors		Potential new interactors
	Total	Recovered	
	phys./gen.	phys. genet.	
Bim1p–HPM	6/57	1 —	Rpb2p, Rpl12A/Bp, Rpl22Ap, Rps25A/Bp, Rps29Ap, Rps5p, YGR161C–Cp
Cdc20p–HPM	12/3	6 —	Bub3p, Cct4p, Cct6p, Cct7p, Cct8p, Hef3p, Ilv6p, Pnc1p, Rfa1p
Chk1p–HPM	16/0	— —	Act1p, Car2p, Gpd2p, Hht1p, Hht2p, Htb2p, Htb1p, Htz1p, Pnc1p
Cla4p–HPM	15/77	—	Rpl17Bp, Pbp1p, Pre8p, Rpl36Ap, Rpl36Bp, Rpl7Ap, Rpl17Ap, Rpl7Bp, Rpp2Ap, Rps2p, Sec23p, Skm1p, Rpl19Bp, YBR225Wp, Yhb1p Rpl19Ap
Dbf2p–HPM	27/9	3	Dbf20p, Adh5p, Caf20p, Car2p, Cdc33p, Emi2p, Mob1p* Gfa1p, Gly1p, Gpd2p, Hsp42p, Ilv6p, Pnc1p, Pro1p, Rib4p, Sec23p, Shm2p, Snf1p, Trp3p, Tub2p

## 2.4.2 Screening for Interactions

Having established the relative reproducibility of TAP–MudPIT and the comparability of the HPM tag to other available bipartite affinity tags, we set out to address three issues. First, we wished to determine what fraction of TAP–MudPIT experiments yield usable results. Second, we hoped to determine whether the parallel application of MudPIT to numerous baits would enable us to cull nonspecific contaminants by comparing protein identifications across multiple experiments. Third, we wanted to test whether it will be feasible for an investigator in a cell biology laboratory to work at the scale needed to dissect a biological pathway or process by systematic application of MudPIT to a few dozen gene products. To address these questions, we screened for new protein–protein interactions in a test set of 25 gene products involved in transcription and progression through mitosis. Table 2.2 summarizes the results and gives an overview of potential new interactors. The

**Table 2.3b Potential New Interactors for a Test Set of HPM Tagged Proteins.** Samples were prepared and analyzed as described in 2.3. Column “Known interactors—Total” lists the number of physical/genetic interactions reported for the bait in the combined GRID/MIPS/YPD databases. “Known interactors—Recovered” represents the number of known physical/genetic interactors experimentally retrieved in this study. Partners marked “\*” are reported to interact physically as well as genetically. Column “Potential new interactors” contains all gene products identified by TAP–MudPIT, which are not listed as known interactors and are recovered in association with less than 20% of the baits analyzed ( $n = 22$ ).

Gcn5p–HPM	99/12	18	Ngg1p*	Ade3p, Eft2p, Eft1p, Gfa1p, Glc7p, Msn4p, Ppz2p, Rpl16Ap, Rpp2Ap, Rpp2Bp, Rps25Ap, Rps25Bp, Rps29Bp, Sds22p, Sod2p, Tfc1p, Trp3p, Tub2p, Ura7p, YCR082Wp, Yhb1p, Ypi1p, YPL047Wp, YPL137Cp, Ysh1p
Glc7p–HPM	177/9	28	Ppz2p*, Ppz1p*, Reg1p*	Abf1p, Ade16p, Ade17p, Ahp1p, Bmh1p, Bmh2p, Ccr4p, Cka2p, Eno1p, Fun21p, Gal83p, Hsp60p, Imp2p, Mor1p, Pdc1p, Pgc1p, Pol2p, Rpp2Ap, Snf1p, Sol1p, Sol2p, YBR225Wp, YDR474Cp, YER158Cp, YGR237Cp, YHR097Cp, YPL137Cp
Ino4p–HPM	52/0	1	—	Act1p, Mdn1p, Pmd1p, Xrs2p

primers to tag CDC5 and ESS1 while TAP–MudPIT experiments for Bir1p–HPM and Nbp1p–HPM resulted in little or no recovery of the tagged baits themselves. Of the 21 “successful” purifications that yielded sequence coverage for the tagged bait, 20 of the experiments (95%) yielded interacting proteins that are either true binding partners validated by other direct approaches, probable binding partners that display genetic interaction with the bait, or candidate binding partners that were found in association with only one bait. The Pho2p–HPM experiment yielded ‘hits’ only from proteins that were found associated with other, unrelated baits or were otherwise deemed to be likely contaminants.

The set of bait proteins evaluated in this study overlaps considerably with the Ho et al. (2002) effort. Figure 2.4 compares the retrieval of physical interactors for 13 gene products used as baits in both studies. Notably, in each case our approach identified at least as many or more of the previously–known binding partners of the

**Table 2.4c Potential New Interactors for a Test Set of HPM Tagged Proteins.** Samples were prepared and analyzed as described in 2.3. Column “Known interactors—Total” lists the number of physical/genetic interactions reported for the bait in the combined GRID/MIPS/YPD databases. “Known interactors—Recovered” represents the number of known physical/genetic interactors experimentally retrieved in this study. Partners marked “\*” are reported to interact physically as well as genetically. Column “Potential new interactors” contains all gene products identified by TAP–MudPIT, which are not listed as known interactors and are recovered in association with less than 20% of the baits analyzed ( $n = 22$ ).

Lte1p–HPM	48/12	5	—	Ade4p, Aro2p, Asc1p, Asn1p, Bcy1p, Bmh1p, Caf20p, Car2p, Cdc33p, Eft2p, Eft1p, Emi2p, Eno1p, Eno2p, Flo8p, Gad1p, Glk1p, Glt1p, Gly1p, Gpm1p, Gua1p, Hef3p, Hem1p, Hsp60p, Ilv6p, Lpd1p, Mkt1p, Nfs1p, Pbi2p, Pdc1p, Pkg1p, Pnc1p, Pro1p, Rax2p, Rib4p, Rpl23Ap, Rpl23Bp, Rps23Ap, Rps23Bp, Rps29Ap, Rps29Bp, Rps5p, Sec23p, Sec24p, Shm2p, Sod1p, Tpi1p, Tps3p, Vps1p, YDR348Cp, Yhb1p, YHL021Cp
Mad2p–HPM	11/10	2	—	Apl4p, Caf20p, Eno1p, Eno2p, Pdc1p, Pkg1p, Rrb1p, Trx2p, Ura7p, YOR283Wp
Mcd1p–HPM	17/8	3	Smc1p*, Trf4p	Bdf1p, Csm1p, Nuf2p, Not5p, Pom152p, Srm1p, Stu2p, YBL005W–Ap, YDR170W–Ap, YMR045Cp, YNL284C–Bp, YNL284C–Ap, YMR046Cp
Pds1p–HPM	4/1	1	Esp1p*	Azr1p, Ire1p, Mss1p, Swi3p
Pds5p–HPM	0/1	—	Mcd1p	Aro4p, Chs5p, Hal5p, Kem1p, Mss1p, Pbp1p
Pho2p–HPM	4/1	—	—	Rpl35Bp, Rpl35Ap, Rps5p, YBL005W–Ap, YDR170W–Ap, YDR261W–Bp, YGR161C–Cp, YJR026Wp, YOL103W–Ap, YML040Wp, YLR256W–Ap, YLR227W–Ap, YLR157C–Ap, YJR028Wp, YMR045Cp, YNL284C–Bp

13 bait proteins. For eight of the baits, Ho et al. (2002) identified more putative interacting partners. However, since Ho et al. (2002) utilized single–step affinity purification of overproduced bait protein, additional interactions revealed only in that study should be considered as tentative, pending verification by independent methods.

The second issue that we addressed was the feasibility of using a filtering approach to cull nonspecific contaminants from the list of proteins identified in each

**Table 2.5d Potential New Interactors for a Test Set of HPM Tagged Proteins.** Samples were prepared and analyzed as described in 2.3. Column “Known interactors—Total” lists the number of physical/genetic interactions reported for the bait in the combined GRID/MIPS/YPD databases. “Known interactors—Recovered” represents the number of known physical/genetic interactors experimentally retrieved in this study. Partners marked “\*” are reported to interact physically as well as genetically. Column “Potential new interactors” contains all gene products identified by TAP–MudPIT, which are not listed as known interactors and are recovered in association with less than 20% of the baits analyzed ( $n = 22$ ).

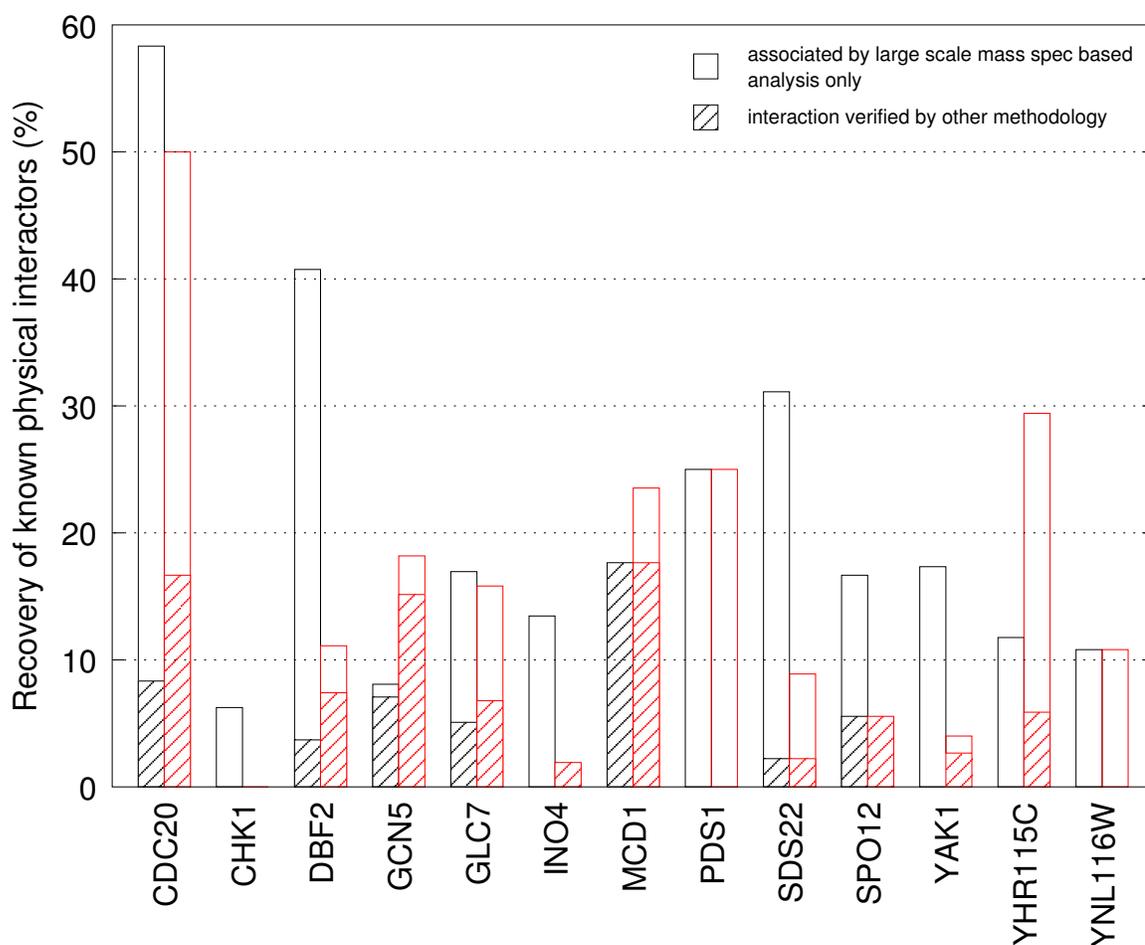
Pho4p-HPM	11/1	—	—	Ade16p, Ade3p, Ape3p, Aro2p, Aro4p, Asn1p, Bbc1p, Bcy1p, Cct4p, Cct8p, Cdc33p, Cdc73p, Chs5p, Dbp2p, Dbp3p, Dig1p, Eap1p, Eft2p, Eft1p, Fas1p, Fun12p, Glk1p, Glt1p, Gly1p, Gua1p, Hef3p, Hom3p, Hrb1p, Hsp60p, Imd4p, Kem1p, Kri1p, Lys21p, Lys20p, Myo5p, Nfs1p, Nma1p, Nop1p, Nop58p, Nsr1p, Pab1p, Rpa135p, Rpa34p, Rpl11Bp, Rpl11Ap, Rpl12Bp, Rpl12Ap, Rpl16Ap, Rpl23Ap, Rpl23Bp, Rpl24Ap, Rpl24Bp, Rpl26Bp, Rpl26Ap, Rpl29p, Rpl34Ap, Rpl34Bp, Rpl35Bp, Rpl35Ap, Rpl36Ap, Rpl36Bp, Rpl38p, Rpl43Bp, Rpl43Ap, Rpl5p, Rpl7Ap, Rpl7Bp, Rpp1Ap, Rpp2Ap, Rpp2Bp, Rps12p, Rps19Bp, Rps19Ap, Rps2p, Rps23Ap, Rps23Bp, Rps25Ap, Rps25Bp, Rps27Bp, Rps27Ap, Rps29Ap, Rps29Bp, Rps5p, Rps7Ap, Rps7Bp, Rps9Ap, Rps9Bp, Rrb1p, Rrp5p, Rsp5p, Sec23p, Ses1p, Shm2p, Sik1p, Sin3p, Snf1p, Srm1p, Ste11p, Ste50p, Stm1p, Tsr1p, Tub1p, Tub2p, Tub3p, Ura7p, Utp7p, Vip1p, Vps1p, Vrp1p, YAR075Wp, YBL101W–Bp, YGR161W–Bp, YFL002W–Ap, YDR210W–Bp, YDR034C–Dp, YCL019Wp, YDR261W–Bp, YGL068Wp, YHR121Wp, YIL137Cp, YMR045Cp, YMR050Cp, YMR237Wp, YNL054W–Bp
-----------	------	---	---	---

TAP–MudPIT experiment. The idea is that nonspecific proteins should show up in a high fraction of experiments, whereas specific interactors should only show up in one or a small number of experiments (depending upon the degree of functional relatedness of the tagged genes in the query set). We found that proteins that were identified in five or more TAP–MudPIT experiments tended to have a high codon

**Table 2.6e Potential New Interactors for a Test Set of HPM Tagged Proteins.** Samples were prepared and analyzed as described in 2.3. Column “Known interactors—Total” lists the number of physical/genetic interactions reported for the bait in the combined GRID/MIPS/YPD databases. “Known interactors—Recovered” represents the number of known physical/genetic interactors experimentally retrieved in this study. Partners marked “\*” are reported to interact physically as well as genetically. Column “Potential new interactors” contains all gene products identified by TAP–MudPIT, which are not listed as known interactors and are recovered in association with less than 20% of the baits analyzed ( $n = 22$ ).

Rtt102p–HPM	2/0	—	—	Aro4p, Arp7p, Arp9p, Fyv6p, Gsy2p, Hsl1p, Hta2p, Hta1p, Htl1p, Ldb7p, Nfi1p, Nfs1p, Npl6p, Rim1p, Rpl35Bp, Rpl35Ap, Rpl36Ap, Rpl36Bp, Rpl43Bp, Rpl43Ap, Rps2p, Rps29Bp, Rrb1p, Rsc1p, Rsc2p, Rsc3p, Rsc4p, Rsc58p, Rsc6p, Rsc8p, Rsc9p, Sfh1p, Snf12p, Snf2p, Snf5p, Snf6p, Sth1p, Swi1p, Swi3p, Taf14p, YFL049Wp, YHR097Cp
Sds22p–HPM	45/0	4	—	Nip100p, Ppz1p, Snf1p, Stu1p, Vps8p, YBL010Cp
Snf2p–HPM	164/13	11	—	Chs5p, Pab1p, Rpl11Bp, Rpl11Ap, Rpl16Ap, Rpl26Ap, Rpl26Bp, Rpl34Ap, Rpl34Bp, Rpl35Bp, Rpl35Ap, Rpl36Ap, Rpl36Bp, Rps12p, Rps2p, Rtt102p, Sth1p, Stm1p, YDL053Cp, YGR161C–Cp
Spo12p–HPM	18/5	1	—	Act1p, Ado1p, Ahp1p, Ald6p, Azr1p, Bmh1p, Cpr1p, Cys3p, Eft2p, Eft1p, Eno1p, Eno2p, Gpm1p, Hsp12p, Hsp42p, Hxk2p, Pdc1p, Pgi1p, Pkg1p, Rhr2p, Rps12p, Rps19Bp, Rps19Ap, Tif2p, Tif1p, Trp3p, Trx2p, Yhb1p, YNL134Cp, YPL257W–Bp
Yak1p–HPM	75/0	3	—	Caf20p, Glt1p, Gly1p, Hef3p, Kem1p, Nfs1p, Rib4p, YJL206Cp
YHR115Cp–HPM	17/0	8	—	Dbp3p, Gcd11p, Jip5p, Mkt1p, Sec16p, YBL101W–Bp, YLR410W–Bp, YGR161W–Bp, YFL002W–Ap, YDR210W–Bp, YDR034C–Dp, YCL019Wp, YJR026Wp, YOL103W–Ap, YML040Wp, YLR256W–Ap, YLR227W–Ap, YLR157C–Ap, YJR028Wp

adaptation index (Sharp and Li 1987), which is a rough measure of abundance (Jansen et al. 2003, data not shown). Based on this correlation, we automatically considered proteins found in more than five experiments to be probable contaminants. A similar filtering approach was employed by Gavin et al. (2002) and Ho



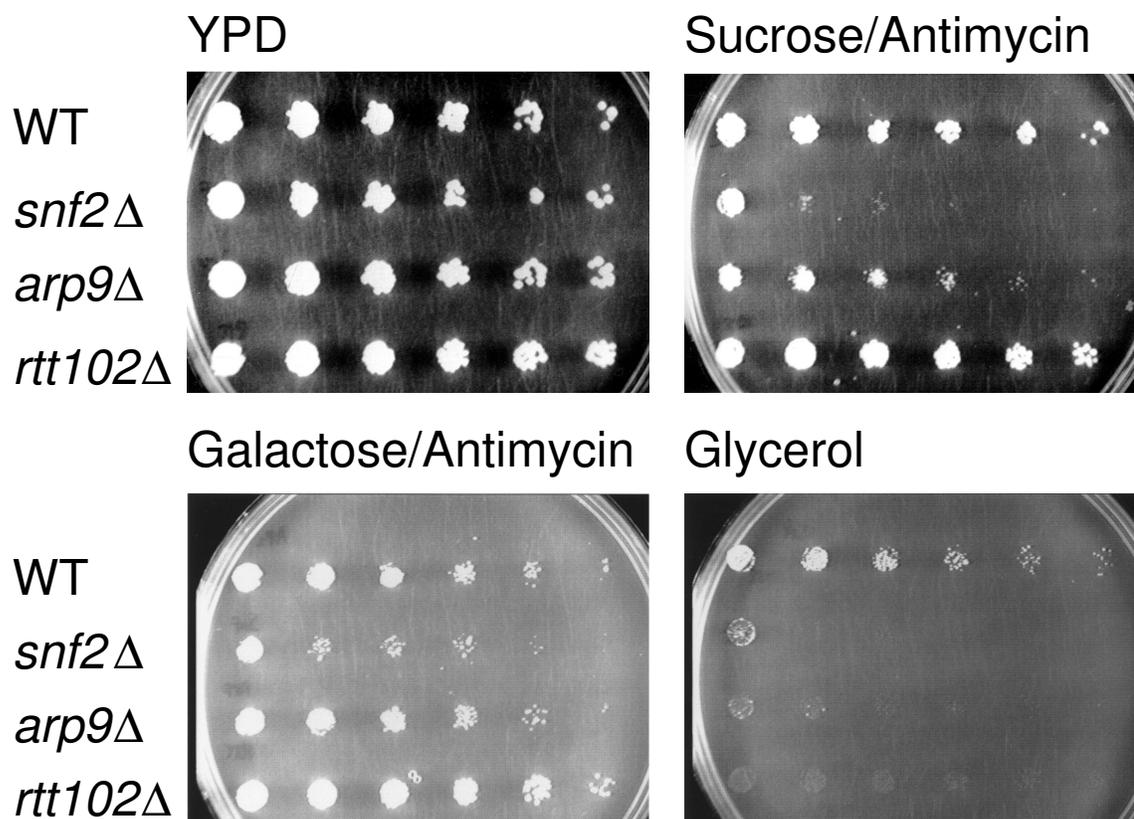
**Figure 2.4 Comparison with Ho et al. (2002).** Comparison of the data set presented here (red) with that of Ho et al. (2002) (black). ORFs listed were used as bait proteins in both studies. Bars represent the percentage of previously known interacting partners (as reported in MIPS CYGD, GRID and YPD) that was recovered in each experiment. Note that the set of interacting partners listed in these databases includes those reported by Ho et al. (2002). Empty bars represent percentage of gene products reported as interactors only by large scale mass spectrometric analysis whereas hatched bars represent interactions established or verified by other methods.

et al. (2002), but since their data-sets were much larger they were able to employ lower thresholds.

To showcase the possibility of identifying new potential interacting partners in any given TAP-MudPIT experiment, we analyzed in more detail our results for Snf2p-HPM. Snf2p is a subunit of the Swi/Snf complex and founding member of the

ATP-dependent family of chromatin remodeling factors (Fry and Peterson 2001). TAP-MudPIT analysis of Snf2p-HPM yielded eight of the nine known members of this complex (Cote et al. 1994; Henry et al. 1994; and Cairns et al. 1998; Arp7p, Arp9p, Snf5p, Snf6p, Swi1p, Swi3p, Snf12p, Taf14p; missing: Snf11p) as well as YFL049W, a protein of unknown function reported to copurify with Snf2p via its interaction with Snf5p (Gavin et al. 2002). A prominent Snf2p-HPM copurifying protein that was not commonly retrieved by other baits was Rtt102p, a protein of unknown function, whose inactivation results in a slight increase in Ty1 retrotransposon mobility (Scholes et al. 2001). To check whether the interaction of Snf2p with Rtt102p was reciprocal, we tagged the Rtt102p locus with sequences encoding the HPM epitope, and performed TAP-MudPIT analysis for Rtt102p-HPM. This experiment yielded all of the subunits of the Swi/Snf chromatin remodeling complex that copurified with Snf2p-HPM (see above), as well as all subunits of the RSC chromatin remodeling complex (Scholes et al. 2001; Npl6p, Rsc1p, Rsc2p, Rsc3p/Rsc30p, Rsc4p, Rsc58p, Rsc6p, Rsc8p, Rsc9p, Sfh1p, Sth1p). YFL049W copurified with Rtt102p-HPM as well as with Snf2p-HPM, further strengthening the case that it is a bona fide Swi/Snf component. These results suggest that Rtt102p, like Arp7p and Arp9p (Cairns et al. 1998; and Peterson et al. 1998), is specifically associated with the Swi/Snf and RSC chromatin remodeling complexes, and may be an integral component of both.

Knockouts of Swi/Snf complex members show reduced growth on sucrose/antimycin, galactose/antimycin and glycerol (Peterson et al. 1998). When tested for growth on these carbon sources, a *rtt102* $\Delta$  strain grew similar to wild type on glucose, sucrose/antimycin and galactose/antimycin, but exhibited a severe growth phenotype on glycerol (see fig. 2.5), further supporting a functional Rtt102p-Swi/Snf



**Figure 2.5 An *rtt102*Δ Strain Partially Recapitulates the Phenotype of Mutants Lacking Swi/Snf Complex.** “WT” is W303 *pep4*Δ::*TRP1*, *bar1*Δ::*HIS3* (RJD 415). “*snf2*Δ” is RJD 415, *snf2*Δ::*HIS3* (RJD 2566). “*arp9*Δ” is RJD 415, *arp9*Δ::*HIS3* (RJD 2567). “*rtt102*Δ” is RJD 415, *rtt102*Δ::*HIS3* (RJD 2568). Media compositions are: 1% yeast extract, 2% peptone and 2% final concentration of glucose, sucrose, galactose or glycerol. Sucrose and galactose containing media were supplemented with 1 μg/ml antimycin.

connection.

## 2.5 Discussion

A key goal of proteomics research is to identify and characterize protein interaction networks. Several approaches have been taken to achieve this goal, including genome-wide two-hybrid analyses and protein chip-based approaches (Uetz et al. 2000; Ito et al. 2001; and Zhu et al. 2001). A limitation of both of these methods

is that they primarily reveal binary interactions. Large-scale mass spectrometric analyses of affinity-purified protein complexes have been reported by two different groups (Gavin et al. 2002; and Ho et al. 2002). Whereas this approach bypasses some of the key limitations of two-hybrid and protein chip assays, the efforts reported so far were based on gel separation of purified proteins, which both greatly increased the number of mass spectrometry runs required to analyze each bait and limited the dynamic range to proteins that could be stained and visualized on the same gel. Indeed, both efforts were carried out in an industrial context that can not be readily adapted to a conventional molecular/cellular biology laboratory. We believe this is an important issue, because unlike the genomic sequence, the protein interactions that exist in a cell or organism are not a finite and bounded set that can be determined as a complete “reference” knowledge set. Rather, their most important feature is that they change as a function of intracellular and extracellular signals and learning how they change is essential for probing the cellular processes of interest. Thus, to characterize fully the protein interaction networks in a cell and their dynamic changes over time, it will be necessary to perform multiple analyses under different conditions and in different genotypes. In this sense, mass spectrometry-based proteomics resembles microarray-based transcriptomics. This fact underscores the need for simple, reproducible, rapid, portable (i. e. can be performed outside of a specialized mass spectrometry environment), yet powerful methods for exploring protein interaction networks.

We show here that a combination of double affinity purification and multidimensional capillary chromatography in line to mass spectrometry (TAP-MudPIT) fulfills these criteria. TAP-MudPIT can be applied to rapidly identify interacting proteins for any given bait in a single mass spectrometry analysis. Using this

approach, a single investigator working with a single mass spectrometer and performing the complete protocol from affinity purification to data analysis can readily screen 20 samples per month (i. e. 20 different baits or one bait evaluated under 20 different conditions). Thus, it is feasible for a single investigator to perform, in a reasonable time frame, a thorough analysis of a focused collection of baits that define a particular organelle, pathway, or process.

It should also be noted that in addition to protein identification, the TAP–MudPIT approach enables the parallel analysis of posttranslational modifications (Cheeseman et al. 2002).

Although an exhaustive analysis of every one of the 22 TAP–MudPIT experiments that we performed (21 from the original collection of baits plus Rtt102p) is beyond the scope of this paper, we wish to highlight several interesting points. First, our analysis of Swi2p/Snf2p identified a new interacting partner, Rtt102p, which is remarkable given the large body of work that has already been performed on this extensively characterized protein and its interacting partners. Second, we uncovered TRF4 as a candidate partner of the cohesin Mcd1p/Scc1p. Trf4p was originally reported to function as an alternative DNA polymerase that mediates sister chromatid cohesion (Wang et al. 2000), but this proposal has been the subject of controversy following the report that Trf4p can catalyze polymerization of poly(A) tails on mRNA transcripts (Saitoh et al. 2002). Third, Bub3p was found as a Cdc20p–associated protein and Mcd1p/Scc1p was found as a Pds5p–associated protein. Although these pairs of proteins were already known to function together in mitotic checkpoint signaling and sister chromatid cohesion, respectively, a physical association of the yeast proteins has not been reported. Finally, in addition to Trf4p, Mcd1p/Scc1p retrieved the Csm1p subunit of monopolin and the Nuf2p

subunit of the Tid3p/Nuf2p/Spc24p/Spc25p centromere-binding complex (Janke et al. 2001). Both interactions are excellent candidates to subserve a role in chromosome segregation given the known functions of the proteins involved.

Analysis of Rtt102p, identified here as a Swi2p/Snf2p interactor, illustrated the power of this system for making fast and simple first-order interaction validation. This was accomplished by a reciprocity test, in which Rtt102p was shown to specifically retrieve Swi2p and other known components of the Swi/Snf complex. Because this is an independent determination, it provides a more convincing confirmation for an interaction than a mere repetition of the initial measurement. The experiment also illustrates how TAP-MudPIT can be used for directed interaction “walks” (Seol et al. 2001), in this case showing that Rtt102p also interacts with, or is a component of, the RSC chromatin remodeling complex.

Whereas TAP-MudPIT is sufficiently robust to be applied in a nonspecialized environment, two substantial problems remain to be addressed. First, the interpretation of the data that is generated would benefit from improvement. The combination of `2to3` (Sadygov et al. 2002), `SEQUEST` (Eng et al. 1994) and `DTASelect` (Tabb et al. 2002) enables analysis and display of raw mass spectrometrical data. What are missing, however, are tools that simplify interpretation of the massive amount of data generated by the analysis of even a protein interaction network of even modest size. In particular, separating good candidates for novel interaction partners from the contaminating chaff is a major challenge. We followed the approach used by Gavin et al. (2002) and Ho et al. (2002), by excluding from consideration any protein that was found associated with more than 20% of the baits analyzed (the comparable thresholds were 3% in Ho et al. (2002), 3.5% in Gavin et al. 2002). When applied to the proteins found in all three independent Gcn5p-HPM TAP-

MudPIT analyses shown in fig. 2.3, our filter threshold retains only the previously known interactors and the potential new Gcn5p-interacting protein YCR082Wp. A problem with excluding candidates by this criterion is that we were not using an unbiased reference data set. Since the proteins that we analyzed are all involved in either transcription or mitosis, it is possible that some true interacting proteins were improperly excluded.

The complete data set contains a total of 464 potential interactions passing the requirement of being associated with less than 20% of the baits analyzed. However, this subset includes ribosomal, cytoskeletal and other proteins, that, due to their abundance, have a high probability of being contaminants. Discarding Ty-Element related proteins and applying a filter that allows a maximum CAI of 0.6 eliminates these problematic candidates and reduces the number of potential new interaction partners identified to 279.

In addition to “post hoc” approaches, honing the purification protocol and making it more stringent may lessen the problem posed by contaminating proteins. However, this comes at the possible expense of disrupting specific interactions. When analyzing a single bait under varying conditions, optimizing the purification may greatly improve the specificity of the purification, but optimization becomes a daunting task when dealing with multiple baits.

The second major problem arises from the databases used to biologically annotate the gene products identified by MudPIT. Given the amount of data produced by a MudPIT experiment, machine readability of data bases is of great value. Unfortunately, of the data bases used in this study only the regularly updated data in SGD and MIPS CYGD are readily accessible in an automated manner (`ftp`). GRID data can be manually downloaded in a tab delimited file, but YPD does not

allow any such access, and thus requires manual merging of its annotation data into a computationally annotated data set.

As more and more large-scale analyses are performed, an issue that looms large for the future is how to evaluate the quality of the datasets. Even relatively small-scale analyses like the one reported here are prone to produce false positives (e. g., the large number of ribosomal proteins classified as potential interactors for Pho4p in table 2.2). As a specific example of this problem, consider Adh1p (alcohol dehydrogenase). Adh1p is annotated in YPD as a protein in complex with Gcn5p and Snf2p, because Adh1p was reported to copurify with these proteins in TAP experiments using Spt15p and Med2p (Gcn5p) or Enp1p (Snf2p) as bait proteins (Gavin et al. 2002). However, given that we found Adh1p associated with 86 % of our baits, it is most likely a common contaminant that nevertheless cleared the filter imposed by Gavin et al. (2002). An important challenge is to generate databases that express the likelihood that a protein-protein interaction is relevant based on the number of independent analyses (and methods) upon which the conclusion is based.

In conclusion, we report the application of TAP-MudPIT—tandem-affinity purification coupled with multidimensional capillary chromatography in line to mass spectrometry—to identify binding partners for a set of 22 budding yeast proteins involved in gene regulation or progression through mitosis. Our analysis uncovered 102 previously known and 279 potential physical interactions. TAP-MudPIT is simple, rapid, reproducible and can be carried out in a traditional cell biology laboratory. The simplicity and power of this method enables a depth of analysis that will facilitate thorough characterization of protein interaction networks.

## 2.6 Acknowledgements

This work was supported by Howard Hughes Medical Institute (R. J. D.), Beckman Institute (B. J. W.), National Institute of Health grant RR11823 to J. R. Y. and W. H. M. and Department of Energy grant DE-FG03-01ER63231 to B. J. W. and R. J. D. We thank Erin O'Shea (UCSF) for the gift of RJD 2067. J. G. thanks all members of the Yates lab (The Scripps Research Institute) and Eberhard Dürre (Sidney Kimmel Cancer Center) for introduction into mass spectrometry and SEQUEST and D. A. Gold, E. E. Griffin and A. Mah (all Caltech) for critical reading of the manuscript.

## 2.7 Note Added in Proof

The reproducibility of MudPIT applied to whole cell extracts was recently reported by Washburn et al. (2003).

## 2.8 References

- Breitkreutz, B. J., Stark, C. and Tyers, M. (2003). The GRID: The General Repository for Interaction Datasets. *Genome Biol*, 4(3):R23.
- Cairns, B. R., Erdjument-Bromage, H., Tempst, P., Winston, F. and Kornberg, R. D. (1998). Two actin-related proteins are shared functional components of the chromatin-remodeling complexes RSC and SWI/SNF. *Mol Cell*, 2(5):639–51.

- Cheeseman, I. M., Anderson, S., Jwa, M., Green, E. M. and Kang, J. et al. (2002). Phospho-regulation of kinetochore-microtubule attachments by the Aurora kinase Ipl1p. *Cell*, 111(2):163–72.
- Cheeseman, I. M., Brew, C., Wolyniak, M., Desai, A. and Anderson, S. et al. (2001). Implication of a novel multiprotein Dam1p complex in outer kinetochore function. *J Cell Biol*, 155(7):1137–45.
- Cherry, J. M., Adler, C., Ball, C., Chervitz, S. A. and Dwight, S. S. et al. (1998). SGD: *Saccharomyces* genome database. *Nucleic Acids Res*, 26(1):73–9.
- Cordingley, M. G., Callahan, P. L., Sardana, V. V., Garsky, V. M. and Colonno, R. J. (1990). Substrate requirements of human rhinovirus 3C protease for peptide cleavage *in vitro*. *J Biol Chem*, 265(16):9062–5.
- Cote, J., Quinn, J., Workman, J. L. and Peterson, C. L. (1994). Stimulation of GAL4 derivative binding to nucleosomal DNA by the yeast SWI/SNF complex. *Science*, 265(5168):53–60.
- Deckert, J. and Struhl, K. (2001). Histone acetylation at promoters is differentially affected by specific activators and repressors. *Mol Cell Biol*, 21(8):2726–35.
- Eberharter, A., Sterner, D. E., Schieltz, D., Hassan, A. and Yates III, J. R. et al. (1999). The ADA complex is a distinct histone acetyltransferase complex in *Saccharomyces cerevisiae*. *Mol Cell Biol*, 19(10):6621–31.
- Eng, J. K., McCormack, A. L. and Yates III, J. R. (1994). An approach to correlate tandem mass-spectral data of peptides with amino-acid sequences in a protein database. *J Am Soc Mass Spectr*, 5(11):976–989.
- Estruch, F. and Carlson, M. (1993). Two homologous zinc finger genes identified by multicopy suppression in a SNF1 protein kinase mutant of *Saccharomyces cerevisiae*. *Mol Cell Biol*, 13(7):3872–81.

- Evan, G. I., Lewis, G. K., Ramsay, G. and Bishop, J. M. (1985). Isolation of monoclonal antibodies specific for human c-myc proto-oncogene product. *Mol Cell Biol*, 5(12):3610–6.
- Fry, C. J. and Peterson, C. L. (2001). Chromatin remodeling enzymes: Who's on first?. *Curr Biol*, 11(5):R185–97.
- Garrels, J. I. (1996). YPD—A database for the proteins of *Saccharomyces cerevisiae*. *Nucleic Acids Res*, 24(1):46–9.
- Gavin, A. C., Bosche, M., Krause, R., Grandi, P. and Marzioch, M. et al. (2002). Functional organization of the yeast proteome by systematic analysis of protein complexes. *Nature*, 415(6868):141–7.
- Giaever, G., Chu, A. M., Ni, L., Connelly, C. and Riles, L. et al. (2002). Functional profiling of the *Saccharomyces cerevisiae* genome. *Nature*, 418(6896):387–91.
- Grant, P. A., Schieltz, D., Pray-Grant, M. G., Yates III, J. R. and Workman, J. L. (1998). The ATM-related cofactor Tra1 is a component of the purified SAGA complex. *Mol Cell*, 2(6):863–7.
- Graumann, J., Dunipace, L. A., Seol, J. H., McDonald, W. H. and Yates III, J. R. et al. (2004). Applicability of tandem affinity purification MudPIT to pathway proteomics in yeast. *Mol Cell Proteomics*, 3(3):226–37.
- Henry, N. L., Campbell, A. M., Feaver, W. J., Poon, D. and Weil, P. A. et al. (1994). TFIIF–TAF–RNA polymerase II connection. *Genes Dev*, 8(23):2868–78.
- Ho, Y., Gruhler, A., Heilbut, A., Bader, G. D. and Moore, L. et al. (2002). Systematic identification of protein complexes in *Saccharomyces cerevisiae* by mass spectrometry. *Nature*, 415(6868):180–3.
- Ito, H., Fukuda, Y., Murata, K. and Kimura, A. (1983). Transformation of intact yeast cells treated with alkali cations. *J Bacteriol*, 153(1):163–8.

- Ito, T., Chiba, T., Ozawa, R., Yoshida, M. and Hattori, M. et al. (2001). A comprehensive two-hybrid analysis to explore the yeast protein interactome. *Proc Natl Acad Sci USA*, 98(8):4569–74.
- Janke, C., Ortiz, J., Lechner, J., Shevchenko, A. and Shevchenko, A. et al. (2001). The budding yeast proteins Spc24p and Spc25p interact with Ndc80p and Nuf2p at the kinetochore and are important for kinetochore clustering and checkpoint control. *EMBO J*, 20(4):777–91.
- Jansen, R., Bussemaker, H. J. and Gerstein, M. (2003). Revisiting the codon adaptation index from a whole-genome perspective: Analyzing the relationship between gene expression and codon occurrence in yeast using a variety of models. *Nucleic Acids Res*, 31(8):2242–51.
- Link, A. J., Eng, J., Schieltz, D. M., Carmack, E. and Mize, G. J. et al. (1999). Direct analysis of protein complexes using mass spectrometry. *Nat Biotechnol*, 17(7):676–82.
- Longtine, M. S., McKenzie III, A., Demarini, D. J., Shah, N. G. and Wach, A. et al. (1998). Additional modules for versatile and economical PCR-based gene deletion and modification in *Saccharomyces cerevisiae*. *Yeast*, 14(10):953–61.
- McCormack, A. L., Schieltz, D. M., Goode, B., Yang, S. and Barnes, G. et al. (1997). Direct analysis and identification of proteins in mixtures by LC/MS/MS and database searching at the low-femtomole level. *Anal Chem*, 69(4):767–76.
- McDonald, W. H., Ohi, R., Miyamoto, D. T., Mitchison, T. J. and Yates III, J. R. (2002). Comparison of three directly coupled hplc ms/ms strategies for identification of proteins from complex mixtures: Single-dimension lc-ms/ms, 2-phase mudpit, and 3-phase mudpit. *Int J Mass Spectrom*, 219(1):245–251.

