Protein Science (1994), *3*:2435-2446. Cambridge University Press. Printed in the USA.

High-sensitivity sequencing of large proteins: Partial structure of the rapamycin-FKBP12 target

HEDIYE ERDJUMENT-BROMAGE,[1] MARY LUI,[1] DAVID M. SABATINI,[2] SOLOMON H. SNYDER,[2] AND PAUL TEMPST[1]

[1] Molecular Biology Program, Memorial Sloan-Kettering Cancer Center, New York, New York 10021
[2] Department of Neuroscience, Johns Hopkins University School of Medicine, Baltimore, Maryland 21205

(RECEIVED August 11, 1994; ACCEPTED September 30, 1994)

Abstract

We report on studies leading to refinements of various steps of the protein internal sequencing process. Specifically, the developments comprise (1) higher-sensitivity chemical sequencing through background reduction; (2) improved peptide recovery from rapid in situ digests of nanogram amount, nitrocellulose-bound proteins; and (3) accurate UV spectroscopic identification of Trp- and Cys-containing peptides. In addition, we describe strategies for 2-dimensional liquid chromatographic peptide isolation from complex mixtures and a multi-analytical approach to peptide sequence analysis (Edman sequencing, matrix-assisted laser desorption mass spectrometry, and UV spectroscopy). Both strategies were applied in tandem to the primary structural analysis of a gel-purified, 250-kDa protein (mammalian target of rapamycin-FKBP12 complex), available in low picomolar quantities only. More than 300-amino acids worth of sequence was obtained in mostly uninterrupted stretches, several containing Trp, Cys, His, and Ser. That information has allowed the matching of a biological function of a mammalian protein to a yeast gene product with a well-characterized mutant phenotype. The results also demonstrate that extended chemical sequencing analysis (e.g., 26 successive amino acids) is now feasible, starting with initial yields well below 1 pmol.

Keywords: chemical sequencing; in situ proteolysis; liquid chromatography; matrix-assisted laser-desorption mass spectrometry; rapamycin; ultraviolet spectroscopy; 250-kDa protein

Protein sequencing is an integral part of basic health and life sciences. For decades, direct covalent analysis was the only way to get primary structure information, a tedious task involving several proteolytic digests and exhaustive analysis (Ambler, 1963). During the past 10 years, however, a major new role has been to provide partial sequences to assist in cloning of the corresponding genes (Hunkapiller et al., 1984; Tempst et al., 1990). Cloned DNA can be readily analyzed and the entire protein primary structure deduced; in addition, it facilitates studies on transcriptional regulation of the gene and of structure/function and cellular interactions of its product. With a large number of genes sequenced and repositories rapidly expanding (Dujon et al., 1994; Wilson et al., 1994), partial protein sequences have also allowed, with increasing frequency, the matching of biological function (or regulation) to a specific database entry (Merino et al., 1993; Dou et al., 1994; Fumagalli et al., 1994). The focus of protein sequencing has therefore shifted again, to questions on protein associations in the cell. After discovery of one or more components of a functional complex, known or presumed to form during various processes (e.g., signal transduction, cell cycle and differentiation, immunosuppression, vesicle targeting, among others), the question typically arises with which other proteins they might interact (Liu et al., 1991; Gout et al., 1993; Sollner et al., 1993). Because many of the targets are only available in minute quantities, it is imperative that analytical studies be carried out at the highest levels of sensitivity.

Despite current high performance of polypeptide analysis instruments, positive results are not always and not universally obtained. Because this relates primarily to loss and contamination of the analyte during preparation, high-resolution 1D and

Reprint requests to: Paul Tempst, Memorial Sloan-Kettering Cancer Center, 1275 York Avenue, New York, New York 10021; e-mail: p-tempst@mskcc.org.

Abbreviations: ATZ, anilinothiazolinone; AUFS, absorption unit full scale; BLAST, basic local alignment search tool; DPU, diphenylurea; DTT, dithiothreitol; 1D, 1-dimensional; 2D, 2-dimensional; FKBP, FK506 binding protein; G6PD, glucose-6-phosphate dehydrogenase; GST, glutathione S-transferase; ID, inner diameter; IY, initial yield; Lys-C, endoproteinase Lys-C; MALDI-MS, matrix-assisted laser-desorption ionization mass spectrometry; MALDI-TOF MS, matrix-assisted laser desorption ionization time-of-flight mass spectrometry; MeCN, acetonitrile; *m/z*, mass to charge; NC, nitrocellulose; OD, optical density; pe, pyridylethyl; PITC, phenylisothiocyanate; PTH, phenylthiohydantoin; PVDF, polyvinylidene difluoride; PVP, polyvinylpyrrolidone; RP-HPLC, reversed-phase high performance liquid chromatography; RTX-100, hydrogenated triton X-100; RY, repetitive yield; TFA, trifluoroacetic acid; TW, tween.

2D gel electrophoresis (all or not followed by electroblotting) as a final fractionation step has largely overcome these problems. Another major obstacle in microsequencing is the frequent occurrence of posttranslationally modified N-termini. This condition renders the affected proteins inaccessible ("blocked") for cyclic chemical degradation. Blocked proteins must be fragmented, and resulting peptides separated and individually analyzed to obtain sequence information. Because an estimated 80% of all soluble proteins in higher eukaryotic cells are N-α-acetylated (Brown & Roberts, 1976), digests are almost a universal requirement for successful protein analysis. On the positive side, partial sequences from different parts of the protein facilitate cloning experiments, especially when a PCR-based approach is taken. Different digest techniques for gel-fractionated proteins have been developed (Aebersold et al., 1987; Eckerskorn & Lottspeich, 1989; Fernandez et al., 1992; Rosenfeld et al., 1992; Williams et al., 1993), all of which have since been adapted for use at the 10–20 pmol level.

Most recently, multi-analytical approaches have been taken to characterization of peptide covalent structure (Hess et al., 1993; Wong et al., 1993; Geromanos et al., 1994). Specific strategies are primarily determined by the intended use of the sequence information. For example, gene cloning experiments require several stretches of uninterrupted sequence. Unfortunately, ambiguous calls during the first few cycles (from contaminations), gaps in the sequence (e.g., presence of W, C, H, S, modified residues), and inability to identify the C-terminal Lys or Arg (of tryptic peptides) are quite common with automated chemical sequencing at low picomole levels (initial yield below 2 pmol); sequences can also be mixed. Accurate mass measurements have now greatly facilitated interpretation/deconvolution of such incomplete/complex data.

For protein identifications ("typing") by database searches, sequence data do not need to be perfect, but multiple stretches increase the level of confidence. Identifications thus obtained can be further substantiated by "mass-matching" of several more peptides with predicted proteolytic fragments (Sollner et al., 1993; Sogaard et al., 1994). Theoretically, accurate masses of several peptides from a single protein could be sufficient to search entire databases for such matches. Several algorithms have been written (Henzel et al., 1993; Mann et al., 1993; Pappin et al., 1993; Yates et al., 1993) but require unfailing high accuracy of the mass analyses and the combination of many peptides, a feat that is easier accomplished with standard proteins than with minute amounts of an "unknown." Instead, to allow a wider error margin in the mass searches, short stretches of experimental sequence can be added as a restriction (Mann et al., 1993). Also, an obvious shortcoming of mass-aided protein "typing" are interspecies sequence differences. It seems therefore, that Edman chemical sequencing and mass spectrometry will continue to be paired for state-of-the-art sequencing applications. Because mass analysis is about 2 orders of magnitude more sensitive than Edman sequencing (Hunt et al., 1992; Chait et al., 1993; Wong et al., 1993), the latter is the limiting factor in combined approaches to peptide analysis. More sensitive peptide sequencing tools and more efficient peptide preparation techniques are therefore needed.

Previously, we have described small-scale Aebersold-type in situ digests (Tempst et al., 1990), microbore (1.0-mm ID) RP-HPLC peptide purification (Elicone et al., 1994), improvements of commercial sequencers for low picomole level analysis (Tempst & Riviere, 1989; Erdjument-Bromage et al., 1993), and combined chemical sequencing/MALDI-MS strategies for peptide analysis (Sollner et al., 1993; Geromanos et al., 1994). Here, we describe simple improvements of existing instruments and protocols, making sample preparation more efficient and sequence analysis more sensitive and accurate than hitherto reported. We have applied the newly developed tools as part of a comprehensive multi-analytical/preparative approach to extensive sequence analysis, at the low picomole level, of a 250-kDa mammalian protein, the target of the immunosuppressant drug rapamycin.

Results and discussion

Improvements of sequencer performance for low picomole peptide analysis

Commercial automated sequencers can be optimized and improved to operate at levels of performance beyond those stated by the manufacturer. The specific situation in our laboratory is such that we have made low picomole sequence analysis of peptides (<30 amino acids) the top priority. Sensitivity in sequencing is determined by, among other factors, high-sensitivity chromatography of PTH-amino acids. Signals of as little as 100 fmol can be observed in real time (i.e., on stripchart recorders), provided the UV detector (OD_{269}) is set continuously at 0.001 absorption units full scale; we will henceforth refer to this setting as 1 mAUFS. When trying to do this on an Applied Biosystems 477A-120A system, 2 major problems become apparent: considerable fluctuations of the baseline and major chemical background peaks. Although it is relatively easy to cope with the first problem (Tempst & Riviere, 1989; Applied Biosystems, 1993a; Atherton et al., 1993; Erdjument-Bromage et al., 1993), background reduction requires a more elaborate set of adjustments.