- Mewes, H. W., Albermann, K., Heumann, K., Liebl, S. and Pfeiffer, F. (1997). MIPS: A database for protein sequences, homology data and yeast genome information. *Nucleic Acids Res*, 25(1):28–30.
- Mewes, H. W., Frishman, D., Guldener, U., Mannhaupt, G. and Mayer, K. et al. (2002). MIPS: A database for genomes and protein sequences. *Nucleic Acids Res*, 30(1):31–4.
- Peng, J., Schwartz, D., Elias, J. E., Thoreen, C. C. and Cheng, D. et al. (2003). A proteomics approach to understanding protein ubiquitination. *Nat Biotechnol*, 21(8):921–6.
- Peterson, C. L., Zhao, Y. and Chait, B. T. (1998). Subunits of the yeast SWI/SNF complex are members of the actin-related protein (ARP) family. *J Biol Chem*, 273(37):23641–4.
- Pollard, K. J. and Peterson, C. L. (1997). Role for ADA/GCN5 products in antagonizing chromatin-mediated transcriptional repression. *Mol Cell Biol*, 17(11):6212–22.
- Pray-Grant, M. G., Schieltz, D., McMahon, S. J., Wood, J. M. and Kennedy, E. L. et al. (2002). The novel SLIK histone acetyltransferase complex functions in the yeast retrograde response pathway. *Mol Cell Biol*, 22(24):8774–86.
- Rigaut, G., Shevchenko, A., Rutz, B., Wilm, M. and Mann, M. et al. (1999). A generic protein purification method for protein complex characterization and proteome exploration. *Nat Biotechnol*, 17(10):1030–2.
- Sadygov, R. G., Eng, J., Durr, E., Saraf, A. and McDonald, H. et al. (2002). Code developments to improve the efficiency of automated MS/MS spectra interpretation. *J Proteome Res*, 1(3):211–5.

- Saitoh, S., Chabes, A., McDonald, W. H., Thelander, L. and Yates III, J. R. et al. (2002). Cid13 is a cytoplasmic poly(A) polymerase that regulates ribonucleotide reductase mRNA. *Cell*, 109(5):563–73.
- Sanders, S. L., Jennings, J., Canutescu, A., Link, A. J. and Weil, P. A. (2002). Proteomics of the eukaryotic transcription machinery: identification of proteins associated with components of yeast TFIID by multidimensional mass spectrometry. *Mol Cell Biol*, 22(13):4723–38.
- Scholes, D. T., Banerjee, M., Bowen, B. and Curcio, M. J. (2001). Multiple regulators of Ty1 transposition in *Saccharomyces cerevisiae* have conserved roles in genome maintenance. *Genetics*, 159(4):1449–65.
- Seol, J. H., Shevchenko, A., Shevchenko, A. and Deshaies, R. J. (2001). Skp1 forms multiple protein complexes, including RAVE, a regulator of V-ATPase assembly. *Nat Cell Biol*, 3(4):384–91.
- Sharp, P. M. and Li, W. H. (1987). The codon adaptation index—a measure of directional synonymous codon usage bias, and its potential applications. *Nucleic Acids Res*, 15(3):1281–95.
- Tabb, D. L., McDonald, W. H. and Yates III, J. R. (2002). DTASelect and Contrast: Tools for assembling and comparing protein identifications from shotgun proteomics. *J Proteome Res*, 1(1):21–6.
- Treger, J. M., Magee, T. R. and McEntee, K. (1998). Functional analysis of the stress response element and its role in the multistress response of *Saccharomyces cerevisiae*. *Biochem Biophys Res Commun*, 243(1):13–9.
- Uetz, P., Giot, L., Cagney, G., Mansfield, T. A. and Judson, R. S. et al. (2000). A comprehensive analysis of protein–protein interactions in *Saccharomyces cerevisiae*. *Nature*, 403(6770):623–7.

- Walker, P. A., Leong, L. E., Ng, P. W., Tan, S. H. and Waller, S. et al. (1994). Efficient and rapid affinity purification of proteins using recombinant fusion proteases. *Biotechnology (N Y)*, 12(6):601–5.
- Wang, Z., Castano, I. B., De Las Penas, A., Adams, C. and Christman, M. F. (2000). Pol kappa: A DNA polymerase required for sister chromatid cohesion. *Science*, 289(5480):774–9.
- Washburn, M. P., Ulaszek, R. R. and Yates III., J. R. (2003). Reproducibility of quantitative proteomic analyses of complex biological mixtures by multidimensional protein identification technology. *Anal Chem*, 75(19):5054–61.
- Winzeler, E. A., Shoemaker, D. D., Astromoff, A., Liang, H. and Anderson, K. et al. (1999). Functional characterization of the *S. cerevisiae* genome by gene deletion and parallel analysis. *Science*, 285(5429):901–6.
- Zhu, H., Bilgin, M., Bangham, R., Hall, D. and Casamayor, A. et al. (2001). Global analysis of protein activities using proteome chips. *Science*, 293(5537):2101–5.

# Appendices

# A Multiubiquitin Chain Receptors Define a Layer of Substrate Selectivity in the Ubiquitin–Proteasome System

This chapter constitutes a further example of the use of multidimensional protein identification technology for the analysis of moderately complex polypeptide mixtures as resulting from affinity purification of protein complexes, in this case the proteasome from *Saccharomyces cerevisiae*. J. G.'s contribution to this work encompasses advice on MudPIT compatible experiment design, MudPIT and data analysis as well as data presentation using RAYzer (see section 2.3.9). The copyright for the presented material, published as

Verma, R., Oania, R., Graumann, J. and Deshaies, R. J. (2004). Multiubiquitin chain receptors define a layer of substrate selectivity in the ubiquitin–proteasome system. *Cell*, 118(1):99–110.

is held by Cell Press and is used with permission. Supplementary material referred to can be found at <http://download.cell.com/supplementarydata/cell/118/1/99/DC1/index.htm>.

## A.1 Abstract

Recruitment of ubiquitinated proteins to the 26S proteasome lies at the heart of the ubiquitin–proteasome system (UPS). Genetic studies suggest a role for the multiubiquitin chain binding proteins (MCBPs) Rad23 and Rpn10 in recruitment, but biochemical studies implicate the Rpt5 ATPase. We addressed this issue by analyzing degradation of the ubiquitinated Cdk inhibitor Sic1 (UbSic1) *in vitro*. Mutant

*rpn10* $\Delta$  and *rad23* $\Delta$  proteasomes failed to bind or degrade UbSic1. Although Rpn10 or Rad23 restored UbSic1 recruitment to either mutant, rescue of degradation by Rad23 uncovered a requirement for the VWA domain of Rpn10. *In vivo* analyses confirmed that Rad23 and the multiubiquitin binding domain of Rpn10 contribute to Sic1 degradation. Turnover studies of multiple UPS substrates uncovered an unexpected degree of specificity in their requirements for MCBPs. We propose that recruitment of substrates to the proteasome by MCBPs provides an additional layer of substrate selectivity in the UPS.

## A.2 Introduction

Proteolysis by the UPS is required for the maintenance of cellular homeostasis (Hershko and Ciechanover 1998; and Pickart and Cohen 2004). Proteins destined to be degraded by the proteasome are marked for elimination by the covalent attachment of ubiquitin (Ub). The C terminus of Ub is linked by an isopeptide bond to the  $\alpha$  amino group of a lysine residue in the substrate. A multiubiquitin (multiUb) chain is formed by attachment of successive Ubs, primarily to the Lys48 residue of the distal-most Ub tethered to the substrate. Once the multiUb chain contains at least four Ubs, it can bind the proteasome and serve as a signal for degradation (Chau et al. 1989; and Thrower et al. 2000). Following specific binding, the ubiquitinated substrate is unfolded, deubiquitinated, and translocated by the 19S regulatory “cap” of the 26S proteasome into the 20S protease core, where it is proteolyzed to peptide remnants (Hershko and Ciechanover 1998; Verma et al. 2002; and Yao and Cohen 2002).

Recognition of multiUb chains by the proteasome is central to Ub-selective degradation. The receptor(s) that mediates this process has thus been sought intensively. Over the past decade, three different classes of proteins have been advanced as candidate receptors that link Ub conjugates to the proteasome for degradation. Rpn10 was the first protein that was shown to bind selectively to polyubiquitin (polyUb) chains. Because Rpn10 is a bona fide stoichiometric subunit of the 26S proteasome, it was proposed that Rpn10 is the multiUb chain receptor (Deveraux et al. 1994). However, even though proteasomal proteolysis is essential, Rpn10 is dispensable for life in budding yeast (Fu et al. 1998; and Van Nocker et al. 1996). Indeed, only one UPS substrate, Ub-proline- $\beta$ -galactosidase (Ub-Pro- $\beta$ -gal, or the related substrate Ub<sup>V76</sup>-Valine- $\beta$ -gal), has been shown to be stabilized in *rpn10* $\Delta$  cells, and, paradoxically, Ub-Pro- $\beta$ -gal turnover does not require the Ub binding domain of Rpn10 (Fu et al. 1998). Additionally, Rpn10 assembled into 26S proteasomes does not crosslink to a chemically reactive tetraubiquitin chain (Lam et al. 2002), and recombinant Rpn10 inhibits proteolysis in frog extracts (Deveraux et al. 1995). Taken together, these observations raised doubts as to whether Rpn10 functioned in the context of the 26S proteasome to recruit ubiquitinated substrates for degradation (Pickart and Cohen 2004).

Attention was thus diverted to a second group of proteins exemplified by Rad23 and Dsk2. These proteins each contain a Ub-like domain (UbL) that binds the proteasome (Elsasser et al. 2002; Saeki et al. 2002b; and Schaubert et al. 1998) and UBA domains that bind multiUb chains (Rao and Sastry 2002; and Wilkinson et al. 2001). However, the role of Rad23 and Dsk2 in guiding multiUb chain-bearing substrates to the proteasome is equally controversial. Budding and fission yeast *rad23* $\Delta$  and *dsk2* $\Delta$  mutants accumulate reporter substrates and high molecular

weight Ub conjugates, supporting a positive role for these proteins in the UPS (Chen and Madura 2002; Funakoshi et al. 2002; Rao and Sastry 2002; Saeki et al. 2002a; and Wilkinson et al. 2001). However, *rad23Δrpn10Δ* double mutants are proficient in bulk turnover of short-lived proteins (Lambertson et al. 1999). Additionally, overexpression of Dsk2 or Rad23 in mammalian and yeast cells typically inhibits substrate turnover by the 26S proteasome (Kleijnen et al. 2000; and Ortolan et al. 2000) but can apparently stimulate turnover in some contexts (Funakoshi et al. 2002). Indeed, a key limitation to the argument that Rad23 and Dsk2 serve as substrate receptors is that such a role has never been directly demonstrated. In the only direct test so far of the hypothesis that Rad23 acts as a receptor that links substrates to the proteasome, it was shown that recombinant Rad23 actually inhibits substrate turnover by purified 26S proteasome *in vitro* (Raasi and Pickart 2003). Similar results have been reported for Rpn10 (Deveraux et al. 1995). In light of the lack of conclusive, direct evidence that Rad23 serves as a receptor to guide ubiquitinated substrates to the proteasome, other functions have been sought for this protein. Bioinformatics has revealed that the UBA domain is conserved in a number of enzymes of the UPS, including E2s, E3s, and Ub proteases (Ubps) (Hofmann and Bucher 1996). Some members of the latter class, such as Ubp14, bind polyUb chains and cleave them (Amerik et al. 1997). Although binding of Rad23 to Ub conjugates did not cause cleavage of the Ub chain, it did inhibit Ub chain assembly (Ortolan et al. 2000) as well as disassembly (Hartmann-Petersen et al. 2003; and Raasi and Pickart 2003), suggesting that Rad23 may promote degradation by serving as a shield that retards deubiquitination of substrates that are en route to the proteasome (Pickart and Cohen 2004).

To complicate matters further, a third candidate receptor (S6'/Rpt5) has re-

cently been identified based on UV crosslinking of a tetra-Ub chain to purified 26S proteasomes (Lam et al. 2002). Rpt5 is a member of the AAA ATPase family of enzymes, with an as yet undefined multiUb chain binding domain. A putative receptor function for Rpt5 is appealing based on precedent from other systems. The related AAA ATPases of bacterial compartmentalized proteases contribute to enzyme specificity by directly binding to short peptide degrons within substrates (Flynn et al. 2003), and the mammalian AAA ATPase p97/Cdc48 promotes turnover of I $\kappa$ B by binding directly to multiubiquitin chains (Dai and Li 2001). However, a functional role for S6'/Rpt5 in recruiting ubiquitinated substrates to the proteasome has not been validated yet by either biochemical or genetic studies.

The studies summarized above highlight several key unresolved issues. For example, what is the nature of the primary gateway through which proteins targeted by the numerous cellular ubiquitin ligases are recognized by the proteasome and sent to meet their final fate? Is there a single gateway (e. g., Rpt5) or multiple gateways (e. g., Rad23, Rpn10, and other Ub binding proteins)? If the latter, do the gateways function in parallel or in series? Are all ubiquitinated substrates processed in the same manner, or is there an additional layer of substrate specificity downstream of the ubiquitin ligases? In this work, we employ a combination of *in vitro* reconstitution and *in vivo* turnover assays to address these questions.

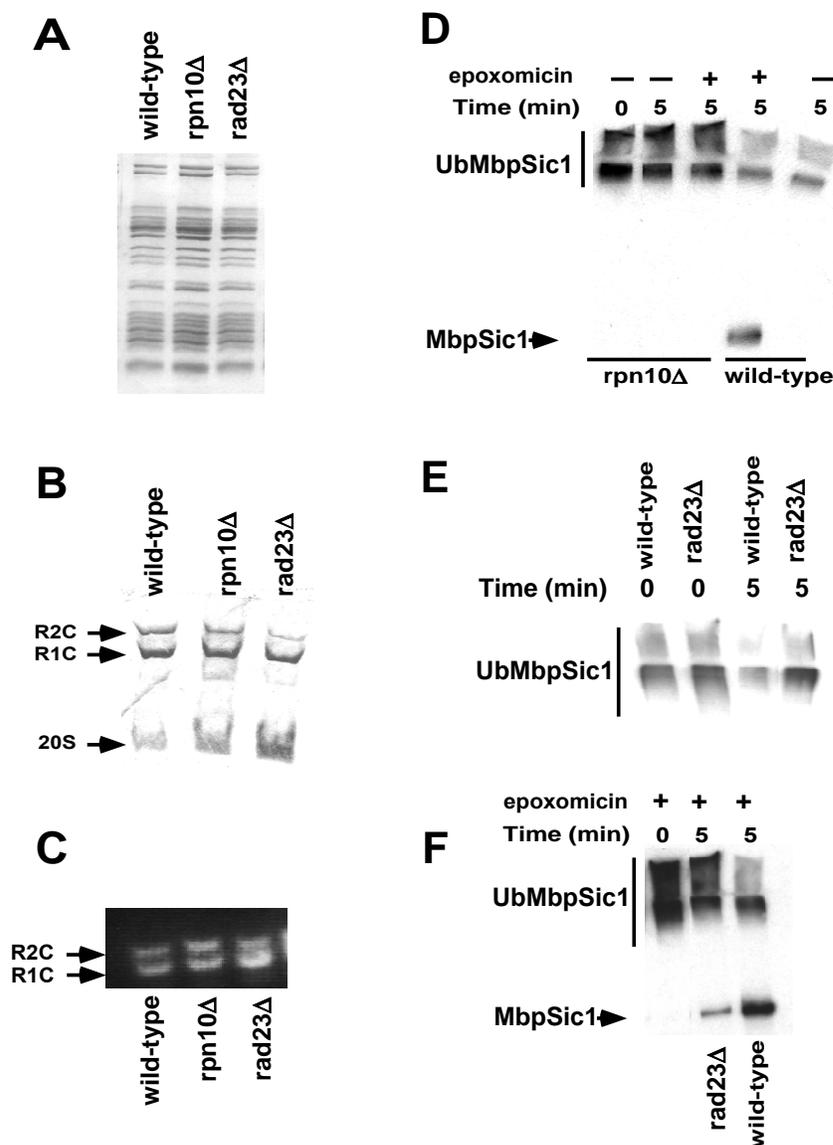
## A.3 Results and Discussion

### A.3.1 Intact 26S Proteasomes Can Be Isolated from *rpn10* $\Delta$ and *rad23* $\Delta$ Mutants

To address the molecular basis for substrate recruitment by the 26S proteasome, we employed a system that recapitulates the selective ubiquitination and degradation of budding yeast S-Cdk inhibitor Sic1 using purified components (Verma et al. 2001). The chromosomal locus that encodes PRE1, a subunit of the 20S core, was tagged with the Flag epitope in wild-type, *rpn10* $\Delta$ , and *rad23* $\Delta$  mutant cells. 26S proteasomes were purified by single-step affinity chromatography on anti-Flag beads as described (Verma et al. 2000, also see section A.4). The data in Figure A.1A demonstrate that subunit composition, as visualized by SDS-PAGE, was essentially the same for 26S proteasomes purified from wild-type and mutant cells. This result was corroborated by MudPIT mass spec analysis ((Link et al. 1999); see Supplemental Table S2 at <http://www.cell.com/cgi/content/full/118/1/99/DC1>). Assembly was also normal as determined by Coomassie blue staining (Figure A.1B) and in-gel peptidase assay of purified proteasomes separated on native gels (Figure A.1C). Some decrease in the doubly capped particle (R2C) with concomitant increase in 20S was seen for the mutants, particularly *rad23* $\Delta$ .

### A.3.2 *rpn10* $\Delta$ and *rad23* $\Delta$ 26S Proteasomes Are Defective at Degrading Ubiquitinated Sic1

The protein degradation activity of the wild-type and mutant 26S proteasomes was assessed by incubation with a ubiquitinated maltose binding protein-Sic1 chimera



**Figure A.1 Structural and Functional Characterization of 26S Proteasomes Isolated from *rpn10Δ* and *rad23Δ* Mutants by Affinity Chromatography.** Extracts from wild-type and mutant yeast strains expressing PRE1FH (Supplemental Table S1) were incubated with anti-Flag M2 resin. Bound proteins were eluted with Flag peptide and analyzed by (A) SDS-PAGE and Coomassie blue staining; (B) native gel (nondenaturing) electrophoresis and Coomassie blue staining; or (C) nondenaturing electrophoresis and incubation with a fluorogenic peptide substrate (Verma et al. 2000). (D) *rpn10Δ* 26S are completely defective in the degradation and deubiquitination of UbMbpSic1. UbMbpSic1 was incubated at 30°C with 26S proteasomes isolated from either wild-type or *rpn10Δ* cells. Degradation reactions (lanes 2 and 5) were set up and analyzed by SDS-PAGE followed by immunoblotting with anti-Sic1 polyclonal antibody as described in section A.4. For assessing deubiquitination (lanes 3 and 4), the 26S proteasome preparations were preincubated with 100 μM epoxomicin for 45 min at 30°C before incubation with UbMbpSic1. 26S proteasomes isolated from *rad23Δ* mutants were partially defective in (E) degradation and (F) deubiquitination of UbMbpSic1. Analysis was performed as described for *rpn10Δ* proteasomes in (D).

(UbMbpSic1), which was prepared as described (Seol et al. 1999). Degradation was monitored by loss of high molecular weight Sic1, which typically migrates at the top of a 7.5% gel and is also observed in the stacker (Verma et al. 2000, 2001). Whereas wild-type 26S proteasomes degraded UbMbpSic1 rapidly, *rpn10* $\Delta$  26S proteasomes were completely defective (compare lanes 2 and 5 with lane 1, Figure A.1D), and *rad23* $\Delta$  proteasomes were largely but not completely defective (Figure A.1E). The strength of these defects was surprising given the reported mild phenotype of *rpn10* $\Delta$  mutants (Fu et al. 1998; and Van Nocker et al. 1996). To confirm these unexpected results by a different method, we also evaluated whether *rpn10* $\Delta$  and *rad23* $\Delta$  proteasomes were deficient in Rpn11-dependent substrate deubiquitination (DUB) activity (Verma et al. 2002; and Yao and Cohen 2002). A block in Rpn11 DUB activity leads to a block in degradation. Rpn11 activity is assayed in the presence of the 20S core protease inhibitor epoxomicin, which results in conversion of ubiquitinated substrate to an unmodified protein (MbpSic1; lane 4, Figure A.1D; Verma et al. 2002). We presumed that, concomitant with its deubiquitination by Rpn11, MbpSic1 was translocated into the lumen of the 20S core but was not degraded due to the presence of epoxomicin. This hypothesis is supported by the observation that MbpSic1 formed upon incubation with proteasomes *in vitro*—but not naive MbpSic—was specifically coprecipitated with 20S subunits (see Supplemental Figure S1 at Cell web site). As was observed in the degradation assay, *rpn10* $\Delta$  proteasomes were completely deficient in deubiquitination of MbpSic1 (Figure A.1D, lanes 3 and 4), whereas *rad23* $\Delta$  proteasomes were largely but not completely defective (Figure A.1F). Because it is easier to visualize the accumulation of deubiquitinated Sic1 as opposed to the disappearance of ubiquitinated Sic1 to evaluate proteasome function, we sometimes used the DUB assay in lieu of

the degradation assay in subsequent experiments.

### A.3.3 Restoration of Activity by Recombinant Rpn10 and Rad23

Although *rpn10* $\Delta$  and *rad23* $\Delta$  proteasomes appeared to be fairly normal by multiple physical and functional criteria (Figure A.1), it remained possible that they were indirectly and/or irreversibly compromised by the absence of either of these proteins. To address this possibility, we performed add-back experiments using recombinant Gst-Rpn10 and Gst-Rad23 purified from *E. coli* (Supplemental Figure S2A). Strikingly, deubiquitination (Figure A.2B) and degradation (Figure A.2A) activities comparable to wild-type levels were obtained upon adding back Gst-Rpn10 to *rpn10* $\Delta$  proteasomes. The effect of Gst-Rpn10 was exquisitely dosage sensitive. Very low levels (30–60 nM) were sufficient to rescue *rpn10* $\Delta$  proteasomes but had little effect on wild-type proteasomes. However, at a concentration (120 nM) just  $\approx$  1.5- to 2-fold in molar excess over wild-type proteasomes, inhibition was observed, and at  $\approx$  3- to 4-fold molar excess (300 nM), inhibition was complete. Essentially the same effect was seen if Gst-Rpn10 was cleaved with thrombin to remove Gst (data not shown).

The ability of Gst-Rpn10 to rescue *rpn10* $\Delta$  proteasomes allowed us to map the domains of Rpn10 required for complementation. Mutational analysis of RPN10 in prior studies has demonstrated that the N-terminal domain of Rpn10 (also called the von Willebrand A or VWA domain, Whittaker and Hynes 2002) is required for conferring resistance to amino acid analogs and Ub-Pro- $\beta$ -gal degradation (Fu et al. 1998). The C terminus contains the conserved LAMALRL multiUb chain recognition motif that constitutes part of the UIM domain and that is also required



for binding UbMbpSic1 (Supplemental Figure S2C). No phenotype has ever been linked to this domain, even though it constitutes the multiUb chain recognition domain of Rpn10. As shown in Figure A.2D, either point mutation (first five amino acids of the recognition motif mutated; Gst-N5rpn10) or deletion of the UIM domain (Gst-VWARpn10 or UIM<sup>-</sup>) destroyed Rpn10 activity, underscoring the requirement for the UIM domain of Rpn10 for UbMbpSic1 degradation. To our knowledge, this is the first functional assay in which a direct requirement for the UIM has been demonstrated.

We next investigated the ability of recombinant Rad23 to complement the partial defect in DUB activity observed with *rad23*Δ 26S proteasomes. The results in Figure A.2C demonstrate that bacterially expressed Gst-Rad23 was functional and rescued the DUB defect. As observed for Rpn10, optimal rescue by Gst-Rad23 was highly concentration dependent. Efficient restoration of activity was observed at 40 nM, but high concentrations of Gst-Rad23 actually inhibited the basal activity of *rad23*Δ proteasomes. A recent study using wild-type 26S proteasomes supplemented with a 3-fold molar excess of Rad23 concluded that Rad23 has an inhibitory function in proteolysis (Raasi and Pickart 2003). Likewise, previous reports documented an inhibitory role for Rpn10 *in vitro* (Deveraux et al. 1995). However, our observations indicate that both Rad23 and Rpn10 actually promote protein degradation by the proteasome—at least when the substrate is UbSic1—but that for both proteins it is essential to use mutant proteasome preparations to identify the optimal dose, because these proteins inhibit degradation even when present in only modest stoichiometric excess over the 26S proteasome.

Our results caused us to wonder why Rad23 present in *rpn10*Δ proteasomes and Rpn10 present in *rad23*Δ proteasomes did not provide sufficient activity to sustain

normal rates of UbMbpSic1 turnover. Do these proteins operate in parallel as redundant substrate-targeting factors to sustain a maximal rate of Sic1 turnover, or might they act in series? One simple explanation is that Rad23 is normally present at only substoichiometric levels in 26S proteasome preparations, such that there was not enough to sustain UbMbpSic1 turnover in the absence of Rpn10. This contention is consistent with SDS-PAGE/microsequence analysis of purified yeast proteasomes (Glickman et al. 1998), immunoblot analysis of purified mammalian proteasomes (Raasi and Pickart 2003), and the very low sequence coverage observed for Rad23 in our MudPIT experiments (Supplemental Table S2). Likewise, immunoblotting experiments revealed that Rpn10 was present in *rad23* $\Delta$  proteasomes at one-third to one-half the levels observed in wild-type 26S proteasomes (Supplemental Figure S3). Significantly, addition of just 30 nM Rpn10 rescued the defective DUB activity of *rad23* $\Delta$  26S proteasomes (Figure A.2C), arguing that Rpn10 and Rad23 can act redundantly to sustain UbMbpSic1 deubiquitination and turnover, and the action of Rpn10 was not dependent upon Rad23.

#### A.3.4 Redundant Roles for Rad23 and the UIM Domain of Rpn10 in Sustaining UbSic1 Degradation

Crossrescue of *rad23* $\Delta$  26S proteasomes by Rpn10 encouraged us to investigate if the reverse was true, i. e., could addition of Rad23 restore activity to *rpn10* $\Delta$  26S proteasomes? Surprisingly, although recombinant Gst-Rad23 was fully functional in restoring activity to *rad23* $\Delta$  26S proteasomes (Figure A.2C), it rescued *rpn10* $\Delta$  26S proteasomes weakly (Figure A.2D). Because the requirement for Rpn10 function for *in vivo* turnover of the synthetic reporter substrate Ub-Pro- $\beta$ -gal mapped to the

N-terminal VWA domain of Rpn10 (Fu et al. 1998), we wondered whether Rad23 would rescue *rpn10* $\Delta$  proteasomes in the presence of the VWA domain of Rpn10. Remarkably, although Gst-VWARpn10 (UIM domain deleted) and Gst-N5rpn10 (mutant UIM) by themselves were inactive, the combination of either protein with GstRad23 restored full activity to *rpn10* $\Delta$  proteasomes (Figure A.2D). Taken together, these observations support two important conclusions about the functions of Rpn10 and Rad23. First, the Ub binding domains of Rpn10 and Rad23 do not need to act sequentially. Instead, there exists a functional redundancy between Rad23 (see below) and the Rpn10 UIM domain, suggesting that they function in parallel pathways to sustain degradation of Sic1. Second, the VWA domain of Rpn10 was required for Rad23 to promote optimal rates of UbSic1 proteolysis. This was also observed with Dsk2, another UbL-UBA domain protein like Rad23 (Funakoshi et al. 2002). Although rescue was weak, there was clearly an enhancement in activity when the Rpn10 VWA domain and Dsk2 were added together (Figure 2DA.2, lanes 11 and 14). It could be that Dsk2 is less potent than Rad23 because it has only one UBA domain, and Rad23 has two. Indeed, Dsk2 bound less UbMbpSic1 than Rad23 (Supplemental Figure S2C). Since Rpn10 functions to enhance the weak complementation by Rad23 (and Dsk2), we propose the term “facilitator” for Rpn10.

### A.3.5 Both the UBA and the UbL Regions of Rad23 Are Required for Function

Rescue of *rad23* $\Delta$  26S proteasomes by recombinant Rad23 allowed us to assess the relative contributions of both its Ub chain binding (UBA) and proteasome binding (UbL) regions. As predicted by prior studies (Schauber et al. 1998; and Wilkinson et al. 2001), a mutant protein (shown in Supplemental Figure S2B) lacking the UbL

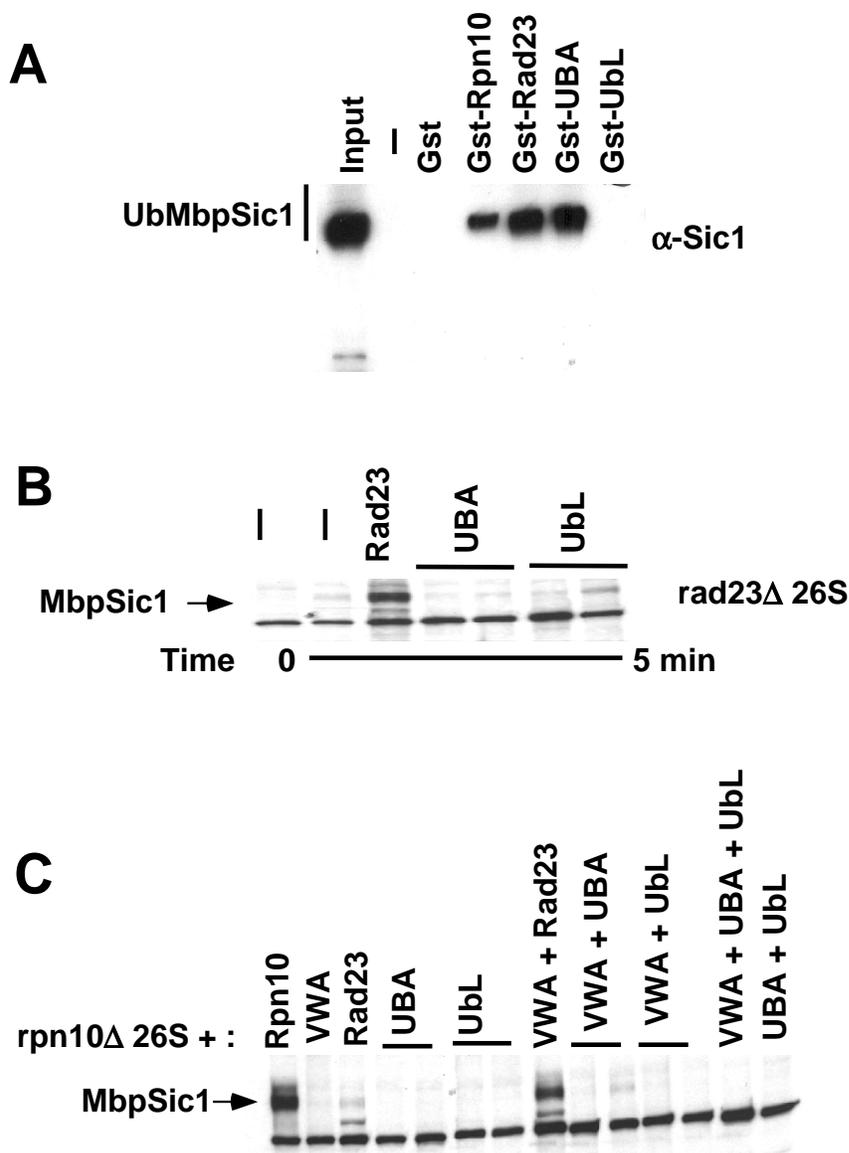
but containing both UBA domains bound UbMbpSic1 (Figure A.3A), whereas the reciprocal construct that contains the UbL domain but lacks both UBA domains selectively bound 26S proteasomes (Supplemental Figure S2D). However, neither the UbL nor UBA segments sustained robust rescue of *rad23* $\Delta$  (Figure A.3B) or *rpn10* $\Delta$  (Figure A.3C) 26S proteasomes.

### A.3.6 Rad23 and the UIM Domain of Rpn10 Link UbSic1 to the Proteasome

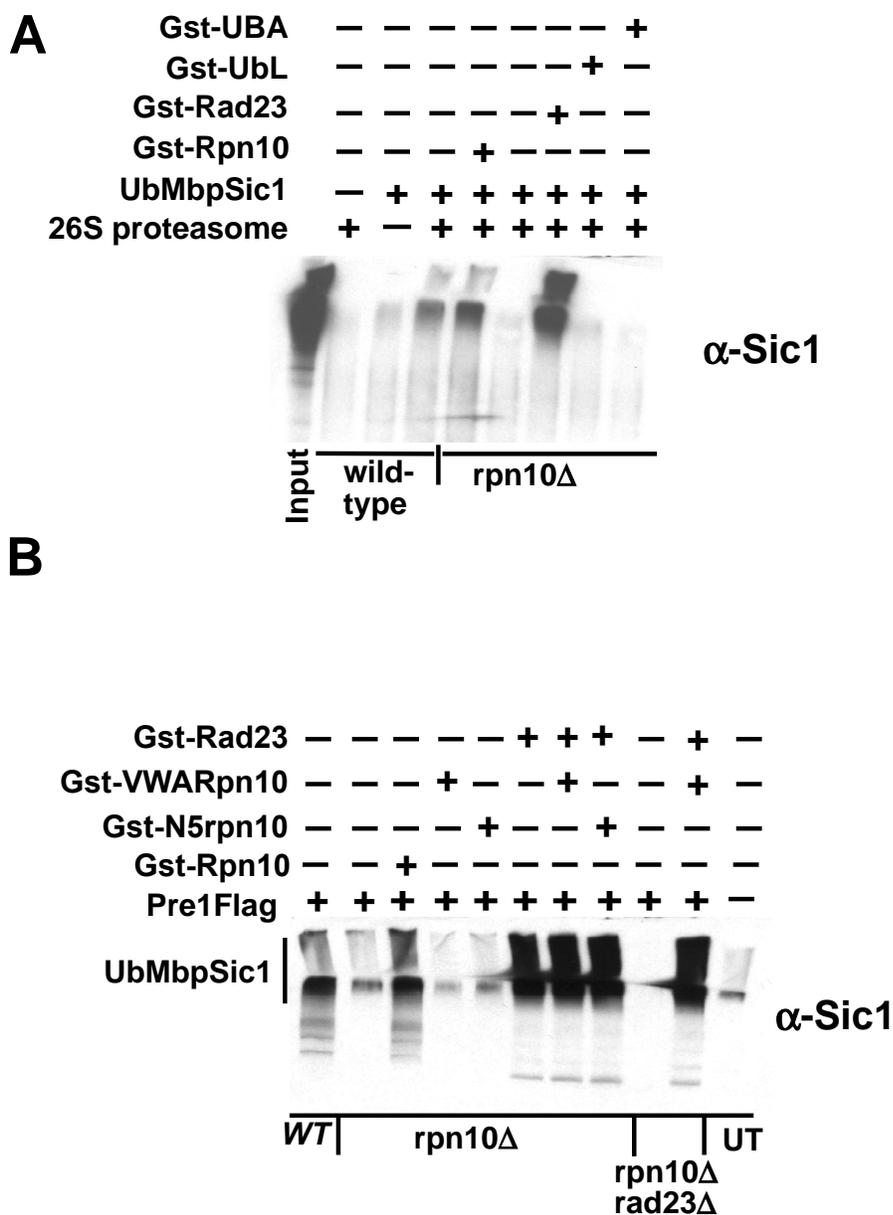
The ability of the UBA domain of Rad23 and the UIM domain of Rpn10 to bind multiUb chains (Figures A.3A and Supplemental S2C) suggested that the redundant function provided by these elements is to target UbSic1 to the proteasome for degradation. To address this hypothesis, the substrate binding capacities of wild-type and *rpn10* $\Delta$  26S proteasomes were investigated by incubating UbMbpSic1 (in the presence of inhibitors of deubiquitination and degradation) with 26S proteasomes immobilized on anti-Flag beads (Figure A.4A). Wild-type 26S proteasomes bound UbMbpSic1 whereas *rpn10* $\Delta$  26S proteasomes displayed little or no binding activity. Gst-Rpn10 efficiently rescued the substrate binding defect of *rpn10* $\Delta$  proteasomes (Figure A.4), but Gst-VWARpn10 and Gst-N5rpn10 did not (Figure A.4B), underscoring that this recruitment activity required the UIM domain. Gst-Rad23 bound *rpn10* $\Delta$  proteasomes in a UbL-dependent manner (Supplemental Figure S2D) and endowed them with enhanced substrate binding activity (Figure A.4).

### A.3.7 Rpn10 VWA Domain Facilitates the Degradation-Promoting Activity of Rad23

Surprisingly, although the VWA domain of Rpn10 was required for optimal proteo-



**Figure A.3 Complementation of *rad23* $\Delta$  Proteasomes Requires Both the Ub Binding UBA Domains and the Proteasome Binding UbL Domain of Rad23.** (A) The UBA domains bind UbMbpSic1. Purified Gst and Gst fusion proteins (1  $\mu$ g each) bound to glutathione beads were incubated with UbMbpSic1, after which the input (20% of total) and bound material (33% of total) were fractionated by SDS-PAGE and visualized by immunoblotting with anti-Sic1 serum. Note that Gst-UBA lacks the UbL domain but contains both UBA domains found in Rad23, whereas Gst-UblL is the reciprocal molecule lacking both UBA domains (Rao and Sastry 2002). (B) Rescue of *rad23* $\Delta$  26S proteasomes by Rad23. Deubiquitination reactions were set up using *rad23* $\Delta$  26S proteasomes and UbMbpSic1 in the presence or absence of Gst-Rad23 (80 nM), Gst-UBA (80 and 40 nM respectively), or Gst-UblL (80 and 40 nM), respectively, as described in the legend to Figure A.1D. (C) Rescue of *rpn10* $\Delta$  26S DUB defect by full-length Rad23 and Gst-VWA. Deubiquitination reactions were assayed by incubation of UbMbpSic1 with *rpn10* $\Delta$  26S proteasomes in the presence or absence of various Gst-fusion proteins as described above.



**Figure A.4 26S Proteasomes from *rpn10Δ* Are Defective in Binding UbMbpSic1.** (A and B) The binding defect of *rpn10Δ* 26S proteasomes can be rescued by either recombinant Rpn10 or Rad23. Extracts from wild-type (WT), *rpn10Δ*, and *rpn10Δrad23Δ* cells expressing PRE1FH (Supplemental Table S1) or untagged PRE1 (UT) were bound to anti-Flag M2 resin in the presence of ATP and washed with buffer containing ATP as described for 26S purification (section A.4). Resin-immobilized 26S proteasomes were then incubated with 1 mM phenanthroline, 2.5  $\mu$ M Ub aldehyde, 100  $\mu$ M MG132, 1 mM ATP, and 5 mM  $MgCl_2$  in the absence or presence of the various Gst-fusion proteins on ice for 60 min. UbMbpSic1 was then added, and, after 90 min incubation at 4°C, the bound fraction was washed and analyzed by SDS-PAGE and immunoblotting for Sic1. In (A), 5% of input and 25% of the bound fractions were loaded.

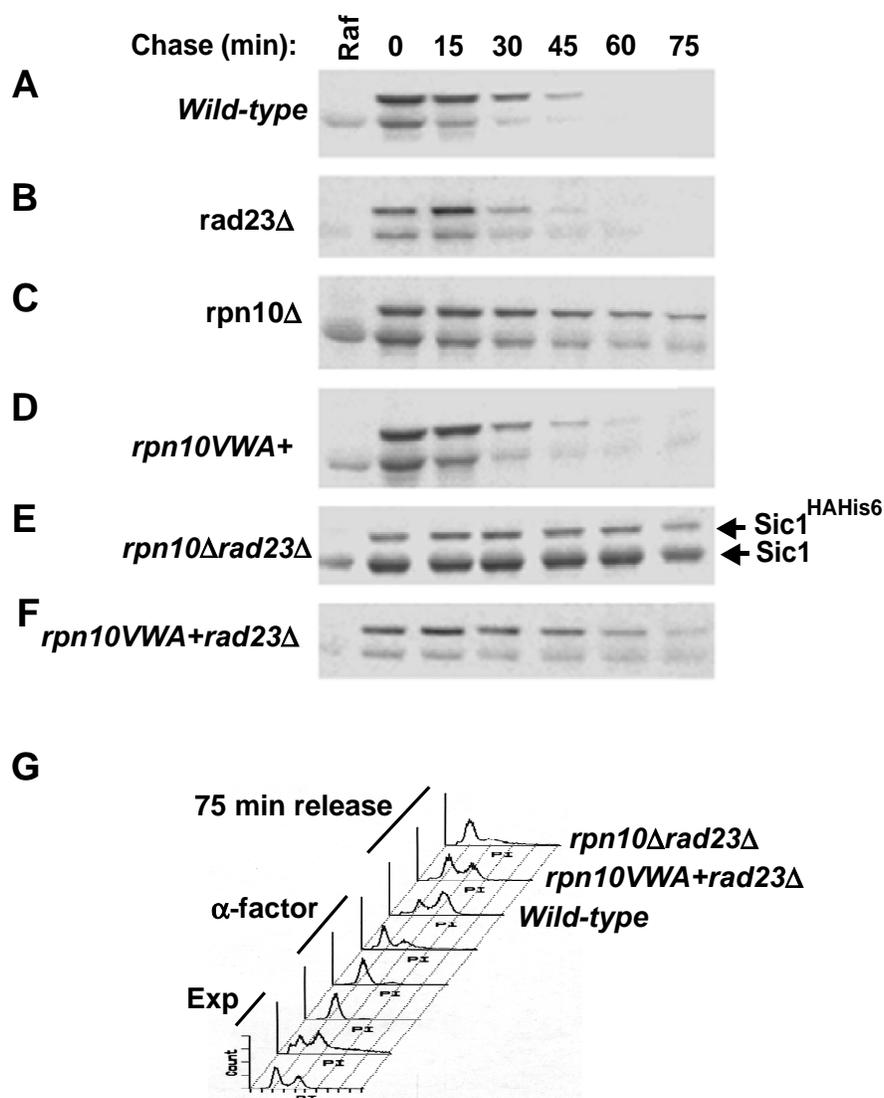
lysis-promoting activity of Rad23 (Figure A.2D), it was not required for Rad23-dependent tethering of UbMbpSic1 to the proteasome (Figure A.4). Thus, binding is not a reliable surrogate assay for degradation. We conclude that the VWA domain acts downstream of Rad23 and enables proteasome bound, ubiquitinated substrate to engage productively with the degradation machinery. Owing to its additional facilitator function encoded within the VWA domain, we suggest that the term facilitator be applied to Rpn10 to distinguish it from substrate receptors such as Rad23. A widespread role for Rpn10 as a substrate receptor facilitator is suggested by the findings that deletion of RPN10 in *Drosophila* results in pupal lethality (Szlanka et al. 2003), and its downregulation by RNAi causes G2/M phase arrest in *Trypanosoma brucei* (Li and Wang 2002). Given that yeast *rpn10* $\Delta$  mutants are viable, we surmise that either Rad23, Dsk2, or other substrate receptors retain sufficient function to sustain life (note the weak albeit detectable activity of Rad23 in the absence of Rpn10VWA; Figure A.2D, lane 10), or other proteins provide a facilitator function *in vivo* that is redundant with that of Rpn10's VWA domain.

### A.3.8 Both RPN10 and RAD23 Contribute to Sic1 Turnover *In Vivo*

The *in vitro* assays indicate important roles for Rpn10 and Rad23 in Sic1 turnover. To date, all studies on these mutants *in vivo* have relied either on artificial substrates (Van Nocker et al. 1996); indirect readouts for degradation, such as steady state analysis (Wilkinson et al. 2001); or a substrate (Clb2) whose degradation is subject to indirect regulation via cell cycle checkpoints (Lambertson et al. 1999). Thus, to monitor Sic1 degradation *in vivo*, we evaluated turnover during the appropriate cell cycle phase. Wild-type and mutant cells were arrested in G1 with  $\alpha$  factor and then

released synchronously into the cell cycle (Figure A.5). Both GAL1-expressed and endogenous Sic1 are normally degraded at the G1/S boundary (Verma et al. 1997). As shown in Figure A.5, both GAL1-expressed and endogenous Sic1 tapered off by 45 min as cells entered S phase. Based on our reconstitution experiments, we reasoned that Sic1 might be targeted for degradation *in vivo* by either Rad23 or the UIM domain of Rpn10. Indeed, whereas Sic1 was degraded with normal kinetics in *rad23* $\Delta$  and in a mutant lacking the UIM domain of Rpn10 (*rpn10VWA+*), significant stabilization was observed in an *rpn10VWA+ rad23* $\Delta$  double mutant. As expected from the facilitator role played by the VWA domain in the operation of other receptor pathways *in vitro*, Sic1 was significantly more stable in *rpn10* $\Delta$  than in *rpn10VWA+* cells. Additionally, failure to promptly degrade Sic1 correlated with a reduced rate of entry into S phase, as shown for the *rpn10* $\Delta rad23$  $\Delta$  mutant (Figure A.5), which remained in G1 phase 75 min after release from  $\alpha$  factor. Degradation of Sic1 is essential for entry into S phase (Verma et al. 1997). Delayed entry into S phase and residual turnover of Sic1 in *rpn10* $\Delta rad23$  $\Delta$  cells indicate that there must exist a third receptor pathway (possibly Dsk2, Figure A.2D) by which Sic1 can engage the proteasome and be degraded, albeit at a greatly reduced rate.

Since the *rpn10* $\Delta rad23$  $\Delta$  double mutant displayed unexpectedly strong stabilization of Sic1, the growth phenotype of this mutant was reassessed. It has been reported that these mutants are cold sensitive at 13°C (Lambertson et al. 1999). However, we observed a severe growth defect even at 25°C (Supplemental Figure S4), which was exacerbated in synthetic medium. Consistent with the *in vitro* and *in vivo* data presented here and elsewhere (Fu et al. 1998), the slow growth phenotypes of the double mutant were linked to the absence of the VWA domain of



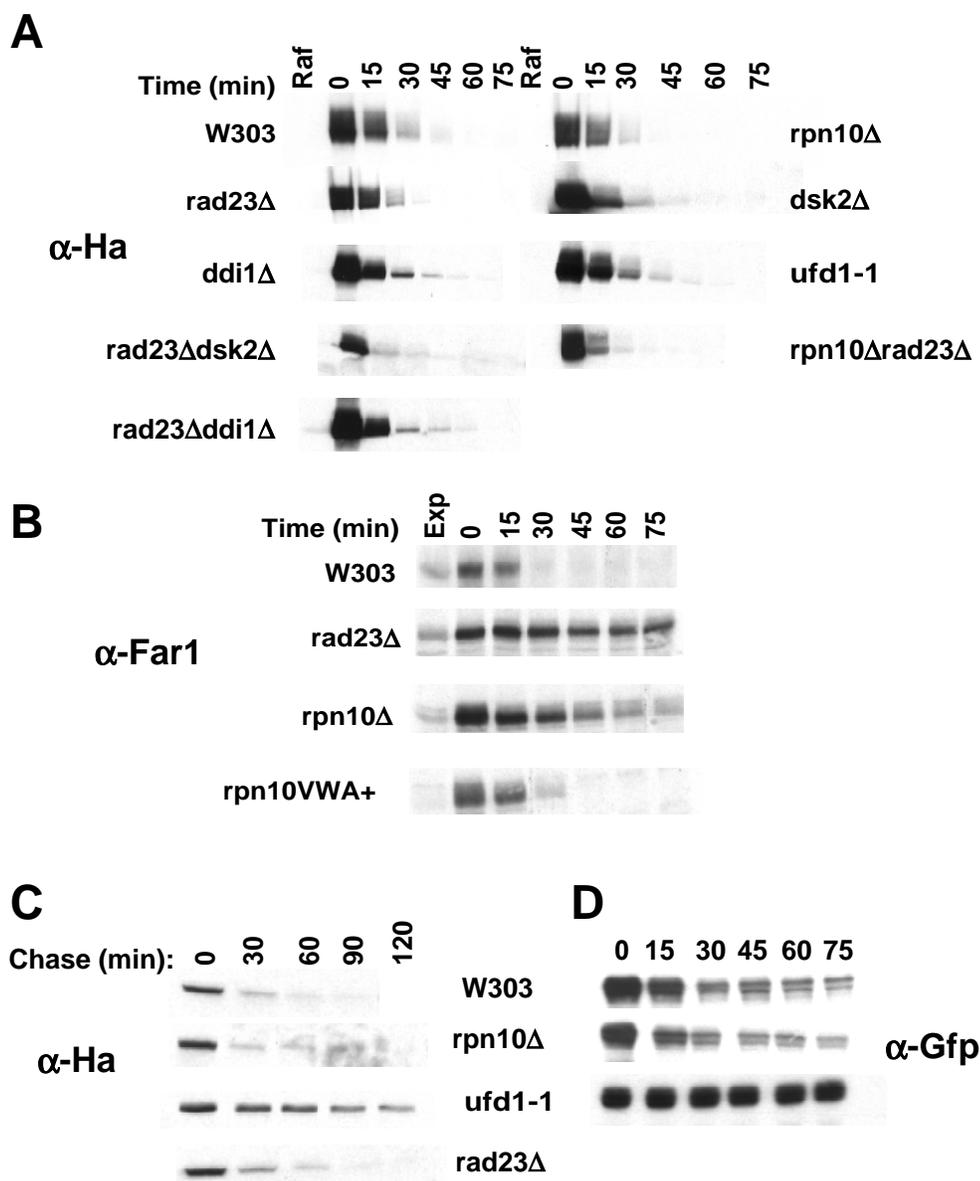
**Figure A.5 Rpn10 UIM Domain and Rad23 Serve Redundant Roles in Sic1 Turnover *In Vivo*.** (A–F) Wild-type and mutant cells (Supplemental Table S1) expressing a GAL1-driven, epitope-tagged (HaHis6) allele of SIC1 in addition to endogenous untagged SIC1 were arrested with  $\alpha$  factor and released synchronously into the cell cycle at 25°C (except *rpn10Δrad23Δ*, which were released at 30°C because they grew poorly at 25°C). Extracts were prepared at the indicated time points and analyzed by SDS-PAGE followed by immunoblotting with anti-Sic1 serum that detects both the endogenous and the epitope-tagged versions of Sic1. (G) Wild-type, *rpn10VWA rad23Δ*, and *rpn10Δ rad23Δ* cells collected at the indicated time points were evaluated for cell cycle distribution by flow cytometry.

RPN10 (Supplemental Figure S4).

### A.3.9 Specificity in the Requirement for Different MCBPs for *In Vivo* Turnover of UPS Substrates

To address the generality of our observations, we next tested whether the relative contributions of Rad23 and Rpn10 to Sic1 degradation would hold true for another physiological substrate of the UPS—the G1 cyclin Cln2 (Deshaies et al. 1995). HA-tagged Cln2 expressed from the GAL1 promoter was rapidly degraded in G1 phase cells and unlike Sic1 was not stabilized in *rpn10* $\Delta$ , *rad23* $\Delta$ , or *rpn10* $\Delta$ *rad23* $\Delta$  mutants. This prompted us to look at its turnover in additional MCBP mutants. As shown by the data in Figure A.6A, mutations in the genes encoding the UBA domain-containing putative targeting factors Ddi1, Dsk2 (Saeki et al. 2002a), and the UT3 domain-containing Ufd1 (Ye et al. 2003) had no effect on Cln2 turnover. From this analysis, we conclude that an as yet unknown receptor or set of receptors, possibly including Rpt5, functions to link ubiquitinated Cln2 to the proteasome.

Whereas Sic1 is a substrate of the E3 Ub ligase SCF<sup>Cdc4</sup> (Seol et al. 1999), Cln2 is an SCF<sup>Grr1</sup> substrate (Seol et al. 1999; and Skowyra et al. 1999). To determine if the identity of the ubiquitin ligase influenced the different receptor dependencies exhibited by Sic1 and Cln2, we examined the turnover of the SCF<sup>Cdc4</sup> substrate Far1 (Henchoz et al. 1997) and the SCF<sup>Grr1</sup> substrate Gic2 (Jaquenoud et al. 1998). Far1 is a G1 cyclin–Cdk inhibitor, and Gic2 is an effector of the Cdc42 cell polarity regulator. In both cases, turnover of the endogenous protein was examined during G1 phase, when Far1 and Gic2 are normally degraded (Jaquenoud et al. 1998; also



**Figure A.6 UPS Substrates Have Differential Requirements for Multiubiquitin Chain Receptors *In Vivo*.** For experiments shown in panels (A)–(D), aliquots of cells of the indicated genotypes were withdrawn at various times after initiation of chase (min), and whole cell lysates were fractionated by SDS–PAGE and immunoblotted with the indicated antibodies. (A) Wild–type and mutant cells expressing Ha epitope–tagged Cln2 from the GAL1 promoter were grown in YP raffinose at 30°C, and expression of Cln2–Ha was induced with 2% galactose at 25°C for 90 min. Induction was terminated and chase was initiated by transfer of cells to YP–2% dextrose. (B) To monitor turnover of Far1, wild–type and mutant cells were arrested with  $\alpha$  factor for 3 h at 25°C, and the chase period was initiated by release into fresh medium in the absence of  $\alpha$  factor, which results in rapid downregulation of Far1 message (see <http://www.yeastgenome.org/> for expression analysis) (C) The stability of CPY\*HA was monitored upon initiating a chase period by adding 100  $\mu$ g/ml cycloheximide to wild–type and mutant cultures at 25°C. (D) Cycloheximide chase was done as described in (C) to monitor turnover of Deg1–Gfp.

see <http://www.yeastgenome.org/>). In contrast to Sic1, Far1 degradation was impeded more in *rad23* $\Delta$  than in *rpn10* $\Delta$  mutants (Figure A.6B). Meanwhile, Gic2 mimicked Sic1 and not Cln2 in that it was strongly stabilized in *rpn10* $\Delta$  cells (Supplemental Figure S5A). Additionally, Clb2, an APC substrate (Harper et al. 2002), also mimicked Sic1 (Supplemental Figure S5). Thus, no simple rule could be formulated that relates a ubiquitinated substrate's dependency upon a targeting receptor to the identity of its E3.

In addition to proteolysis of regulatory proteins, the UPS is also required for the degradation of misfolded proteins. Secretory pathway proteins that fail to fold properly in the ER are retrotranslocated into the cytosol and degraded by the 26S proteasome in a process called ER-associated degradation (ERAD, Tsai et al. 2002). The Cdc48/Ufd1/Npl4 complex is required for ERAD and recognizes membrane-associated Ub conjugates via the UT3 domains of Ufd1/Cdc48 (Ye et al. 2003). The ERAD substrate CPY\* is stabilized in mutants defective in individual subunits of the Cdc48/Ufd1/Npl4 complex (Jarosch et al. 2002; Figure A.6C). To determine if ERAD substrates are “handed off” to proteasomal receptors following their extraction from the membrane by Cdc48/Ufd1/Npl4 (Flierman et al. 2003), we evaluated the turnover of CPY\* in *rpn10* $\Delta$  and *rad23* $\Delta$  mutants. Surprisingly, no stabilization was observed (Figure A.6C). These data suggest that Cdc48/Ufd1/Npl4 may shepherd the extracted CPY\* directly to the proteasome or deliver it to Rpt5 or an as yet unknown receptor.

The Cdc48/Ufd1 complex binds specifically to K48-linked polyUb chains via the UT3 domain (Ye et al. 2003) and also participates in degradation of non-ERAD substrates such as cytosolic Ub<sup>V76</sup>-V- $\beta$ -galactosidase (Johnson et al. 1995) and spindle disassembly factors Cdc5 and Ase1 (Cao et al. 2003). We monitored the

turnover of the cytoplasmic Deg1-Gfp, which contains the degradation signal from the transcriptional repressor MAT $\alpha$ 2. This fusion substrate is interesting because, although it is soluble, it is ubiquitinated by enzymes resident in the ER membrane (Swanson et al. 2001). As shown in Figure A.6D, Deg1-Gfp was stabilized in *ufd1-1*. However, like the ERAD substrate CPY\*, Deg1-Gfp was not stabilized in *rpn10 $\Delta$*  mutants.

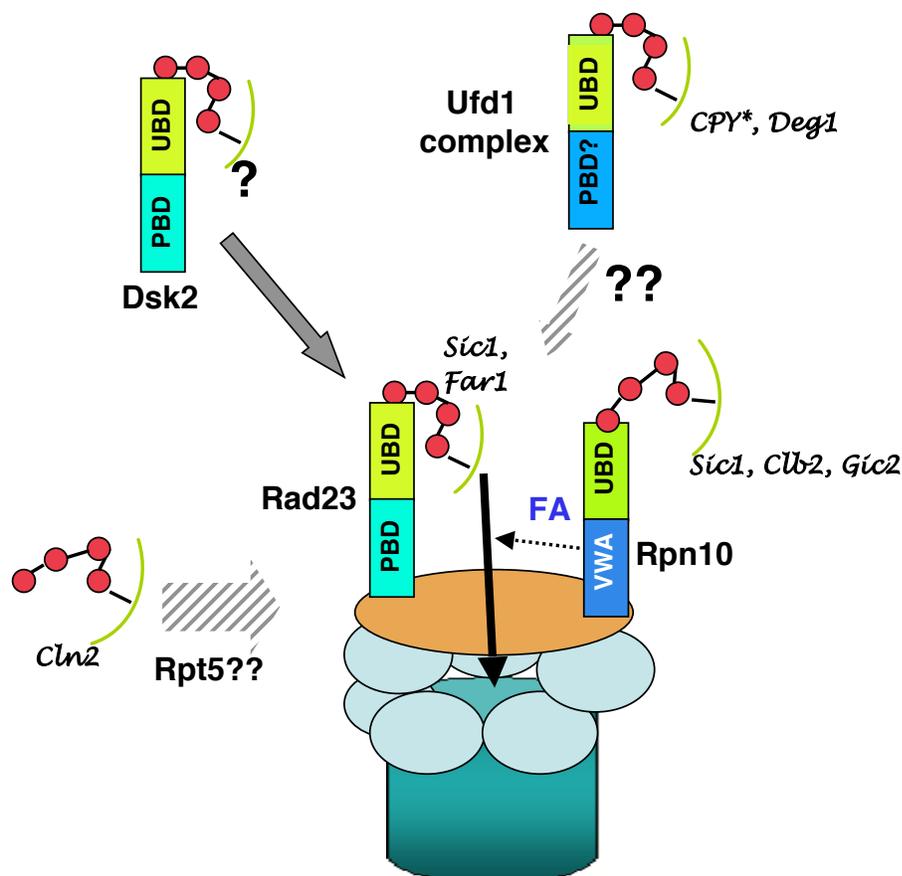
### A.3.10 Natural versus Synthetic Substrates of the UPS

An important principle emerges from considering the targeting requirements observed for physiological versus synthetic substrates. Reporter substrates such as Ub-Pro- $\beta$ -gal, Ub<sup>V76</sup>-V- $\beta$ -gal, and Ub<sup>V76</sup>-V-DHFR exhibit simultaneous dependence on multiple putative receptor pathways, including Rpn10, Rad23, and Cdc48/Ufd1 (Johnson et al. 1995; Rao and Sastry 2002; and Xie and Varshavsky 2002; see also Supplemental Figure S5D). This simultaneous dependence suggests that these factors typically serve nonredundant, possibly even sequential (Chen and Madura 2002) roles in degradation. By contrast, none of the physiological substrates examined in this study (including Far1, Sic1, Gic2, Cln2, CPY\*, and Clb2) exhibited an equivalently broad dependence on multiple putative receptor pathways. Thus, although synthetic substrates have proved very useful for defining components of the UPS system, we caution that their turnover may not be reflective of typical physiologic mechanisms, and, thus, general conclusions about the mechanism/specificity of the UPS should be rooted in the study of physiological substrates.

### A.3.11 One Universal Targeting Signal with Multiple Receptors

It is commonly thought that specificity in substrate turnover by the UPS lies at the level of ubiquitin chain assembly controlled by E2, E3, and isopeptidase enzymes. Our findings, however, lead to the unexpected conclusion that proteasome-targeting pathways downstream of the ubiquitin ligases exhibit a surprising degree of substrate specificity. A scheme that graphically summarizes our key proposals is depicted in Figure A.7. Rpn10, Rad23, Dsk2, and possibly Ufd1/Cdc48 and Rpt5 are envisioned to comprise distinct receptor pathways that link ubiquitinated substrates to the proteasome. It is important to note that there are no functional data indicating that either Ufd1/Cdc48 or Rpt5 recruits ubiquitinated substrates to the proteasome. However, others have suggested a receptor function for Rpt5 based on crosslinking data (Lam et al. 2002), and we suggest a receptor activity for Ufd1/Cdc48 as a working hypothesis in light of data reported here and elsewhere (Flierman et al. 2003; and Ye et al. 2003).

Some substrates, like Sic1 and Clb2, are recruited to the proteasome and degraded in a manner that depends strongly on the receptor and/or facilitator (FA) functions of the proteasome subunit Rpn10, whereas others, such as Far1, show a weaker dependence on Rpn10 and a correspondingly stronger dependence on Rad23. Yet other substrates such as CPY\* and Deg1-Gfp appear to bypass Rpn10 entirely but depend on a complex containing Ufd1 and Cdc48. (It has been reported that Far1 degradation also depends upon Cdc48 using a novel G1-specific td allele (Fu et al. 2003), but we have not observed a defect in Far1 turnover in *cdc48-3* or *ufd1-1* mutants; data not shown). Finally, at least one substrate, Cln2, does not depend upon any known receptor pathway. However, our data on Sic1 underscore that it is



**Figure A.7 Hypothetical Model for Physiological Targeting Pathways that Deliver Ubiquitinated Substrates to the 26S Proteasome.** The schematic shows the 20S proteolytic core capped by the base, which comprises a hexameric ring of the AAA ATPases (Rpt1–Rpt6, depicted as light blue ovals) and the PC repeat containing proteins Rpn1 and Rpn2 (collectively depicted as a beige oval). Rad23 and Rpn10 associate with the proteasome via the Rpn1/Rpn2 subunits to deliver substrates tethered to their Ub binding domains (UBD), including Far1, Sic1, Gic2, and Clb2. Deubiquitination and degradation of substrates delivered by Rad23 requires a facilitator activity (FA) encoded within the VWA domain of Rpn10. Dsk2, a UBA domain containing protein like Rad23, is postulated to also deliver substrates to the same entry port used by Rad23, but the identity of these substrates remains unknown. Ufd1-containing complexes that contain Cdc48 are proposed to deliver ERAD and non-ERAD substrates such as CPY\*, Deg1, and Cdc5 to the proteasome, but the putative proteasome binding domain (PBD) and docking site employed by this complex remain unknown. Ubiquitinated Cln2 is targeted for degradation by a pathway that remains unknown but does not require the activity of Rpn10, Rad23, Dsk2, or Ufd1. It is possible that Cln2 gains access to the proteasome via the putative Rpt5 gateway or an unknown receptor or utilizes multiple receptor pathways in a highly redundant manner.

important to distinguish “dependency” from “involvement.” Rad23 can be involved in Sic1 turnover (as evidenced by the fact that Sic1 was unstable in *rpn10*<sup>VWA</sup> but

was stabilized in *rpn10<sup>VWA</sup>rad23Δ*), even though Sic1 turnover does not normally depend upon Rad23 (as evidenced by rapid Sic1 turnover in a *rad23Δ* mutant). Thus, Cln2 may not depend upon the known receptors, because it can be targeted by multiple receptors in a highly redundant manner, or because it arrives at the proteasome by a distinct route involving Rpt5 or an unknown receptor. Yet other targeting strategies are likely to exist, given that ubiquitin ligases such as Parkin, Ufd4, and Hul5 can bind directly to the proteasome (Demand et al. 2001; Sakata et al. 2003; Xie and Varshavsky 2002; and Leggett et al. 2002). Interesting challenges for the future will be to determine how many receptor pathways exist, to sort out the mechanism underlying the allocation of substrates to different receptor pathways, and to determine whether individual receptor pathways are differentially regulated to modulate the repertoire of proteins degraded by the UPS in response to specific signals.

Our data indicate that a putative receptor activity intrinsic to Rpt5 (Lam et al. 2002) by itself is insufficient to target UbSic1 for degradation in a defined *in vitro* system. Moreover, our *in vivo* analysis implies that an Rpt5-mediated targeting mechanism would appear to be insufficient to sustain normal rates of degradation *in vivo* for seven of eight UPS substrates characterized in this study. What, then, is the role of Rpt5 in substrate targeting? It is possible that Rpt5 serves as the primary conduit by which a subset of unstable proteins poorly represented in this study (but possibly including Cln2) gains access to the proteasome. On the other hand, we favor the notion that Rpt5 serves as a central conduit that gathers together substrates delivered by different receptor pathways (Rpn10, Rad23, and Cdc48/Ufd1) and positions them for subsequent unfolding, deubiquitination, and translocation. This latter possibility calls to mind translocation of secretory precursors cross the

ER membrane, where the primary signal peptide-mediated targeting step is carried out by upstream receptors such as Signal Recognition Particle, following which the signal peptide is transferred to the Sec61 channel to enable precursor translocation across the membrane. An analogous two-step recognition system may operate in *E. coli*, where the SspB protein functions as a specificity factor for the AAA ATPase ClpX, enhancing degradation of *ssrA*-tagged substrates (Levchenko et al. 2000).

### A.3.12 Note Added in Proof

While this manuscript was under review, Elsasser et al. (2004) reported that Rad23 and Rpn10 can tether autoubiquitinated Cdc34 to 26S proteasome. In a second publication, Medicherla et al. (2004) reported that *rad23Δdsk2Δ* mutants are defective in CPY\* turnover. Medicherla et al. (2004) also reported that Deg1-GFP is degraded normally in *ufd1-1*, a result that conflicts with our Figure A.6D. We do not know the reason for this discrepancy.

### A.3.13 Acknowledgements

We thank H. Rao, H. Yokosawa, H. Fu, M. Funakoshi, K. Madura, R. Vierstra, R. Hampton, E. Johnson, M. Peter, S. Reed, Y. Zheng, and D. Kellogg for yeast strains, expression plasmids, and antibodies. We thank the members of the Deshaies lab for helpful discussions; and A. Varshavsky and T. Mayor for critically reading the manuscript. We gratefully acknowledge the Caltech Flow Cytometry Cell Sorting Facility and R. Diamond for performing flow cytometry; and the Caltech Protein

Expression Facility and P. Snow for expression of the G1 Cdk and SCF complexes in insect cells. This work was supported by the Howard Hughes Medical Institute.

## A.4 Experimental Procedures

### A.4.1 Yeast Strains and Extract Preparation

Yeast strains used in this study are listed in Supplemental Table S1.

For turnover analysis of UPS substrates, wild-type and mutant cells grown to an  $OD_{600\text{ nm}}$  of 0.2–0.3 were processed as described in Figure A.4 and Figure A.5. Cells were harvested by centrifugation and drop frozen in liquid nitrogen. They were thawed and washed in a buffer containing 50 mM Tris (pH 7.5); 10 mM EDTA; 20 mM NaF; 0.05% azide; 5 mM NEM; 1 mM PMSF; 0.5 mM AEBSF; and 1× protease inhibitor cocktail containing pepstatin, chymostatin, aprotinin, and leupeptin at 5  $\mu\text{g}/\text{ml}$ . Glass beads (Sigma, 425–600  $\mu\text{m}$ , acid washed) equal in volume to the cell pellet were added, and the cell pellets were plunged into boiling water for 3 min after brief vortexing. Cells were then resuspended at uniform concentration (26.7 OD units/ml) in 1× SDS sample buffer and vortexed in a ThermoSavant Fast-Prep at 4°C for 45 s at the maximum speed setting (6.5). Vortexed cell pellets were boiled again for 3 min and aliquots resolved by SDS-PAGE. Ponceau S staining was done after transfer to nitrocellulose membrane to assess uniformity of protein levels across the gel. The blot was incubated with the appropriate antibody and processed using ECL.

#### A.4.2 Degradation and Deubiquitination Assays

Ubiquitinated MbpSic1 substrate (Seol et al. 1999) and affinity-purified 26S proteasomes (Verma et al. 2000) were prepared essentially as described. Degradation and deubiquitination assays ( $\approx 300$  nM substrate,  $\approx 100$  nM proteasome, incubated at  $30^\circ\text{C}$  for 5 min) were conducted as described previously (Verma et al. 2002).

#### A.4.3 Preparation of Gst-Fusion Proteins

Gst-fusion proteins were expressed in BL21/pLysS according to standard procedures. Proteins were eluted from glutathione resin with 50 mM Tris (pH 8.8), 50 mM NaCl, 5 mM DTT, 1 mM EDTA, and 40 mM glutathione at  $4^\circ\text{C}$  for 3 h and then dialyzed against buffer containing 25 mM Tris (pH 7.5), 100 mM NaCl, and 15% glycerol. Aliquots were drop frozen in liquid  $\text{N}_2$  and stored at  $-70^\circ\text{C}$ .

#### A.4.4 FACS Analysis

Yeast cells were processed for flow cytometry as described (Verma et al. 1997).

### A.5 References

Amerik, A. Y., Swaminathan, S., Krantz, B. A., Wilkinson, K. D. and Hochstrasser, M. (1997). *In vivo* disassembly of free polyubiquitin chains by yeast Ubp14 modulates rates of protein degradation by the proteasome. *EMBO J*, 16(16):4826–38.

- Cao, K., Nakajima, R., Meyer, H. H. and Zheng, Y. (2003). The AAA-ATPase Cdc48/p97 regulates spindle disassembly at the end of mitosis. *Cell*, 115(3):355–67.
- Chau, V., Tobias, J. W., Bachmair, A., Marriott, D. and Ecker, D. J. et al. (1989). A multiubiquitin chain is confined to specific lysine in a targeted short-lived protein. *Science*, 243(4898):1576–83.
- Chen, L. and Madura, K. (2002). Rad23 promotes the targeting of proteolytic substrates to the proteasome. *Mol Cell Biol*, 22(13):4902–13.
- Dai, R. M. and Li, C. C. (2001). Valosin-containing protein is a multi-ubiquitin chain-targeting factor required in ubiquitin-proteasome degradation. *Nat Cell Biol*, 3(8):740–4.
- Demand, J., Alberti, S., Patterson, C. and Hohfeld, J. (2001). Cooperation of a ubiquitin domain protein and an E3 ubiquitin ligase during chaperone/proteasome coupling. *Curr Biol*, 11(20):1569–77.
- Deshaies, R. J., Chau, V. and Kirschner, M. (1995). Ubiquitination of the G1 cyclin Cln2p by a Cdc34p-dependent pathway. *EMBO J*, 14(2):303–12.
- Deveraux, Q., Ustrell, V., Pickart, C. and Rechsteiner, M. (1994). A 26 S protease subunit that binds ubiquitin conjugates. *J Biol Chem*, 269(10):7059–61.
- Deveraux, Q., van Nocker, S., Mahaffey, D., Vierstra, R. and Rechsteiner, M. (1995). Inhibition of ubiquitin-mediated proteolysis by the Arabidopsis 26 S protease subunit S5a. *J Biol Chem*, 270(50):29660–3.
- Elsasser, S., Chandler-Militello, D., Muller, B., Hanna, J. and Finley, D. (2004). Rad23 and Rpn10 serve as alternative ubiquitin receptors for the proteasome. *J Biol Chem*, 279(26):26817–22.

- Elsasser, S., Gali, R. R., Schwickart, M., Larsen, C. N. and Leggett, D. S. et al. (2002). Proteasome subunit Rpn1 binds ubiquitin-like protein domains. *Nat Cell Biol*, 4(9):725–30.
- Flierman, D., Ye, Y., Dai, M., Chau, V. and Rapoport, T. A. (2003). Polyubiquitin serves as a recognition signal, rather than a ratcheting molecule, during retrotranslocation of proteins across the endoplasmic reticulum membrane. *J Biol Chem*, 278(37):34774–82.
- Flynn, J. M., Neher, S. B., Kim, Y. I., Sauer, R. T. and Baker, T. A. (2003). Proteomic discovery of cellular substrates of the ClpXP protease reveals five classes of ClpX-recognition signals. *Mol Cell*, 11(3):671–83.
- Fu, H., Sadis, S., Rubin, D. M., Glickman, M. and van Nocker, S. et al. (1998). Multiubiquitin chain binding and protein degradation are mediated by distinct domains within the 26 S proteasome subunit Mcb1. *J Biol Chem*, 273(4):1970–81.
- Fu, X., Ng, C., Feng, D. and Liang, C. (2003). Cdc48p is required for the cell cycle commitment point at Start via degradation of the G1-CDK inhibitor Far1p. *J Cell Biol*, 163(1):21–6.
- Funakoshi, M., Sasaki, T., Nishimoto, T. and Kobayashi, H. (2002). Budding yeast Dsk2p is a polyubiquitin-binding protein that can interact with the proteasome. *Proc Natl Acad Sci USA*, 99(2):745–50.
- Glickman, M. H., Rubin, D. M., Fried, V. A. and Finley, D. (1998). The regulatory particle of the *Saccharomyces cerevisiae* proteasome. *Mol Cell Biol*, 18(6):3149–62.
- Harper, J. W., Burton, J. L. and Solomon, M. J. (2002). The anaphase-promoting complex: It's not just for mitosis any more. *Genes Dev*, 16(17):2179–206.

- Hartmann-Petersen, R., Hendil, K. B. and Gordon, C. (2003). Ubiquitin binding proteins protect ubiquitin conjugates from disassembly. *FEBS Lett*, 535(1–3):77–81.
- Henchoz, S., Chi, Y., Catarin, B., Herskowitz, I. and Deshaies, R. J. et al. (1997). Phosphorylation- and ubiquitin-dependent degradation of the cyclin-dependent kinase inhibitor Far1p in budding yeast. *Genes Dev*, 11(22):3046–60.
- Hershko, A. and Ciechanover, A. (1998). The ubiquitin system. *Annu Rev Biochem*, 67:425–79.
- Hofmann, K. and Bucher, P. (1996). The UBA domain: A sequence motif present in multiple enzyme classes of the ubiquitination pathway. *Trends Biochem Sci*, 21(5):172–3.
- Jaquenoud, M., Gulli, M. P., Peter, K. and Peter, M. (1998). The Cdc42p effector Gic2p is targeted for ubiquitin-dependent degradation by the SCFGrr1 complex. *EMBO J*, 17(18):5360–73.
- Jarosch, E., Taxis, C., Volkwein, C., Bordallo, J. and Finley, D. et al. (2002). Protein dislocation from the ER requires polyubiquitination and the AAA-ATPase Cdc48. *Nat Cell Biol*, 4(2):134–9.
- Johnson, E. S., Ma, P. C., Ota, I. M. and Varshavsky, A. (1995). A proteolytic pathway that recognizes ubiquitin as a degradation signal. *J Biol Chem*, 270(29):17442–56.
- Kleijnen, M. F., Shih, A. H., Zhou, P., Kumar, S. and Soccio, R. E. et al. (2000). The hPLIC proteins may provide a link between the ubiquitination machinery and the proteasome. *Mol Cell*, 6(2):409–19.

- Lam, Y. A., Lawson, T. G., Velayutham, M., Zweier, J. L. and Pickart, C. M. (2002). A proteasomal ATPase subunit recognizes the polyubiquitin degradation signal. *Nature*, 416(6882):763–7.
- Lambertson, D., Chen, L. and Madura, K. (1999). Pleiotropic defects caused by loss of the proteasome-interacting factors Rad23 and Rpn10 of *Saccharomyces cerevisiae*. *Genetics*, 153(1):69–79.
- Leggett, D. S., Hanna, J., Borodovsky, A., Crosas, B. and Schmidt, M. et al. (2002). Multiple associated proteins regulate proteasome structure and function. *Mol Cell*, 10(3):495–507.
- Levchenko, I., Seidel, M., Sauer, R. T. and Baker, T. A. (2000). A specificity-enhancing factor for the ClpXP degradation machine. *Science*, 289(5488):2354–6.
- Li, Z. and Wang, C. C. (2002). Functional characterization of the 11 non-ATPase subunit proteins in the trypanosome 19 S proteasomal regulatory complex. *J Biol Chem*, 277(45):42686–93.
- Link, A. J., Eng, J., Schieltz, D. M., Carmack, E. and Mize, G. J. et al. (1999). Direct analysis of protein complexes using mass spectrometry. *Nat Biotechnol*, 17(7):676–82.
- Medicherla, B., Kostova, Z., Schaefer, A. and Wolf, D. H. (2004). A genomic screen identifies Dsk2p and Rad23p as essential components of ER-associated degradation. *EMBO Rep*, 5(7):692–7.
- Ortolan, T. G., Tongaonkar, P., Lambertson, D., Chen, L. and Schaubert, C. et al. (2000). The DNA repair protein rad23 is a negative regulator of multi-ubiquitin chain assembly. *Nat Cell Biol*, 2(9):601–8.