The approach to minimize chemical background must be 3-fold: (1) eliminate impurities in chemicals and solvents; (2) reduce the amounts of chemicals and solvents that pass through cartridge and flask; and (3) increase efficiency of byproduct extractions. The difficult part is not so much devising hardware or cycle changes to this end, but to do so without jeopardizing initial and repetitive yields. Two years ago, a report by Totty et al. (1992) on accelerated high-sensitivity sequencing appeared in the literature. The authors advocated the use of a miniaturized cartridge (holding 9-mm discs) in which the sample is concentrated on a smaller surface area; as a result, consumption of reagents is considerably reduced. Soon thereafter, the small cartridge became commercially available (Applied Biosystems, 1993b). However, the cycles recommended by the manufacturer were significantly different from those published. In the current study, we have analyzed various critical steps of "microcartridge" peptide sequencing in a 477A instrument, with the focus exclusively on improved sensitivity and not on increased speed.

Coupling and cleavage

Although the cartridge volume has now been cut in half (30 μL to 15 μL), Edman reagents are still delivered in vast excess over the 5 pmol (or less) of peptide on the disc. We evaluated the effects of reduced coupling and lower PITC concentrations on IY, RY, lag, and background, all done with 10 pmol of a test peptide (Table 1). Previously, we determined the opti-

Table 1. *Model 477A (Applied Biosystems) background reduction and effects on sequencing performance*[a]

Rxn. crt. ID (m)	Temp. (°C)	Coupling PITC (%)	Coupling Time (s)	Cleavage (TFA) Deliv. (s)	Cleavage (TFA) Pause (s)	ATZ extract	Ala (@5 pmol)	RY (%)	Lag (@10%)	DPU (W units) (pmol)	Background (@D, S, I arbitrary)	No. expt. # (n)
12	48	5	3 × 400	mtr	350	cont	4.4	92.0	13	5.4	4+	3
12	48	1	3 × 400	mtr	350	cont	4.5	88.5	13	3.3	2+	3
12	48	0.5	3 × 400	mtr	350	cont	4.5	88.0	13	2.0	1+	3
9	48	5	3 × 400	1	350	cont	4.2	89.3	17	4.3	3+	2
9	48	5	3 × 400	2	350	cont	0	NA	NA	NA	NA	1
9	48	5	3 × 400	1	350	pulse	3.8	91.6	16	4.3	2+	1
9	48	5	2 × 400	1	350	pulse	2.8	89.2	29	1.4	1+	1
9	52	5	2 × 400	1	350	pulse	4.1	89.6	15	4.8	3+	1
9	48	5	1 × 400	1	350	pulse	2.8	85.5	47	0.9	–	1
9	48	5	3 × 400	1	250	pulse	4.2	91.5	16	4.8	3+	1
9	48	5	3 × 400	1	175	pulse	2.8	87.2	28	1.8	1+	1
9	48	1	3 × 400	1	250	pulse	3.9	89.5	21	1.6	1+	2
9	48	1	3 × 400	1	290	pulse	4.0	91.5	16	1.6	1+	5
9	48	0.5	3 × 400	1	350	pulse	2.6	88.4	20	1.0	1+	1

[a] During all experiments, 1.5 mg polybrene (containing 0.1 mg NaCl) was applied onto the 9-mm glass-fiber filter disc (double amounts on the 12-mm discs), followed by 2 filter cycles. TMA vapor and PITC/heptane were delivered with regulator settings at, respectively, 0.5 psi and 1.5 psi; the cartridge inlet line has an inner diameter of 0.3 mm and is 31 cm long. Line purges (see text) were always done when using the 9-mm cartridge, but not with the 12-mm one. Test peptide was 10 pmol PEPEPII: ISCWAQIGKEPITFEHINYERVSDR. No extended cleavage times (i.e., "Pro-cycles") were used at cycle 11. Background corrected yields of Ala (in cycle 5) are given; RY values are averaged from background-corrected values of Ile (cycles 7, 12, 17) and Glu (cycles 10, 15, 20). Cumulative lag (at cycle 10) is the % ratio of Lys 10/Lys 9 (background-corrected values). Size of DPU peaks, averaged over cycles 6–25, are expressed in "pmol Trp" units. Background peaks eluting in the positions of PTH-Asp, -Ser, or -Ile have been scored on an arbitrary scale (highest values represent most background). Other abbreviations: Rxn. crt. ID, reaction cartridge inner diameter; Deliv., delivery; No. expt., number of independent experiments; mtr, metered TFA delivery; cont, continuous solvent flow.

mal amount of polybrene applied to the glass-fiber disc to be about 25 $\mu g/mm^2$, because a further, 2-fold reduction caused unacceptable washout (Erdjument-Bromage et al., 1993). Therefore, we have used 1.5 mg polybrene on the 9-mm filter throughout. Cutting back on coupling time (3 to 2 to 1 × 400 s) resulted in lower background (DPU; background peaks around PTH-Asp, -Ser, and -Ile), but also in higher lag (Table 1). Double coupling at elevated temperature (52 °C) reduced lag but increased background again. The use of 1% (or even 0.5%) PITC as coupling reagent, in conjunction with a standard cartridge, caused the background to go down, with a moderate reduction in repetitive yields. In the case of the small cartridge, however, RY was not affected at all when using 1% PITC (Table 1).

As originally pointed out by Totty et al. (1992), reduced cleavage times may improve recoveries of labile amino acid derivatives, the reason being that once ATZ-amino acids have been formed, they should not sit around in TFA needlessly and deteriorate. When coupling with 5% PITC, cleavage could be shortened to 250 s without adverse affects on repetitive yields; 290-s cleavage (at 48 °C) proved to be optimal in combination with reduced PITC (1%) coupling (Table 1). Under those conditions, we did observe slight increases in recoveries of PTH-Ser, -Thr, -Trp, and -peCys. So far, we have failed to get "pressurized cleavage" (150 s at 50 °C under argon pressure [Totty et al., 1992]) to work. The problem could be related to instrument-specific features.

Byproduct and ATZ-amino acid extractions

A problem with small discs, as originally identified by Totty et al. (1992), is channeling of extraction solvents through the center of the disc. The remedy suggested by the authors is to interrupt deliveries with three 5-s pulses of argon (termed "pulsed extraction"). Although we have found this to be indeed useful, extraction efficiencies can be further improved by inserting 10-s pauses before the argon pulses, allowing better diffusion toward the perimeter, and by additionally washing with butyl chloride (S3) between the heptane and ethyl acetate. As shown in Table 1, pulsed byproduct/ATZ extraction results in a better repetitive yield and overall reduced background; levels of DPU, however, remained the same. The modified ATZ extraction (see reaction cycle SMA-1 in Table 2) required adjustments of drying times in the flask before and after aqueous TFA conversion (the conversion cycle is taken from Tempst and Riviere [1989] and is not shown here).

Reagents and solvents

Reagents should not be allowed to age for more than a month and should not be kept on the instrument for longer than a week. We do not add DTT to solvents S1, 2, and 3; reduction in RY does not outweigh the chromatographic disturbances associated with the presence of reducing agent. Aging of chemicals is not a serious problem on busy instruments (>120 cycles per week). However, mostly overlooked is the fact that, between deliveries, slugs of reagent are left sitting in the teflon lines connecting bottles with valve blocks. Because these lines (several feet in length; dead volumes ranging from 117 to 187 μL) are pervious to air, and because some chemicals (e.g., PITC and TFA) can be trapped there for the duration of a full cycle (40 min), the slug of liquid should be purged to waste prior to delivery of "fresh" reagent from the reservoir. This is particularly impor-

Table 2. *Model 477A (Applied Biosystems) modified reaction cycle*[a]

Step	Function	Time (s)	Step	Function	Time (s)
1	Prep R2	6	36	Pause	10
2	Load R2	15	37	Argon dry	5
3	Block flush	10	38	Load S2	7
4	Prep R2	6	39	Deliver S2	20
5	Deliver R2	20	40	Pause	10
6	Load R1	4	41	Argon dry	5
7	Block flush	10	42	Deliver S2	20
8	Prep R1	6	43	Pause	10
9	Deliver R1	1	44	Argon dry	5
10	Argon dry	30	45	Deliver S2	20
11	Deliver R2	400	46	Pause	10
12	Block flush	5	47	Argon dry	75
13	Load R1	4	48	Prep R3	10
14	Block flush	10	49	Load R3	7
15	Prep R1	6	50	Block flush	10
16	Deliver R1	1	51	Deliver R3	1
17	Argon dry	30	52	Argon dry	3
18	Deliver R2	400	53	Pause	290
19	Block flush	5	54	Load S2	6
20	Load R1	4	55	Block flush	10
21	Block flush	10	56	Argon dry	90
22	Prep R1	6	57	Prep transfer	10
23	Deliver R1	1	58	Deliver S1	7
24	Argon dry	30	59	Pause	5
25	Deliver R2	400	60	Transfer w/ S3	15
26	Argon dry	90	61	Pause	20
27	Load S1	6	62	Transfer w/ argon	5
28	Deliver S1	10	63	Transfer w/ S3	12
29	Pause	10	64	Pause	15
30	Argon dry	5	65	Transfer w/ argon	30
31	Deliver S1	10	66	End transfer	1
32	Pause	10	67	Deliver S3	15
33	Argon dry	5	68	Argon dry	5
34	Load S3	5	69	Deliver S3	15
35	Deliver S3	15	70	Argon dry	90

[a] S1 = heptane; S2 = ethyl acetate; S3 = butyl chloride; R1 = 1% PITC in heptane; R2 = 12.5% TMA in water; R3 = TFA; temperature of reaction = 48 °C. Cycle name = SMA-1; cycle length = 70 steps; runtime = 40 min, 57 s; cartridge holds 9-mm disc.

tant for PITC and TFA; with the 9-mm cartridge, only small volumes of these reagents are consumed, and all of it would be "line-volume." Purging (or "line-flushing") is done by loading and flushing reagents in and out of the valve blocks to waste (see reaction cycle SMA-1 in Table 2). The combined effect of small cartridge and line purges is illustrated in Table 1; with 1% PITC coupling, RY is up by 3% and background reduced by more than 50%. In the many sequencing applications done in our laboratory since the changes were implemented, background has consistently been below this level.