- Pickart, C. M. and Cohen, R. E. (2004). Proteasomes and their kin: Proteases in the machine age. *Nat Rev Mol Cell Biol*, 5(3):177–87.
- Raasi, S. and Pickart, C. M. (2003). Rad23 ubiquitin-associated domains (UBA) inhibit 26 S proteasome-catalyzed proteolysis by sequestering lysine 48-linked polyubiquitin chains. *J Biol Chem*, 278(11):8951–9.
- Rao, H. and Sastry, A. (2002). Recognition of specific ubiquitin conjugates is important for the proteolytic functions of the ubiquitin-associated domain proteins Dsk2 and Rad23. *J Biol Chem*, 277(14):11691–5.
- Saeki, Y., Saitoh, A., Toh-e, A. and Yokosawa, H. (2002a). Ubiquitin-like proteins and Rpn10 play cooperative roles in ubiquitin-dependent proteolysis. *Biochem Biophys Res Commun*, 293(3):986–92.
- Saeki, Y., Sone, T., Toh-e, A. and Yokosawa, H. (2002b). Identification of ubiquitin-like protein-binding subunits of the 26S proteasome. *Biochem Biophys Res Commun*, 296(4):813–9.
- Sakata, E., Yamaguchi, Y., Kurimoto, E., Kikuchi, J. and Yokoyama, S. et al. (2003). Parkin binds the Rpn10 subunit of 26S proteasomes through its ubiquitin-like domain. *EMBO Rep*, 4(3):301–6.
- Schauber, C., Chen, L., Tongaonkar, P., Vega, I. and Lambertson, D. et al. (1998). Rad23 links DNA repair to the ubiquitin/proteasome pathway. *Nature*, 391(6668):715–8.
- Seol, J. H., Feldman, R. M., Zachariae, W., Shevchenko, A. and Correll, C. C. et al. (1999). Cdc53/cullin and the essential Hrt1 RING-H2 subunit of SCF define a ubiquitin ligase module that activates the E2 enzyme Cdc34. *Genes Dev*, 13(12):1614–26.

- Skowyra, D., Koepp, D. M., Kamura, T., Conrad, M. N. and Conaway, R. C. et al. (1999). Reconstitution of G1 cyclin ubiquitination with complexes containing SCFGrr1 and Rbx1. *Science*, 284(5414):662–5.
- Swanson, R., Locher, M. and Hochstrasser, M. (2001). A conserved ubiquitin ligase of the nuclear envelope/endoplasmic reticulum that functions in both ER-associated and Matalpha2 repressor degradation. *Genes Dev*, 15(20):2660–74.
- Szlanka, T., Haracska, L., Kiss, I., Deak, P. and Kurucz, E. et al. (2003). Deletion of proteasomal subunit S5a/Rpn10/p54 causes lethality, multiple mitotic defects and overexpression of proteasomal genes in *Drosophila melanogaster*. *J Cell Sci*, 116(Pt 6):1023–33.
- Thrower, J. S., Hoffman, L., Rechsteiner, M. and Pickart, C. M. (2000). Recognition of the polyubiquitin proteolytic signal. *EMBO J*, 19(1):94–102.
- Tsai, B., Ye, Y. and Rapoport, T. A. (2002). Retro-translocation of proteins from the endoplasmic reticulum into the cytosol. *Nat Rev Mol Cell Biol*, 3(4):246–55.
- Van Nocker, S., Sadis, S., Rubin, D. M., Glickman, M. and Fu, H. et al. (1996). The multiubiquitin-chain-binding protein Mub1 is a component of the 26S proteasome in *Saccharomyces cerevisiae* and plays a nonessential, substrate-specific role in protein turnover. *Mol Cell Biol*, 16(11):6020–8.
- Verma, R., Annan, R. S., Huddleston, M. J., Carr, S. A. and Reynard, G. et al. (1997). Phosphorylation of Sic1p by G1 Cdk required for its degradation and entry into S phase. *Science*, 278(5337):455–60.
- Verma, R., Aravind, L., Oania, R., McDonald, W. H. and Yates III, J. R. et al. (2002). Role of Rpn11 metalloprotease in deubiquitination and degradation by the 26S proteasome. *Science*, 298(5593):611–5.

- Verma, R., Chen, S., Feldman, R., Schieltz, D. and Yates, J. et al. (2000). Proteasomal proteomics: Identification of nucleotide-sensitive proteasome-interacting proteins by mass spectrometric analysis of affinity-purified proteasomes. *Mol Biol Cell*, 11(10):3425–39.
- Verma, R., McDonald, H., Yates III, J. R. and Deshaies, R. J. (2001). Selective degradation of ubiquitinated Sic1 by purified 26S proteasome yields active S phase cyclin-Cdk. *Mol Cell*, 8(2):439–48.
- Verma, R., Oania, R., Graumann, J. and Deshaies, R. J. (2004). Multiubiquitin chain receptors define a layer of substrate selectivity in the ubiquitin-proteasome system. *Cell*, 118(1):99–110.
- Whittaker, C. A. and Hynes, R. O. (2002). Distribution and evolution of von Willebrand/integrin A domains: Widely dispersed domains with roles in cell adhesion and elsewhere. *Mol Biol Cell*, 13(10):3369–87.
- Wilkinson, C. R., Seeger, M., Hartmann-Petersen, R., Stone, M. and Wallace, M. et al. (2001). Proteins containing the UBA domain are able to bind to multi-ubiquitin chains. *Nat Cell Biol*, 3(10):939–43.
- Xie, Y. and Varshavsky, A. (2002). UFD4 lacking the proteasome-binding region catalyses ubiquitination but is impaired in proteolysis. *Nat Cell Biol*, 4(12):1003–7.
- Yao, T. and Cohen, R. E. (2002). A cryptic protease couples deubiquitination and degradation by the proteasome. *Nature*, 419(6905):403–7.
- Ye, Y., Meyer, H. H. and Rapoport, T. A. (2003). Function of the p97-Ufd1-Npl4 complex in retrotranslocation from the ER to the cytosol: Dual recognition of nonubiquitinated polypeptide segments and polyubiquitin chains. *J Cell Biol*, 162(1):71–84.

## B MS1, MS2, and SQT—Three Unified, Compact, and Easily Parsed File Formats for the Storage of Shotgun Proteomic Spectra and Identifications

This chapter describes adaptation of the file infrastructure used by `Sequest` (Eng et al. 1994) to the significant number of spectra produced in a MudPIT experiment.<sup>1</sup> `Sequest` was modified to use and produce the described file formats as described in Sadygov et al. (2002). J. G.'s contribution to the presented material was the Perl script `Unitemare` for the conversion of the original `Sequest` file formats into the ones described here and Perl scripting for the data presentation by `show`. This chapter was published as

McDonald, W. H., Tabb, D. L., Sadygov, R. G., MacCoss, M. J. and Venable, J. et al. (2004). MS1, MS2, and SQT—three unified, compact, and easily parsed file formats for the storage of shotgun proteomic spectra and identifications. *Rapid Commun Mass Spectrom*, 18(18):2162–2168.

The Copyright is held by John Wiley and Sons and reprinted here with permission.

### B.1 Abstract

As the speed with which proteomic labs generate data increases along with the scale

---

<sup>1</sup> E. g., 6×6000 spectra for a typical six 2 h chromatography cycles MudPIT experiment on Thermo-Electron's DecaXP ion trap mass spectrometer, acquiring a maximum of three fragmentation spectra after each full scan.

of projects they are undertaking, the resulting data storage and data processing problems will continue to challenge computational resources. This is especially true for shotgun proteomic techniques that can generate tens of thousands of spectra per instrument each day. One design factor leading to many of these problems is caused by storing spectra and the database identifications for a given spectrum as individual files. While these problems can be addressed by storing all of the spectra and search results in large relational databases, the infrastructure to implement such a strategy can be beyond the means of academic labs. We report here a series of unified text file formats for storing spectral data (MS1 and MS2) and search results (SQT) that are compact, easily parsed by both machine and humans, and yet flexible enough to be coupled with new algorithms and data-mining strategies.

## B.2 Introduction

Proteomic technologies are helping to change the scale at which biological experiments can be performed. Unfortunately, they also generate such voluminous data that they can result in a computational quagmire. Shotgun proteomic strategies, in which tandem mass spectra (MS/MS) are collected on mixtures of thousands of peptides, require the collection of tens of thousands of spectra (McCormack et al. 1997). Incorporating multidimensional separation strategies such as MudPIT (Multidimensional Protein Identification Technology) can easily balloon this into hundreds of thousands of spectra (Link et al. 1999; Washburn et al. 2001; Florens et al. 2002; and Peng et al. 2003). Spectra must be stored for identification via database search software such as SEQUEST. In many cases, multiple identification strategies and even

algorithms are applied in the course of a complete analysis (MacCoss et al. 2002; and Gatlin et al. 2000).

Since its inception as the file format recognized by the SEQUEST database search algorithm (Eng et al. 1994), the DTA file has seen widespread use as a format to store individual MS/MS spectra. It and the SEQUEST-generated OUT file were sufficient for experiments producing merely hundreds of spectra. However, as the scope and complexity of these experiments have expanded, the limitations of these files have become increasingly evident; the number of files produced for individual experiments makes directory management problematic, and the storage space wasted in these formats is problematic as well. To deal with some of these limitations we have developed and implemented a new set of unified file formats that are simple, compact, and yet retain their flexibility.

There were several major goals associated with moving towards unified formats. First, unified formats dramatically reduced the number of files required to represent a proteomic data set. This was important because the huge number of files (hundreds of thousands) added to the file servers each week were taxing their stability and exceeded file system limitations. Second, the new formats reduced the amount of storage space used. Many of the individual files were small enough to be below the minimum block size limit for a file (typically 4 or 8). Thus, simply concatenating the files together reduced the total amount of disk space required to store this information. Third, switching to unified file formats enabled greater efficiency in data storage by removing fields that had been repeated in each file and by grouping data that had been distributed to multiple files. Finally, the unified formats were formatted for automated parsing and designed for extensibility. In keeping with this final goal, the formats had to be adaptable to existing programs and able to

accommodate future code developments.

In order to accomplish these goals, we adopted three unified file formats—MS1, MS2, and SQT. All of these store their particular information type for an entire experimental step, e. g., an entire LC/MS/MS experiment or a single salt step from a MudPIT run. The naming convention is simple; the base filename of the instrument-generated file is used with a new extension. For instance, the ThermoFinnigan LCQ file `salt_step.RAW` would generate `salt_step.ms1` and `salt_step.ms2` files and, after database searching, would yield a `salt_step.sqt` file.

The MS1 file contains full-scan data and is used for analyses that require this type of data such as quantitation or measurement of chromatographic efficiency. The MS2 file stores MS/MS data and replaces a folder of thousands of DTA files; it contains all the spectral information necessary for database searching algorithms. Finally, the SQT file unifies the database search results. While initially designed to replace the SEQUEST OUT file, it has proven flexible enough to work quite well with other algorithms used in the lab, e. g., PEP\_PROBE (Sadygov and Yates 2003) and GutenTag (Tabb et al. 2003).

### B.3 Format Descriptions

These formats were intended to store all necessary information in as compact and accessible a format as possible while retaining human legibility. In general, they contain information generic to all records in a header at the start of the file while data for specific spectra are stored individually through the body of the file. Unless specifically noted, all fields are tab delimited within an individual line. A compre-

hensive description of these formats with examples is available on our website (<http://fields.scripps.edu/sequest/unified>).

The MS1 format is the simplest of the three. It contains four types of lines: H, S, I, and [m/z intensity]. The header is defined by a series of H lines. Each line includes a field label and its corresponding value (string, integer, or floating point). The following fields are required in the header: **CreationDate**, **Extractor**, **ExtractorVersion**, and **ExtractorOptions**. The values for these are the date the file was created, the program used, its version, and any specific options used in the program (fig. B.1). Some optional field labels include: **InstrumentType** (ion trap, q-tof, tof-tof, etc.), **InstrumentSN** (serial number), and **Comment** (other information general to the file).

In the body of the file, each full scan in the experiment begins with an S line which contains the scan number. This can be followed by the optional I line which can be used to store any ancillary information such as retention time (**I RTime 33.2**). Next comes a series of [m/z intensity pairs] (space separated) representing the spectral data for that entire full mass scan. This pattern (**S, [m/z intensity]<sup>n</sup>** or **S, I, [m/z intensity]<sup>n</sup>**) is repeated for each full mass scan in the experiment. If necessary, multiple I lines can be used for a given spectrum (e. g., LC retention time).

The format for the MS2 file is similar to the MS1 file except that MS/MS spectra are stored. It shares the H, S, I, and [m/z intensity] lines with the MS1 file but adds two additional lines, Z and D, to store charge-state-dependent information. The header itself has additional field labels such as **IAnalyzer**, to denote a program which does not consider the charge state of the precursor ion (e. g., spectral quality filtering), and **DAnalyzer**, to denote a program that analyzes charge-state-specific

## A

symbol	meaning	generic form		required
H	header	H	[field label] [corresponding value]	yes
S	scan	S	[first scan] [second scan]	yes
I	charge independent analysis	I	[field label] [corresponding value]	
[data]	mass intensity pairs	[m/z]	[intensity]	yes
<b>"H" line required field labels</b>				
<b>field label</b>	<b>description</b>			
CreationDate	The date and time on which this file was created.			
Extractor	The name of the software used to create the MS2 file.			
ExtractorVersion	The version number of the Extractor software.			
ExtractorOptions	The options used in running the Extractor software.			
<b>"H" line optional field labels</b>				
<b>field label</b>	<b>description</b>			
IAnalyzer	If software to conduct charge state-independent analysis of the spectra is used to analyze the MS2 file (such as removing sparse spectra), the name of the program should be recorded in the header.			
IAnalyzerVersion	The version number of the IAnalyzer			
IAnalyzerOptions	The options used for the IAnalyzer			
InstrumentType	The type of mass analyzer used.			
Comment	Remarks, ownership, and copyright information may be included. Multiple comment lines may be included.			
InstrumentSN	The serial number of the mass spectrometer used.			

## B

```

H   CreationDate   07/10/2003
H   Extractor      RAWXtract
H   ExtractorVersion  1.0
H   ExtractorOptions
S   0001   0001
I   RTime    0.01
400.3 828908
400.9 695425
401.9 683908
403.1 556803
405.8 906291
407.2 293691
408.2 2475720
409.4 305360
412.4 324123
413.6 1974477
414.8 2339087
415.6 242656
416.5 1402914
417.4 1635225
418.1 354
419.8 681890
420.5 63
[data continued ...]

```

**Figure B.1 MS1 File Format Description.** (A) General description of required fields and format used in the MS1 file. Both required and optional lines and field descriptions are noted along with a generic pattern for data storage. (B) Example MS1 file and a partial spectrum. Following the H lines of the header each full-scan spectrum begins with an S line denoting its scan number. Next, optional I lines give additional information about that scan such as retention times. Finally, the spectral data are stored in as series of m/z intensity pairs. This pattern of S(I)[m/z intensity]<sup>n</sup> continues until each full-scan spectrum has been represented.

features (e. g., charge–state discrimination or neutral losses off the precursor). Any specific features noted by the `DAnalyzer` programs are annotated in the D line following their specific Z line (see below). One advantage of the MS2 file format is that the file format logs which algorithms have been applied serially to the file.

The MS2 file body is structured similarly to the MS1 except for the addition of the Z and D lines. The description of each spectrum begins with the S line which has fields for the `[start scan]`, `[end scan]`, and `[precursor m/z]`. This can be followed by the optional I line which contains a datum or analytical result that is independent of the charge–state prediction of the precursor, such as a spectral quality score or instructions to the search program not to query this particular spectrum. Next comes the Z line with the `[charge state]` and `[predicted [M+H]+]` fields. The optional D line may follow and can be used to store information specific to the charge state of the preceding Z line. This can include annotations of a particular structural feature that might necessitate the use of a different search algorithm, e. g., neutral loss of phosphoric acid off of the precursor as an indication of a phosphorylated peptide. There can be multiple Z and optional D lines for a given spectrum depending on how well the precursor charge state is able to be discriminated. The spectral data are stored in the same manner as the MS1 file with Z and `[m/z intensity]` stored for every peak in the experimental spectrum. The minimum pattern to represent a spectrum is S, Z, `[m/z intensity]`<sup>n</sup>, but, as previously mentioned, can also contain multiple Z lines and the optional I and D lines to encompass additional information. One obvious advantage of this format over the DTA format is that to have the search algorithm to consider an additional charge state, all one has to do is add another Z line rather than producing a separate DTA file (currently each charge state to be considered has a corresponding DTA file), with all of the `[m/z`

`intensity]`<sup>n</sup> information repeated.

The **SQT** file format is a greater departure from the **SEQUEST OUT** files it replaces. The design aims were to provide the same information as reported in the **OUT** file but to do so in a more compact, more easily parsed format while retaining a degree of human legibility. It is comprised of **H**, **S**, **M**, and **L** lines. The header lines are similar in format to both the **MS1** and **MS2** files, except that it has its own distinct set of field labels and values. Figure B.2 shows the required fields and an example of how they are employed. Invariant information usually stored in each **OUT** file is stored just once in the header of the **SQT** file. The data characterizing a particular identification are stored in a block of lines lower in the file.

Each search result for a spectral entry is denoted by the following generic line pattern: **S**(**M**(**L**)<sup>k</sup>)<sup>n</sup> (fig. B.3). The **S** line contains information specific for that spectrum and search. It is followed by an **M** line which describes a particular matching sequence along with its characteristic scores. Next comes at least one **L** line that notes which protein in the database contains this particular peptide sequence; there can be multiple **L** lines depending on how many proteins within the database contain the matched peptide sequence. The **M** line for the second highest scoring peptide match is followed by its respective **L** line(s). This pattern of **M** and **L** lines continues for as many search results as were set to be stored in the search parameters (typically 5–10). The **SQT** file also allows the inclusion of a column which stores manual evaluation information. The state can be either the default of **U** (unevaluated), **Y** (yes), **N** (no), or **M** (maybe). The inclusion of this field allows the **SQT** file format to store manual validation information that cannot be stored in **OUT** files.

To institute such file format changes we made modifications to several pre-search (extraction and filtering), searching (**SEQUEST**), organization and summary

A

symbol	meaning	generic form		required
H	header	H	[field label] [corresponding value]	yes
S	scan	S	[first scan] [second scan]	yes
I	charge-independent analysis	I	[field label] [corresponding value]	
Z	charge	Z	[charge]	yes
D	charge-dependent analysis	D	[field label] [corresponding value]	
[data]	mass intensity pairs	[m/z]	[intensity]	yes
<b>"H" line required field labels</b>				
<b>field label</b>	<b>description</b>			
CreationDate	The date and time on which this file was created.			
Extractor	The name of the software used to create the MS2 file.			
ExtractorVersion	The version number of the Extractor software.			
ExtractorOptions	The options used in running the Extractor software.			
<b>"H" line optional field labels</b>				
<b>field label</b>	<b>description</b>			
IAnalyzer	If software to conduct charge state-independent analysis of the spectra is used to analyze the MS2 file (such as removing sparse spectra), the name of the program should be recorded in the header.			
IAnalyzerVersion	The version number of the IAnalyzer			
IAnalyzerOptions	The options used for the IAnalyzer			
DAnalyzer	If software to conduct charge state-dependent analysis of the spectra is used to analyze the MS2 file (such as removing possible precursor charge states), the name of the program should be recorded in the header.			
DAnalyzerVersion	The version number of the DAnalyzer			
DAnalyzerOptions	The options used for the DAnalyzer			
SortedBy	If a program is used to sort the spectra, which field are they sorted by?			
InstrumentType	The type of mass analyzer used.			
Comment	Remarks, ownership, and copyright information may be included. Multiple comment lines may be included.			
InstrumentSN	The serial number of the mass spectrometer used.			

B

```

H          CreationDate      unknown
H          Extractor         RAWXtract
H          ExtractorVersion   1.0
H          ExtractorOptions
S  0002    0002    534.04
I  RTime   0.02
Z  2       1067.08
Z  3       1600.12
151.0 22869
153.0 14453
155.1 35721
157.1 17102
158.1 12948
166.3 24733
167.0 103456
168.0 11485
171.0 92158
172.1 62988
173.1 136408
[data continued ...]

```

**Figure B.2 MS2 File Format Description.** (A) General description of required fields and format used in the MS2 file. The format follows the general conventions of the MS1 file format except that MS/MS information is stored and with the addition of the required Z lines and the optional D lines. (B) Example MS2 file and partial spectrum. As with the MS1 file, each spectral description in the MS2 file begins with an S line. Z lines denote which charge states are to be considered for the spectrum. Like the MS1 file, the spectrum itself is represented as a series of [m/z intensity] pairs. Optional I and D lines can be used to store charge-state-independent and charge-state-dependent information, respectively. The general pattern  $S(I)[Z(D)]^k[m/z \text{ intensity}]^n$  continues until all MS/MS spectra are represented.

**A**

symbol	meaning	generic form	required
H	header	H [field label] [corresponding value(s)]	yes
S	spectrum	S [low-scan] [high-scan] [charge] [process time] ...	yes
		... [server] [obs. Mass] [lowest SP] [# sequences match]	
M	match	M [rank by Xcorr] [Rank by Sp] [calculated mass] ...	yes
		... [DeltCN] [Xcorr] [Sp] [matched ions] [expected ions] ...	
		... [sequence matched] [validation status]	
L	locus	L [locus] [description (if contained in database)]	yes
<b>"H" line required field labels</b>			
<b>field label</b>	<b>description</b>		
SQTGenerator	Which program created this SQT file?		
SQTGeneratorVersion	What revision of the program was used to produce this SQT?		
Database	What was the path and filename of the database used to produce this SQT? Multiple databases may be listed here.		
FragmentMasses	Were average or monoisotopic residue masses used to predict the fragment ion masses?		
PrecursorMasses	Where average or monoisotopic residue masses used to predict the precursor ion mass? ("AVG" and "MONO")		
StartTime	When was this file initiated?		
StaticMod	If any nonstandard amino acid masses were used in identification, they must be listed here. If multiple static modifications were used, multiple StaticMod lines should be present.		
DynamicMod	If any dynamic modifications were sought in identification, they must be listed here.		
<b>"H" line optional field labels</b>			
<b>field label</b>	<b>description</b>		
Comment	Remarks, ownership, and copyright information may be included. Multiple comment lines may be included.		
DBSeqLength	How many amino acids appear in the sequence database?		
DBLocusCount	How many protein sequences appear in the sequence database? This field must follow the same rules as DBSeqLength.		
DBMDSum	To ensure that the same database is currently present as when this search was conducted, the MDSSum of the database may be stored in the SQT header. This field must follow the same rules as DBSeqLength.		
SortedBy	If the IDs in this file have been sorted, which field was used?		
Alg-	Other fields may be added to the SQT header. Any algorithm-specific fieldname should begin with "Alg-" to prevent parsing errors. Field names may not include whitespace characters.		

**B**

```

H SQTGenerator SEQUEST
H SQTGeneratorVersion 2.7
H Comment SEQUEST was written by J Eng and JR Yates, III
H Comment SEQUEST ref. J. Am. Soc. Mass Spectrom., 1994, v. 4, p. 976
H Comment SEQUEST ref. Eng, J.K.; McCormack A.L.; Yates J.R.
H Comment SEQUEST is licensed to Finnigan Corp.
H Comment Parallelization Program is run_ms2
H Comment run_ms2 was written by Rovshan Sadygov
H StartTime 07/10/2003, 08:00 PM
H EndTime 07/10/2003, 08:00 PM
H Database /wfs/dbase/nci/MAPdb.nci
H DBSeqLength 901292
H DBLocusCount 1180
H PrecursorMasses AVG
H FragmentMasses AVG
H Alg-PreMassTol 3.000
H Alg-FragMassTol 0.0
H Alg-XCorrMode 0
H StaticMod C=160.139
H DiffMod ST*+=80.000
H Alg-MaxDiffMod 4H Alg-DisplayTop 5
H Alg-IonSeries 0 1 1 0.0 1.0 0.0 0.0 0.0 0.0 0.0 1.0 0.0^M
S 0012 0012 1 1 shamu40 659.57 3716.5 73.8 111330
M 1 1 658.600 0.0000 0.0000 168.4 7 10 S.LSSNGT*.N U
L GP:AF075587_1
M 1 2 656.627 0.0000 0.0000 155.1 7 12 P.AALGSAS*.A U
L SW:DHE3_HUMAN
M 1 3 657.656 0.0000 0.0000 147.4 7 10 K.VTGLST*.R U
L GP:HS511B24_3
M 1 4 656.673 0.0000 0.0000 146.5 8 16 K.PTGGPGGGG.T U
L GP:HUMRCK_1
L PIR2:S22651
L SW:DDX6_HUMAN

```

**Figure B.3 SQT File Format Description.** (A) General description of lines, required fields, and generic format for the SQT file. (B) An example SQT file header with a portion of search results for an MS/MS spectrum. The search results for each spectrum start with an S line which contains the scan numbers and certain other metrics relating to that spectrum. The first M line gives the search results for the highest scoring peptide in the database and is followed by one or more L lines that give locus names for the proteins in the database in which that peptide could be found. M and L line combinations are given for the remaining recorded search results for that spectrum. Results for each searched spectrum are recorded in the general format: S[ML<sup>k</sup>]<sup>n</sup>.

(DTASelect, Tabb et al. 2002), and visualization (results display) programs. For spectral extraction we have the Linux-compiled, `makems2` to extract MS/MS from `a.dat` (ICIS) file into an MS2 format and perform limited charge-state selection and filtering. It is basically a unified format version of the `extractms` program. For spectral extraction directly from RAW files, we have `MSMaker` which extracts both MS and MS/MS spectra, but does only limited charge-state selection. Charge-state selection is done primarily using `2to3u` (Sadygov et al. 2002). `DTASelect` has been designed to accommodate a variety of file formats, including these new ones. For visualization purposes we developed a new CGI, `show` (<http://fields.scripps.edu/sequest/show/index.html>), which gathers information from both the SQT and MS2 files and passes them to an applet version of the `DTASelect` ion display graphical interface. It is also back compatible with DTA, OUT, and a variety of intermediate file formats. Finally, we developed a PERL script, `Unitemare`, which transcodes previously searched DTAs and OUTs into the new unified formats. With the exception of `SEQUEST`, these programs are freely available to academic and other nonprofit groups; see the group website for details (<http://fields.scripps.edu>).

For performance comparisons a single MudPIT cycle was chosen from a six-cycle analysis of a previously described protein mixture (McDonald et al. 2002). These data were extracted either into DTA or MS2 format without filtering and only rudimentary charge-state discrimination. The 2886 nonblank MS/MS spectra generated a total of 5642 DTA files to be searched (as a result of the need to store a separate DTA file if multiple charge states were to be considered). Both formats were searched using `SEQUEST` against the same database using identical settings to generate OUT and SQT files. Disk usage measures were performed either using the Linux `du -b` command or folder properties in WindowsXP.

## B.4 Results and Discussion

One of the primary design goals for these files was to reduce the number of files which must be stored on the hard drive. Clearly, this was accomplished since the MS/MS spectra and search results of a typical MudPIT cycle were able to be stored in two files rather than the 5642 DTAs and their corresponding 5642 OUTs present in the example file. While it is difficult to quantify the impact that this has had on the stability of our file servers, anecdotally we noted a dramatic increase in uptime with a concomitant decrease in errors (primarily network file system, NFS, errors). We experienced this increased stability in spite of going from 1/3 terabytes of data to > 1.5 terabytes of data stored on our two Linux file servers.

A potential problem with aggregating results into a single file is that accessing a specific spectrum or search result can be much slower than having them split out as single files. We tried to address this in two ways. One was through the line tags that preclude the need for complex matching strategies across an entire line. For instance to find a particular spectrum, one need only consider lines starting with the S token. Another is that the files were kept as streamlined as possible. Sorting the files to place the highest scoring identifications and their corresponding spectra to the top of the files (**SQTSort**, <http://fields.scripps.edu/sequest/SQTSort.html>) enabled even faster access to the most relevant data. For speed and even more compactness, we are exploring the possibilities of having indexed binary versions of these file formats. In addition, a conscious decision was made to store only the necessary information and not bloat them with information that, while likely to be useful at some future point, was either present in the initial instrument file or more efficiently stored in a separate file within a given experimental folder.

The next goal was to reduce the total disk space required to store the files. Simply concatenating the spectral and search result files would be predicted to help

**Table B.6a Savings Per File Type**

		Bytes	Saved
MS/MS Data	DTAs	35 115 008	

**Table B.7b Savings Per File Type**

	MS2	10 207 232	70.93 %
Search Results	OUTs	23 355 392	
	SQT	3 858 432	83.48 %

since many of these files were smaller than the block size or minimum file size. This could be seen by measuring the total disk usage for a folder of DTA and OUT files, 58 908 672 bytes total, versus the total usage if all the DTA files and OUT files were put together into two files, 35 729 408 bytes. However, the redundancy of the headers in the OUT files and the file repetitions needed to represent multiple charge states allowed a total saving of about 75 k% for the SQT and MS2 file formats (Table B.8). The largest reduction in disk space requirements was seen in going from the OUT files to the SQT file, an 83 % saving (Table B.6). Similar differences were seen when the files were stored on a `WindowsXP` machine (NTFS formatted partitions) (Table B.8). These savings scaled proportionally to the number of LC/MS/MS runs (data not shown). Since it is now trivial to collect 100s of gigabytes of data in a relatively short period of time, being able to store data in one-quarter of the space is significant, especially for those academic labs that are unable to afford or maintain multiterabyte storage arrays. As newer and faster instruments emerge these considerations will become even more significant; for instance, the next generation ThermoFinnigan linear ion trap (LTQ) collects about four and a half times as many spectra as the LCQ in the same amount of time.

Another advantage we have noted is that programs which have to read through every result in a given experiment, such as `DTASelect`, are able to parse through

**Table B.8a Total Savings**

	Linux	Saved	WinXP	Saved
DTAs and OUTs	58 462 208		59 469 923	

**Table B.9b Total Savings**

Concatenated	35 274 752	39.66 %	36 630 528	38.40 %
MS2 and SQT	14 061 566	75.95 %	14 987 264	74.80 %

these files more quickly. For those same 5642 OUT files, DTASelect required 4.2 s to read through and gather all the necessary information. The corresponding SQT files were parsed in a less than a second. Again, while not substantial for a single file, this parsing can take a great deal of time when the dataset consists of > 50 MudPIT cycles which in turn were searched against multiple databases (Florens et al. 2002). In comparison to their predecessors, the design of the unified file formats makes it relatively easy to develop software to read them, especially when compared to the difficulty of dealing with all of the subtle differences in OUT files produced by different SEQUEST revisions.

These file formats also allowed for a more streamlined, logical, and flexible workflow. First, the MS2 file logs the serial application of multiple programs during the data analysis workflow. After initial extraction of the various files, these could include spectral quality filtering/scoring, charge–state selection, and feature annotation. Which programs have been run is stored as header information and the I and D lines provide the flexibility to annotate specific spectra and/or specific charge states of those spectra. This expansion room without compromising file size is an important feature that was missing from the original DTA file format.

Another workflow advantage can be seen when one wishes to analyze the same data versus multiple databases or for a variety of posttranslational modifications

(e. g., see MacCoss et al. 2002). Instead of having to search copies of the DTA files it becomes quite practical to use symbolic links back to the original MS2 files. Even more dramatic space savings can be realized under such a scenario. Using symbolic links to the MS2 files, six different searches could be performed on our example LC/MS/MS run with a total savings of 94 % over DTAs and OUTs (352 megabytes vs. 22 megabytes). In fact, after all of these searches have been performed, it is possible to collate the various answers back into a single aggregate SQT file. Extensive analysis of complex datasets requires flexible formats to bring these results together into an easily digested final output. The MS1 file format, for instance, could be used to extract full-scan chromatograms of individual peptides for purposes of quantitation or characterizing chromatographic efficiency.

Ongoing discussions seek to standardize file formats for proteomic data, with the ultimate goal of moving towards a common database schema that can be employed globally (Orchard et al. 2003; and Taylor et al. 2003). However, there is an evident need for an intermediate step moving from either the single spectrum or proprietary instrument manufacturer formats to this ultimate goal. Several groups are proposing moving towards common XML (extensible markup language) formats in order to store all the data concerning a particular experiment. There are many advantages to this idea in terms of tools available to deal with XML data, and of course, a common language spoken by all proteomics labs. However, XML files typically spend many bytes on formatting information, potentially increasing rather than decreasing storage capacity requirements. The extensibility to new instruments and experimental strategies possible with XML formatting may not prove an adequate gain for the cost in file size.

The MS1, MS2, and SQT file formats do not meet all of the goals of these upcoming

database and XML standards, and are not intended to substitute for them. However, since they are compact, flexible, easily parsed, and mature in their implementation, we propose that they can serve a very useful role in the proteomic mass spectrometry community. They should be particularly appealing to small-scale labs that are still able to generate large volumes of data, but have necessarily limited computational and storage resources. The ease with which these files can be parsed and the existing suite of tools under continuing development in our group and others make them an appealing platform. The use of tab-delimited text files makes the creation of translation software to produce other formats of data trivial, allowing export to whatever industry standards are ultimately adopted. We propose that they are a viable alternative to an XML-based single file format because of advantages in disk space required, developmental flexibility, and ease in later translation to an industry-standard XML or database format for dissemination and sharing of data.

## B.5 References

- Eng, J. K., McCormack, A. L. and Yates III, J. R. (1994). An approach to correlate tandem mass-spectral data of peptides with amino-acid sequences in a protein database. *J Am Soc Mass Spectr*, 5(11):976–989.
- Florens, L., Washburn, M. P., Raine, J. D., Anthony, R. M. and Grainger, M. et al. (2002). A proteomic view of the *Plasmodium falciparum* life cycle. *Nature*, 419(6906):520–6.
- Gatlin, C. L., Eng, J. K., Cross, S. T., Detter, J. C. and Yates III, J. R. (2000). Automated identification of amino acid sequence variations in proteins by HPLC/microspray tandem mass spectrometry. *Anal Chem*, 72(4):757–63.

- Link, A. J., Eng, J., Schieltz, D. M., Carmack, E. and Mize, G. J. et al. (1999). Direct analysis of protein complexes using mass spectrometry. *Nat Biotechnol*, 17(7):676–82.
- MacCoss, M. J., McDonald, W. H., Saraf, A., Sadygov, R. and Clark, J. M. et al. (2002). Shotgun identification of protein modifications from protein complexes and lens tissue. *Proc Natl Acad Sci USA*, 99(12):7900–5.
- McCormack, A. L., Schieltz, D. M., Goode, B., Yang, S. and Barnes, G. et al. (1997). Direct analysis and identification of proteins in mixtures by LC/MS/MS and database searching at the low-femtomole level. *Anal Chem*, 69(4):767–76.
- McDonald, W. H., Ohi, R., Miyamoto, D. T., Mitchison, T. J. and Yates III, J. R. (2002). Comparison of three directly coupled hplc ms/ms strategies for identification of proteins from complex mixtures: Single-dimension lc-ms/ms, 2-phase mudpit, and 3-phase mudpit. *Int J Mass Spectrom*, 219(1):245–251.
- McDonald, W. H., Tabb, D. L., Sadygov, R. G., MacCoss, M. J. and Venable, J. et al. (2004). MS1, MS2, and SQT—three unified, compact, and easily parsed file formats for the storage of shotgun proteomic spectra and identifications. *Rapid Commun Mass Spectrom*, 18(18):2162–2168.
- Orchard, S., Hermjakob, H. and Apweiler, R. (2003). The proteomics standards initiative. *Proteomics*, 3(7):1374–6.
- Peng, J., Schwartz, D., Elias, J. E., Thoreen, C. C. and Cheng, D. et al. (2003). A proteomics approach to understanding protein ubiquitination. *Nat Biotechnol*, 21(8):921–6.

- Sadygov, R. G., Eng, J., Durr, E., Saraf, A. and McDonald, H. et al. (2002). Code developments to improve the efficiency of automated MS/MS spectra interpretation. *J Proteome Res*, 1(3):211–5.
- Sadygov, R. G. and Yates III, J. R. (2003). A hypergeometric probability model for protein identification and validation using tandem mass spectral data and protein sequence databases. *Anal Chem*, 75(15):3792–8.
- Tabb, D. L., McDonald, W. H. and Yates III, J. R. (2002). DTASelect and Contrast: Tools for assembling and comparing protein identifications from shotgun proteomics. *J Proteome Res*, 1(1):21–6.
- Tabb, D. L., Saraf, A. and Yates III, J. R. (2003). GutenTag: High-throughput sequence tagging via an empirically derived fragmentation model. *Anal Chem*, 75(23):6415–21.
- Taylor, C. F., Paton, N. W., Garwood, K. L., Kirby, P. D. and Stead, D. A. et al. (2003). A systematic approach to modeling, capturing, and disseminating proteomics experimental data. *Nat Biotechnol*, 21(3):247–54.
- Washburn, M. P., Wolters, D. and Yates III, J. R. (2001). Large-scale analysis of the yeast proteome by multidimensional protein identification technology. *Nat Biotechnol*, 19(3):242–7.

## C Analysis of Polyubiquitin Conjugates Reveals that the Rpn10 Substrate Receptor Contributes to the Turnover of Multiple Proteasome Targets

The publication reprinted here represents the use of MudPIT for a complex mixture of low abundant polypeptides: purified polyubiquitin conjugates from *Saccharomyces cerevisiae*. J. G.'s contributed introduction of T. M. and G. T. S. to MudPIT and data analysis. The work was published as

Mayor, T., Lipford, J., Graumann, J., Smith, G. and Deshaies, R. (2005). Analysis of polyubiquitin conjugates reveals that the rpn10 substrate receptor contributes to the turnover of multiple proteasome targets. *Mol Cell Proteomics*, 4(6):741–51.

The copyright is held by the American Society for Biochemistry and Molecular Biology. The reprint here is authorized. The supplementary material referred to can be found at <http://www.mcponline.org/cgi/content/full/M400220-MCP200/DC1>.

### C.1 Abstract

The polyubiquitin receptor Rpn10 targets ubiquitylated Sic1 to the 26S proteasome for degradation. In contrast, turnover of at least one ubiquitin–proteasome system (UPS) substrate, CPY\*, is impervious to deletion of *RPN10*. To distinguish whether *RPN10* is involved in the turnover of only a small set of cell cycle regulators that includes Sic1 or plays a more general role in the UPS, we sought to develop a general

method that would allow us to survey the spectrum of ubiquitylated proteins that selectively accumulate in *rpn10Δ* cells. Polyubiquitin conjugates from yeast cells that express hexahistidine-tagged ubiquitin (H<sub>6</sub>-ubiquitin) were first enriched on a polyubiquitin binding protein affinity resin. This material was then denatured and subjected to IMAC to retrieve H<sub>6</sub>-ubiquitin and proteins to which it may be covalently linked. Using this approach, we identified 127 proteins that are candidate substrates for the 26S proteasome. We then sequenced ubiquitin conjugates from cells lacking Rpn10 (*rpn10Δ*) and identified 54 proteins that were uniquely recovered from *rpn10Δ* cells. These include two known targets of the UPS, the cell cycle regulator Sic1 and the transcriptional activator Gcn4. Our approach of comparing the ubiquitin conjugate proteome in wild-type and mutant cells has the resolving power to identify even an extremely inabundant transcriptional regulatory protein and should be generally applicable to mapping enzyme substrate networks in the UPS.

## C.2 Introduction

In eukaryotic cells, protein degradation plays a critical role in the regulation of a variety of cellular processes including the cell cycle, apoptosis, signal transduction, and gene expression. The ubiquitin-proteasome system (UPS) is the principal pathway that targets proteins for degradation. In this pathway, proteins to be degraded are marked by covalent modification of a lysine residue with an ubiquitin chain. The enzymatic reaction (ubiquitylation) is driven by an ubiquitin-activating enzyme E1, ubiquitin-conjugating enzyme E2, and ubiquitin-ligase E3 (Weissman 2001). The substrate conjugated to the ubiquitin chain is then recognized by the 26S proteasome

and degraded. The exquisite specificity of substrate recognition for ubiquitylation is thought to be determined primarily by E3, which binds specifically to substrate (Orlicky et al. 2003; and Wu et al. 2003). The budding yeast genome encodes about 50 putative ubiquitin–ligases,<sup>1</sup> whereas metazoans may have more than 400 (Semple 2003). Because each ubiquitin–ligase presumably can target several substrates, ubiquitylation represents one of the main posttranslational modifications in the cell. Therefore, deciphering the network of enzyme–target interactions in the UPS will be a major undertaking.

To be recognized by the proteasome, a substrate–linked ubiquitin chain must assemble through lysine 48 (Lys<sup>48</sup>) of ubiquitin (Chau et al. 1989). By contrast, mono–ubiquitin linkages and multiubiquitin chains linked via the alternative lysine 63 (Lys<sup>63</sup>) of ubiquitin regulate multiple pathways by nonproteolytic means, including DNA repair (Hoege et al. 2002), chromatin topology, and vesicle trafficking (Hicke 2001). In the past few years, several proteins that recognize specifically Lys<sup>48</sup>–linked chains have been identified. Rpn10, a stoichiometric component of the 26S proteasome, was the first protein shown to bind polyubiquitin chains (Deveraux et al. 1994). Rpn10 harbors two characterized domains: the amino–terminal von Willebrand A (VWA) domain that mediates proteasome association and the carboxyl–terminal ubiquitin–interacting motif (UIM) domain. The UIM is also present in other proteins involved in the ubiquitin pathway and endocytosis (Hofmann and Falquet 2001). Based on its ability to bind to the proteasome and to ubiquitylated proteins, Rpn10 was predicted to be the major proteasome receptor for ubiquitylated substrates. However, deletion of *RPN10* in budding yeast is

---

<sup>1</sup> T. Mayor and R. J. Deshaies, unpublished data.

not lethal, indicating that other proteins must act as proteasome receptors (Fu et al. 1998). Rad23 and Dsk2 belong to a second group of proteins that interact with the proteasome via an amino-terminal ubiquitin-like domain and contain a carboxyl-terminal polyubiquitin binding motif, the ubiquitin-associated (UBA) domain. There is evidence suggesting that both proteins can act as proteasome receptors (Wilkinson et al. 2001; and Rao and Sastry 2002). There is also other evidence that suggests these two proteins may play an alternative role in protecting ubiquitylated substrates from deubiquitylation activity and in promoting or inhibiting multiubiquitylation of substrates (Kim et al. 2004; Ortolan et al. 2000; Raasi and Pickart 2003; and Hartmann-Petersen et al. 2003). Whereas the physiological functions of ubiquitin binding proteins remain to be fully elucidated, a recent study showed that mutations in *RPN10*, *RAD23*, or *UFD1* (Ufd1 is a member of a protein complex that may also act as a proteasome substrate receptor) selectively impair the turnover of distinct substrates of the UPS (Verma et al. 2004). This surprising finding implies that different targeting mechanisms are used by the proteasome to degrade specific subsets of substrates. Certain UPS substrates (Sic1, Clb2, and Gic2) but not others (CPY\* and the Deg1 degron of Mata2) are strongly influenced by Rpn10 (Verma et al. 2004). This suggested that a restricted class of UPS substrates, possibly short-lived regulators of the cell cycle and its efferent pathways, is targeted to the proteasome by Rpn10.

Here, we employ a new method for ubiquitin conjugate affinity purification to identify proteins that accumulate as ubiquitylated species in yeast cells that lack Rpn10. Our analysis greatly expands the role that Rpn10 plays in protein turnover *in vivo*. By applying the approach described here, it should be possible to systematically identify the constellation of substrates targeted to the proteasome by each

individual receptor pathway in *Saccharomyces cerevisiae*.

## C.3 Experimental Procedures

### C.3.1 Yeast Strains and Plasmids

All *S. cerevisiae* strains used in this study are listed in supplemental Table 1. Strain RJD 2997 was generated by integrating the plasmid RDB 1848, which contains the coding sequences for H<sub>6</sub>-ubiquitin flanked by the GPD constitutive promoter and PGK terminator sequences (Mayor and Deshaies 2005), into the *TRP1* locus. Control strain RJD 2998 was obtained by integrating the empty vector into the *TRP1* locus. Mutant *rpn10*Δ was retrieved from the Yeast Deletion Library (Open Biosystems) and back crossed into the W303 background. Gcn4-Myc9 was previously described (Chi et al. 2001). S288C strains with TAP-tagged genes were retrieved from the Yeast TAP-Fusion Library (Open Biosystems).

The H<sub>8</sub>-ubiquitin coding sequence was placed between the GPD constitutive promoter and PGK terminator sequences in pRS 316 (RDB 1851). A pair of primers (5'-GCGGATCCATGAGAGGTAGTCACCACCATCATCACCATCATCACGGTGGTATGCAGATTTTCG-3' and 5'-GAGCTCGAGACCACCTCTTAGCCTTAGCAC-3') was used to amplify by PCR yeast ubiquitin (the first repeat of the *UBI4* locus). The PCR fragment was digested with BamHI and XhoI and ligated into the yeast expression vector pG-1 (digested with BamHI and Sall; Schena et al. 1991). An EcoRI-NaeI fragment containing H<sub>8</sub>-ubiquitin was then ligated into pRS 316 (digested with EcoRI and SmaI).

### C.3.2 Immobilization of Polyubiquitin Binding Proteins

GST–Rad23 and GST–Dsk2 were generous gifts from H. Kobayashi and H. Yokosawa, respectively. Fusion proteins were expressed in BL21(DE3)/pLysS and purified using glutathione–Sepharose resin. 10 mg of GST–Dsk2p and 20 mg of GST–Rad23 were separately coupled to 1.5 ml of resin volume of CNBr–activated Sepharose 4B (Amersham Biosciences) in 100 mM NaHCO<sub>3</sub>, pH 8.3, 0.5 M NaCl. Coupled resin was stored at 4°C in a 50% slurry with 100 mM Tris–HCl, pH 8.0, 0.5 M NaCl, 0.02% NaN<sub>3</sub>.

### C.3.3 Two–step Purification

Cells were grown in 6 l of YPD medium (2% peptone, 1% yeast extract, 2% dextrose) at 25°C to an A<sub>600 nm</sub> of 1.5. Cells were washed with 1/6 volume of ice–cold TBS followed by 1/30 volume of ice–cold TBS with 1 mM 1,10–phenanthroline, 10 mM iodoacetamide. Cells were lysed using a One Shot Cell Disrupter (Constant Systems) at 30000 lb/inch<sup>2</sup> in 40 ml of lysis buffer (300 mM NaCl, 0.5% Triton X–100, 50 mM sodium phosphate, pH 8.0, 0.5 mM AEBSF, 5 µg/ml aprotinin, 5 µg/ml chymostatin, 5 µg/ml leupeptin, 1 µg/ml pepstatin A, 1 mM 1,10–phenanthroline, 10 mM iodoacetamide). Lysate (typically 1.5 g of protein) was cleared by centrifugation at 4°C in a Sorvall SS34 for 20 min at 14,000 rpm. 2 mg each of GST–Rad23 and GST–Dsk2 coupled to Sepharose (preequilibrated with lysis buffer) were added to the clarified lysate and mixed for 90 min at 4°C. The resin was then washed with 40 ml of lysis buffer, further mixed for 15 min with 20 ml of 50 mM sodium phosphate, pH 8.0, 2 M NaCl, and washed once with 20 ml

of 50 mM sodium phosphate, pH 8.0, 2 M NaCl and twice with 20 ml of 50 mM sodium phosphate, pH 8.0, 300 mM NaCl, 0.1 % Triton X-100. Elution was performed at room temperature with two successive incubations with 1 ml of urea buffer (UB: 8 M urea, 100 mM  $\text{NaH}_2\text{PO}_4$ , 10 mM Tris-HCl, pH 8.0), and imidazole was added to a final concentration of 20 mM. Eluate was then mixed with 125  $\mu\text{l}$  of nickel magnetic bead slurry (Promega V8565, prewashed in the UB) for 60 min on a rotating wheel. The beads were washed with 1 ml of UB and mixed for 15 min with UB supplemented with 0.5 % SDS. The beads were then washed with 1 ml of UB with 0.5 % Triton X-100 and mixed for 15 min with another 1 ml of UB with 0.5 % Triton X-100. The last procedure was repeated using UB only.

To generate peptides for MS-based sequencing, we performed the tryptic digest directly on the beads. The beads were incubated with 500  $\mu\text{l}$  of UB with 3 mM Tris-(2-carboxyethyl)phosphine (T-CEP) for 20 min and then for another 15 min following addition of iodoacetamide to 11 mM. The buffer volume was reduced to 75  $\mu\text{l}$  by removing excess liquid, and 0.2  $\mu\text{g}$  of endoproteinase Lys-C (Roche) was added. Beads were incubated at 37°C with intermittent shaking for 5 h. Dilution buffer (225  $\mu\text{l}$  of 100 mM Tris-HCl, pH 8.0, 1.33 mM  $\text{CaCl}_2$ ) was then added followed by 1  $\mu\text{g}$  of trypsin (Roche Applied Science), and the beads were further incubated with intermittent shaking for 16 h at 37°C. The supernatant was carefully collected, and formic acid was added to a final concentration of 5 %.

#### C.3.4 MS and Data Analysis

The proteolytically digested sample was further processed for multidimensional chromatography coupled in-line to ESI-MS as described previously (Graumann et

al. 2004). As a variation to the chromatography program, samples were stepped off the strong cation exchanger phase of the triphasic column using 12.5 %, 20 %, 30 %, 40 %, and 100 % buffer C (500 mM ammonium acetate, 5 % ACN, 0.1 % formic acid). Centroided fragmentation spectra acquired by Xcalibur 1.3 (ThermoElectron) were evaluated for spectrum quality and charge state using **2to3** (Sadygov et al. 2002) and searched against the translated open reading frames of the *Saccharomyces* Genome Database (SGD Cherry et al. 1998; release time stamp: 07/26/2004; 6860 entries) with **Sequest** (version 27, revision 9; Ref. 24) utilizing unified input and output files (McDonald et al. 2004). Relevant **Sequest** parameters used were: (i) peptide mass tolerance of 3.0 amu, (ii) parent ion masses were treated as monoisotopic, (iii) fragmentation ion masses were treated as averaged, and (iv) a 57.0 amu static modification on cysteines accounted for alkylation. **Sequest** results were filtered using **DTASelect** 1.9 and **Contrast** (Tabb et al. 2002) with the following requirements for peptide and locus identifications considered valid: minimum Xcorrs of 1.8, 2.5, and 3.5 for singly, doubly, and triply charged ions, respectively; a minimum  $\Delta Cn$  of 0.08; and a minimum of two valid peptides per locus.

### C.3.5 Small Scale Cell Extraction, IMAC, and Western Blotting

For direct comparison of protein level in wild-type and *rpn10* $\Delta$  strains, S288C cells were grown in YPD at 25°C until an  $A_{600\text{ nm}}$  of 0.5–1 was reached. An amount of yeast cells corresponding to 4–5  $A_{600}$  was collected, briefly washed with 1 ml of 1 $\times$  TBS, and frozen in liquid nitrogen. Cells were directly resuspended in prewarmed sample buffer, incubated for 2 min at 96°C, lysed with glass beads in a FastPrep 120 (Thermo Savant) for 45 s with a speed setting of 5.5, and incubated for an-

other 4 min at 96°C. For IMAC purification of H<sub>8</sub>-ubiquitin, cells transformed with a *URA3*-based plasmid coding for H<sub>8</sub>-ubiquitin were grown in 100 ml of SD-URA medium (0.67% yeast nitrogen base, 5% dextrose) at 30°C to an A<sub>600 nm</sub> of 1. TCA (20% final) was added directly to the cell culture, and cells were incubated for 10 min on ice and washed with ice-cold 100 mM Tris-HCl (once with pH 8.5, twice with pH 8.0). Cells were resuspended in 0.6 ml of 0.2% SDS, 8 M urea, 100 mM Hepes, pH 8.0, 1 mM 1,10-phenanthroline, 5 mM N-ethylmaleimide (NEM), 0.5 mM AEBSF, 5 μg/ml aprotinin, 5 μg/ml chymostatin, 5 μg/ml leupeptin, 1 μg/ml pepstatin A, and lysed by agitation with glass beads in a FastPrep 120. Glass beads were further washed with 0.6 ml of lysis buffer without SDS, and lysate (containing 0.1% SDS) was cleared 10 min at 14,000 rpm in a microcentrifuge. Imidazole (20 mM final) and nickel magnetic beads (70 μl) were added to 8.5 mg of lysate protein and mixed for 1 h at room temperature. Beads were then washed three times in 0.1% SDS, 8 M urea, 100 mM Hepes, pH 8.0, and proteins were eluted in SDS-PAGE sample buffer supplemented with 1 M imidazole, 4 M urea, 50 mM Hepes, pH 8.0. TAP-tagged proteins were detected using the anti-calmodulin binding peptide antibody (Upstate Biotechnology), ubiquitin with MAB1510 (Chemicon International), Cdc28 with PSTAIR antibody (Santa Cruz Biotechnology), and Gcn4-Myc9 with 9E10 monoclonal antibody.

## C.4 Results

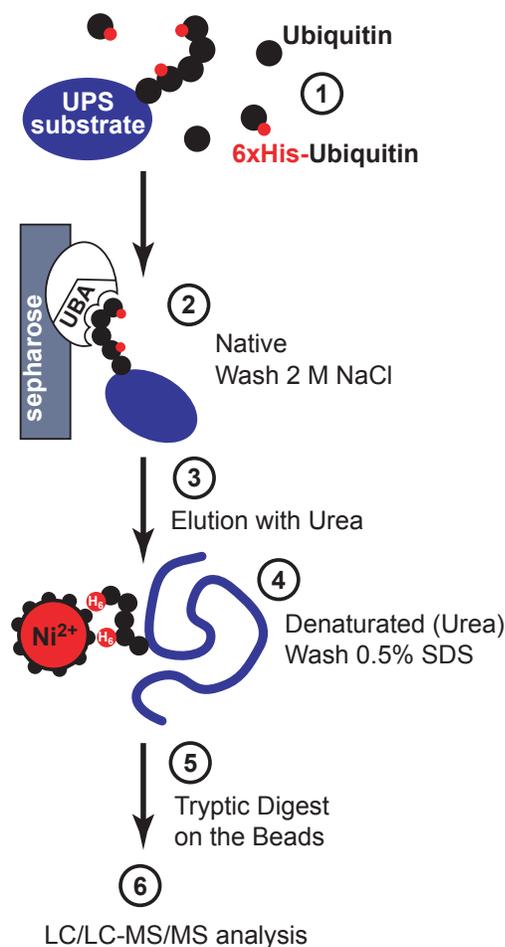
### C.4.1 Two-step Purification of Ubiquitin Conjugates

We performed two-step purification of ubiquitin conjugates (fig. C.1) from cells that

express ubiquitin fused to an amino-terminal hexahistidine tag (H<sub>6</sub>-ubiquitin), and as control we repeated the procedure with cells that do not express H<sub>6</sub>-ubiquitin. In both experiments, the first purification step yielded a similar amount of proteins, whereas the IMAC only recovered appreciable material from the H<sub>6</sub>-ubiquitin strain (fig. C.2A). The signal revealed by silver staining of material fractionated by SDS-PAGE ranged from 50 to 250 Da and produced a spread rather than discrete bands, as expected for a large collection of different proteins conjugated to ubiquitin chains of various lengths. We calculated that the first step in purification recovered about 15% of the polyubiquitin conjugates in the cell (fig. C.2B). Notably, mono-, di-, and triubiquitin species were not recovered. This implies that the UBA domains of Rad23 and Dsk2 were only enriching for proteins conjugated to ubiquitin chains that contained more than three ubiquitins. Because a tetraubiquitin chain is thought to comprise the minimal signal for targeting substrates to the proteasome for degradation (Piotrowski et al. 1997; and Thrower et al. 2000), the UBA affinity step appears to enrich specifically for those ubiquitin conjugates that are proteasome substrates. In the second step, the majority of the ubiquitin conjugates (> 80%) eluted from the first resin were recovered (fig. C.2C). In this experiment, only 25%–30% of the bound material was eluted with sample buffer from the nickel beads (data not shown). Overall, our procedure resulted in a 3000- to 5000-fold enrichment of polyubiquitin conjugates (1500 mg of protein extract resulted in 30–50  $\mu$ g of protein, representing 10% of the polyubiquitin in the cell).

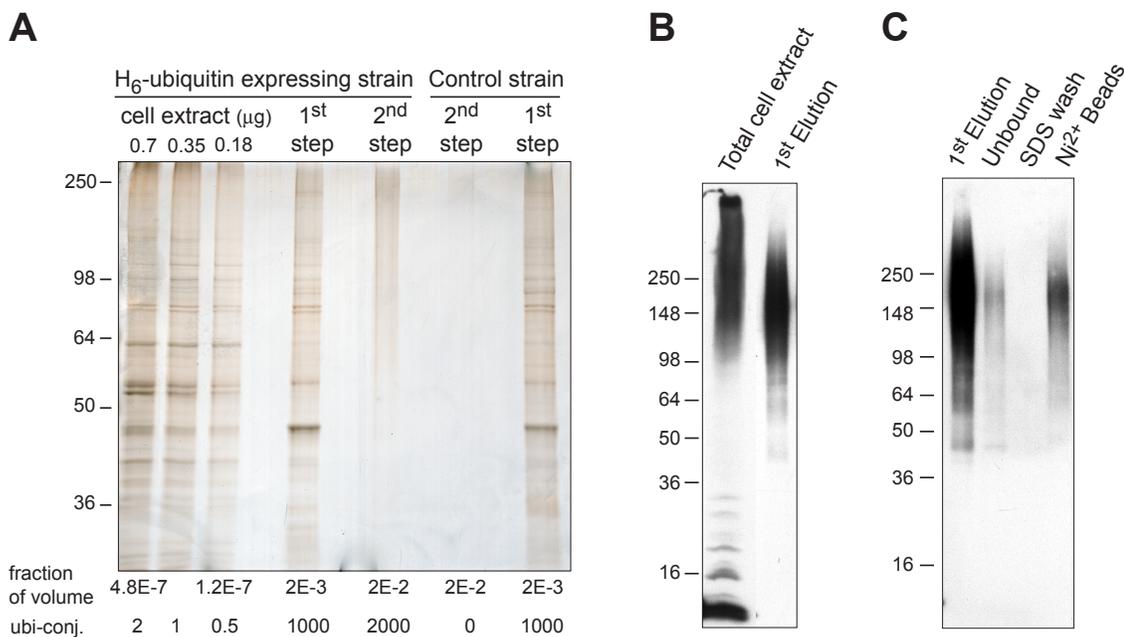
#### C.4.2 MS Analysis

Purified proteins were directly digested on the nickel beads, and the peptide mixture



**Figure C.1 Flow Diagram for the Two-step Purification of Polyubiquitin Conjugates.** Yeast cells that constitutively express ubiquitin modified with an amino-terminal hexahistidine tag are lysed in nondenaturing buffer (1). Polyubiquitin chains are purified using matrices derivatized with the recombinant UBA domain-containing proteins Rad23 and Dsk2. UBA domains bind tightly to multiubiquitin chains, with a preference for chains linked via lysine 48 of ubiquitin (Wilkinson et al. 2001; and Raasi et al. 2004). Contaminant proteins are removed by washes with 2 M NaCl (2), and specifically bound proteins are then eluted in 8 M urea (3) and mixed with nickel magnetic beads (4). In this second purification step, stringent washing conditions (0.5% SDS) are used to remove contaminants. Trypsin is then applied directly to the beads (5), and peptides released from the beads are analyzed by LC/LC-MS/MS (6).

was analyzed by multidimensional LC-MS/MS or MudPIT. *Sequest* and *DTASelect* algorithms were used to analyze the spectra generated by the complex mixture of affinity-purified proteins, and 180 nonredundant proteins were identified (supplemental Table 2). The most abundant protein in our analysis was ubiquitin. Of



**Figure C.2 Two-step Affinity Purification Specifically Enriches for Polyubiquitylated Proteins.** A, SDS-PAGE analysis of the two-step purification. Purifications were performed using the H<sub>6</sub>-ubiquitin-expressing strain or the wild-type control strain that lacks tagged ubiquitin. Aliquots of total cell extract, proteins eluted after the first step (UBA affinity) of the purification, and proteins from the second step (those bound to the nickel magnetic beads) were separated by SDS-PAGE on a 10% polyacrylamide gel and stained with silver. Amounts loaded in comparison to initial volumes are indicated immediately below each lane. Below that, the amount of ubiquitin conjugates for each lane (as estimated by Western blotting, data not shown) is indicated in arbitrary units. B, immunoblotting of the first purification step. Aliquots of total cell extract and the eluate from the UBA domain affinity step (first elution) were separated by SDS-PAGE on a 4–20% polyacrylamide gradient gel and immunoblotted with an anti-ubiquitin antibody. The sample from the first elution is overloaded 10-fold relative to the total cell extract. C, immunoblotting of the IMAC purification step. Equal portions of initial volumes corresponding to proteins that were eluted from the UBA domain matrix, failed to bind the nickel-based matrix (unbound), were washed away with 0.5% SDS (SDS wash), or bound to the nickel beads (Ni<sub>2</sub><sup>+</sup> beads) were processed as in B.

a total of 5347 sequencing events, 457 peptides derived from ubiquitin. This was expected because ubiquitin should be the most prominent protein after the purification. For clarity, we further filtered our data by removing transposon-related genes, duplicated genes, ubiquitin fusion genes, and Rad23 and Dsk2 that leached from the resin used in the first purification step (data not shown). The 127 remaining proteins are listed in Figure C.3. We classified these proteins in different cate-

gories according to their function (fig. C.4A). The majority of identified proteins is involved in metabolism and translation. Several proteins are components of regulated pathways, and several were previously shown to be targets for degradation. These include Ole1, a short lived protein (Braun et al. 2002), Rpo21, which is ubiquitinated by Rsp5 (Huibregtse et al. 1997), and Gdh1 and Mdh2, which were shown previously to be targeted for proteolysis (Minard and McAlister-Henn 1994; and Mazon and Hemmings 1979). Moreover, the list includes proteins for which ubiquitination sites were previously identified; 14 of our 127 proteins were among the 71 identified in the initial global study of ubiquitinated proteins (Peng et al. 2003), and 8 of our 127 proteins were among the 33 found in a screen for membrane-associated ubiquitinated proteins (Hitchcock et al. 2003). Thus, although we identified only  $\approx 2\%$  of the yeast proteome ( $127/\approx 6000$ ), these proteins accounted for 21% of the ubiquitinated proteins identified by Gygi and coworkers (Peng et al. 2003; and Hitchcock et al. 2003), a 10-fold enrichment.

Because our ultimate goal was to compare the pool of ubiquitinated proteins in wild-type and *rpn10* $\Delta$  cells, it was important to assess the variability of the MS analysis. The sample from the two-step purification described above had been split in half after the trypsin digest but prior to the MS analysis. When the second half of the sample was analyzed, we identified 176 proteins (supplemental Table 3). The two LC/LC-MS/MS analyses of the same sample were then compared using the Contrast algorithm (fig. C.4B). More than 80% of the proteins identified in one analysis were found in the other analysis. We noticed that the variability was accounted for mainly by proteins identified by two peptides (as the loss of one peptide identification for a particular protein led to its exclusion from the analysis). When we also took into account proteins identified by only one peptide,

Name	Sequence coverage (%)	Peptide	Name	Sequence coverage (%)	Peptide	Name	Sequence coverage (%)	Peptide
SSA2 <sup>a</sup>	61.5	43	RPS13	19.9	3	TDH1	9.6	3
RPL2A, B	57.1	17	RPS17A, B	19.9	2	ACT1	9.1	2
RPL21A, B	55.0	12	IML2	19.7	10	RPL34A, B	9.1	2
RPS7B	54.2	6	SRO9	19.6	4	SAN1	9.0	2
SSA1 <sup>a</sup>	47.4	32	PMA1 <sup>b</sup>	19.2	14	GPM1	8.9	2
RPL10	43.0	9	CIT2 <sup>a</sup>	17.8	6	STI1	8.8	4
RPL3	42.9	20	GLN1 <sup>a,b</sup>	17.8	6	RPT1	8.8	3
RPL15A	42.2	10	RPL4A, B	17.7	4	UBP6	8.6	5
RPS20 <sup>a,b</sup>	42.1	8	ENO1	17.4	6	NOP4	8.3	4
ERG1 <sup>a,b</sup>	41.7	19	PMA2	16.1	13	HSP82	7.9	7
NCE103	40.7	6	PGK1	16.1	8	UFD2	7.8	6
RPS4A, B <sup>a</sup>	40.6	11	DRE2	16.1	5	GPD2	7.7	2
RPS7A	40.0	5	RPS8A, B	15.5	2	ACS2 <sup>a</sup>	7.6	2
RPL27A, B	37.5	6	BGL2	15.3	4	OLE1	7.6	3
RPS11A, B	37.2	10	SSA4	15.1	12	FAS1	7.5	13
RPL28	36.2	10	RPS6A, B	14.8	4	HSP42	7.5	2
AAH1	36.0	10	HSP150	14.7	3	HSP104 <sup>b</sup>	7.4	5
RPL19A	34.9	11	ENO2	14.6	4	RPB2	7.1	6
VMA7	34.7	2	PNG1	14.6	5	FAA4	7.1	3
RPL8A	34.4	7	RPL11B	14.4	2	YMR210W	6.7	2
TEF1, 2	34.3	9	CBR1	14.0	4	HEF3	6.6	6
RPL15B	33.8	7	TDH2	13.9	4	YOR091W	6.2	2
GDH1 <sup>a</sup>	33.5	15	ADH1	13.2	4	LYS1	6.2	2
RPA190	32.5	49	YLR407W	13.1	2	PHO84 <sup>a</sup>	6.1	3
RPL1A, B	31.3	5	UBC6	12.8	2	RPF2	6.1	2
RPS26A	30.3	3	URA2	12.2	20	CDC48 <sup>a</sup>	5.1	3
ERG11	29.8	17	UBP3	12.1	9	FKS1 <sup>b</sup>	5.0	5
RPS1A, B	29.8	7	RPT2	12.1	4	KCC4	4.9	2
RPL24A, B	29.7	7	SSA3	12.0	9	TKL1	4.9	3
TDH3	29.5	6	MLF3	11.9	3	GAS1	4.5	2
RPS18A, B	28.8	5	ERG5 <sup>a</sup>	11.7	4	SNF1	4.3	2
YEF3	28.7	24	YDJ1	11.7	3	RFC1	3.5	2
RPS27A, B	28.0	2	RPS3 <sup>a</sup>	11.7	2	STP2	3.3	2
SIK1	27.8	9	SAM1	11.5	2	KAR2	3.1	3
SSB1, 2	27.7	12	RPA135	11.2	7	RPO21 <sup>b</sup>	2.8	4
RPL6B	26.7	6	CBF5	10.6	3	KAP123	2.7	2
RPS12	26.6	3	GDH3	10.5	7	GSC2 <sup>a</sup>	2.2	2
YBR071W	25.6	4	RPN1	10.3	10	CRM1	2.2	2
RPS5	25.3	3	RPL6A	10.2	2	RET1	1.6	2
EFT2, 1	25.2	15	HSC82	10.1	8	NUM1	1.2	2
HYP2	24.8	2	RPL32	10.0	2	TIP20	1.0	2
MDH2	21.7	7	VTC4	9.8	8			
RPL18A, B	21.5	5	PRE9 <sup>b</sup>	9.7	2			

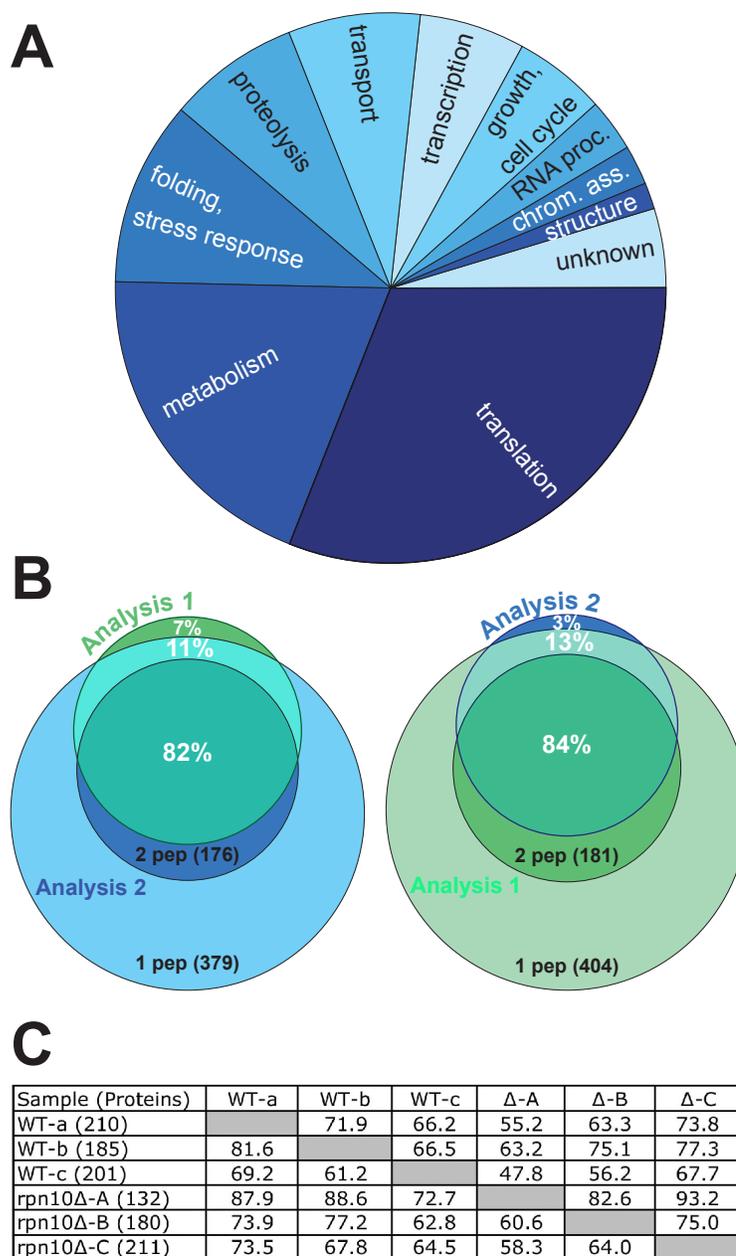
**Figure C.3 Proteins Identified by LC/LC–MS/MS After Two–step Purification of Ubiquitin Conjugates.** “<sup>a</sup>” Ubiquitylated proteins identified by Peng et al. (2003). “<sup>b</sup>” Ubiquitylated proteins identified by Hitchcock et al. (2003).

≈ 95 % of proteins identified by two peptides in either dataset were also identified by at least one peptide in the duplicate analysis (fig. C.4B). This indicated that there was some variation in the data analysis, albeit tolerable, arising from either the HPLC or mass spectrometer. Moreover, proteins defined by our minimum cutoff of two peptides (and thus possibly of low abundance in the purified sample)

were disproportionately susceptible to being overlooked. Because many potential targets of interest might be in the inabundant category, we decided to perform our subsequent analyses in triplicate to ensure the identification of a maximum number of ubiquitin conjugates.

### C.4.3 Impact of the Proteasome Substrate Receptor Rpn10 on the Pool of Ubiquitin Conjugates

Our key motivation for developing proteomic methods to identify ubiquitin conjugates on a global scale was to use the method to identify substrates/targets for ubiquitin ligase and isopeptidase enzymes and other specificity determining factors in the UPS. In particular, we sought to determine the breadth of the impact of Rpn10 on ubiquitin-dependent proteolysis. We reasoned that deletion of *RPN10* would prevent the degradation of substrates dependent on Rpn10 for turnover. These substrates would then accumulate as polyubiquitylated conjugates. To proceed, we collected and analyzed six independent samples; three were obtained from wild-type cells (supplemental Tables 2 and 4) and three others from *rpn10* $\Delta$  cells (supplemental Table 5). We compared the six datasets using the **Contrast** algorithm (fig. C.4C). The variability between the different datasets ( $\approx 30\%$ ) was in general higher than previously observed between two identical samples (fig. C.4B). This was expected because it is essentially impossible to grow cells, lyse cells, and carry out consecutive affinity purification steps in a manner that is perfectly precise. Nevertheless, to identify the candidate targets of Rpn10, proteins represented in any of three *rpn10* $\Delta$  samples but not in any *RPN10* sample were extracted and rank ordered according to sequence coverage of the identified protein (fig. C.5). What



**Figure C.4 Protein Representation and Reproducibility Overview.** A, pie diagram of the identified proteins. Protein functions retrieved from the YPD database (Incyte) were plotted according to their representation in Figure C.3. B, reproducibility of LC/LC-MS/MS analysis. Left, of 181 proteins identified by at least two peptides in Analysis 1 (green circle), 82% were also identified by at least two peptides in Analysis 2 (dark blue circle), 11% were identified by only one peptide in Analysis 2 (light blue circle), and 7% were not recovered in Analysis 2. Right, same as left, except that the diagram indicates the percentage of the 176 proteins from Analysis 2 (two peptide hits) that were identified at various levels of stringency in Analysis 1. C, pairwise analysis of the different samples (wild type and *rpn10* $\Delta$ ). The percentage of proteins from one analysis (row) present in another analysis (column) is indicated. For each analysis, the number of identified proteins is shown in parentheses.

is particularly noteworthy is that the second highest ranked candidate in this subtractive screen of the entire *S. cerevisiae* proteome was the cell cycle regulator Sic1, which is ubiquitylated by the SCF<sup>Cdc4</sup> complex at the G1/S transition (Petroski and Deshaies 2003). We had previously shown that Sic1 degradation is substantially dependent upon Rpn10 (Verma et al. 2004), suggesting that it is likely to accumulate as a ubiquitylated species in *rpn10Δ* cells (an assumption that was not addressed previously but has been validated as described below). Other candidates revealed by this subtractive approach are also known to be targets for ubiquitylation. The transcription factor Gcn4 is targeted for proteolysis after ubiquitylation by SCF<sup>Cdc4</sup> complex (Chi et al. 2001; Meimoun et al. 2000; and Kornitzer et al. 1994), and Aro10, Ald6, Erg3, and Ecm21 were identified as ubiquitylated proteins in a global analysis (Peng et al. 2003). Taken together, these findings suggest that our subtractive approach was sufficiently sensitive to identify critical regulatory targets of the UPS, even those of exceptionally low abundance such as Gcn4, which is estimated to be present at less than 50 molecules per cell (Ghaemmaghami et al. 2003).