Taken together then, although we have not investigated the effects of every variable in automated chemical sequencing, the proposed optimizations did culminate in major background reductions without adverse effects on IY (40% at cycle 5 with a 5–10% error margin) and RY (91.5 ± 2%). It should be mentioned that, although chemical background phenomena are of a general nature, yields are heavily peptide dependent and quite often better (see example below), and sometimes worse, than those listed for the test peptide in Table 1.

Filter conditioning

Discs with fresh polybrene must be precycled, at least 2 filter cycles (SMAFIL) and then SMA-1 cycles (usually 2–6), until the background becomes acceptable. SMAFIL cycles (not shown) are derived from SMA-1 cycles (Table 2) in the same way as Applied Biosystems' FIL cycles are from NORMAL cycles, except that we do not use line-flushing and S3 wash steps are omitted after coupling but extended after the cleavage step to 100 s (in 4 deliveries of 25 s, interrupted 3 times by a 10-s pause and 5-s argon dry). Chemical background will decrease with cycle number until it stagnates at low levels, usually after 10–20 cycles. Because peptides are always sequenced all the way to the C-terminus, or until no more signals are observed, a second sample can be analyzed without reapplying polybrene. This second run is more ideally suited for the most precious (=least material) samples; a little planning is thus in order.

An example of such an experiment is presented in Figure 1. "Raw" chromatograms (i.e., strip chart recordings at 0.5 mAUFS) of 5 cycles from a sequencing experiment (SMA-1 cycles; 1% PITC) on a peptide, obtained from an in situ digest and fractionation on a 1.0-mm-ID RP-HPLC column, are shown; yields were below 1 pmol throughout (23 cycles). As illustrated, signals of 500 fmol or less are quite easily detected. The amounts of Val and Leu, as listed, are corrected for background; total amounts are therefore somewhat higher, as can be easily discerned by visual comparison with the standard. Even with backgrounds of about 200 fmol for several amino acids, identifications are generally unambiguous. As expected for a sequencing run on "used" polybrene (25 previous cycles), the DPU peak was on the order of 500–800 fmol only. The RY for this experiment, calculated from background-corrected values of Val at cycles 7 and 17, was 94.9%; the precise amount loaded for analysis was unknown but estimated to be about 2 pmol (from LC peak height).

Efficient digest/recovery of polypeptides from NC

Since its introduction (Aebersold et al., 1987), in situ digestion of proteins attached to membranes has become a standard tool of protein chemists. Surprisingly, few significant modifications of the original recipe have been suggested, except for the potential use of different enzymes (Tempst et al., 1990; Fernandez et al., 1994) and the inclusion of detergent to promote elution of fragments from PVDF membranes (Fernandez et al., 1992). We have now examined preferred usage of detergents for digests of proteins off NC. A reported undesirable effect of RTX-100, the release of PVP-40 from the PVDF membrane with concomitant disturbance of LC separations, has been remedied, simply, by omitting the blocking step (Fernandez et al., 1994). This facile procedure seems to work equally well with NC digests. However, we have observed a sharp reduction in recoveries when submicrogram quantities of protein are adsorbed onto larger surfaces (>50 mm^2) of "nonblocked" NC (data not shown). Unfortunately, such nonideal conditions are quite common in the applications laboratory, and, in those cases, the PVP-40 step is still required.

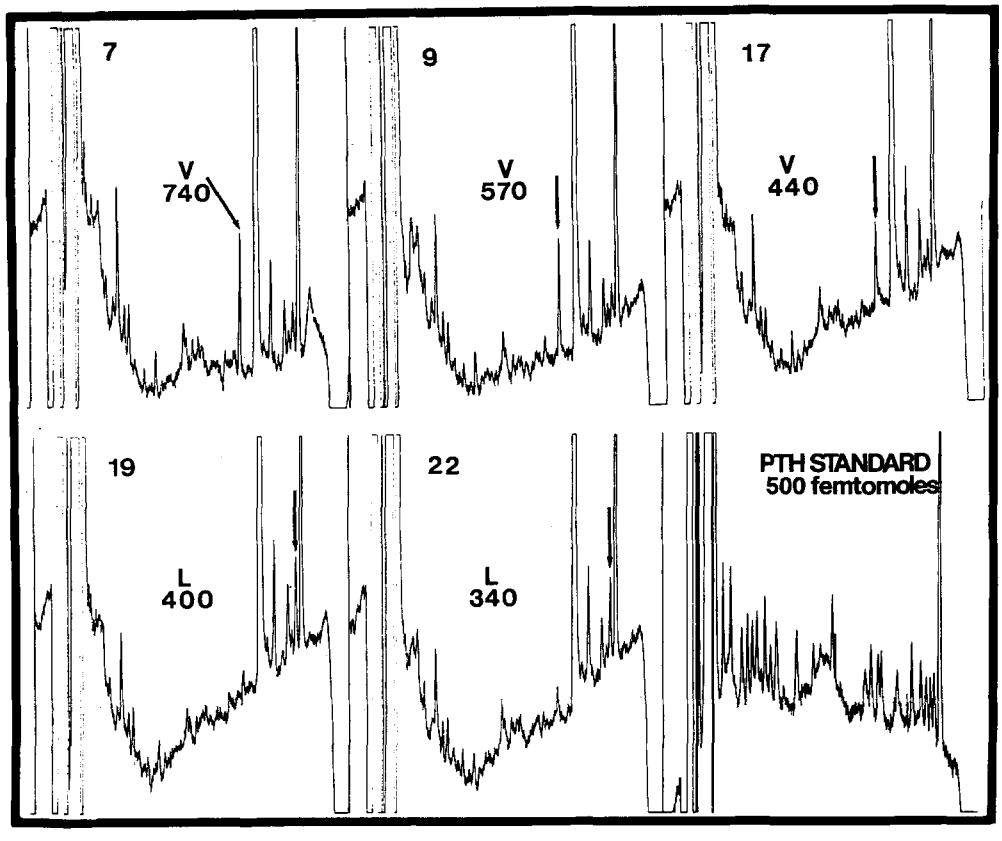

Fig. 1. Chemical peptide sequence analysis. Amino acid sequence analysis of a tryptic peptide, isolated by microbore RP-HPLC from an in situ digest of electrophoretically separated (and blotted) protein *JAES1*. Sequencing conditions were as discussed under Results and discussion and listed under Materials and methods. Starting amount of peptide was estimated to be 2 pmol (from LC peak height). Chromatograms 7, 9, 17, 19, and 22 are shown; full scale corresponds to 0.0005 AUFS. PTH-amino acid peaks are indicated with an arrow and the femtomolar quantities shown. A chromatographic standard (500 fmol of each amino acid in the sequencer flask) is shown at the bottom right; scale of the standard is 0.0005 AUFS at 269 nm.

Detergents and incubation times

An initial survey of detergents led to the identification of TW-80 (0.3%, by volume, in 10% acetonitrile) as the preferred additive to NC digests. The reason is 3-fold: (1) recoveries of tryptic peptides (determined by measuring release of radiolabeled lysozyme, hemoglobin, ovalbumin, G6PD, and β-galactosidase fragments) were increased from 40% to, on the average, 70% (Fig. 2); (2) TW-80 is UV transparent; and (3) unlike RTX-100 and TW-20, TW-80 does not strip PVP-40 off the NC, at least not within a 3-h incubation period. As shown in Figure 2, there is no need to let the digest proceed for 15 h; 1–3 h is sufficient.

Different proteases

We tested several proteases in the presence and absence of detergent and for various incubation times. As predicted from the cleavage specificities, more polypeptide material was released with chymotrypsin and less with endoproteinase Lys-C as compared to the tryptic digests (Fig. 2). For lysine-rich proteins (e.g., hemoglobin), tryptic and Lys-C digests were comparable. We expected, however, that for larger M_r proteins (>100 kDa), Lys-C digests would result in less crowded chromatograms with preferentially bigger size peptides. Digestion (3 h, with detergent) of β-galactosidase (1,021 amino acids; 30 Lys, 47 Arg) yielded recoveries of 31% and 67% from Lys-C and tryptic digestions, respectively. Although the Lys-C digest chromatogram was indeed less convoluted, the hoped for longer fragments were conspicuously absent (data not shown); they probably did not elute.