#### C.4.4 Validation of Rpn10 Targets

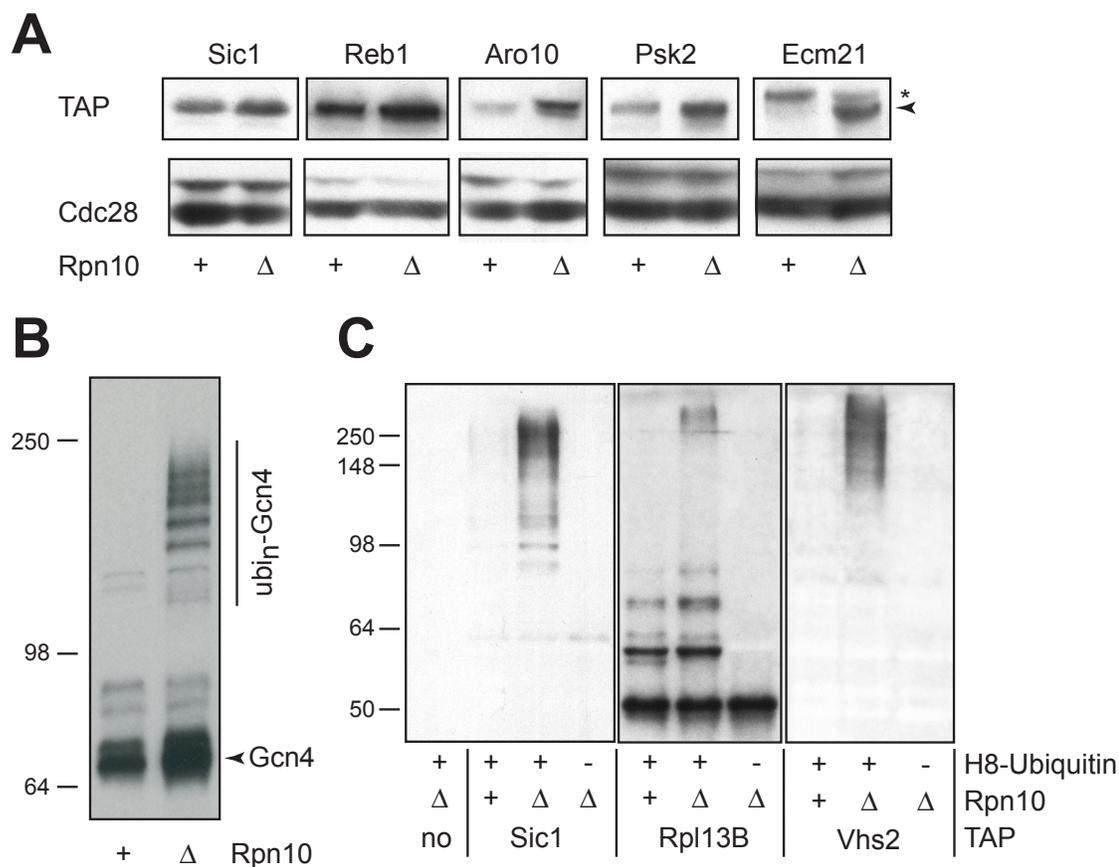
To evaluate the role of Rpn10 in turnover of candidate substrates identified by our MudPIT approach, we assayed several of the proteins from Figure C.5 for abundance and ubiquitylation. First, we compared protein levels in *RPN10* and *rpn10Δ* strains in which the endogenous loci were modified to encode the candidate proteins with TAP tags fused to their C termini (fig. C.6A and fig. C.5). For several candidates, protein levels were elevated in the *rpn10Δ* strain, suggesting that normal turnover of these proteins was Rpn10 dependent. For Gcn4, we employed a well

Name	A	B	C	Total	Validation			Name	A	B	C	Total	Validation		
GCN4	22.4	22.4	32.4	35.9	+	1	1	RPL16B		8.1		8.1			
SIC1		32		32	+	1	1	VTS1			7.5	7.5			
VMA2			24.2	24.2				CPA1		7.5		7.5	-	0	2
PUP3	23.9			23.9				YLL012W			7.2	7.2			
YJR014W		23.2		23.2	-	0	0	FET3		6.8		6.8	-	0	NT
VHS2		17.2	9.6	22.9	+	0	1	LEU1			6.7	6.7			
RPL13A, B		22.6		22.6	+	0	1	MCH4			6.4	6.4			
LYS20			22.4	22.4				PPQ1			5.6	5.6			
RPS29A, B	19.6			19.6				VPS72		4.9		4.9	-	0	2
LYS21			19.5	19.5				LYS2	3.6		3.2	4.8			
PCL1	12.2	12.2	12.5	18.6				SGV1			4.7	4.7			
SEL1			17.6	17.6				ERG3			4.4	4.4			
RPL20A, B		16.3		16.3				MBP1		4.2		4.2	-	0	NT
RPL17A, B		15.8		15.8				REB1		4.1		4.1	+	1	NT
ARO10	7.7	3.3	6	15.3	+	1	2	ILV2			3.9	3.9			
NOG2			14.8	14.8				NSP1		3.9		3.9	-	0	0
TUB1			13.4	13.4				TUB2			3.3	3.3			
TOM22	13.2		13.2	13.2				YOR112W		3		3			
GAT2		10.5	3.8	12.3				SHQ1			2.8	2.8			
RTS3			11.4	11.4				ECM21		2.7		2.7	+	1	NT
DDR48		11.2		11.2				SIR4		2.7		2.7			
TSR1		5.7	4.6	10.3	-	0	0	CHS7		2.5		2.5			
ALK1			10	10				STP1	1.9		1.9	1.9			
UBX7			9.6	9.6				CDC39			1.9	1.9			
SSF2			9.5	9.5				MLP1		1.7		1.7	+	1	NT
NIP1			9.2	9.2				PSK2		1.1		1.1	+	1	NT
ALD6	2.4	8.6		8.6				MDN1		0.3		0.3	+	1	NT

**Figure C.5 Putative Ubiquitylated Proteins Identified in *rpn10Δ* But Not Wild-type Cells.** Proteins listed were identified (by a minimum of two valid peptides) in any of three independent analyses of *rpn10Δ* cells (A, B, and C) but not in any of the three independent analyses of control cells (*RPN10*). Sequence coverage is indicated in percentages for A, B, and C analyses and in the total column (corresponding to the sum of sequence coverage in the three experiments). The final validation status (+ or -) for Rpn10 targets is indicated in the first column. The score for the increase of protein level in *rpn10Δ* and the presence of ubiquitylated species detected after IMAC in *rpn10Δ* are indicated in the middle and last column, respectively. NT, not tested; 0, not validated; 1, validated; 2, ubiquitylated species were detected in both *rpn10Δ* and *RPN10* cells.

characterized allele that encodes a carboxyl-terminal Myc9 tag integrated into the GCN4 locus (Chi et al. 2001). We found that Gcn4 accumulated in *rpn10Δ* extracts, and we could also detect species migrating with a lower mobility that correspond to polyubiquitylated Gcn4 (fig. C.6B).

In addition to inabundant proteins like Gcn4, our analysis also identified highly abundant proteins as such the ribosomal subunit Rpl13B. However, by Western blotting we could not see any increase in the level of Rpl13B in *rpn10Δ* (data not shown). We reasoned that in this case perhaps only a small fraction of the protein pool was targeted for degradation, and thus the overall protein abundance was not altered in



**Figure C.6 Analysis of Candidate Substrates of the Rpn10-dependent Targeting Pathway.** A, proteins whose level was increased in *rpn10Δ*. The chromosomal locus for each candidate investigated was modified to introduce a TAP epitope tag fused to the carboxyl terminus of the encoded protein. For each TAP-tagged candidate shown, equal amounts of proteins from *RPN10* (wild type) and *rpn10Δ* ( $\Delta$ ) cells were fractionated by SDS-PAGE on a 10% polyacrylamide gel and transferred to nitrocellulose. Immunoblotting was performed with anti-calmodulin binding peptide antibody that recognizes the TAP tag and anti-Cdc28 (which served as a loading control). The caret highlights a novel species of Ecm29 that was detected only in *rpn10Δ*. B, ubiquitylated Gcn4 accumulates in *rpn10Δ*. Equal amounts of proteins from *RPN10 GCN4<sup>myc9</sup>* and *rpn10Δ GCN4<sup>myc9</sup>* cells were separated by SDS-PAGE on a 10% polyacrylamide gel and transferred to nitrocellulose. Gcn4-Myc9 was detected using the 9E10 antibody. C, purification of proteins conjugated to H8-ubiquitin. Proteins from strains with the indicated genotypes that bound nickel beads in buffer containing 8 M urea plus 0.1% SDS were loaded onto a 10% polyacrylamide gel and subjected to SDS-PAGE followed by immunoblotting with anti-calmodulin binding peptide antibody.

*rpn10Δ*. To test this, we devised a single-step purification with nickel beads using cells transformed with a plasmid that expressed ubiquitin with an octahistidine tag

fused to the amino terminus (H<sub>8</sub>-ubiquitin). Purified proteins were detected with the TAP tag antibody. An untagged *rpn10*Δ strain that expressed H<sub>8</sub>-ubiquitin was used as a negative control and gave no signal in the Western blot (fig. C.6C). After performing the same procedure with a Sic1-TAP strain, we noticed the distinctive accumulation of high molecular mass Sic1 conjugates in *rpn10*Δ but not in *RPN10* cells that expressed H<sub>8</sub>-ubiquitin (fig. C.6C). No signal was readily detected *rpn10*Δ cells not expressing the H<sub>8</sub>-ubiquitin. Therefore, ubiquitylated Sic1 specifically accumulated in cells lacking Rpn10. Rpl13B showed similar behavior. Although there was some nonspecific binding of unmodified Rpl13B to the nickel beads (lower band present in all three lanes), Rpl13B species that migrated at high molecular masses (> 250 kDa) were exclusively detected in *rpn10*Δ cells that expressed H<sub>8</sub>-ubiquitin. Notably, species modified with one, two, and three ubiquitins were also detected in wild-type cells whenever H<sub>8</sub>-ubiquitin was expressed. However Rpl13B was only detected by MS in samples from *rpn10*Δ cells. Therefore the species modified with one, two, and three ubiquitins that were also present in *RPN10* cells most likely were not enriched in the two-step purification, as is the case for free (i.e., not substrate-linked) mono-, di-, and triubiquitin (fig. C.2B). Vhs2 protein level was also found unaltered in *rpn10*Δ cells (despite its relative low abundance), but ubiquitylated Vhs2 was detected after IMAC of extracts from *rpn10*Δ cells that expressed H<sub>8</sub>-ubiquitin (fig. C.6C).

## C.5 Discussion

In this article, we describe a new approach to the purification and analysis of ubiquitin conjugates in the budding yeast *S. cerevisiae*. Our approach involves two affinity

purification steps. The first step selects for ubiquitin chains that were able to bind recombinant UBA domain-containing proteins and thus were most likely competent to support degradation of attached proteins. In the second step, ubiquitin conjugates that contain H<sub>6</sub>-ubiquitin were enriched by IMAC. Conjugates that survived the two enrichment steps were digested to yield peptides, which were separated by multidimensional chromatography and sequenced by MS/MS. This protocol enabled us to identify a collection of candidate ubiquitin-conjugated proteasome substrates. By performing a “subtractive” comparison of conjugates recovered from wild-type cells versus *rpn10Δ* cells that lack the proteasome substrate receptor Rpn10, we were able to identify a collection of proteins that accumulate selectively in *rpn10Δ* and thus are candidate ligands for Rpn10. This effort revealed that the pool of candidate Rpn10 ligands is much larger than appreciated previously from one-off analyses.

The approach described here differs from prior “proteome-wide” analyses of ubiquitin (Peng et al. 2003; and Hitchcock et al. 2003) and SUMO-conjugated proteins (Wohlschlegel et al. 2004; Zhou et al. 2004; Panse et al. 2004; Rosas-Acosta et al. 2005; Zhao et al. 2004; and Denison et al. 2005) in several important respects. First, we present data on replicate analyses. We found modest variation ( $\approx 17\%$ ) in duplicate MS analyses of a single sample, but significant variations ( $\approx 30\%$ ) when the entire affinity purification and LC/LC-MS/MS analysis were repeated. Performing replicate analyses is thus of considerable importance when comparing the ubiquitin conjugate proteome in different strains (e.g., wild type and *rpn10Δ*) to ensure that any differences seen are due to the mutation under study and are not simply a product of experimental variability. Performing replicate experiments also helps to ensure that an analysis is as thorough as possible. For example, some

candidates that were validated (e.g., Sic1) were only identified in one of three analyses. Indeed, of the candidates for which identification was least robust (Mlp1, Psk2, and Mdn1, each of which was found in only one analysis at  $< 2\%$  sequence coverage), all three were validated as being responsive to Rpn10 function. Thus, we strongly recommend that multidimensional analyses be performed with replicate samples both to minimize false positives and to enhance identification of target proteins.

A second key difference is that we employed a “functional” affinity purification step in tandem with a tag-dependent affinity purification step. By comparison, Gygi and coworkers (Peng et al. 2003; and Hitchcock et al. 2003) employed a single nickel-nitrilotriacetic acid affinity purification step in their analyses of the ubiquitin proteome. The inclusion of a second, function-based affinity step had two important consequences; first, it enabled superior enrichment for ubiquitin-conjugated proteins, and second, it focused our analysis on a particular subset of ubiquitin-conjugated proteins (i.e., those that are candidate substrates for the proteasome). In our hands, single-step purification with H<sub>6</sub>-ubiquitin led to a relatively modest enrichment of ubiquitin conjugates (100- to 200-fold)<sup>2</sup> as compared with the two-step purification (3,000- to 5,000-fold). This is in keeping with our experience that  $\approx 0.5\%$  of total yeast extract proteins bind specifically to IMAC resins. Thus, it is possible that a fair fraction of the proteins identified previously are not bona fide UPS substrates. Importantly, our approach has permitted the identification of even the extremely inabundant UPS substrate Gcn4, which is present at less than 50 molecules per cell (Ghaemmaghami et al. 2003). Consistent with the greater

---

<sup>2</sup> T. Mayor and R. J. Deshaies, unpublished data.

degree of target focus intrinsic to our analysis, we did not identify proteins that are known to be conjugated with a single ubiquitin (e. g., histone H2A, B), nor did we enrich for mono-, di-, or triubiquitin chains (fig. C.2B). Finally, when we searched for peptides of ubiquitin itself that carried the Gly–Gly signature, Lys<sup>48</sup> was found to be the most prominent conjugation site that was recovered (data not shown). Lys<sup>29</sup>, Lys<sup>33</sup>, and Lys<sup>6</sup> were more rarely identified, and modified Lys<sup>63</sup> was not found. These findings suggest that we have established a new approach to identify specifically those proteins that are polyubiquitylated substrates of the proteasome. In the future, other ubiquitin receptors, like proteins containing UIM domains that bind mono-ubiquitylated targets in the endocytic pathway (e. g., Vps27 and Ent1) or ZnF domains that bind selectively to Lys<sup>63</sup>-linked ubiquitin chains (Kanayama et al. 2004), may be used to identify factors in nonproteasomal pathways that are regulated allosterically by ubiquitylation.

Of the more than 120 proteins that we implicated as substrates of the UPS, most function in translational and metabolic pathways, and half of the candidates have high codon adaptation index values ( $> 0.4$ ).<sup>3</sup> Many ribosomal proteins were identified including some that were shown previously to be ubiquitylated, like Rpl28, Rps3, and Rps20 (Peng et al. 2003; and Spence et al. 2000). Because ribosomes are highly abundant and formed by tight macromolecular interactions, we cannot exclude that some of the identified proteins were contaminants. However, it is also possible that some of these candidate substrates might represent biosynthetic intermediates that fail to fold or assemble properly, resulting in their rapid degradation either during or shortly following the completion of translation (Schubert et

---

<sup>3</sup> T. Mayor, J. Graumann, and R. J. Deshaies, unpublished observations.

al. 2000; and Turner and Varshavsky 2000). In the latter case, one would predict that the UPS might have little impact on the total level of the candidate protein and that only a very small fraction of the total protein pool in the cell is ubiquitylated (depending on the fraction of the protein that misfolds or misassembles). This is exactly what we observed for Rpl13B. If a small fraction of Rpl13B fails to assemble properly and is degraded rapidly by the UPS, it could help to explain the presence of many proteins with high codon adaptation index values in our analysis. Thus, the bulk of proteins degraded by the proteasome in yeast cells might correspond to misfolded, damaged, or improperly translated proteins rather than proteins such as cyclins, CDK inhibitors, and transcription factors whose functions are regulated by proteolysis. Further studies will be required to address the important issue of substrate flux through the UPS in yeast. Notably, our method provides a means to identify substrates of the chaperone pathways that enable efficient protein folding and assembly as well as the ubiquitin ligases that target misfolded proteins for degradation by the UPS.

To gain a sense of the quality of our subtractive dataset of conjugates uniquely found in *rpn10Δ* samples, we employed two different assays to evaluate 17 of the 54 candidate Rpn10 substrates. The first and simplest assay was to compare by immunoblotting the level of the candidate protein in wild-type and *rpn10Δ* cells on the assumption that Rpn10 substrates might accumulate to a higher level in *rpn10Δ*. However, we recognized that there may be substrates for which only a small fraction of the total pool is degraded by an Rpn10-dependent pathway, and these substrates might fail this test. Thus, we devised a second assay that measured the level of ubiquitylated candidate protein that was present in wild-type and *rpn10Δ* cells. This second assay allowed us to confirm some candidate proteins (e. g., Rpl13) that

were not validated by the first assay. Ultimately, we were able to confirm that nearly 60 % (10 of 17) of the candidates analyzed are responsive to Rpn10 function. It is important to note that the validation experiments were done with TAP-tagged chromosomal loci (which are in the S288C genetic background), and that the cells were grown in synthetic medium to select for a H<sup>8</sup>-ubiquitin expression plasmid. By contrast, the affinity purification-mass spec analyses were performed with cells of the W303 strain background grown in rich (YPD) medium. Thus, a failure to confirm a candidate should not be construed as definitive evidence that the candidate is not an Rpn10 ligand. Nevertheless, the apparent high rate of false positives underscores that it is critical to carry out secondary analyses to confirm data acquired in multidimensional MS analyses. Future developments, including the implementation of quantification methods and higher stringency biochemical separations, may reduce the experimental variations and false positive rate.

A previous study (Verma et al. 2004) from this laboratory revealed that the proteasome substrate receptors Rpn10 and Rad23 can promote degradation of specific subsets of UPS targets and suggested that Rpn10 targets might be restricted to a small class of UPS substrates. However, that study was based on the piecemeal examination of a handful of UPS targets, and it was not designed to reveal the full spectrum of substrates targeted to the proteasome by a given ubiquitin chain receptor. By using the two-step purification multidimensional MS method described here, we have identified several dozen candidate ligands for an Rpn10-dependent targeting pathway that function in a broad range of processes including metabolism, transcription, translation, nuclear transport, and cell cycle. By applying this approach to mutants lacking other receptors (e.g., *rad23*Δ, *dsk2*Δ), it should be feasible to begin the task of constructing a “linkage map” that reveals the spec-

trum of substrates that are targeted to the proteasome by a specific receptor, which may in turn provide insight into the mechanisms that underlie the allocation of ubiquitylated substrates to different receptor pathways.

## C.6 Acknowledgments

We thank K. R. Yamamoto, H. Kobayashi, and H. Yokosawa for providing reagents and B. M. Padhiar for technical assistance. We thank all current and past members of the Deshaies laboratory, and in particular G. Alexandru, G. Kleiger, and R. Verma, for technical advice, helpful discussions, and support. We thank J. R. Yates III for providing *Sequest*, *2to3*, *DTASelect*, and *Contrast*. R. J. D. and J. G. thank J. R. Yates III, M. MacCoss, and W. H. MacDonald. Without their support, encouragement, and teaching, this work would not have been possible. R. J. D. also thanks R. S. Annan and S. A. Carr for introducing him to the power and possibilities of mass spectrometry.

## C.7 References

- Braun, S., Matuschewski, K., Rape, M., Thoms, S. and Jentsch, S. (2002). Role of the ubiquitin-selective CDC48(UFD1/NPL4) chaperone (segregase) in ERAD of OLE1 and other substrates. *EMBO J*, 21(4):615–21.
- Chau, V., Tobias, J. W., Bachmair, A., Marriott, D. and Ecker, D. J. et al. (1989). A multiubiquitin chain is confined to specific lysine in a targeted short-lived protein. *Science*, 243(4898):1576–83.

- Cherry, J. M., Adler, C., Ball, C., Chervitz, S. A. and Dwight, S. S. et al. (1998). SGD: *Saccharomyces* genome database. *Nucleic Acids Res*, 26(1):73–9.
- Chi, Y., Huddleston, M. J., Zhang, X., Young, R. A. and Annan, R. S. et al. (2001). Negative regulation of Gcn4 and Msn2 transcription factors by Srb10 cyclin-dependent kinase. *Genes Dev*, 15(9):1078–92.
- Denison, C., Rudner, A. D., Gerber, S. A., Bakalarski, C. E. and Moazed, D. et al. (2005). A proteomic strategy for gaining insights into protein sumoylation in yeast. *Mol Cell Proteomics*, 4(3):246–54.
- Deveraux, Q., Ustrell, V., Pickart, C. and Rechsteiner, M. (1994). A 26 S protease subunit that binds ubiquitin conjugates. *J Biol Chem*, 269(10):7059–61.
- Fu, H., Sadis, S., Rubin, D. M., Glickman, M. and van Nocker, S. et al. (1998). Multiubiquitin chain binding and protein degradation are mediated by distinct domains within the 26 S proteasome subunit Mub1. *J Biol Chem*, 273(4):1970–81.
- Ghaemmaghami, S., Huh, W. K., Bower, K., Howson, R. W. and Belle, A. et al. (2003). Global analysis of protein expression in yeast. *Nature*, 425(6959):737–41.
- Graumann, J., Dunipace, L. A., Seol, J. H., McDonald, W. H. and Yates III, J. R. et al. (2004). Applicability of tandem affinity purification MudPIT to pathway proteomics in yeast. *Mol Cell Proteomics*, 3(3):226–37.
- Hartmann-Petersen, R., Hendil, K. B. and Gordon, C. (2003). Ubiquitin binding proteins protect ubiquitin conjugates from disassembly. *FEBS Lett*, 535(1–3):77–81.
- Hicke, L. (2001). Protein regulation by monoubiquitin. *Nat Rev Mol Cell Biol*, 2(3):195–201.

- Hitchcock, A. L., Auld, K., Gygi, S. P. and Silver, P. A. (2003). A subset of membrane-associated proteins is ubiquitinated in response to mutations in the endoplasmic reticulum degradation machinery. *Proc Natl Acad Sci USA*, 100(22):12735–40.
- Hoegel, C., Pfander, B., Moldovan, G. L., Pyrowolakis, G. and Jentsch, S. (2002). RAD6-dependent DNA repair is linked to modification of PCNA by ubiquitin and SUMO. *Nature*, 419(6903):135–41.
- Hofmann, K. and Falquet, L. (2001). A ubiquitin-interacting motif conserved in components of the proteasomal and lysosomal protein degradation systems. *Trends Biochem Sci*, 26(6):347–50.
- Huibregtse, J. M., Yang, J. C. and Beaudenon, S. L. (1997). The large subunit of RNA polymerase II is a substrate of the Rsp5 ubiquitin-protein ligase. *Proc Natl Acad Sci USA*, 94(8):3656–61.
- Kanayama, A., Seth, R. B., Sun, L., Ea, C. K. and Hong, M. et al. (2004). TAB2 and TAB3 activate the NF- $\kappa$ B pathway through binding to polyubiquitin chains. *Mol Cell*, 15(4):535–48.
- Kim, I., Mi, K. and Rao, H. (2004). Multiple interactions of rad23 suggest a mechanism for ubiquitylated substrate delivery important in proteolysis. *Mol Biol Cell*, 15(7):3357–65.
- Kornitzer, D., Raboy, B., Kulka, R. G. and Fink, G. R. (1994). Regulated degradation of the transcription factor Gcn4. *EMBO J*, 13(24):6021–30.
- Mayor, T. and Deshaies, R. J. (2005). Two-Step affinity purification of multiubiquitylated proteins from *Saccharomyces cerevisiae*. *Methods Enzymol*, 399:385–92.

- Mayor, T., Lipford, J., Graumann, J., Smith, G. and Deshaies, R. (2005). Analysis of polyubiquitin conjugates reveals that the rpn10 substrate receptor contributes to the turnover of multiple proteasome targets. *Mol Cell Proteomics*, 4(6):741–51.
- Mazon, M. J. and Hemmings, B. A. (1979). Regulation of *Saccharomyces cerevisiae* nicotinamide adenine dinucleotide phosphate-dependent glutamate dehydrogenase by proteolysis during carbon starvation. *J Bacteriol*, 139(2):686–9.
- McDonald, W. H., Tabb, D. L., Sadygov, R. G., MacCoss, M. J. and Venable, J. et al. (2004). MS1, MS2, and SQT—three unified, compact, and easily parsed file formats for the storage of shotgun proteomic spectra and identifications. *Rapid Commun Mass Spectrom*, 18(18):2162–2168.
- Meimoun, A., Holtzman, T., Weissman, Z., McBride, H. J. and Stillman, D. J. et al. (2000). Degradation of the transcription factor Gcn4 requires the kinase Pho85 and the SCF(CDC4) ubiquitin–ligase complex. *Mol Biol Cell*, 11(3):915–27.
- Minard, K. I. and McAlister-Henn, L. (1994). Glucose–induced phosphorylation of the MDH2 isozyme of malate dehydrogenase in *Saccharomyces cerevisiae*. *Arch Biochem Biophys*, 315(2):302–9.
- Orlicky, S., Tang, X., Willems, A., Tyers, M. and Sicheri, F. (2003). Structural basis for phosphodependent substrate selection and orientation by the SCFCdc4 ubiquitin ligase. *Cell*, 112(2):243–56.
- Ortolan, T. G., Tongaonkar, P., Lambertson, D., Chen, L. and Schaubert, C. et al. (2000). The DNA repair protein rad23 is a negative regulator of multi–ubiquitin chain assembly. *Nat Cell Biol*, 2(9):601–8.

- Panse, V. G., Hardeland, U., Werner, T., Kuster, B. and Hurt, E. (2004). A proteome-wide approach identifies sumoylated substrate proteins in yeast. *J Biol Chem*, 279(40):41346–51.
- Peng, J., Schwartz, D., Elias, J. E., Thoreen, C. C. and Cheng, D. et al. (2003). A proteomics approach to understanding protein ubiquitination. *Nat Biotechnol*, 21(8):921–6.
- Petroski, M. D. and Deshaies, R. J. (2003). Redundant degrons ensure the rapid destruction of Sic1 at the G1/S transition of the budding yeast cell cycle. *Cell Cycle*, 2(5):410–1.
- Piotrowski, J., Beal, R., Hoffman, L., Wilkinson, K. D. and Cohen, R. E. et al. (1997). Inhibition of the 26 S proteasome by polyubiquitin chains synthesized to have defined lengths. *J Biol Chem*, 272(38):23712–21.
- Raasi, S., Orlov, I., Fleming, K. G. and Pickart, C. M. (2004). Binding of polyubiquitin chains to ubiquitin-associated (UBA) domains of HHR23A. *J Mol Biol*, 341(5):1367–79.
- Raasi, S. and Pickart, C. M. (2003). Rad23 ubiquitin-associated domains (UBA) inhibit 26 S proteasome-catalyzed proteolysis by sequestering lysine 48-linked polyubiquitin chains. *J Biol Chem*, 278(11):8951–9.
- Rao, H. and Sastry, A. (2002). Recognition of specific ubiquitin conjugates is important for the proteolytic functions of the ubiquitin-associated domain proteins Dsk2 and Rad23. *J Biol Chem*, 277(14):11691–5.
- Rosas-Acosta, G., Russell, W. K., Deyrieux, A., Russell, D. H. and Wilson, V. G. (2005). A universal strategy for proteomic studies of SUMO and other ubiquitin-like modifiers. *Mol Cell Proteomics*, 4(1):56–72.

- Sadygov, R. G., Eng, J., Durr, E., Saraf, A. and McDonald, H. et al. (2002). Code developments to improve the efficiency of automated MS/MS spectra interpretation. *J Proteome Res*, 1(3):211–5.
- Schena, M., Picard, D. and Yamamoto, K. R. (1991). Vectors for constitutive and inducible gene expression in yeast. *Methods Enzymol*, 194:389–98.
- Schubert, U., Anton, L. C., Gibbs, J., Norbury, C. C. and Yewdell, J. W. et al. (2000). Rapid degradation of a large fraction of newly synthesized proteins by proteasomes. *Nature*, 404(6779):770–4.
- Semple, C. A. (2003). The comparative proteomics of ubiquitination in mouse. *Genome Res*, 13(6B):1389–94.
- Spence, J., Gali, R. R., Dittmar, G., Sherman, F. and Karin, M. et al. (2000). Cell cycle-regulated modification of the ribosome by a variant multiubiquitin chain. *Cell*, 102(1):67–76.
- Tabb, D. L., McDonald, W. H. and Yates III, J. R. (2002). DTASelect and Contrast: Tools for assembling and comparing protein identifications from shotgun proteomics. *J Proteome Res*, 1(1):21–6.
- Thrower, J. S., Hoffman, L., Rechsteiner, M. and Pickart, C. M. (2000). Recognition of the polyubiquitin proteolytic signal. *EMBO J*, 19(1):94–102.
- Turner, G. C. and Varshavsky, A. (2000). Detecting and measuring cotranslational protein degradation *in vivo*. *Science*, 289(5487):2117–20.
- Verma, R., Oania, R., Graumann, J. and Deshaies, R. J. (2004). Multiubiquitin chain receptors define a layer of substrate selectivity in the ubiquitin–proteasome system. *Cell*, 118(1):99–110.
- Weissman, A. M. (2001). Themes and variations on ubiquitylation. *Nat Rev Mol Cell Biol*, 2(3):169–78.

- Wilkinson, C. R., Seeger, M., Hartmann-Petersen, R., Stone, M. and Wallace, M. et al. (2001). Proteins containing the UBA domain are able to bind to multi-ubiquitin chains. *Nat Cell Biol*, 3(10):939–43.
- Wohlschlegel, J. A., Johnson, E. S., Reed, S. I. and Yates III, J. R. (2004). Global analysis of protein sumoylation in *Saccharomyces cerevisiae*. *J Biol Chem*, 279(44):45662–8.
- Wu, G., Xu, G., Schulman, B. A., Jeffrey, P. D. and Harper, J. W. et al. (2003). Structure of a beta-TrCP1–Skp1–beta-catenin complex: Destruction motif binding and lysine specificity of the SCF(beta-TrCP1) ubiquitin ligase. *Mol Cell*, 11(6):1445–56.
- Zhao, Y., Kwon, S. W., Anselmo, A., Kaur, K. and White, M. A. (2004). Broad spectrum identification of cellular small ubiquitin-related modifier (SUMO) substrate proteins. *J Biol Chem*, 279(20):20999–1002.
- Zhou, W., Ryan, J. J. and Zhou, H. (2004). Global analyses of sumoylated proteins in *Saccharomyces cerevisiae*. Induction of protein sumoylation by cellular stresses. *J Biol Chem*, 279(31):32262–8.

## D The WD40 Protein Caf4p is a Component of the Mitochondrial Fission Machinery and Recruits Dnm1p to Mitochondria

This chapter represents a further example of the use of MudPIT to analyze the polypeptide mixtures of moderate complexity resulting from affinity purified protein complexes. It was published as

Griffin, E. E., Graumann, J. and Chan, D. C. (2005). The WD40 protein Caf4p is a component of the mitochondrial fission machinery and recruits Dnm1p to mitochondria. *J Cell Biol*, 170(2):237–48.

The copyright is held by The Rockefeller University Press. This is an authorized reprint. Supplementary material is present at <http://www.jcb.org/cgi/content/full/jcb.200503148/DC1>.

### D.1 Abstract

The mitochondrial division machinery regulates mitochondrial dynamics and consists of Fis1p, Mdv1p, and Dnm1p. Mitochondrial division relies on the recruitment of the dynamin-related protein Dnm1p to mitochondria. Dnm1p recruitment depends on the mitochondrial outer membrane protein Fis1p. Mdv1p interacts with Fis1p and Dnm1p, but is thought to act at a late step during fission because Mdv1p is dispensable for Dnm1p localization. We identify the WD40 repeat protein Caf4p as a Fis1p-associated protein that localizes to mitochondria in a Fis1p-dependent manner. Caf4p interacts with each component of the fission apparatus: with Fis1p

and Mdv1p through its NH<sub>2</sub>-terminal half and with Dnm1p through its COOH-terminal WD40 domain. We demonstrate that *mdv1*Δ yeast contain residual mitochondrial fission due to the redundant activity of Caf4p. Moreover, recruitment of Dnm1p to mitochondria is disrupted in *mdv1*Δ *caf4*Δ yeast, demonstrating that Mdv1p and Caf4p are molecular adaptors that recruit Dnm1p to mitochondrial fission sites. Our studies support a revised model for assembly of the mitochondrial fission apparatus.

## D.2 Introduction

Mitochondria are dynamic organelles that undergo fusion and fission. These processes intermix the mitochondria within cells and control their morphology. In addition to controlling mitochondrial shape, recent studies have also implicated components of the fission machinery in regulation of programmed cell death (Frank et al. 2001; Fannjiang et al. 2004; and Jagasia et al. 2005). Genetic approaches in *Saccharomyces cerevisiae* have identified *DNM1*, *FIS1*, and *MDV1* as components of the mitochondrial fission pathway (Shaw and Nunnari 2002). Dnm1p and its mammalian homologue Drp1 are members of the extensively studied dynamin family of large, oligomeric GTPases. Although the precise mechanism remains controversial, dynamins may couple GTP hydrolysis to a conformational constriction that causes membrane scission (Praefcke and McMahon 2004). In yeast cells, Dnm1p dynamically localizes to dozens of puncta that are primarily associated with mitochondria (Otsuga et al. 1998; Bleazard et al. 1999; Sesaki and Jensen 1999; and Legesse-Miller et al. 2003). A subset of these puncta are sites of future fission.

The assembly of functional Dnm1p complexes on mitochondria is a critical issue in understanding the mechanism of mitochondrial fission. The mitochondrial outer membrane protein Fis1p is required for the formation of normal Dnm1p puncta on mitochondria. In *fis1* $\Delta$  cells, Dnm1p puncta are primarily cytosolic or form abnormally large aggregates on mitochondria (Otsuga et al. 1998; Bleazard et al. 1999; Sesaki and Jensen 1999; and Legesse-Miller et al. 2003). Mdv1p interacts with Fis1p through its NH<sub>2</sub>-terminal half and with Dnm1p through its COOH-terminal WD40 domain. However, Mdv1p appears dispensable for Dnm1p assembly on mitochondria because *mdv1* $\Delta$  cells show little or no change in Dnm1p localization, even though mitochondrial fission is disrupted (Fekkes et al. 2000; Tieu and Nunnari 2000; Tieu et al. 2002; and Cervený and Jensen 2003). These observations have led to two important features of a recently proposed model for mitochondrial fission (Shaw and Nunnari 2002; Tieu et al. 2002; and Osteryoung and Nunnari 2003). First, Fis1p acts to assemble and distribute Dnm1p on mitochondria in an Mdv1p-independent step. Second, Mdv1p acts downstream of Dnm1p localization to stimulate membrane scission. An alternative model proposes that Dnm1p marks the site of mitochondrial fission and recruits Fis1p and Mdv1p into an active fission complex (Shaw and Nunnari 2002; Tieu et al. 2002; and Osteryoung and Nunnari 2003). Again, in this model Mdv1p functions downstream of Dnm1p localization.

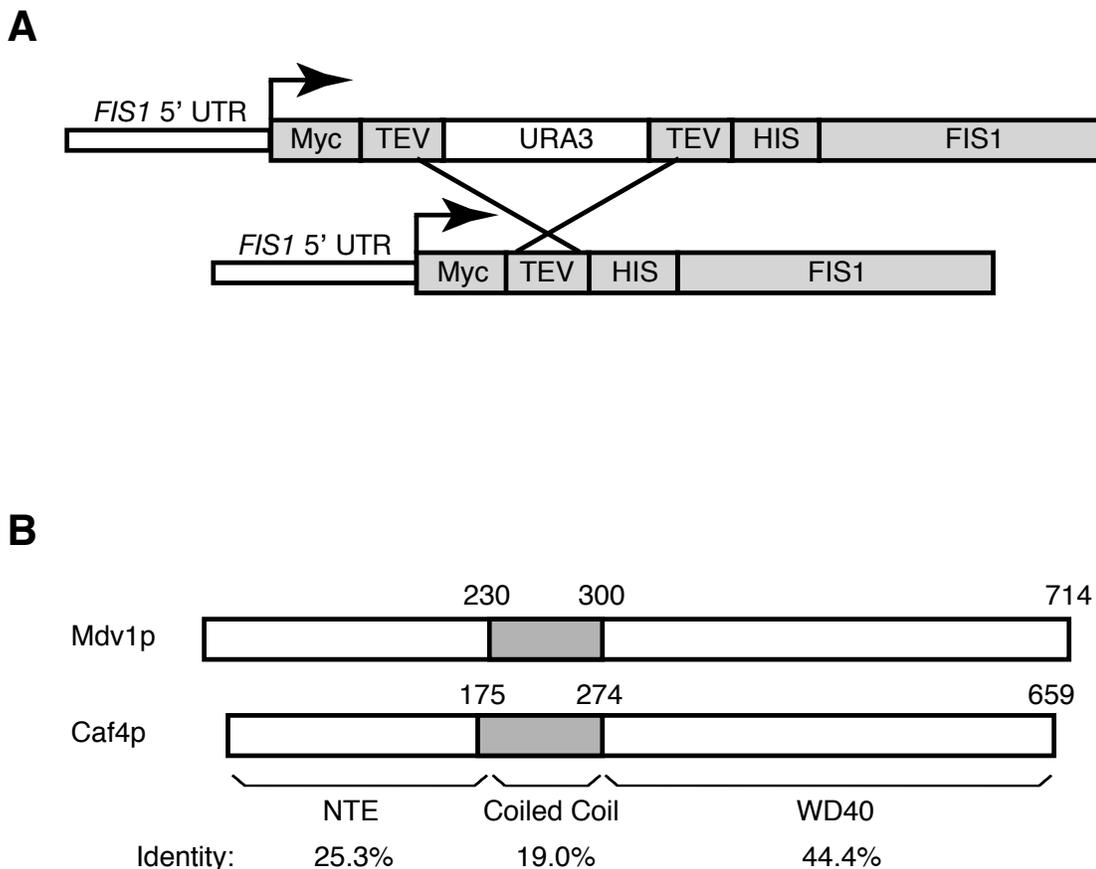
Despite extensive efforts, however, there is no evidence that Fis1p can interact directly with Dnm1p. We speculated that there may be an additional component of the mitochondrial fission pathway required for the Fis1p-dependent assembly of Dnm1p puncta on mitochondria. Because a genome-wide screen for mitochondrial morphology mutants (Dimmer et al. 2002) did not yield obvious candidates, we used a biochemical approach to identify additional components of the mitochondr-

ial fission machinery. Using immunopurification and mass spectrometry, we have identified the WD40 repeat protein Caf4p as a Fis1p-interacting protein. Caf4p localizes to mitochondria and associates with Fis1p, Mdv1p, and Dnm1p. Moreover, we show that *mdv1* $\Delta$  cells are only partially deficient in mitochondrial fission due to the redundant activity of Caf4p. Importantly, Caf4p mediates recruitment of Dnm1p puncta to mitochondria in *mdv1* $\Delta$  yeast. Inclusion of *CAF4* significantly clarifies the current models for mitochondrial fission.

## D.3 Results

### D.3.1 Caf4p is Associated with Fis1p

To identify Fis1p-associated proteins by multidimensional protein identification technology (MudPIT) (Dimmer et al. 2002), we constructed a yeast strain containing endogenous Fis1p with an NH<sub>2</sub>-terminal tandem affinity tag (fig. D.1A). NH<sub>2</sub>-terminal tagging is necessary because *FIS1* is nonfunctional when COOH-terminally tagged (unpublished data). We first designed a recombination cassette containing 9 $\times$ Myc/TEV/*URA3*/TEV/His<sub>8</sub> modules (fig. D.1A). After targeted integration into the *FIS1* locus, spontaneous and precise recombination between the flanking  $\approx$  50 bp tobacco etch virus (TEV) protease sites excises *URA3*. This strategy was used to generate a yeast strain (DCY 1557) that expresses a functional Fis1p with an NH<sub>2</sub>-terminal 9 $\times$ Myc/TEV/His<sub>8</sub> tag (M<sub>9</sub>TH-Fis1p) from the endogenous locus.



**Figure D.1 Construction of M<sub>9</sub>TH-*FIS1* and Caf4p/Mdv1p Alignment.** (A) A 9×Myc-TEV-*URA3*-TEV-His<sub>8</sub> cassette was PCR amplified with *FIS1*-targeting primers and integrated in-frame into the NH<sub>2</sub>-terminus of *FIS1*. Pop-out of the *URA3* cassette by recombination between flanking TEV sites yielded M<sub>9</sub>TH-*FIS1* under the control of the endogenous *FIS1* promoter. UTR: untranslated region. (B) Schematic of Mdv1p and Caf4p. The NH<sub>2</sub>-terminal extension (NTE), coiled-coil, and WD40 regions are shown with percent identity. Overall identity is 37 and overall similarity is 57.

Tandem affinity-purified M<sub>9</sub>TH-Fis1p was subjected to MudPIT analysis in two independent experiments (see materials and methods). Fis1p was identified in both experiments (61.3% coverage, 14 unique peptides; 58.7% coverage, 9 unique peptides). Mdv1p, a previously identified member of the mitochondrial fission pathway and a known Fis1p-interacting protein, was also identified in both experiments (21.1% coverage, 12 unique peptides; 10.2% coverage, 5 unique peptides). These

data confirmed that our MudPIT procedure could preserve and identify Fis1p complexes relevant to mitochondrial fission. Dnm1p was not observed in either dataset, in agreement with previous immunoprecipitation experiments (Mozdy et al. 2000). The complete datasets are presented in Table S1 (available at <http://www.jcb.org/cgi/content/full/jcb.200503148/DC1>).

Interestingly, peptides derived from the WD40 repeat protein Caf4p were identified in both Fis1p MudPIT experiments (24.4% coverage, 9 unique peptides; 8.5% coverage, 3 unique peptides). *CAF4* (YKR036C) was first identified in a yeast two-hybrid screen for CCR4p-interacting proteins (Liu et al. 2001). CCR4p is a central component of the CCR4–NOT transcriptional regulator and cytosolic deadenylase complex (Denis and Chen 2003). Caf4p is the nearest homologue of Mdv1p in *S. cerevisiae* (38% identity and 57% similarity), and the two proteins show extensive sequence identity throughout their lengths (fig. D.1B). Both proteins share a unique NH<sub>2</sub>-terminal extension (NTE; 25.3% identity), a central coiled-coil (CC) domain (19% identity) and a COOH-terminal WD40 repeat domain (44.4% identity). The Caf4p CC scores significantly more weakly ( $\approx 0.3$  probability) than the Mdv1p coiled coil ( $\approx 1.0$  probability) in the MultiCoil prediction program (Wolf et al. 1997).

### D.3.2 Caf4p Interacts with Components of the Mitochondrial Fission Machinery

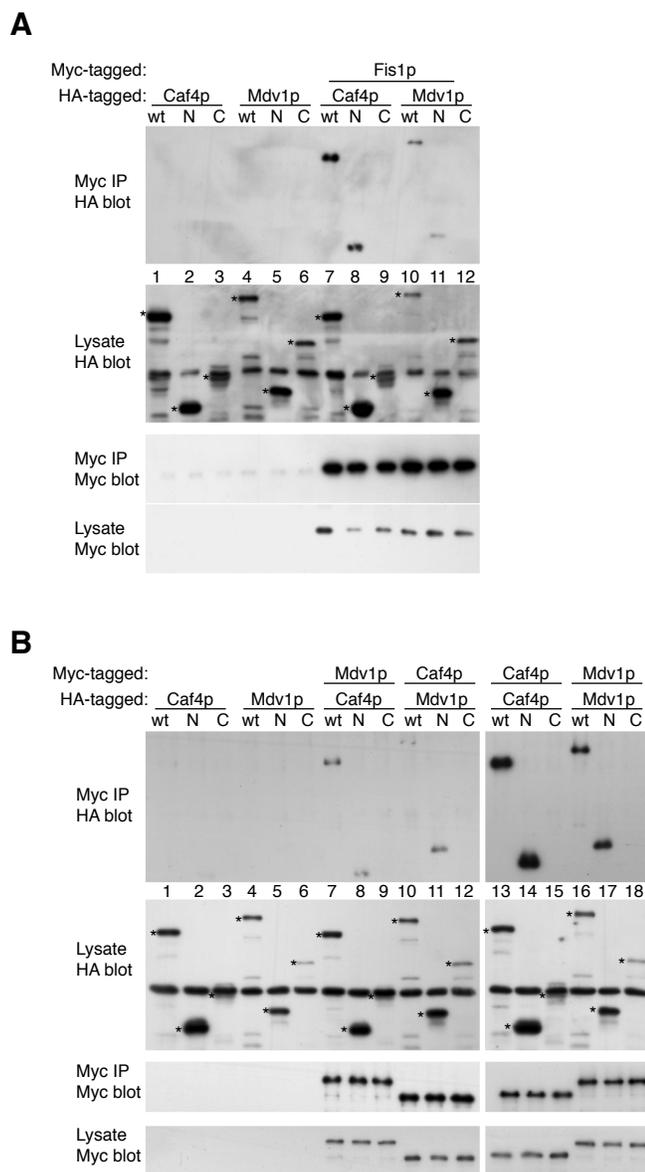
We sought independent confirmation of the physical interaction between Fis1p and Caf4p. For immunoprecipitation experiments, Caf4p–HA or Mdv1p–HA were expressed from their endogenous promoters in strains carrying chromosomal M<sub>3</sub>TH–*FIS1* (3×Myc/TEV/His<sub>8</sub>–*FIS1*) and deleted for *CAF4* or *MDV1*, respectively. When

M<sub>3</sub>TH–Fis1p was immunoprecipitated,  $\approx 5\%$  of both Caf4p–HA and Mdv1p–HA coprecipitated (fig. D.2A, lanes 7 and 10).

Previous yeast two–hybrid analysis determined that the NTE/CC region of Mdv1p (residues 1–300) is responsible for its interaction with Fis1p (Tieu et al. 2002). We detected the same interaction by coimmunoprecipitation (fig. D.2A, lane 11). Additionally, we found that the analogous region of Caf4p (residues 1–274) also interacted with Fis1p (fig. D.2A, lane 8). A shorter Caf4p fragment lacking the majority of the predicted coiled coil (residues 1–250) interacted equally well with Fis1p (unpublished data). In contrast, Fis1p did not bind to the COOH–terminal regions of either Mdv1p or Caf4p (fig. D.2A, lanes 9 and 12). These data suggest that both Caf4p and Mdv1p likely interact with Fis1p through a common mechanism involving the NTE domain.

We also used a yeast two–hybrid assay to analyze the interaction of Caf4p and Mdv1p with Fis1p and Dnm1p (Table D.8). Full–length Caf4p and an NTE/CC fragment of Caf4p interacted strongly with the cytosolic portion of Fis1p (residues 1–128), consistent with our immunoprecipitation data. Similar interactions were observed between Fis1p and both full–length Mdv1p and the NTE/CC region of Mdv1p, as has been previously reported (Tieu et al. 2002; and Cerveny and Jensen 2003). The WD40 domain of both Mdv1p and Caf4p interacted strongly with Dnm1p. However, full–length Mdv1p interacted more weakly and an interaction between full–length Caf4p and Dnm1p was not detected. These results suggest that the interaction of the WD40 domain with Dnm1p is regulated and may be inhibited by the NH<sub>2</sub>–terminal region of Caf4p and Mdv1p.

We also detected homotypic and heterotypic interactions between Caf4p and Mdv1p. Approximately 5% of Caf4p–HA and Caf4p–N–HA (residues 1–274), but



**Figure D.2 Caf4p and Mdv1p Coimmunoprecipitation Experiments.** (A) Yeast carrying the indicated HA- and Myc-tagged constructs were lysed and immunoprecipitated with an anti-Myc antibody. Total lysates (labeled “Lysate”) and immunoprecipitated samples (labeled “Myc IP”) were analyzed by immunoblotting with anti-Myc (9E10) and anti-HA (12CA5) antibodies as indicated. The expression constructs were: Caf4p wt (residues 1–659), Caf4p N (residues 1–274), Caf4p C (residues 275–659), Mdv1 wt (residues 1–714), Mdv1p N (residues 1–300), and Mdv1p C (residues 301–714). The yeast backgrounds were: (A) wild-type, lanes 1–6; *caf4* $\Delta$  M<sub>3</sub>TH-*FIS1*, lanes 7–9; *mdv1* $\Delta$  M<sub>3</sub>TH-*FIS1*, lanes 10–12; (B) wild-type, lanes 1–6; *caf4* $\Delta$  MDV1-HTM, lanes 7–9; *mdv1* $\Delta$  CAF4-HTM, lanes 10–12; CAF4-HTM, lanes 13–15; MDV1-HTM, lanes 16–18. (B) Yeast carrying the indicated HA- and Myc-tagged constructs were immunoprecipitated and analyzed as in A. Immunoprecipitated samples were loaded at 10 (A) and 20 (B) equivalents of the lysate samples. HA-tagged proteins in the lysate are marked with an asterisk. The HA-tagged Caf4p C polypeptide comigrates with a background band in the total lysate blot probed with HA antibody.

**Table D.8 Caf4p and Mdv1p Interact With Dnm1p and Fis1p in a Yeast Two-hybrid Assay.** Caf4p, Mdv1p, Fis1p, and Dnm1p fragments were scored for growth (+), no growth (−), or poor growth (weak) on adenine-deficient plates. All constructs showed no growth when paired with empty activation domain or DNA-binding domain vector. Binding domain fusions are listed across the top of the table and activation domain fusions are listed down the left. Caf4p and Mdv1p N and C fragments are defined in fig. D.2.

		Fis1p	Dnm1p	Caf4p			Mdv1p		
				wt	N	C	wt	N	C
Fis1p		−	−	+	+	−	+	+	−
Dnm1p		−	+	−	−	−	−	−	−
Caf4p	wt	+	−	−	−	−	−	weak	−
	N	+	−	−	−	−	−	−	−
	C	−	+	−	−	−	−	−	−
Mdv1p	wt	+	+	−	+	−	+	+	−
	N	+	−	−	+	−	+	+	−
	C	−	+	−	−	−	−	−	−

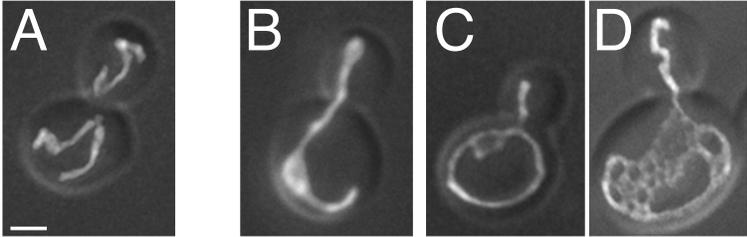
not Caf4-C-HA (residues 275–659), coimmunoprecipitate with full-length Caf4p-HTM (fig. D.2B, lanes 13–15). A similar level of Mdv1p-HA and Mdv1p-N-HA (residues 1–300), but not Mdv1-C-HA (residues 301–714), coimmunoprecipitated with Mdv1p-HTM (fig. D.2B, lanes 16–18). When Caf4p-HTM was precipitated,  $\approx 1\%$  of Mdv1p-HA and Mdv1p-N-HA, but not Mdv1p-C-HA, coprecipitated (fig. D.2B, lanes 10–12). Similarly, when Mdv1p-HTM was precipitated,  $\approx 1\%$  of Caf4p-HA and Caf4p-N-HA, but not Caf4p-C-HA, coprecipitated (fig. D.2B, lanes 7–9). Moreover, the NTE/CC regions of Caf4p and Mdv1p interact in the two-hybrid assay (Table D.8). Therefore, Caf4p interacts with all three members of the fission pathway, with the NH<sub>2</sub>-terminal region mediating interactions with Fis1p, Mdv1p, and homotypic interactions with Caf4p.

### D.3.3 Caf4p is Involved in Mitochondrial Division

Given that Caf4p interacts with Fis1p, Mdv1p, and Dnm1p, we hypothesized that

Caf4p, like Mdv1p, is a component of the mitochondrial division apparatus. *caf4* $\Delta$  yeast, however, display normal mitochondrial morphology, with tubular mitochondria evenly dispersed around the cell cortex (fig. D.3). Wild-type mitochondrial morphology was also observed at elevated temperatures and on carbon sources other than dextrose (glycerol or galactose; unpublished data). This observation is not surprising, given that *CAF4* was not identified in a genome-wide screen of deletion strains for mitochondrial morphology mutants (Dimmer et al. 2002).

We next tested whether *caf4* $\Delta$  cells show synthetic defects in mitochondrial morphology when other components of the fission machinery are deleted. Yeast defective in mitochondrial fission display net-like mitochondrial morphology due to unopposed mitochondrial fusion (Bleazard et al. 1999; and Sesaki and Jensen 1999). These mitochondrial nets can have a spread morphology (fig. D.33, C and D), or they can collapse to one side of the cell (fig. D.3B). Although *FIS1*, *DNM1*, and *MDV1* are all involved in mitochondrial fission, we found that *mdv1* $\Delta$  cells have a distribution of mitochondrial profiles that can be readily distinguished from both *fis1* $\Delta$  and *dnm1* $\Delta$  cells (fig. D.3). In rich dextrose medium, almost all *fis1* $\Delta$  or *dnm1* $\Delta$  cells (93 % and 90 %, respectively) contain collapsed mitochondrial nets. In contrast, less than half of *mdv1* $\Delta$  cells contain collapsed nets, with the majority displaying a spread net morphology. The spread nets range in morphology from interconnected tubules with several loops (fig. D.3C) to networks with complex fenestrations (fig. D.3D). *mdv1* $\Delta$  *dnm1* $\Delta$  cells behave identically to *dnm1* $\Delta$  cells, with > 90 % collapsed nets in dextrose (fig. D.3). This observation indicates that the *dnm1* $\Delta$  collapsed net phenotype is epistatic to the *mdv1* $\Delta$  spread net phenotype. In rich galactose medium (unpublished data), a greater portion of all strains contain spread nets, but again *mdv1* $\Delta$  cells have a higher percentage of cells with spread



Genotype	Wild-type	Collapsed net	Spread net
Wild-type	100	0	0
<i>caf4</i> $\Delta$	100	0	0
<i>mdv1</i> $\Delta$	0	45.5	54.5
<i>dnm1</i> $\Delta$	0	90	10
<i>fis1</i> $\Delta$	0	92.5	7.5
<i>mdv1</i> $\Delta$ <i>caf4</i> $\Delta$	0	91.5	8.5
<i>mdv1</i> $\Delta$ <i>dnm1</i> $\Delta$	0	91.5	8.5

**Figure D.3** *CAF4* Regulates Mitochondrial Morphology. Strains expressing mitochondrially targeted GFP were grown in YP dextrose to mid-log phase and fixed. The percentage of cells ( $n = 400$ ) with mitochondria having wild-type (A), collapsed net (B), or spread net morphology (C and D) is tabulated. The spread net phenotype encompasses a distribution of morphologies ranging from simple structures containing one or two loops (C) to complexly fenestrated mitochondria with dozens of loops (D). For both wild-type and *caf4* $\Delta$  strains, the wild-type category includes 1% fragmented cells. Bar, 1  $\mu$ m.

nets (80%) compared with *fis1* $\Delta$  (45.5%), *dnm1* $\Delta$  (53%), or *mdv1* $\Delta*dnm1* $\Delta$  cells (40.5%). These results agree with a previous report that *mdv1* $\Delta$  cells have more spread nets compared with *dnm1* $\Delta$  cells in galactose medium (Cervený et al. 2001). However, this study found that the *mdv1* $\Delta$  spread net phenotype is epistatic to the *dnm1* $\Delta$  collapsed net phenotype (Cervený et al. 2001). The reason for this$

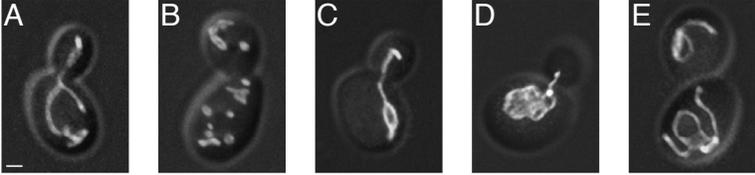
discrepancy is unclear, but we note the *mdv1* $\Delta$  morphology is most distinct in dextrose cultures.

Most interestingly, we found that *mdv1* $\Delta$  *caf4* $\Delta$  cells have mitochondrial net distributions indistinguishable from either *dnm1* $\Delta$  cells or *fis1* $\Delta$  cells. Deletion of *CAF4* in *mdv1* $\Delta$  cells markedly shifts the distribution to one composed almost entirely of collapsed mitochondrial nets (> 90% in dextrose, fig. D.3). Our results support a model in which partial reduction of mitochondrial fission results in predominantly spread mitochondrial nets, and complete loss of fission eventually results in collapse of the nets. That is, *mdv1* $\Delta$  cells retain residual mitochondrial fission, whereas *mdv1* $\Delta$  *caf4* $\Delta$  cells are devoid of fission, similar to *dnm1* $\Delta$ , *fis1* $\Delta$ , or *mdv1* $\Delta$  *dnm1* $\Delta$  cells. An analogous situation appears to exist in mammalian cells, in which weak Drp1 dominant-negative alleles cause the formation of spread nets, whereas strong dominant-negative alleles cause nets to collapse (Cervený et al. 2001).

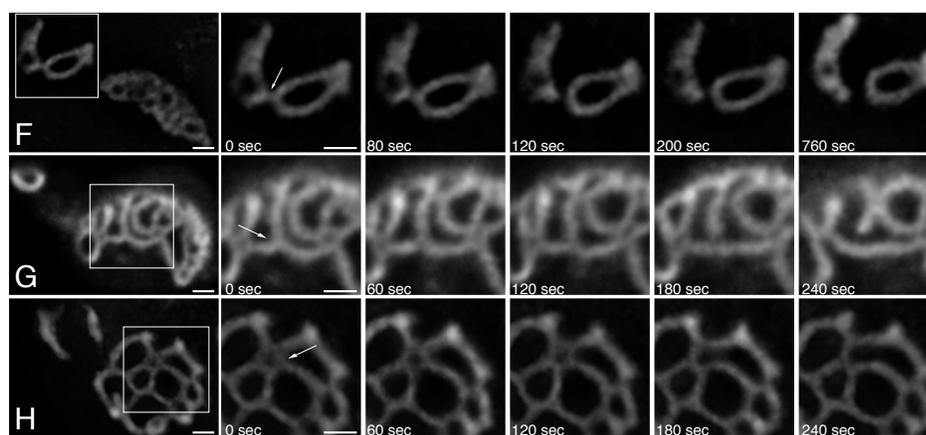
We tested this model by reanalyzing mitochondrial morphologies in the presence of latrunculin A, which disrupts the actin cytoskeleton. Disruption of the actin cytoskeleton leads to rapid fragmentation of the mitochondrial network due to ongoing mitochondrial fission (Boldogh et al. 1998; and Jensen et al. 2000). Latrunculin A treatment rapidly resolves a fraction of collapsed nets into spread nets (Jensen et al. 2000; and Cervený et al. 2001), and allows a closer examination of the degree of connectivity in mitochondrial nets. Similarly, in mammalian cells, collapsed mitochondrial nets induced by overexpression of dominant-negative Drp1 can be spread by the microtubule-depolymerizing agent nocodazole (Smirnova et al. 2001). Both wild-type and *caf4* $\Delta$  yeast treated with latrunculin A show mitochondrial fragmentation (fig. D.4). 80% of *mdv1* $\Delta$  cells treated with latrunculin A contain partial

mitochondrial nets (fig. D.4E, partial net) that are less interconnected and have fewer fenestrations than the collapsed or spread nets that predominate in latrunculin A-treated *dnm1* $\Delta$  or *fis1* $\Delta$  cells. 95% of latrunculin A-treated *mdv1* $\Delta$  *caf4* $\Delta$  cells show either collapsed nets or highly fenestrated spread nets, a profile indistinguishable from that in *dnm1* $\Delta$  or *fis1* $\Delta$  cells (fig. D.4). Thus, after disruption of the actin cytoskeleton, *mdv1* $\Delta$  yeast display a distribution of mitochondrial morphologies that suggest an incomplete defect in mitochondrial fission. In contrast, *mdv1* $\Delta$  *caf4* $\Delta$  yeast have mitochondrial morphologies similar to that in *fis1* $\Delta$  and *dnm1* $\Delta$  yeast. We conclude that *CAF4* mediates low levels of mitochondrial fission in *mdv1* $\Delta$  cells.

We next monitored the mitochondrial network in *mdv1* $\Delta$  cells by time-lapse microscopy to assess the levels of mitochondrial fission. In pilot experiments, we found that free mitochondrial ends produced by fission events in *mdv1* $\Delta$  cells were rapidly involved in fusion events, making unambiguous documentation of fission difficult. Because latrunculin A reduces the levels of fusion and thereby should prolong the presence of free mitochondrial ends, we monitored mitochondrial dynamics in latrunculin A-treated *mdv1* $\Delta$  cells carrying the outer membrane marker OM45-GFP. In 8 out of 10 *mdv1* $\Delta$  cells, we observed at least one fission event in a 30 min recording period (fig. D.4, F-H; videos 1 and 2, available at <http://www.jcb.org/cgi/content/full/jcb.200503148/DC1>). Due to the complexity and rapid rearrangements of the mitochondrial networks in these cells (see videos 1 and 2), these numbers likely underestimate the actual levels of fission. In contrast, no fission events were observed in 8 *mdv1* $\Delta$  *caf4* $\Delta$  cells. We conclude that the ability of *CAF4* to mediate mitochondrial fission events contributes significantly to the spread net morphology of *mdv1* $\Delta$  cells.



Genotype	LatA treatment	Wild-type	Fragments and short tubules	Collapsed net	Spread net	Partial net
Wild-type	-	95	5	0	0	0
	+	8	92	0	0	0
<i>caf4</i> Δ	-	95	4.5	0	0	0.5
	+	8.5	91.5	0	0	0
<i>mdv1</i> Δ	-	0	0	45.5	50.5	4
	+	0	0	5.5	14.5	80
<i>dnm1</i> Δ	-	0	0	96	3.5	0.5
	+	0	0	32.5	63.5	4
<i>fis1</i> Δ	-	0	0	96	4	0
	+	0	0	36.5	58.5	5
<i>mdv1</i> Δ <i>caf4</i> Δ	-	0	0	95	4	1
	+	0	0	38.5	57	4.5
<i>mdv1</i> Δ <i>dnm1</i> Δ	-	0	0	95	4.5	0.5
	+	0	0	26	70.5	3.5



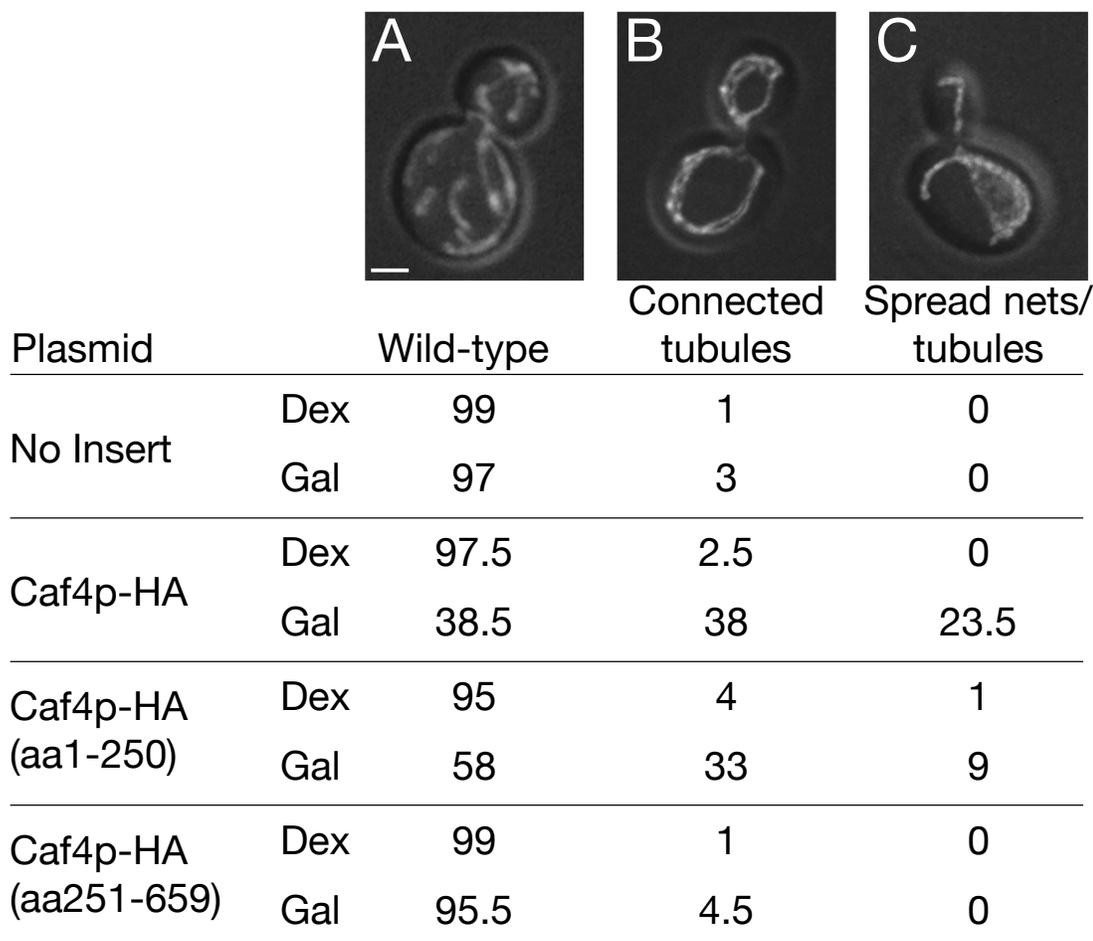
**Figure D.4** *CAF4* Mediates Residual Fission in *mdv1*Δ Cells. Top: mid-log cultures grown in YP dextrose were treated for 60 min with 200 μM latrunculin A (+) or vehicle (-). For each strain, 200 cells were scored into the following phenotypic categories: wild-type (A), fragments and short tubules (B), collapsed net (C), spread net (D), or partial net (E). Numbers shown are percentages. The fragments and short tubules category encompasses a range of morphologies from completely fragmented (as shown in B) to a mixture of fragments and short tubules. (F-H) Still images from time-lapse movies showing fission events in *mdv1*Δ yeast treated with 200 μM latrunculin A. The boxed area in the first frame is magnified in the subsequent sequence of five images. Arrows indicate fission events. Mitochondria were visualized with the outer membrane marker OM45-GFP. Bars, 1 μm.

### D.3.4 Mitochondrial Fission is Blocked by Overexpression of Caf4p or Caf4p Fragments

Because overexpression of Mdv1p or Mdv1p fragments inhibits mitochondrial fission Smirnova et al. (2001), we next tested the effects of Caf4p overproduction. Caf4p-HA under the control of the GalL promoter was expressed  $\approx 20$  times above endogenous levels in rich galactose medium (unpublished data). Spread mitochondrial nets formed in 23.5% cells (fig. D.5C). An additional 38% of cells had an intermediate phenotype that we termed “connected tubules,” consisting of a completely interconnected mitochondrial network in which no tubular ends were detected (fig. D.5B). Overexpression of an NH<sub>2</sub>-terminal fragment that interacts with Fis1p (residues 1–250; unpublished data) had a similar effect (9% spread nets, 33% connected tubules; fig. D.5), suggesting that the formation of mitochondrial net-like structures may result from a dominant-negative effect on Fis1p function. A similar distribution of mitochondrial phenotypes resulted from 20-fold overproduction of Mdv1p-HA (7.5% spread nets and 24.5% interconnected tubules) and an Mdv1p-HA NH<sub>2</sub>-terminal fragment (5% spread nets and 39% interconnected tubules; unpublished data). These data confirm that Caf4p interacts with the mitochondrial fission apparatus.

### D.3.5 Full Bypass Suppression of *fzo1* $\Delta$ Requires Loss of Both *MDV1* and *CAF4*

Yeast fission mutants are able to suppress the glycerol growth defect of cells deficient in mitochondrial fusion (Bleazard et al. 1999). Indeed, *MDV1* was originally identified because of its ability to suppress the glycerol growth defect of strains



**Figure D.5 Caf4p Overexpression Blocks Mitochondrial Fission.** Wild-type yeast (DCY 1979) carrying the pRS 416 GalI vector with no insert, full-length CAF4-HA, CAF4-HA N, or CAF4-HA C were grown in rich dextrose or galactose media for 180 min and fixed. Cells were scored into the following phenotypic categories: wild-type (A), connected tubules (B), or spread nets with tubules (C). Numbers shown are percentages ( $n = 200$ ). Overexpression in galactose cultures was estimated to result in 20-fold greater expression than endogenous levels by Western blots of serially diluted lysates (not depicted). Bar, 1  $\mu$ m.

carrying temperature-sensitive *fzo1* or *mgm1* alleles (Fekkes et al. 2000; Mozdy et al. 2000; Tieu and Nunnari 2000; and Cervený et al. 2001). Deletion of *MDV1* has previously been reported to suppress the glycerol growth defect of *fzo1* $\Delta$  cells less efficiently than deletion of *DNM1* (Cervený et al. 2001). To further test our hypothesis that *mdv1* $\Delta$  cells have only a partial loss of mitochondrial fission, we compared

the efficiencies with which the *mdv1* $\Delta$  and *dnm1* $\Delta$  mutations suppress the glycerol growth defect of *fzo1* $\Delta$  cells. Diploids were sporulated, genotyped, and scored by serial dilution for their ability to grow on glycerol plates relative to dextrose plates (fig. D.6). As expected, all wild-type and no *fzo1* $\Delta$  spores grew on glycerol plates. Of 17 *mdv1* $\Delta$  *fzo1* $\Delta$  spores tested, 7 showed no detectable growth on glycerol and an additional 4 spores grew very poorly, with  $< 1\%$  cell survival on glycerol. Only 3 of the 6 remaining spores showed  $> 20\%$  survival on glycerol. More than half of *dnm1* $\Delta$  *fzo1* $\Delta$  spores grew robustly on glycerol plates, with between 20% and 50% cell survival. Most importantly, the triple mutant *mdv1* $\Delta$  *caf4* $\Delta$  *fzo1* $\Delta$  spores grew much more robustly than the *mdv1* $\Delta$  *fzo1* $\Delta$  spores, with all spores growing on glycerol and the majority between 20% and 50% cell survival. The markedly enhanced bypass suppression of *fzo1* $\Delta$  by *mdv1* $\Delta$  *caf4* $\Delta$  double mutations compared with the *mdv1* $\Delta$  mutation provides genetic evidence that *mdv1* $\Delta$  cells retain residual mitochondrial fission due to the activity of Caf4p.

### D.3.6 Caf4p Localizes to Mitochondria in a Fis1p-dependent Manner

We next sought to determine the subcellular localization of Caf4p. Caf4p was detected in highly purified mitochondrial preparations (Sickmann et al. 2003), and a Caf4p-GFP fusion generated in a genome-wide analysis localizes to mitochondria (Sickmann et al. 2003). We confirmed the mitochondrial localization of Caf4p-GFP, but did not study it further because the GFP fusion protein was not functional when expressed from the *CAF4* locus (unpublished data). We instead used immunofluorescence to localize Myc-tagged versions of Caf4p and Mdv1p (termed Caf4p-HTM and Mdv1p-HTM) that are expressed from the endogenous locus and are functional.

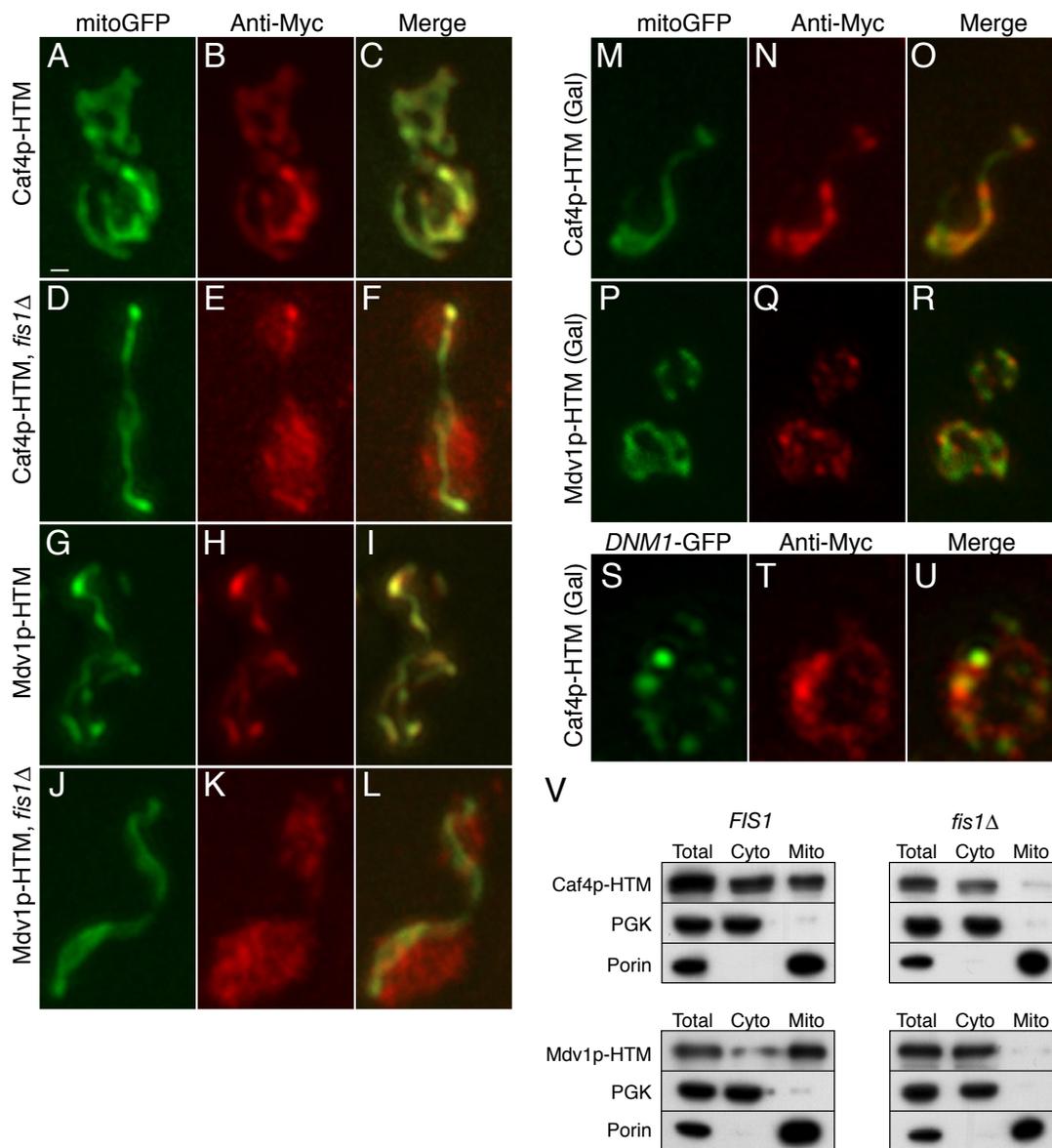


predominantly in the cytosol (fig. D.7, D–F and J–L). In some cells, however, low levels of residual localization to mitochondria could be discerned (e. g., fig. D.7, D–F). In *fis1* mutant yeast, overexpressed GFP–Mdv1p is diffusely cytosolic but also retains some localization to mitochondria (Tieu and Nunnari 2000; and Tieu et al. 2002). Together, these data indicate that the normal mitochondrial localization of both Caf4p and Mdv1p depends largely on Fis1p, although some low levels of residual localization can occur in the absence of Fis1p.

We also evaluated the localization of Caf4p–HTM by subcellular fractionation. We found a significant portion of both Caf4p and Mdv1p in the mitochondrial pellet (fig. D.7V). Mdv1p had previously been shown to be present in mitochondrial fractions (Fekkes et al. 2000; Tieu and Nunnari 2000; and Cervený et al. 2001). However, in *fis1*Δ yeast both proteins behave as cytosolic proteins (fig. D.7V). These data support our immunofluorescence studies and confirm that Mdv1p and Caf4p localize to mitochondria through their association with Fis1p.

### D.3.7 Caf4p Recruits Dnm1p–GFP to Mitochondria

To understand the mechanism of mitochondrial fission, it is crucial to elucidate how Dnm1p is recruited to mitochondria. Given that Mdv1p associates with both Fis1p and Dnm1p, it is puzzling that Dnm1p assembly on mitochondria shows little or no dependence on Mdv1p (Fekkes et al. 2000; Mozdy et al. 2000; Tieu and Nunnari 2000; Tieu et al. 2002; and Cervený and Jensen 2003). With the identification of Caf4p as a component of the fission machinery, we reexamined this issue. We constructed a fully functional Dnm1p–GFP allele and analyzed its localization pattern using deconvolution microscopy (Table D.9). Similar to previous reports



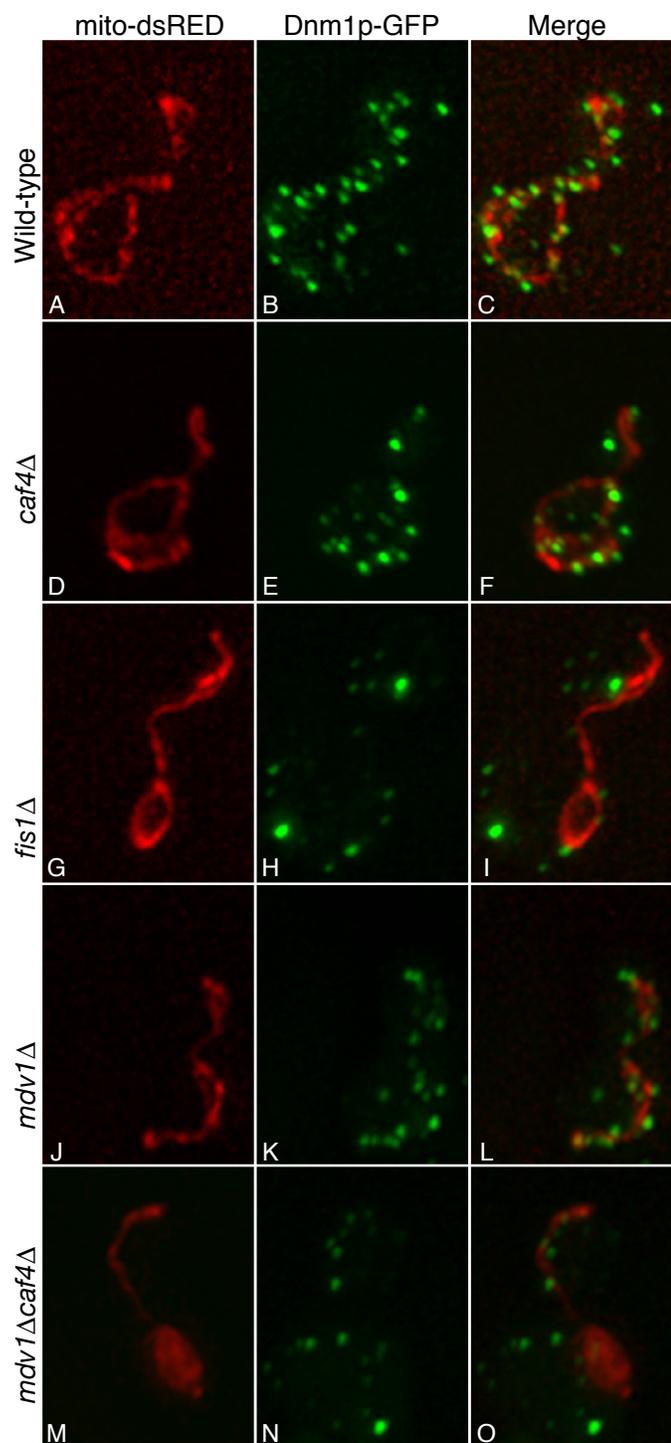
**Figure D.7 Mitochondrial Localization of Caf4p and Mdv1p Requires Fis1p.** Immunofluorescence (red, middle panels) was used to localize Myc-tagged Caf4p (Caf4p-HTM; A-F and M-U) and Mdv1p (Mdv1p-HTM; G-L and P-R) in wild-type (A-C, G-I, and M-U) and *fis1Δ* cells (D-F and J-L). Caf4p-HTM and Mdv1p-HTM are expressed from the endogenous loci and are functional. Mitochondria were labeled with mitochondrially targeted GFP (A-R, left, green). The majority of Dnm1p-GFP puncta colocalize with Caf4p-HTM (S-U). Overlays of the two signals are shown in the merged images (right). Note that both Caf4p and Mdv1p localize to mitochondria in wild-type cells, but are diffusely cytosolic in *fis1Δ* cells. Cells were grown in YP dextrose (A-L) or YP galactose (M-U). Representative maximum intensity projections of deconvolved z-stacks are shown. Bar, 1  $\mu$ m. (V) Caf4p-HTM and Mdv1p-HTM were analyzed by subcellular fractionation. The total cell lysate (Total), high-speed supernatant (Cyto), and mitochondrial pellet (Mito) were analyzed by Western blot with an anti-Myc antibody in wild-type (left) and *fis1Δ* (right) yeast. PGK (3-phosphoglycerate kinase) is a cytosolic marker, and porin is a mitochondrial outer membrane marker.