Ultraviolet spectroscopic detection of tryptophan and peCys-containing peptides

Previously, we reported on the use of diode-array detection to identify Trp-, Tyr-, and peCys-containing peptides by real-time, 4-wavelength UV absorbance monitoring (OD 214, 253, 277, 297) of column eluates (Tempst et al., 1990). Tryptophan has a unique codon and its presence in a peptide sequence facilitates design of oligonucleotide primers with low sequence degeneracy. In addition, Trp and Cys residues are generally well-conserved among related proteins and hence pivotal in delineating possible interspecies and/or protein family relationships when searching sequence databases. However, Trp is relatively difficult to identify when peptide sequencing analysis must be done with subpicomole level signals.

Peptides selectively containing either Trp (and no Tyr or peCys), or Tyr (no W, pC) or peCys (no W, Y) are easily identified as such (Table 3). In cases where two or all three of these

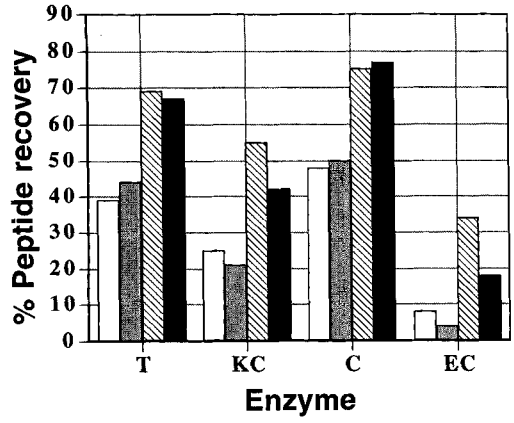

Fig. 2. Recovery of peptides by in situ enzymatic digestion of NC-bound proteins. One microgram amounts each (augmented with 1–5 ng/100,000 cpm of the [^{125}I]-radiolabeled form) of lysozyme (14 kDa), hemoglobin (16 kDa), ovalbumin (45 kDa), G6PD (55 kDa), and β-galactosidase (116 kDa) were electrophoresed (12.5% T polyacrylamide gel), electroblotted onto NC, stained with Ponceau S, and counted (γ-counter). In situ enzymatic digests were carried out for 1 h (open and striped bars) or 15 h (gray and black bars), in the presence (striped and black bars) or absence (open and gray bars) of TW-80, and the supernatant (combined from digest buffer and wash solution) was counted; for further details, see Materials and methods. Recoveries, averaged from the 5 different proteins (3 independent experiments each), are shown as ratios of counts released versus starting amounts (on the NC). Enzymes (1 μg amounts): T, trypsin; KC, endoproteinase Lys-C; C, chymotrypsin; EC, endoproteinase Glu-C.

Table 3. *RP-HPLC with multiple wavelength detection of peptides containing Trp (W), Tyr (Y), or peCys (pC)*[a]

Peptide	Relative peak height (%)					
	A_{253}	A_{277}	A_{297}	W	Y	pC
pCPSPKTPVNFNNFQ	100	12	2	–	–	1
QNpCDQFEK	100	14	1	–	–	1
GNLWATGHF	45	100	28	1	–	–
ILLQKWE	43	100	26	1	–	–
YEVKMDAEF	33	100	3	–	1	–
TGQAPGFTYTDANK	38	100	2	–	1	–
YSLEPSSPSHWGQLPTP	45	100	21	1	1	–
GITWKEETLMEYLENPK	42	100	24	1	1	–
EDWKKYEKYR	40	100	23	1	2	–
YEDWKKYEKYR	37	100	19	1	3	–
Insulin β chain/4VP	100	39	4	–	2	2
Insulin α chain/4VP	100	32	3	–	2	4
DST peptide (25 amino acids)	100	73	23	1	–	1
PepepII (25 amino acids)	100	100	20	1	1	1

[a] Peptides (20 pmol each, or less) were chromatographed on a Vydac C4 (2.1 × 250 mm) column at a flow of 0.1 mL/min; the HPLC system and conditions are described under the Materials and methods. Peak heights on chromatograms, produced by monitoring at different wavelengths, are expressed in %, relative to the tallest peak. Total number of W, Y, or pC present in each peptide are listed. Sequences of bovine insulin α and β chains are taken from the SWISS and PIR databases; PepepII: ISpCWAQIGKEPITFEHINYERVSDR; DST peptide: DLF NAAFVSpCWSELNEDQQDELIR. Insulin was reduced with β-mercaptoethanol and reacted with 4-vinylpyridine (4VP) prior to HPLC.

aromatic residues are present in a single peptide (in various ratios), and due to their overlapping UV spectra, precise identification has proven more difficult. We have now analyzed a number of such possible combinations and calculated ratios of peak heights at the different wavelengths. The results, as presented in Table 3, can be used as empirical guidelines when inspecting multiple-wavelength stripchart recordings. For peptide chromatography at the 5–10-pmol level, this can be done with sufficient accuracy by measuring peak heights with a ruler. Note that the data as listed in Table 3 are valid only when using a chromatographic mobile phase with acidic pH (i.e., 0.1% TFA); UV spectra of Tyr and Trp shift dramatically with pH (Fasman, 1989). Although the proposed method may not allow one to distinguish between a peptide containing Trp only and one containing Trp and 1 or more Tyr residues, it does provide unequivocal evidence for the presence of either Trp or peCys or a combination thereof. Inability to predict the presence of Tyr is not a major drawback because this residue is easily identified by chemical sequencing.

High-sensitivity sequence analysis of large M_r proteins

The simple improvements of existing instrumentation and protocols described in this report have enabled us to realize sequencing objectives leading to discovery and/or functional understanding of several low-abundant proteins (Andrews et al., 1993; Sollner et al., 1993; Polyak et al., 1994). Sequencing studies become progressively more complicated and tedious with increasing molecular weight of the proteins, largely as the result of complex HPLC peptide profiles, and the need for repurification with associated reduction in yields. Although large proteins, such as the ryanodine receptor and inositol triphosphate receptor, have been (partially) sequenced in the past (Marks et al., 1990a, 1990b), these studies were done with well over 100 pmol of starting material, rather substantial amounts by current "high-sensitivity" standards.

Mammalian target protein of the FKBP12-rapamycin complex

Rapamycin, originally identified as an antifungal agent, is currently of major interest for its potency in preventing graft rejection, after organ transplants, for example (reviewed by Sehgal, 1993). Several immunosuppressants, including rapamycin, block T-cell response to antigen. They do so by binding to small soluble proteins called immunophilins. By itself, this interaction does not explain the immunosuppressant effects. Instead, the drug (rapamycin)-immunophilin (FKBP12) complex binds to another target protein (Kunz & Hall, 1993). In yeast, rapamycin causes cell cycle arrest also by binding to an FKBP12 homologue. Through genetic selection, 2 yeast genes have been identified that confer rapamycin resistance when mutated. It has been suggested, but not shown biochemically, that the products of these genes are the targets of the FKBP-rapamycin complex (Kunz et al., 1993). Previously, we have identified, through crosslinking and affinity purification procedures, 2 rat brain proteins (250 and 35 kDa) that interact with FKBP12 only in the presence of rapamycin (Sabatini et al., 1994). An estimated 10 pmol (2.5 μg) was obtained of the larger molecule (p250) after gel electrophoresis and electroblotting. Here, we describe our

approach to extensive sequence analysis of this biologically important protein.

Peptide purification and spectrometric analysis

Affinity-purified p250 was fractionated by SDS-PAGE, transferred to NC, and digested with trypsin using the modified protocol as described above (and under Materials and methods). The resulting peptides were separated by narrow-bore (2.1-mm ID; Vydac 214TP52 C4) reversed-phase LC; more than 90 peak fractions were collected (Fig. 3; top panel). Chromatographic data, produced with a diode array detector, were inspected for clues as to the presence of Trp and peCys residues in the peptides. From the complexity of the OD_{214} pattern and the anticipated presence of trypsin autolytic fragments, we assumed that most peaks would have to be repurified by a second round of (microbore) HPLC to yield peptides sufficiently homogeneous for sequence analysis. Even so, MALDI-MS was used to assess peak purity and a few peptides were selected for direct analysis at this stage (e.g., T76; Figs. 3, 4). Others, exhibiting multiple *m/z* peak values, were rechromatographed on an SGE ODS-2 C18 microbore (1.0-mm ID) column of different selectivity (Elicone et al., 1994), as shown for T63 and T83 in Figure 3 (middle panels). All repurified fractions were analyzed (5% of the sample) by a second round of MALDI-MS (Fig. 3, bottom panels) in anticipation of chemical sequencing. From peak height comparison to external calibrants (14 peptides ranging in size from 9 to 26 amino acids; see Elicone et al., 1994) separated on 2.1-mm and 1.0-mm columns, we estimated that recoveries during rechromatography were on the order of 50–70%. We have

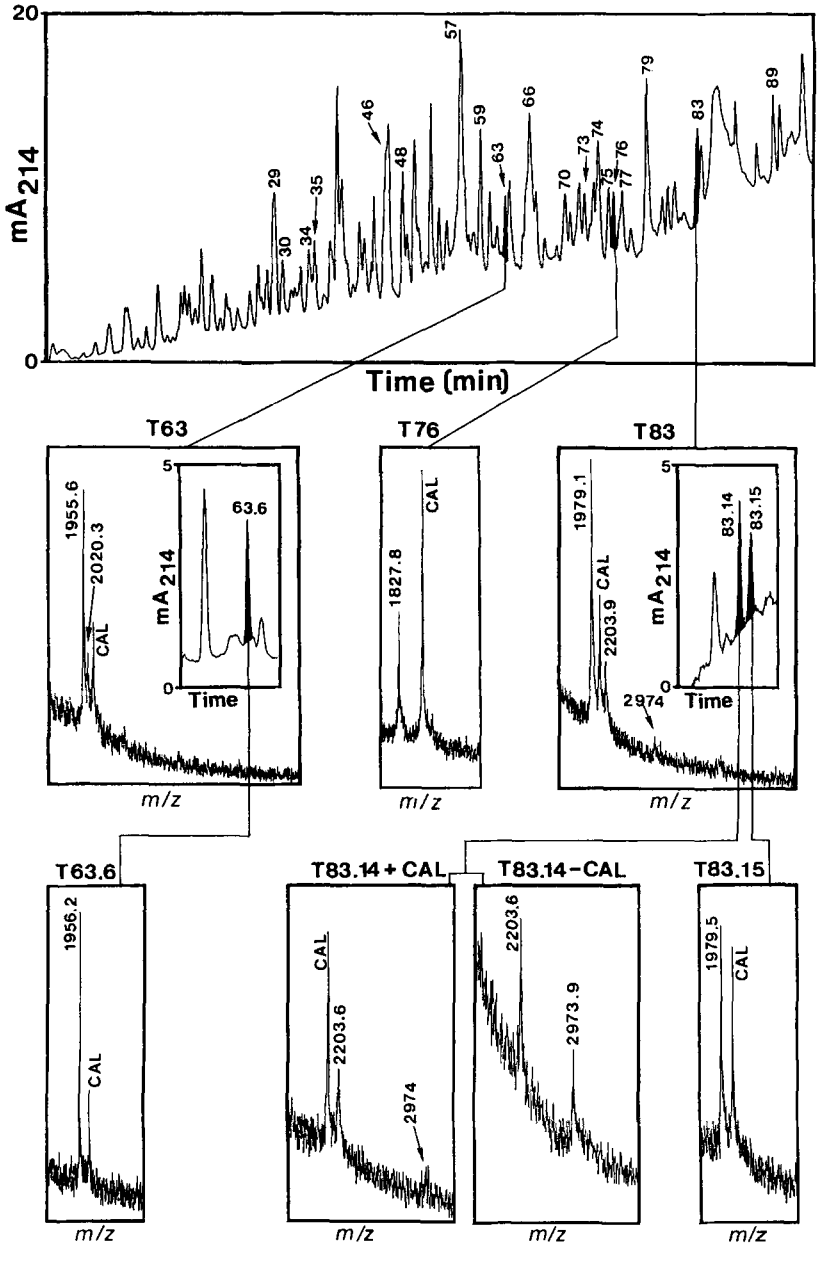

Fig. 3. Peptide purification/mass analysis strategy. The 250-kDa protein "Rapamycin-and-FKBP-Target" was digested in situ (on NC) with trypsin, and peptide fragments were separated by "2.1-mm-bore" RP-HPLC (C4; details under Materials and methods). The resulting chromatogram is shown in the top panel, with a time scale from 18 to 68 min. Labeled peaks contained peptides that were eventually used for successful sequencing experiments (see alignments in Fig. 5). Aliquots (2–4%) of peak fractions were analyzed by MALDI-MS (for details, see Materials and methods) as illustrated for peaks T63, T76, and T83 (middle panels); *m/z* values are listed for all peaks, except for the calibrant peptide (designated CAL; *m/z* = 2,109.45). Fractions exhibiting more than 1 *m/z* peak value were rechromatographed on a "1.0-mm-bore" RP-HPLC column (C18; see Materials and methods) as shown in the insets of panels "T63" and "T83"; only the relevant sections of the secondary chromatograms are given (40–45 min and 54–59 min for T63 and T83, respectively). Repurified peptide peaks (4% aliquots) were then subjected to a new round of mass analyses (bottom panels). Mass spectra in the presence (+CAL) or absence (−CAL) of the calibrant peptide are shown for peak T83.14. Detailed sequence analysis data for peptides T83.14a, T83.14b, T63.6, and T76 are given in Figure 4.

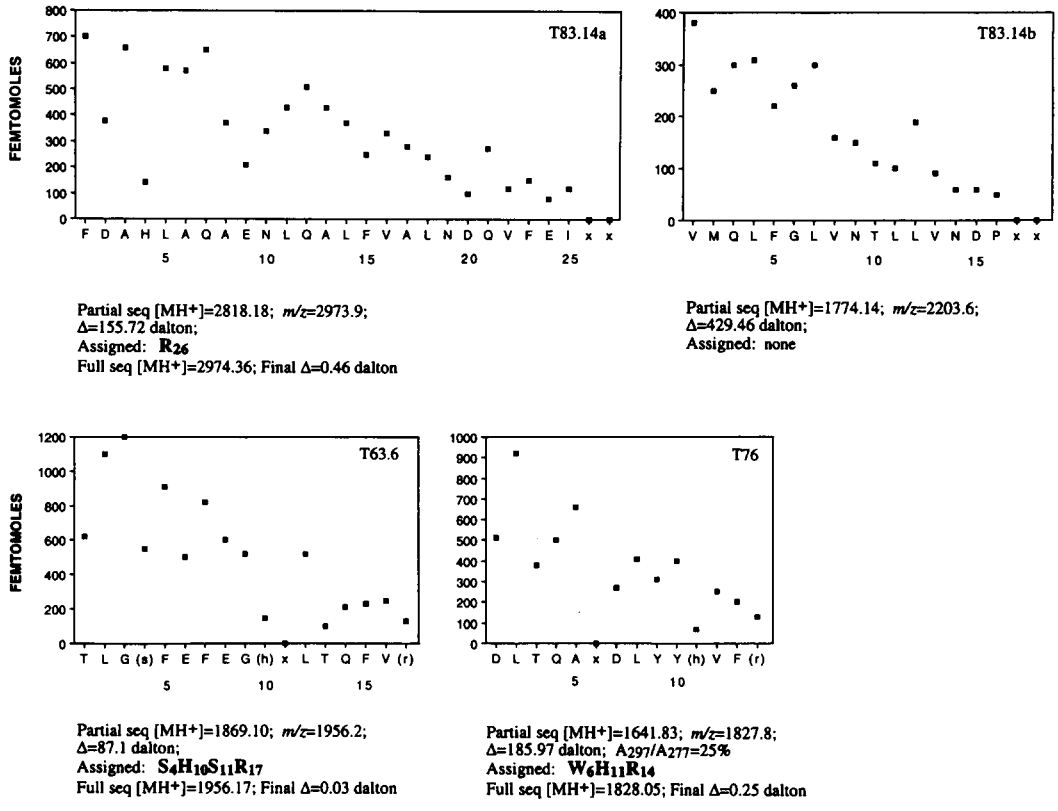

Fig. 4. Peptide structural analysis by combined chemical sequencing/laser-desorption mass spectrometry. Selected peak fractions from the chromatographic separations shown in Figure 3 were analyzed by chemical sequencing and MALDI-MS (see also Fig. 3, middle and bottom panels). Symbols (used in sequencing): amino acids are given in the single letter annotation; those printed in lower case (and in parentheses) were assigned with a lower level of confidence; x = no amino acid assigned. Recoveries (in fmol) of the PTH-amino acids are plotted versus cycle numbers. Partial seq [MH^+] denotes the mass values calculated by summing the average isotopic masses of all amino acids that were identified (positively and tentatively), to which the mass of a single proton [H^+] was added. *m/z* is the experimental mass (by MALDI-MS) of the peptide. A_{297}/A_{277} gives the absorbance ratio as determined by diode-array UV-detection; values >20% are indicative for the presence of tryptophan (see Table 3). Δ indicates the difference between experimental mass and calculated mass of the partial sequence. "Assigned" lists the amino acids (and position in the sequence) assigned to the gaps, and low-confidence calls, as to yield the best possible match between calculated and experimental masses, with the restriction that the C-terminal residue must be either Lys or Arg to satisfy the trypsin specificity requirements. Full seq [MH^+] lists the theoretical average isotopic mass of the final peptide sequence (plus 1 proton); "final Δ" denotes the difference between this value and the above listed *m/z*.

previously shown that losses can be traced primarily to collection, storage, and handling, rather than to the secondary column (Elicone et al., 1994). Large and/or hydrophobic peptides are affected the most in this regard, but such losses can be largely avoided by acidification (Erdjument-Bromage et al., 1993). In the project discussed here, fractions were always acidified (10% TFA final concentration) before rechromatography and recoveries were about the same for all peptides. However, we do not have sufficient data at this time to generalize these findings, and different results could very well be obtained in a more systematic study.