**Table D.9 Quantification of Dnm1–GFP Puncta Localization.** Dnm1p puncta were scored for colocalization with mitochondrially localized DsRed in deconvolved images. For each genotype, 140 budded cells were analyzed by scoring Dnm1p–GFP spots in both the mother and bud, and the average is presented with the SD in parentheses.

	Mitochondrial	Cytosolic
Wild-type	16.9 ( $\pm 5.5$ )	3.3 ( $\pm 2.1$ )
<i>caf4</i> $\Delta$	15.4 ( $\pm 5.2$ )	5.2 ( $\pm 2.6$ )
<i>mdv1</i> $\Delta$	13.7 ( $\pm 5.0$ )	5.1 ( $\pm 3.0$ )
<i>fis1</i> $\Delta$	4.9 ( $\pm 2.7$ )	9.6 ( $\pm 4.3$ )
<i>mdv1</i> $\Deltacaf4\Delta$	4.8 ( $\pm 2.5$ )	10.4 ( $\pm 3.9$ )

(Otsuga et al. 1998), Dnm1p–GFP is found predominantly in puncta associated with mitochondria (average 16.9 mitochondrial vs. 3.3 cytosolic puncta per cell; Table D.9 and fig. D.7, A–C). Deletion of *CAF4* or *MDV1* alone had little effect on this localization (15.4 mitochondrial vs. 5.2 cytosolic and 13.7 mitochondrial vs. 5.1 cytosolic per cell, respectively; Table D.9 and fig. D.8, D–I). In all these strains, the Dnm1p puncta are relatively uniform in size and intensity.

In contrast, *fis1* $\Delta$  mutants showed dramatic defects, with the majority of the puncta now cytosolic (4.9 mitochondrial vs. 9.6 cytosolic) (Table D.9 and fig. D.8, J–L). As has been previously noted, a small fraction of Dnm1p still localizes to mitochondria in *fis1* $\Delta$  cells (Tieu et al. 2002; and Cervený and Jensen 2003), suggesting that Dnm1p may be recruited by a second pathway, perhaps through an intrinsic affinity for mitochondrial lipids or an unidentified mitochondrial binding partner. Importantly, a similar defect in Dnm1p localization was found in *mdv1* $\Delta$  *caf4* $\Delta$  cells (4.8 mitochondrial vs. 10.4 cytosolic per cell) (Table D.9 and fig. D.8, M–O). In both *fis1* $\Delta$  and *mdv1* $\Delta$  *caf4* $\Delta$  cells, Dnm1p–GFP forms a few large aggregates and numerous less intense puncta. Similar results were obtained using immunofluorescence against a Dnm1p–HTM protein (unpublished data). These data clearly demonstrate that either Caf4p or Mdv1p is sufficient for effective recruitment of



**Figure D.8 Fis1p Mediates Dnm1p-GFP Localization Through Either Mdv1p or Caf4p.** The localization of Dnm1p-GFP (middle, green) was compared to mito-DsRed (left, red) in yeast of the indicated genotype. Merged images are shown on the right. Representative maximum intensity projections of deconvolved z-stacks are shown.

Dnm1p to mitochondria, and that Caf4p is essential for Mdv1p-independent recruitment of Dnm1p by Fis1p.

## D.4 Discussion

### D.4.1 *CAF4* and *MDV1* Perform Similar Functions in Mitochondrial Fission

By applying affinity purification and mass spectrometry to Fis1p, we have identified Caf4p as a novel component of the mitochondrial fission machinery. Our biochemical and genetic characterization indicate that *CAF4* functions in the same manner as *MDV1* in mitochondrial fission. Biochemically, both proteins interact with Fis1p and Dnm1p. Caf4p and Mdv1p share a common domain architecture comprised of an NTE, a central CC, and a COOH-terminal WD40 repeat. The NH<sub>2</sub>-terminal regions mediate oligomerization and association with Fis1p, whereas the COOH-terminal WD40 regions mediate interactions with Dnm1p. In addition, both Caf4p and Mdv1p localize to mitochondria in a Fis1p-dependent manner.

Genetically, both *MDV1* and *CAF4* act positively in the mitochondrial fission pathway. *mdv1*Δ cells are dramatically compromised for mitochondrial fission, but a residual level of fission is mediated by *CAF4*. This residual fission activity is revealed by the observation that *mdv1*Δ yeast have a less severe mitochondrial morphology defect compared with *fis1*Δ or *dnm1*Δ yeast. In contrast, *mdv1*Δ *caf4*Δ yeast display predominantly collapsed mitochondrial nets, identical to those seen in *fis1*Δ and *dnm1*Δ cells. Time-lapse imaging of mitochondria in *mdv1*Δ cells indeed reveals a residual level of fission that is absent from *mdv1*Δ *caf4*Δ

cells. These results directly support our conclusion that the morphology differences between *mdv1* $\Delta$  cells versus *mdv1* $\Delta$  *caf4* $\Delta$ , *fis1* $\Delta$ , and *dnm1* $\Delta$  cells are primarily due to differences in fission rates. It is also possible that the proposed role of Dnm1p in cortical distribution of mitochondria may contribute in part to the morphological differences (Otsuga et al. 1998). The *mdv1* $\Delta$  mutation acts as a weak suppressor of the glycerol growth defect in *fzo1* $\Delta$  cells. The *mdv1* $\Delta$  *caf4* $\Delta$  double mutation suppresses this phenotype much more efficiently. Based on these physical interaction and genetic data, we conclude that Caf4p likely acts in a similar manner to Mdv1p to promote mitochondrial fission.

Why are there two proteins that appear to perform similar and partially redundant roles in mitochondrial fission? This question is particularly intriguing because *caf4* $\Delta$  yeast have normal mitochondrial morphology, indicating that disruption of Caf4p does not cause a major loss of mitochondrial fission. First, *CAF4* may play a more important role in mitochondrial fission under conditions not yet tested. Second, the presence of two proteins mediating interactions between Fis1p and Dnm1p would increase the ability of cells to accurately regulate the rate of mitochondrial fission. The heterotypic and homotypic interactions between Caf4p and Mdv1p may provide an additional layer of regulation. Finally, Caf4p may have an additional function in another pathway. Previous two-hybrid studies have implicated Caf4p in the CCR4-NOT complex, which is thought to be involved in regulation of transcription and/or mRNA processing (Liu et al. 2001).

#### D.4.2 A Revised Model for Mitochondrial Fission

The current models for mitochondrial fission propose that Mdv1p acts late in the

fission pathway. One model proposes a two-step pathway in which Fis1p first recruits Dnm1p, in an Mdv1p-independent manner. Mdv1p then acts as a molecular adaptor at a postrecruitment step, along with Fis1p, to promote fission by Dnm1p (Shaw and Nunnari 2002; Tieu et al. 2002; and Osteryoung and Nunnari 2003). A second model also proposes that Mdv1p acts after Dnm1p recruitment to organize an active fission complex (Cervený and Jensen 2003).

Our study reveals a new role for Mdv1p and Caf4p early in mitochondrial fission. Fis1p recruits Dnm1p to mitochondrial fission complexes through Mdv1p or Caf4p, which act as molecular adaptors. This revised model is strongly supported by our demonstration that Dnm1p recruitment in *mdv1* $\Delta$  yeast depends on Caf4p function. In the absence of both Mdv1p and Caf4p, Fis1p is unable to recruit Dnm1p.

Although Mdv1p and Caf4p clearly act early in the fission pathway, there is evidence that at least Mdv1p has a subsequent role in the activation of fission, as previously proposed (Shaw and Nunnari 2002; Tieu et al. 2002; and Cervený and Jensen 2003). In *caf4* $\Delta$  cells, Mdv1p recruits Dnm1p to fission complexes, and fission occurs at apparently normal levels. However, in *mdv1* $\Delta$  cells, Caf4p is similarly able to recruit Dnm1p to fission complexes, but mitochondrial fission is severely compromised. Therefore, Mdv1p and Caf4p can independently recruit Dnm1p, but complexes recruited by Mdv1p appear to be more highly active. These observations suggest that Dnm1p recruitment by itself is insufficient for fission to occur. Indeed, studies of Dnm1p dynamics indicates that most Dnm1p puncta do not result in fission (Legesse-Miller et al. 2003). Our identification of Caf4p as part of the fission machinery clarifies the early steps in mitochondrial fission. Future studies will need to define the additional steps beyond Dnm1p recruitment necessary for fission.

## D.5 Materials and Methods

### D.5.1 Media and Yeast Genetic Techniques

Yeast strains are listed in Table S1. Standard genetic techniques and yeast media were used. SC and YP media supplemented with either 2% dextrose, 3% glycerol, 2% raffinose, or 2% galactose were prepared as described previously (Guthrie and Fink 1991). YJG 12 and DCY 1557 are in the W303 background. All other strains are in the S288C background. *fis1::KanMX6*, *mdv1::KanMX6*, *caf4::KanMX6*, and *dnm1::KanMX6* are derived from the *MATa* deletion library (Open Biosystems).

### D.5.2 Plasmid Construction

The M<sub>9</sub>TH cassette was generated as follows. Primers Eg 258 (see Table S3, available at <http://www.jcb.org/cgi/content/full/jcb.200503148/DC1>) and Eg 259 were used to amplify *URA3* from pRS 416 (Stratagene). Eg 260 and Eg 4, an *FZO1* reverse primer, were used to amplify a TEV/His<sub>8</sub> module from EG 704 (pRS 414+9×Myc/TEV/His<sub>8</sub>-*FZO1*). The 3' end of the *URA3* product overlaps by 18 bp with the 5' end of the TEV/His<sub>8</sub> product. This overlap allows them to anneal together and be amplified in a second PCR with the primers Eg 258 and Eg 4. The *URA3*/TEV/His<sub>8</sub> product was cloned into pRS 403 as an EcoRV/SalI fragment (which removes all *FZO1* sequence), resulting in EG 928. 9×Myc/TEV was amplified with Eg 256 and Eg 260 from EG 704 and fused to the 5' end of *URA3* (Eg 258/259 product) by mixing and amplifying with Eg 256 and Eg 259. The

resulting product was cloned into EG 928 as an EcoRV/EcoRI fragment, yielding EG 940 (pRS 403+9×Myc/TEV/*URA3*/TEV/His<sub>8</sub>). EG 940 was converted to pRS 403+3×Myc/TEV/*URA3*/TEV/His<sub>8</sub> by digesting with Xba1, yielding EG 957.

To construct HA-tagged versions of *CAF4* and *CAF4* fragments, *CAF4* sequences were PCR amplified from *end3Δ* genomic DNA (Open Biosystems). First, the *CAF4* 3' untranslated region (UTR) was amplified with the primers Eg 313 and Eg 314 and cloned as a KpnI/SalI fragment into pRS 416, resulting in pRS 416+*CAF4* 3' UTR. 3×HA was amplified with Eg 327 and Eg 328 and cloned as a SalI/XhoI fragment into the SalI site to generate pRS 416+3×HA/*CAF4* 3' UTR. The *CAF4* 5' UTR was cloned as a SacI/SpeI fragment using Eg 312 and Eg 317, resulting in pRS 416+*CAF4* 5' UTR/3×HA/3' UTR. Full-length *CAF4* was amplified with Eg 316 and Eg 315 and cloned as a SpeI/XhoI fragment into the SpeI/SalI sites, resulting in EG 1041. *CAF4* N (residues 1–274) and C (residues 275–659) were amplified with Eg 316/Eg 353 and Eg 315/Eg 352, respectively, and cloned as SpeI/XhoI fragments, resulting in EG 1045 and EG 1043. Four independent clones encoded glutamine at residue 110 and arginine at residue 111. Full-length *CAF4*–HA was able to complement *caf4Δ* in *caf4Δ mdv1Δ* yeast, indicating that it is functional.

To construct HA-tagged versions of *MDV1* and *MDV1* fragments, *MDV1* sequences were amplified by PCR from *end3Δ* genomic DNA. First, the *MDV1* 3' UTR was amplified with the primers Eg 323 and Eg 324 and cloned as a SacI/SalI fragment into pRS 416. A 3×HA cassette was added as described for *CAF4*–HA, resulting in the plasmid pRS 416+3×HA–*MDV1* 3' UTR. The *MDV1* 5' UTR was amplified with primers Eg 320 and Eg 322 and cloned as a SacII/SpeI fragment, resulting in pRS 416+*MDV1* 5' UTR/3×HA/3' UTR. Full-length *MDV1* was amplified using primers Eg 109 and Eg 321 and cloned as a SpeI/XhoI fragment into the

SpeI/SalI sites, resulting in EG 1047. *MDV1* N (residues 1–300) and C (residues 301–714) were amplified with Eg 323/Eg 326 and Eg 321/325, respectively, and cloned as SpeI/XhoI fragments, resulting in EG 1051 and EG 1049. Full-length *MDV1*-HA complemented the mitochondrial morphology defects in *mdv1* $\Delta$  cells.

The galactose-inducible Caf4p expression vectors EG 1133 (Caf4p-HA), EG 1135 (Caf4p-HA, residues 251–659), and EG 1136 (Caf4p-HA, residues 1–250) were generated by replacing the *CAF4* 5' UTR in EG 1041, EG 1043, and EG 1045 with a SacI/ClaI GalL promoter fragment from p 413 GalL (Mumberg et al. 1994) containing a start codon inserted between the XbaI and EcoRI sites.

pRS 403+GPD/mito-GFP (EG 686) was generated by first cloning the GPD promoter from p 413 GPD (Mumberg et al. 1995) as a SacI (blunt)/SpeI fragment into the SmaI/SpeI sites of pRS 403 (Stratagene), yielding EG 128. Next, a HindIII (blunt)/NotI mito-GFP fragment from pYES-mtGFP (Westermann and Neupert 2000) was inserted into EG 128 linearized with SpeI (blunt)/NotI. pRS 403+GPD/mito-DsRed (EG 823) was generated by subcloning DsRed into the BamHI and NotI sites of EG 686, replacing GFP with DsRed. OM45 was PCR amplified with primers Eg 151 and Eg 154 and cloned as an XhoI/XbaI fragment with an XbaI/BamHI GFP fragment into the XhoI/BamHI sites of pRS 416, yielding pRS 416+OM45-GFP (EG 252).

### D.5.3 Yeast Strain Construction

An M<sub>9</sub>TH-*FIS1* strain was generated by amplifying the 9×Myc/TEV/*URA3*/TEV/His<sub>8</sub> cassette from EG 943 (pRS 403-9×Myc/TEV/*URA3*/TEV/His<sub>8</sub>) with the *FIS1*-targeting primers Eg 261 and Eg 262 and transforming YJG 12. *URA3*+ transfor-

ants were screened by PCR for correct integration (2 out of 8 positive), grown overnight in YPD to allow for loss of *URA3*, and plated on 5-FOA plates. Colonies were screened by Western blotting for expression of M<sub>9</sub>TH-Fis1p (9 out of 16 positive). This strain displayed wild-type morphology in 64% of cells and moderate defects in the remaining cells. The same strategy was used to generate M<sub>3</sub>TH-*FIS1* from the pRS 403-3×Myc/TEV/*URA3*/TEV/His<sub>8</sub> template (EG 957) for subsequent experiments in the S288C background. This strain (DCY 2192) displayed wild-type morphology in 89% of cells and mild defects in the remaining cells. DCY 2192 was crossed to *mdv1*Δ and *caf4*Δ strains (Open Biosystems MATa deletion library) and sporulated to generate M<sub>3</sub>TH-*FIS1 mdv1*Δ (DCY 2302) and M<sub>3</sub>TH-*FIS1 caf4*Δ (DCY 2305).

*fzo1::HIS5* was generated by transformation with a *HIS5* (*Saccharomyces kluyveri*) fragment amplified with the *FZO1* targeting primers Eg 9 and Eg 10. mito-GFP was integrated to the *leu2*Δ0 locus by transformation with NarI-digested EG 686 (pRS 403+GPD/mito-GFP). mito-DsRed was integrated to the *leu2*Δ0 locus by transformation with HpaI-digested EG 823 (pRS 403+GPD/mito-DsRed). *dnm1*Δ::*HIS5* was generated by transformation with a *HIS5* (*S. kluyveri*) fragment amplified with the *DNM1*-targeting primers Eg 57 and Eg 58.

Chromosomal *CAF4*-HTM was generated by transformation of DCY1979 with a His<sub>8</sub>/2TEV/9×Myc/*HIS5* cassette (Westermann and Neupert 2000) amplified with the *CAF4* targeting primers Eg 284 and Eg 285. Chromosomal *MDV1*-HTM was generated transformation with the same cassette amplified with *MDV1* targeting primers Eg 80 and Eg 81. Both *CAF4*-HTM and *MDV1*-HTM are functional because 70% of *CAF4*-HTM *mdv1*Δ yeast display spread mitochondrial nets and 95% of *MDV1*-HTM yeast cells display wild-type mitochondrial morphology.

*DNM1*-GFP was generated by amplifying GFP/*HIS5* from pKT 128 (Sheff and Thorn 2004) with Eg 342 and Eg 343. This product was transformed into DCY 1626 (wild-type yeast with mito-DsRed) to generate DCY 2370. DCY 2370 was crossed to *fis1* $\Delta$  and *mdv1* $\Delta$  *caf4* $\Delta$  strains to generate strains DCY 2404 (*DNM1*-GFP *fis1* $\Delta$ ), DCY 2414 (*DNM1*-GFP *caf4* $\Delta$ ), DCY 2417 (*DNM1*-GFP *mdv1* $\Delta$ ), and DCY 2418 (*DNM1*-GFP *mdv1* $\Delta$  *caf4* $\Delta$ ).

#### D.5.4 Tandem Affinity Purification MudPIT

Pellets from 2 l cultures ( $OD_{600\text{ nm}} \approx 1.5$ ) grown in YPD were prepared essentially as described previously for HPM tag Dual-Step affinity purification (Sheff and Thorn 2004), with the following modifications. Fungal protease inhibitors were used (Sigma-Aldrich) and lysates were cleared at 20 kg for 15 min. Cleavage from 9E10 beads was performed with GST-TEV protease for 3 h at RT. The second affinity step was performed with 40  $\mu$ l Magne-His beads (Promega). Samples were proteolytically digested and analyzed by multidimensional chromatography in-line with a Deca XP ion trap mass spectrometer (ThermoElectron) as described previously (Mayor et al. 2005). Samples were released stepwise from the strong cation exchanger phase of the triphasic capillary columns as reported previously (Mayor et al. 2005).

#### D.5.5 Immunoprecipitation

*CAF4*-HA (EG 1041), *CAF4*-HA residues 1-274 (EG 1043), and *CAF4*-HA residues 275-659 (EG 1045) were expressed in strains DCY 1979 (wild-type) and DCY 2305

( $M_3TH-FIS1\ caf4\Delta$ ).  $MDV1$ -HA (EG 1047),  $MDV1$ -HA residues 1–300 (EG 1049), and  $MDV1$ -HA residues 301–714 (EG 1051) were expressed in DCY1979 or DCY 2302 ( $M_3TH-FIS1\ mdv1\Delta$ ). Cultures were grown in selective SD media and harvested at  $OD_{600\text{ nm}} \approx 0.8$ . 20  $OD_{600\text{ nm}}$  units of cells were lysed with glass beads (40 s with a vortex mixer, 4 times) in 500  $\mu\text{l}$  ice-cold lysis buffer (50 mM Tris, pH 7.4, 150 mM NaCl, 0.1 mM EDTA, and 0.2% Triton X-100) in the presence of Fungal protease inhibitors (Sigma-Aldrich). Lysates were cleared by centrifuging 5 min at 5 krpm and 15 min at 14 krpm. At this point, a total lysate sample was taken. 400  $\mu\text{l}$  of cleared lysate was mixed with a 20  $\mu\text{l}$  bead volume of 9E10-conjugated protein A-Sepharose beads (Sigma-Aldrich) for 90 min. Beads were washed four times with 1 ml washes of lysis buffer. Precipitate was eluted with 100  $\mu\text{l}$  SDS buffer at 95°C for 5 min. SDS-PAGE Western blotting was performed with 9E10 hybridoma supernatant (anti-Myc) or 12CA5 ascites fluid (anti-HA).

### D.5.6 Yeast Two-hybrid Assay

pGAD vectors were transformed into PJ69-4 $\alpha$ . pGBDU vectors were transformed into PJ69-4a (James et al. 1996). Indicated vectors were mated on YPD plates using two transformants for each vector (totaling four matings per combination). Diploids were selected by replica plating to SD-leu-ura plates. Interactions were assayed by replica plating to SD-leu-ura-lys-ade and incubating for 4 d at 30°C.

### D.5.7 Mitochondrial Morphology Analysis

Mitochondrially targeted GFP (mito-GFP) was used to monitor mitochondrial morphology. DCY1979 (wild-type), DCY1945 ( $caf4\Delta$ ), DCY1984 ( $caf4\Delta\ mdv1\Delta$ ),

DCY 2009 (*fis1* $\Delta$ ), DCY 2128 (*mdv1* $\Delta$ ), DCY 2155 (*mdv1* $\Delta$  *dnm1* $\Delta$ ), and DCY 2312 (*dnm1* $\Delta$ ) were grown overnight, diluted 1:20 into fresh medium, and grown for 4 h at 30°C. Cultures were fixed by adjusting cultures to 3.7% formaldehyde and incubated 10 min at 30°C. Cells were washed 4 times with 1 ml PBS and scored for mitochondrial morphology. For *CAF4* overexpression studies, plasmids p416 GalL/*CAF4*-HA (EG 1133), p416 GalL/*CAF4*-HA residues 251–659 (EG 1135), or p416 GalL/*CAF4*-HA residues 1–250 (EG 1137) were transformed into DCY 1979. Cultures were grown overnight in selective SRaff and diluted 1:20 in fresh YPD or YPGal and grown 3 h at 30°C. Samples were taken for Western analysis and the remaining culture was fixed as described above.

For latrunculin A treatment, overnight YPD cultures were diluted 1:20 in fresh YPD and grown for 3 h. Cultures were then treated for 1 h at 30°C with 200  $\mu$ M latrunculin A in or with an equivalent amount of vehicle (DMSO). Cultures were then fixed as described above.

For time-lapse imaging, overnight SGal cultures were diluted 1:20 in fresh YPGal and grown for 3 h. Cells were pelleted, resuspended in fresh media, and embedded in 1% low melting point agarose containing 200  $\mu$ M latrunculin A.

### D.5.8 Bypass Suppression Assay

DCY 2002 and DCY 2343 were sporulated and dissected onto YPD plates. Spores were picked, grown overnight in 3 ml YPD at 30°C, pelleted, and resuspended to  $OD_{600\text{ nm}} \approx 1.0$  in YP. 3  $\mu$ l of 1:5 serial dilutions were spotted on YPD and YPGlycerol and grown at 30°C for 2 and 4 d, respectively, to determine the fraction of cells that grow on glycerol. Genotypes were determined by PCR.

### D.5.9 Differential Centrifugation

Yeast strains *CAF4*-HTM (DCY 2055), *CAF4*-HTM *fis1* $\Delta$  (DCY 2094), *MDV1*-HTM (DCY 2053), and *MDV1*-HTM *fis1* $\Delta$  (DCY 2097) were grown in YPD and harvested at  $OD_{600\text{ nm}} \approx 1.2$ . 100 OD units of cells were spheroplasted with zymolyase and lysed in a small clearance Dounce homogenizer (0.6 M sorbitol and 100 mM Tris, pH 7.4). The lysate was spun twice at 2.9 krpm for 5 min. An aliquot of the second supernatant was saved as the total lysate sample. The second supernatant was spun at 10 krpm for 10 min, and an aliquot of the supernatant was saved as the cytosol sample. The pellet was resuspended and spun again at 10 krpm for 10 min. An aliquot of the final pellet was saved as the mitochondrial pellet. Equal cell equivalents were loaded for Western blot analysis. The difference in porin intensity between the total and mitochondrial fractions most likely results from fewer obscuring proteins in the mitochondrial fraction.

### D.5.10 Imaging

Images were acquired on a microscope (Axiovert 200M; Carl Zeiss MicroImaging, Inc.) using a 100 $\times$  Plan-Apochromat, NA 1.4, oil-immersion objective. Z-stack images (between 0.1 and 0.2  $\mu\text{m}$  intervals for still images and between 0.3 and 0.4  $\mu\text{m}$  intervals for time-lapse images) were collected at RT with an ORCA-ER camera (Hamamatsu), controlled by AxioVision 4.2 software. Images were collected at either 30 or 40 s intervals for 30 min for time-lapse experiments. Iterative deconvolutions were performed with Axiovision 4.2 and maximum intensity projections were generated with AxioVision 4.2 for still images and Image J for time-lapse images.

Fluorescent images in Figs. D.3–D.5 were overlaid with differential interference contrast images (set at 50 % opacity) in Adobe Photoshop CS.

### D.5.11 Immunofluorescence

Cells were processed for immunofluorescence essentially as described previously (Guthrie and Fink 1991) with the following modifications. Cultures were fixed for 15 min with 3.7 % formaldehyde. Tween 20 (0.5 %) was included in blocking buffer (PBS, 1 % BSA) during a 15 min block step. Cells were stained with 9E10 hybridoma supernatant and a Cy3-conjugated anti-mouse secondary antibody. Washes after primary and secondary incubations were 5 min with blocking buffer, 5 min with blocking buffer containing 0.5 % Tween 20, and two 5 min washes with blocking buffer. All incubations were performed at RT. GelMount (Biomed) was used as mounting medium to preserve fluorescence.

### D.5.12 Online supplemental material

Table S1 lists proteins identified in MudPIT experiments with M<sub>9</sub>TH-Fis1p. Table S2 shows yeast strains. Table S3 lists primer sequences. Videos 1 and 2 show mitochondrial fission in *mdv1Δ* yeast. Mitochondria were monitored by the mitochondrial outer membrane marker OM45-GFP. Arrows highlight a subset of fission events. Online supplemental material available at <http://www.jcb.org/cgi/content/full/jcb.200503148/DC1>.

## D.6 Acknowledgments

We thank T.R. Suntoke, S.A. Detmer, and H. Chen for helpful comments on the manuscript; and L. Lytle for technical assistance. J. Copeland constructed some of the two-hybrid plasmids. We thank Dr. Ray Deshaies for stimulating discussions on mass spectrometry.

E.E. Griffin was supported by a National Institutes of Health training grant (NIH GM07616) and a Ferguson fellowship. MudPIT analysis was performed in the mass spectrometry facility of the laboratory of R.J. Deshaies (Investigator, Howard Hughes Medical Institute, Caltech). This facility is supported by the Beckman Institute at Caltech and by a grant from the Department of Energy to R.J. Deshaies and B.J. Wold. J. Graumann is supported by R.J. Deshaies through Howard Hughes Medical Institute funds. D.C. Chan is a Bren Scholar and Beckman Young Investigator. This research was supported by the National Institutes of Health (GM62967).

## D.7 References

Bleazard, W., McCaffery, J. M., King, E. J., Bale, S. and Mozdy, A. et al. (1999).

The dynamin-related GTPase Dnm1 regulates mitochondrial fission in yeast. *Nat Cell Biol*, 1(5):298–304.

Boldogh, I., Vojtov, N., Karmon, S. and Pon, L. A. (1998). Interaction between

mitochondria and the actin cytoskeleton in budding yeast requires two integral mitochondrial outer membrane proteins, Mmm1p and Mdm10p. *J Cell Biol*, 141(6):1371–81.

- Cervený, K. L. and Jensen, R. E. (2003). The WD-repeats of Net2p interact with Dnm1p and Fis1p to regulate division of mitochondria. *Mol Biol Cell*, 14(10):4126–39.
- Cervený, K. L., McCaffery, J. M. and Jensen, R. E. (2001). Division of mitochondria requires a novel DMN1-interacting protein, Net2p. *Mol Biol Cell*, 12(2):309–21.
- Denis, C. L. and Chen, J. (2003). The CCR4–NOT complex plays diverse roles in mRNA metabolism. *Prog Nucleic Acid Res Mol Biol*, 73:221–50.
- Dimmer, K. S., Fritz, S., Fuchs, F., Messerschmitt, M. and Weinbach, N. et al. (2002). Genetic basis of mitochondrial function and morphology in *Saccharomyces cerevisiae*. *Mol Biol Cell*, 13(3):847–53.
- Fannjiang, Y., Cheng, W. C., Lee, S. J., Qi, B. and Pevsner, J. et al. (2004). Mitochondrial fission proteins regulate programmed cell death in yeast. *Genes Dev*, 18(22):2785–97.
- Fekkes, P., Shepard, K. A. and Yaffe, M. P. (2000). Gag3p, an outer membrane protein required for fission of mitochondrial tubules. *J Cell Biol*, 151(2):333–40.
- Frank, S., Gaume, B., Bergmann-Leitner, E. S., Leitner, W. W. and Robert, E. G. et al. (2001). The role of dynamin-related protein 1, a mediator of mitochondrial fission, in apoptosis. *Dev Cell*, 1(4):515–25.
- Griffin, E. E., Graumann, J. and Chan, D. C. (2005). The WD40 protein Caf4p is a component of the mitochondrial fission machinery and recruits Dnm1p to mitochondria. *J Cell Biol*, 170(2):237–48.
- Guthrie, C. and Fink, G. (1991). *Guide to Yeast Genetics and Molecular Biology*, pages 933. Academic Press, San Diego, CA.

- Jagasia, R., Grote, P., Westermann, B. and Conradt, B. (2005). DRP-1-mediated mitochondrial fragmentation during EGL-1-induced cell death in *C. elegans*. *Nature*, 433(7027):754–60.
- James, P., Halladay, J. and Craig, E. A. (1996). Genomic libraries and a host strain designed for highly efficient two-hybrid selection in yeast. *Genetics*, 144(4):1425–36.
- Jensen, R. E., Hobbs, A. E., Cervený, K. L. and Sesaki, H. (2000). Yeast mitochondrial dynamics: Fusion, division, segregation, and shape. *Microsc Res Tech*, 51(6):573–83.
- Legesse-Miller, A., Massol, R. H. and Kirchhausen, T. (2003). Constriction and Dnm1p recruitment are distinct processes in mitochondrial fission. *Mol Biol Cell*, 14(5):1953–63.
- Liu, H. Y., Chiang, Y. C., Pan, J., Chen, J. and Salvatore, C. et al. (2001). Characterization of CAF4 and CAF16 reveals a functional connection between the CCR4-NOT complex and a subset of SRB proteins of the RNA polymerase II holoenzyme. *J Biol Chem*, 276(10):7541–8.
- Mayor, T., Lipford, J., Graumann, J., Smith, G. and Deshaies, R. (2005). Analysis of polyubiquitin conjugates reveals that the rpn10 substrate receptor contributes to the turnover of multiple proteasome targets. *Mol Cell Proteomics*, 4(6):741–51.
- Mozdy, A. D., McCaffery, J. M. and Shaw, J. M. (2000). Dnm1p GTPase-mediated mitochondrial fission is a multi-step process requiring the novel integral membrane component Fis1p. *J Cell Biol*, 151(2):367–80.

- Mumberg, D., Muller, R. and Funk, M. (1994). Regulatable promoters of *Saccharomyces cerevisiae*: comparison of transcriptional activity and their use for heterologous expression. *Nucleic Acids Res*, 22(25):5767–8.
- Mumberg, D., Muller, R. and Funk, M. (1995). Yeast vectors for the controlled expression of heterologous proteins in different genetic backgrounds. *Gene*, 156(1):119–22.
- Osteryoung, K. W. and Nunnari, J. (2003). The division of endosymbiotic organelles. *Science*, 302(5651):1698–704.
- Otsuga, D., Keegan, B. R., Brisch, E., Thatcher, J. W. and Hermann, G. J. et al. (1998). The dynamin-related GTPase, Dnm1p, controls mitochondrial morphology in yeast. *J Cell Biol*, 143(2):333–49.
- Praefcke, G. J. and McMahon, H. T. (2004). The dynamin superfamily: Universal membrane tubulation and fission molecules?. *Nat Rev Mol Cell Biol*, 5(2):133–47.
- Sesaki, H. and Jensen, R. E. (1999). Division versus fusion: Dnm1p and Fzo1p antagonistically regulate mitochondrial shape. *J Cell Biol*, 147(4):699–706.
- Shaw, J. M. and Nunnari, J. (2002). Mitochondrial dynamics and division in budding yeast. *Trends Cell Biol*, 12(4):178–84.
- Sheff, M. A. and Thorn, K. S. (2004). Optimized cassettes for fluorescent protein tagging in *Saccharomyces cerevisiae*. *Yeast*, 21(8):661–70.
- Sickmann, A., Reinders, J., Wagner, Y., Joppich, C. and Zahedi, R. et al. (2003). The proteome of *Saccharomyces cerevisiae* mitochondria. *Proc Natl Acad Sci USA*, 100(23):13207–12.

- Smirnova, E., Griparic, L., Shurland, D. L. and van der Bliek, A. M. (2001). Dynamin-related protein Drp1 is required for mitochondrial division in mammalian cells. *Mol Biol Cell*, 12(8):2245–56.
- Tieu, Q. and Nunnari, J. (2000). Mdv1p is a WD repeat protein that interacts with the dynamin-related GTPase, Dnm1p, to trigger mitochondrial division. *J Cell Biol*, 151(2):353–66.
- Tieu, Q., Okreglak, V., Naylor, K. and Nunnari, J. (2002). The WD repeat protein, Mdv1p, functions as a molecular adaptor by interacting with Dnm1p and Fis1p during mitochondrial fission. *J Cell Biol*, 158(3):445–52.
- Westermann, B. and Neupert, W. (2000). Mitochondria-targeted green fluorescent proteins: convenient tools for the study of organelle biogenesis in *Saccharomyces cerevisiae*. *Yeast*, 16(15):1421–7.
- Wolf, E., Kim, P. S. and Berger, B. (1997). MultiCoil: A program for predicting two- and three-stranded coiled coils. *Protein Sci*, 6(6):1179–89.

## About this Document

This document was typeset in Computer Modern (Knuth 1986), using ConTeXt (Hagen 2001), a modern L<sup>A</sup>T<sub>E</sub>X–analogous (Lamport 1994) macro package around the document typesetting language pdfT<sub>E</sub>X (Thanh 1998), an extension to T<sub>E</sub>X by Knuth (1984) capable of producing Adobe’s *portable document format* (pdf, Adobe Systems Incorporated 2004).

J. G.’s experience in maintaining parallel manuscript versions in pdfT<sub>E</sub>Xed high quality (submission to review) on the one hand and MicroSoft Word–readable format (collaborator interaction, final submission for the majority of biological journals) on the other hand was unpleasant (Graumann et al. 2004, chapter 2, was written with two code bases). ConTeXt’s capability of interpreting *extensible markup language* (XML, Yergeau et al. 2004) in combination with the XML import functionality provided by OpenOffice (<http://www.openoffice.org>) present a way out of source duplicity: pdfT<sub>E</sub>Xed (ConT<sub>E</sub>Xted) and OpenOffice documents (exportable to MicroSoft Word) from one XML source.

J. G.’s ongoing C<sub>o</sub>nT<sub>E</sub>X<sub>M</sub>L project seeks to provide a set of XML entities suitable for scientific publication as well as corresponding stylesheets mapping the resulting XML code into ConT<sub>E</sub>Xt and OpenOffice. This document is entirely coded in XML, implying that the ConT<sub>E</sub>Xt–mapping style sheet is fairly complete at the time of writing. The OpenOffice specific stylesheet is present in rudiments and still under development.

## References

- Adobe Systems Incorporated (2004). *PDF Reference—Adobe Portable Document Format, Version 1.6.* , 5th edition.
- Graumann, J., Dunipace, L. A., Seol, J. H., McDonald, W. H. and Yates III, J. R. et al. (2004). Applicability of tandem affinity purification MudPIT to pathway proteomics in yeast. *Mol Cell Proteomics*, 3(3):226–37.
- Hagen, H. (2001). *ConTeXt, The Manual.* PRAGMA ADE, Hasselt, The Netherlands.
- Knuth, D. E. (1986). *Computers & Typesetting, Volume E: Computer Modern Typefaces.* Addison–Wesley Longman Publishing Co., Boston, MA, USA.
- Knuth, D. E. (1984). *Computers & Typesetting, Volume A: The TeXbook.* Addison–Wesley Longman Publishing Co., Boston, MA, USA.
- Lamport, L. (1994). *LaTeX, a Document Preparation System.* Addison–Wesley, 2nd edition.
- Thanh, H. T. (1998). The pdfTeX Program. *Cahiers GUTenberg*, pages 197–210.
- Yergeau, F., Cowan, J., Bray, T., Paoli, J. and Sperberg-McQueen, C. M. et al. (2004). XML 1.1, W3C Recommendation.