Peptide sequence analysis/assignment

In view of the limited amount of starting material and the need for 2D RP-HPLC, care was taken to optimize the sequencing instruments before each analysis, using previously defined pass/fail criteria (Erdjument-Bromage et al., 1993). As shown in Figure 4, several sequencing runs, using SMA-1 reaction cycles (see text and Materials and methods), were carried out with initial yields well below 1 pmol. Gaps in the sequences occurred and had to be filled by careful matching of the data with those obtained from mass spectrometry and UV spectroscopy. Still, only 23 of 40 sequencing runs eventually gave conclusive results. Examples of peptide sequencing at the femtomole level, using combined analytical techniques, are given in Figure 4 (with corresponding mass spectra shown in Fig. 3).

As expected from the mass analyses, peak T83.14 yielded 2 sequences that could be readily deconvoluted by differences in signal size. The longest peptide (T83.14a) was sequenced for 25 cycles with 700 fmol IY and a secondary sequence (T83.14b; 380 fmol IY) could be reliably called for 16 cycles (Fig. 4). The mass summed from the identified residues in the longest peptide was 155.72 Da less than 1 of the 2 experimentally obtained masses (*m/z*). The difference about equals the average molecular mass of Arg (156.19 Da), 1 of the 2 obligatory residues at the C-terminus of a tryptic peptide. The sequence could thus be considered complete and the deconvolution accurate. An interesting observation with regard to peptide mixture T83.14a (major)

plus T83.14b (minor) is the poor ionization/desorption of the major component during MALDI-MS. In fact, adequate peak signal could only be obtained in the absence of calibrant (see Fig. 3). The calibrant served to determine the accurate mass of T83.14b, which was then used, in turn, to calibrate a secondary mass analysis experiment (T83.14-CAL in Fig. 3, bottom panels).

In the case of peptide T76 (Fig. 4), 14 cycles of interpretable sequence were obtained but with no identification ("gap") of residue 6 and only tentative assignments of His (70-fmol yield) at position 11 and Arg (120 fmol) at 14. Assuming authenticity of H_{11}/R_{14}, the summed mass was 185.97 Da less than the experimental *m/z* value. Because Trp has an average isotopic mass of 186.21 Da, and because the A_{297}/A_{277} absorbance ratio also indicated the presence of Trp, this residue could be positively assigned to position 6. In the same way, Ser was assigned to position 11 in peptide T63.6, as the summed average isotopic masses of positively and tentatively (S_4, 550 fmol; H_{10}, 150 fmol; R_{17}, 130 fmol) assigned amino acids were 87.10 Da less than the recorded *m/z* value, or 0.02 Da different from the average mass of Ser (87.08 Da).

Alignment of mammalian p250-derived peptides with yeast TOR

Although partial homologies with dozens of different proteins were obtained, none of the 23 peptide sequences could be perfectly matched to any entries in the sequence databases (PIR, SwissProt, translated GenBank). However, a particular protein turned up repeatedly in lower stringency searches, namely yeast TOR2 (Kunz et al., 1993). Eighteen of the 23 peptides of rat p250 could be aligned to the yeast sequence with various degrees of similarity (Fig. 5); 5 peptide sequences of the rat protein proved to be unique (sequences listed in legend to Fig. 5).

The C-terminal halves of the homologous yeast proteins, TOR1 and TOR2, are highly conserved and thought to contain a lipid kinase domain (Cafferkey et al., 1993; Helliwell et al., 1994). Thus, as one might expect, the rat p250 tryptic peptides (T83.5, T76, T57.4, T83.14b, and T35a) that align with se-

Fig. 5. Alignment of rat p250-derived peptides with the predicted amino acid sequence of yeast TOR2. The peptide sequences obtained from microsequencing were aligned with TOR2 amino acid sequence using the NCBI BLAST program (Altschul et al., 1990) with manual adjustments. The interrupted TOR2 sequence is listed on the top line and the corresponding p250 peptides are listed below. A single line indicates an identical amino acid and a double dot a chemically conserved substitution. The sequences of 5 peptides could not be readily aligned to TOR2. Their sequences are: EMQKPQWR, MLGHLVSNAP, GLAHQLASPGLTTLP, AYGAMVXXSTLKE, VLTLWTVPGVQDVLK.

quences in the C-terminal half of TOR2 show the highest degree of conservation to the yeast protein. For example, peptide T76 has an 11-amino acid stretch, including a Trp, of absolute identity to the corresponding yeast sequence. However, the other tryptic peptides match the TOR2 sequence in various well-distributed locations, indicating that the relatedness of mammalian and yeast target proteins of FKBP12–rapamycin is not restricted to a single functional (recognition and/or activity) domain, a fact that was later confirmed by cloning and sequencing of the rat cDNA (Sabatini et al., 1994). Isolation and characterization of this rather large cDNA clone (8.6 kb) was greatly facilitated by the availability of 23 stretches of mostly uninterrupted sequence.

Conclusions

Internal sequencing of gel-purified proteins is a multistep process, consisting of proteolysis, peptide purification, and sequence analysis. The procedure becomes increasingly more difficult with bigger proteins and smaller amounts of starting material. We have introduced a number of simple improvements to existing instrumentation and protocols, making peptide preparations more efficient and sequence analysis more sensitive and accurate than hitherto reported.

Specifically, we have described studies leading to (1) enhanced sensitivity of automated chemical peptide sequencing through background reduction but with undiminished yields; (2) improved recovery of peptides from rapid in situ digests of nanogram amount NC-bound proteins through TW-80-promoted elution, without disruptive effects on UV monitoring of LC column eluates; and (3) more accurate UV spectroscopic identification of Trp- and Cys-containing peptides. Used in combination with HPLC systems specifically configured for micropreparative peptide isolation (Elicone et al., 1994), the newly developed tools allowed extensive primary structural analysis (330 amino acids) of low picomole quantities of a gel-purified, 250-kDa protein, the target of rapamycin–FKBP12 in mammalian cells.

Two strategies of technology integration were critical for the successful outcome of that study. First, we used a combination of chemical sequencing, MALDI-MS, and UV spectroscopy for accurate covalent analysis of peptides. Second, the peptide repurification strategy was determined by exhaustive mass spectrometric screening of column fractions after the first round of HPLC. In this way, mixtures and artifacts could be weeded out, saving valuable sequencing time and expenditure while avoiding pointless and wasteful rechromatography of already homogeneous fractions. A word of caution, though: MALDI-MS is not a quantitative technique and unexplained (i.e., not as yet understood) suppression effects occurred when analyzing peptide mixtures. Therefore, in a small but not unimportant number of cases, apparently single mass to charge maxima were obtained from heterogeneous preparations, mostly leading to ill-fated "mixed" sequencing.

The results presented here demonstrate that a substantial number of long, uninterrupted stretches of sequence can be obtained from minute amounts of gel-purified, large-size proteins. Covalent structural information thus obtained has allowed, unequivocally, the matching of a biological function of a mammalian protein to a yeast gene product with a characterized mutant phenotype.

Materials and methods

Materials

Sequencer reagents, solvents, and polybrene were obtained from Applied Biosystems (Foster City, California), acetonitrile from Burdick & Jackson (Muskegon, Wisconsin), TFA from Pierce (Rockford, Illinois), 4-vinylpyridine from Aldrich (Milwaukee, Wisconsin), β-mercaptoethanol from BioRad (Richmond, California), PVP-40 from Sigma (St. Louis, Missouri), Ponceau S from Fluka (Ronkonkoma, New York), all detergents from Calbiochem (San Diego, California) and NC from Schleicher & Schuell (Keene, New Hampshire); all were of the highest possible quality/purity available.

Peptides listed in Table 3 were chemically synthesized, HPLC purified, and quantitated by amino acid composition analysis, following published procedures (Casteels et al., 1994). Reduction and pyridyl ethylation were done as described under "proteolysis" in the text below; derivatized peptides were HPLC purified. Lysozyme, hemoglobin, ovalbumin, G6PD, β-galactosidase, insulin, and chymotrypsin were purchased from Sigma. Stock solutions were made in water and quantitated by amino acid analysis, except for β-galactosidase (dissolved in 1% SDS and quantitated by weight). Trypsin and endoproteinase Glu-C were "sequencing grade" from Boehringer (Indianapolis, Indiana); *Achromobacter lyticus* endoproteinase Lys-C (lysyl endopeptidase) was obtained from Wako Chemicals (Richmond, Virginia). All other chemicals, supplies, and equipment were as indicated elsewhere under Materials and methods.

Purification of protein p250, mammalian target of rapamycin/FKBP12, has been described elsewhere (Sabatini et al., 1994). Briefly, rat brain extract was passed over a heparin-agarose (Sigma) column, concentrated, and incubated with GST-FKBP12 fusion protein in the presence of glutathione-agarose (Sigma) and rapamycin (gift from Wyeth-Ayerst, Philadelphia, Pennsylvania). Beads were extensively washed, and the remaining bound protein eluted in SDS-PAGE sample buffer.

Radioiodination

Standard proteins (50 μg/10 μL 1% SDS) were iodinated using Iodo-beads (Pierce) and $Na^{125}I$ (0.4 mCi; Amersham, Arlington Heights, Illinois) according to the manufacturer's instructions. Protein was separated from unincorporated label by 2 passages over prepacked PD-10 Sephadex G-25 columns (Pharmacia, Piscataway, New Jersey), or more, to yield a solution of labeled protein with over 99% TCA-precipitable counts. The final preparation was aliquoted and stored at −70 °C. Counting was always done using an LKB Wallac 1272 Clinigamma instrument (Pharmacia).

In situ proteolysis

Standard proteins or affinity-purified p250 were fractionated by SDS-PAGE, electroblotted onto an NC membrane, visualized by Ponceau S staining, excised from the blot, and subjected to in situ proteolysis (Aebersold et al., 1987; Tempst et al., 1990), with modifications. Typically, digests were carried out with 1 μg of enzyme in 25 μL 100 mM NH_4HCO_3/10% acetonitrile/0.3% TW-80, for 3 h at 37 °C. During optimization studies, TW-80 was omitted or replaced with TW-20 or RTX-100 in some experiments, and incubation times varied from 1 to 15 h.

After incubation, NC pieces were washed once with digest buffer (equal volume) and supernatants were pooled and counted (radiolabeled proteins). In the case of the p250 digest, the resulting peptide mixture was reduced and *S*-alkylated with, respectively, 0.1% β-mercaptoethanol and 0.3% 4-vinylpyridine and fractionated by 2D RP-HPLC. An enzyme blank was done on an equally sized strip of NC cut from a blank area of the same blot.

RP-HPLC

For the primary separations, a 2.1-mm Vydac 214TP52 C4 column (The Separations Group, Hesperia, California) was used with gradient elution at a flow rate of 100 μL/min. HPLC solvents and system configuration were exactly as described (Tempst et al., 1990), with improved dead volume reduction through the use of glass capillary tubing (C. Elicone & P. Tempst, unpubl.). Identification of Trp-containing peptides was done by manual ratio analysis of absorbances at 297 and 277 nm, monitored in real time with an Applied Biosystems (Foster City, California) model 1000S diode-array detector, and using the indicative ratios listed in Table 3. Detector analog signals (4 channels: OD_{214}, 253, 277, 297) were registered with 2 Kipp & Zonen (VWR, Piscataway, New Jersey) model BD 41 dual-pen stripchart recorders; the OD_{214} signal was also relayed to a model 970 A/D converter (PE Nelson, Cupertino, California), and chromatograms were analyzed and plotted using PE Nelson Turbochrom 3 (version 3.2) software. Fractions were collected by hand, kept on ice for the duration of the run, and then stored at −70 °C before repurification and/or analysis. Repurifications (second dimension LC) were carried out on a 1.0-mm SGE ODS-2 C18 (Scientific Glass Engineering, Austin, Texas) column using the same solvent system but at a flow rate of 30 μL/min; the microbore HPLC assembly has been described in detail elsewhere (Elicone et al., 1994). Samples were always acidified (10% TFA final concentration) and then diluted 2-fold with 0.1% TFA before rechromatography.

Mass spectrometry

Peptide mass analysis was carried out by MALDI-TOF MS using a Vestec (Houston, Texas) LaserTec Research instrument, with a 337-nm output nitrogen laser and a 1.2-m flight tube, and operated according to published principles (Beavis et al., 1991; Hillenkamp et al., 1991); 28 kV ion acceleration and 4.3 kV multiplier voltage were used. Tektronix (Beaverton, Oregon) model 2225 single-channel analog (50 MHz, real time) and model TDS520 dual-channel digitizing (500 MHz:500 megasampling/second averaging) oscilloscopes were connected in parallel to the detector for data acquisition; digitized spectra were downloaded to a ZEOS 486 33-MHz computer. The instrument also has a built-in video camera for real-time inspection of the laser beam impact on the target area. Typically the target was "scanned" until the spot was found that yielded best desorption/ionization. Laser power and number of acquisitions were adjusted interactively as judged from optimal deflections of specific maxima on the scope. The *m/z* spectra were generated from the time-of-flight files using the GRAMS (Galactic Ind., Salem, New Hampshire) data analysis software.

Steel probe tips ("pins") were cleaned by sonication in 5% acetic acid for 5 min, followed by washing with water and acetonitrile, and air drying. α-Cyano-4-hydroxy-cinnamic acid (ACCA) was used as matrix (obtained as a ready-to-use solution from Linear Scientific, Reno, Nevada). Working stocks of calibrant peptide "Apid" (MH^+ = 2,109.45) were stored at −20 °C in 10-μL aliquots (50 pmol/μL of 0.1% TFA/water) for 1-time use; just prior to use, this solution was serially diluted to 1 pmol and 0.1 pmol/μL by addition of 2% TFA in 33% MeCN and kept on ice. In this order, matrix (1 μL), analyte (1 μL), and calibrant mixture (0.5 μL) were spotted on the pin, using separate pipette tips, and mixed in situ by pipetting up and down (3 times). Samples were then air-dried at room temperature for 30 min. Every sample was analyzed twice, in the presence and absence of calibrant. For further details on the practical aspects of MALDI-MS experiments (data acquisition and analysis), we refer to an earlier publication (Geromanos et al., 1994).

Automated chemical sequencing

Purified peptides were sequenced with the aid of an automated instrument (model 477A, Applied Biosystems), operated according to the principles outlined by Hewick et al. (1981). Throughout, we used SMA-1 cycles (Table 2) and a reaction cartridge holding 9-mm discs, except where noted; further details can be found under Results and and discussion in the legends to Tables 1 and 2. Stepwise liberated PTH-amino acids were identified using an "on-line" 120A HPLC system, equipped with a PTH C18 (2.1 × 220-mm; 5-μm particle size) column (Applied Biosystems) and optimized for subpicomole PTH-amino acid analysis (Tempst & Riviere, 1989). Instruments were maintained and performance tested using previously established procedures (Erdjument-Bromage et al., 1993). Peptide-containing column fractions were always supplemented with neat TFA (to give a final concentration of 10%) before loading onto the sequencing disc to increase recoveries.

Computer analysis

Peptide average isotopic masses were summed from the identified residues (including the presumed ones) using ProComp version 1.2 software (obtained from Dr. P.C. Andrews, University of Michigan, Ann Arbor, Michigan). Peptide sequences were compared to entries in various sequence databases using the National Center for Biotechnology Information (NCBI) BLAST program (Altschul et al., 1990). Lower stringency alignments between all peptides and selected proteins were done using the Lipman–Pearson algorithm, available in the "Lasergene" software package (DNASTAR, Madison, Wisconsin).

Acknowledgments

We thank Scott Geromanos for instructing us in the use of MALDI-MS, Chris Elicone and Mike Powell for custom instrument assembly, and Lynne Lacomis for expert help with sequencing applications. The work was supported by an Irma T. Hirshl Career Scientist Award to P.T., a US Public Health Service grant DA-00266 and Research Scientist Award DA-00074 to S.H.S., and training grant GM-07309 to D.M.S. The Sloan-Kettering Protein Sequencing Lab is supported by NCI core grant 5 P30 CA08748-29.

References

Aebersold RH, Leavitt J, Saavedra RA, Hood LE, Kent SBH. 1987. Internal amino acid sequence analysis of protein separated by one- or two-

dimensional gel electrophoresis after in situ protease digestion on nitrocellulose. *Proc Natl Acad Sci USA 84*:6970–6974.
Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. 1990. Basic local alignment search tool. *J Mol Biol 215*:403–410.
Ambler RP. 1963. The amino acid sequence of *Pseudomonas* cytochrome *c*-551. *Biochem J 89*:349–377.
Andrews NC, Erdjument-Bromage H, Davidson MB, Tempst P, Orkin SH. 1993. Erythroid transcription factor NF-E2 is a haematopoietic-specific basic-leucine zipper protein. *Nature 362*:722–728.
Applied Biosystems. 1993a. Solvent B2—Improved separation system for the identification of PTH-tryptophan. *User Bulletin No 59*.
Applied Biosystems. 1993b. Use of the microcartridge for protein sequencing. *User Bulletin No 56*.
Atherton D, Fernandez J, DeMott M, Andrews L, Mische M. 1993. Routine protein sequence analysis below ten picomoles: One sequencing facility's approach. In: Hogue-Angeletti R, ed. *Techniques in protein chemistry IV*. San Diego: Academic Press. pp 409–418.
Beavis RC, Chaudhary T, Chait BT. 1991. α-Cyano-4-hydroxycinnamic acid as a matrix for matrix-assisted laser desorption mass spectrometry. *Org Mass Spectrom 27*:156–158.
Brown JL, Roberts WK. 1976. Evidence that approximately eighty per cent of the soluble proteins from Ehrlich ascites cells are *N*-acetylated. *J Biol Chem 251*:1009–1014.
Cafferkey R, Young PR, McLaughlin MM, Bergsma DJ, Koltin Y, Sathe GM, Faucette L, Eng W, Johnson RK, Livi GP. 1993. Dominant missense mutations in a novel yeast protein related to mammalian phosphatidylinositol 3-kinase and VPS34 abrogate rapamycin cytotoxicity. *Mol Cell Biol 13*:6012–6023.
Casteels, P, Romagnolo J, Castle M, Casteels-Josson K, Erdjument-Bromage H, Tempst P. 1994. Biodiversity of *apidaecin*-type peptide antibiotics: Prospects of manipulating the antibacterial spectrum and combating acquired resistance. *J Biol Chem 269*:26107–26115.
Chait BT, Wang R, Beavis RC, Kent SBH. 1993. Protein ladder sequencing. *Science 262*:89–92.
Dou S, Zeng X, Cortes P, Erdjument-Bromage H, Tempst P, Honjo T, Vales LD. 1994. The recombination signal sequence-binding protein RBP-2N functions as a transcriptional repressor. *Mol Cell Biol 14*:3310–3319.
Dujon B, et al. (109 authors). 1994. Complete DNA sequence of yeast chromosome XI. *Nature 369*:371–378.
Eckerskorn C, Lottspeich F. 1989. Internal amino acid sequence analysis of proteins separated by gel electrophoresis after tryptic digestion in the polyacrylamide matrix. *Chromatographia 28*:92–94.
Elicone C, Lui M, Geromanos S, Erdjument-Bromage H, Tempst P. 1994. Microbore reversed-phase high-performance liquid chromatographic purification of peptides for combined chemical sequencing/laser-desorption mass spectrometric analysis. *J Chromatogr 676*:121–137.
Erdjument-Bromage H, Geromanos S, Chodera A, Tempst P. 1993. Successful peptide sequencing with femtomole level PTH-analysis: A commentary. In: Hogue-Angeletti R, ed. *Techniques in protein chemistry IV*. San Diego: Academic Press. pp 419–426.
Fasman GD. 1989. *Practical handbook of biochemistry and molecular biology*. Boca Raton, Florida: CRC Press.
Fernandez J, Andrews L, Mische SM. 1994. An improved procedure for enzymatic digestion of polyvinylidene difluoride-bound proteins for internal sequence analysis. *Anal Biochem 218*:112–117.
Fernandez J, DeMott M, Atherton D, Mische SM. 1992. Internal protein sequence analysis: Enzymatic digestion for less than 10 μg of protein bound to polyvinylidene difluoride or nitrocellulose membranes. *Anal Biochem 201*:255–264.
Fumagalli S, Totty NF, Hsuan JJ, Courtneidge SA. 1994. A target for Src in mitosis. *Nature 368*:871–874.
Geromanos S, Casteels P, Elicone C, Powell M, Tempst P. 1994. Combined Edman-chemical and laser-desorption mass spectrometric approaches to micro peptide sequencing: Optimization and applications. In: Crabb JW, ed. *Techniques in protein chemistry V*. San Diego: Academic Press. pp 143–150.
Gout I, Dhand R, Hiles ID, Fry MJ, Panayotou G, Das P, Truong O, Totty NF, Hsuan J, Booker GW, Campbell ID, Waterfield MD. 1993. The GTPase dynamin binds to and is activated by a subset of SH3 domains. *Cell 75*:25–36.
Helliwell SB, Wagner P, Kunz J, Deuter-Reinhard M, Henriquez R, Hall MN. 1994. TOR1 and TOR2 are structurally and functionally similar but not identical phosphatidylinositol kinase homologues in yeast. *Mol Biol Cell 5*:105–118.
Henzel WJ, Billeci TM, Stults JT, Wong S, Grimley C, Watanabe C. 1993. Identifying proteins from two-dimensional gels by molecular mass searching of peptide fragments in protein sequence databases. *Proc Natl Acad Sci USA 90*:5011–5015.
Hess D, Covey TC, Winz R, Brownsey RW, Aebersold R. 1993. Analytical and micropreparative peptide mapping by high performance liquid chromatography/electrospray mass spectrometry of proteins purified by gel electrophoresis. *Protein Sci 2*:1342–1351.
Hewick RM, Hunkapiller MW, Hood LE, Dreyer WJ. 1981. A gas–liquid solid phase peptide and protein sequenator. *J Biol Chem 256*:7990–7997.
Hillenkamp F, Karas M, Beavis RC, Chait BT. 1991. Matrix-assisted laser desorption/ionization mass spectrometry of biopolymers. *Anal Chem 63*:1193A–1203A.
Hunkapiller M, Kent S, Caruthers M, Dreyer W, Firca J, Giffin C, Horvath S, Hunkapiller T, Tempst P, Hood L. 1984. A microchemical facility for the analysis and synthesis of genes and proteins. *Nature 310*:105–111.
Hunt DF, Henderson RA, Shabanowitz J, Sakaguchi K, Michel H, Sevilir N, Cox AL, Appella E, Engelhard VH. 1992. Characterization of peptides bound to the class I MHC molecule HLA-A2.1 by mass spectrometry. *Science 255*:1261–1266.
Kunz J, Hall MN. 1993. Cyclosporin A, FK506 and rapamycin: More than just immunosuppression. *Trends Biochem Sci 18*:334–338.
Kunz J, Henriquez R, Schneider U, Deuter-Reinhard M, Movva NR, Hall MN. 1993. Target of rapamycin in yeast, TOR2, is an essential phosphatidyl kinase homolog required for G_1 progression. *Cell 73*:585–596.
Liu J, Farmer JD, Lane WS, Friedman J, Weissman I, Schreiber SL. 1991. Calcineurin is a common target of cyclophilin–cyclosporin A and FKBP–FK506 complexes. *Cell 66*:807–815.
Mann M, Hojrup P, Roepstroff P. 1993. Use of mass spectrometric molecular weight information to identify proteins in sequence databases. *Biol Mass Spectrom 22*:338–345.
Marks AR, Fleischer S, Tempst P. 1990a. Surface topography analysis of the ryanodine receptor/junctional channel complex based on proteolysis sensitivity mapping. *J Biol Chem 265*:13143–13149.
Marks AR, Tempst P, Chadwick C, Riviere LR, Fleischer S, Nadal-Ginard B. 1990b. Smooth muscle and brain inositol-1,4,5-triphospate receptors are structurally and functionally similar. *J Biol Chem 265*:20719–20722.
Merino A, Madden KR, Lane WS, Champoux JJ, Reinberg D. 1993. DNA topoisomerase I is involved in both repression and activation of transcription. *Nature 365*:227–232.
Pappin DJC, Hojrup P, Bleasby AJ. 1993. Rapid identification of proteins by peptide-mass fingerprinting. *Curr Biol 3*:327–332.
Polyak K, Lee MH, Erdjument-Bromage H, Koff A, Roberts JM, Tempst P, Massague J. 1994. Cloning of $p27^{Kip1}$, a cyclin-dependent kinase inhibitor and a potential mediator of extracellular antimitogenic signals. *Cell 78*:59–66.
Rosenfeld J, Capdevielle J, Guillemot JC, Ferrara P. 1992. In-gel digestion of proteins for internal sequence analysis after one- or two-dimensional gel electrophoresis. *Anal Biochem 203*:173–179.
Sabatini DM, Erdjument-Bromage H, Lui M, Tempst P, Snyder SH. 1994. RAFT1: A mammalian protein that binds to FKBP12 in a rapamycin-dependent fashion and is homologous to yeast TORs. *Cell 78*:35–43.
Sehgal SN. 1993. Immunosuppressive profile of rapamycin. *Ann NY Acad Sci 696*:1–8.
Sogaard M, Tani K, Ye RR, Geromanos S, Tempst P, Kirchhausen T, Rothman JE, Sollner T. 1994. A rab protein is required for the assembly of SNARE complexes in the docking of transport vesicles. *Cell 78*:937–948.
Sollner T, Whiteheart SW, Brunner M, Erdjument-Bromage H, Geromanos S, Tempst P, Rothman JE. 1993. SNAP receptors implicated in vesicle targeting and fusion. *Nature 362*:318–324.
Tempst P, Link AJ, Riviere LR, Fleming M, Elicone C. 1990. Internal sequence analysis of proteins separated on polyacrylamide gels at the submicrogram level: Improved methods, applications and gene cloning strategies. *Electrophoresis 11*:537–553.
Tempst P, Riviere L. 1989. Examination of automated polypeptide sequencing using standard phenyl isothiocyanate reagent and subpicomole high performance liquid chromatographic analysis. *Anal Biochem 183*: 290–300.
Totty NF, Waterfield MD, Hsuan JJ. 1992. Accelerated high-sensitivity microsequencing of proteins and peptides using a miniature cartridge. *Protein Sci 1*:1215–1224.
Williams K, Kobayashi R, Lane W, Tempst P. 1993. Internal amino acid sequencing: Observations from four different laboratories. *ABRF News 4*(4):7–12.
Wilson R, et al. (46 authors). 1994. 22.2 Mb of contiguous nucleotide sequence from chromosome III of *C. elegans*. *Nature 368*:32–38.
Wong SW, Grimley C, Padua JH, Bourell JH, Henzel WJ. 1993. Peptide mapping of 2-D gel proteins by capillary HPLC. In: Hogue-Angeletti R, ed. *Techniques in protein chemistry IV*. San Diego: Academic Press. pp 371–378.
Yates JR III, Speicher S, Griffin PR, Hunkapiller T. 1993. Peptide mass maps: A highly informative approach to protein identification. *Anal Biochem 214*:397–408.