Journal of Chromatography, 359 (1986) 213–220
Elsevier Science Publishers B.V., Amsterdam — Printed in The Netherlands

CHROMSYMP. 815

MICROSCALE ELUCIDATION OF AN N-LINKED GLYCOSYLATION SITE BY COMPARATIVE HIGH-PERFORMANCE LIQUID CHROMATOGRAPHIC PEPTIDE MAPPING

JAMES J. L'ITALIEN
Molecular Genetics, Inc., Minnetonka, MN 55343 (U.S.A.)

SUMMARY

The identification of specific sites of post-translational modifications of polypeptides is an important component in understanding protein structure–function relationships. This report describes the application of specific enzymatic methods of N-linked oligosaccharide removal (*endo* H and *endo* F) to the rapid microscale identification of specific sites of N-linked glycosylation. Following deglycosylation the protein is subjected to proteolytic digestion and comparative high-performance liquid chromatographic peptide mapping with an unmodified protein digest. Peptides from which oligosaccharide(s) have been removed exhibit a measurable increase in retention time on reversed-phase high-performance liquid chromatography and can be clearly identified. Once identified, the peptide can be subjected to direct protein microsequence analysis to elucidate the specific site of glycosylation. As illustrated here, these methods are compatible with microscale protein purification by sodium dodecyl sulphate-polyacrylamide gel electrophoresis followed by electroelution and are applicable to the 100-pmol range.

INTRODUCTION

One of the principal aims in modern approaches to protein structural analysis is the localization of sites of post-translational modification with minimal amounts of polypeptide[1]. While many types of post-translational modification have been observed in polypeptides, perhaps the most common modifications in mature eukaryotic polypeptides are sites of glycosylation[2]. Carbohydrate residues can be linked to polypeptides through the side-chain hydroxyl group of serine or threonine (O-linked glycosylation, ref. 3). In addition, carbohydrate may be linked through the amide group of asparagine (N-linked glycosylation) but this is restricted to Asn which exists in the sequence X-Asn-X-Thr/Ser where X can be any amino acid[4]. In general, N-linked carbohydrates are branched oligosaccharides of more than seven sugar residues, and may be classified as either high-mannose asparagine-linked carbohydrate (containing seven to twelve residues) or as complex asparagine-linked carbohydrate (usually larger and endcapped with sialic acid)[5]. Specific methods exist for the enzymatic removal of N-linked high-mannose oligosaccharides (*endo*-β-acetylglucosa-

0021-9673/86/$03.50 © 1986 Elsevier Science Publishers B.V.

minidase H, ref. 6) and N-linked complex oligosaccharides (*endo-β*-acetylglucosaminidase F, ref. 7). Although these enzymes have proven useful in the study of glycoproteins, their use has generally been limited to glycoproteins available in large amounts. Thus, there is a need for a rapid microscale method for the identification of N-linked glycosylation sites for structure–function studies of glycopolypeptides. High-performance liquid chromatographic (HPLC) peptide purification, especially comparative peptide mapping, has proven to be a useful method for identifying subtle differences between polypeptides[8,9]. This paper describes the application of specific enzymatic methods of N-linked carbohydrate removal to the rapid identification of specific sites of N-linked glycosylation of comparative HPLC peptide mapping and micromethods of direct polypeptide sequence analysis.

EXPERIMENTAL

Enzymes used in this study were *endo-β*-N-acetylglucosaminidase H (Miles Laboratories, Elkhart, IN, U.S.A.), *endo-β*-N-acetylglucosaminidase F (New England Nuclear, Boston, MA, U.S.A.), and trypsin-TPCK (Worthington Biochemicals, Freehold, NJ, U.S.A.). All chemicals used were of reagent grade unless otherwise stated. Water was purified by a Milli-Q water system (Millipore, Bedford, MA, U.S.A.). The glycoprotein used in this study was isolated from an animal herpesvirus by a two-step procedure involving immunoaffinity chromatography and electroelution of the protein following sodium dodecyl sulphate-polyacrylamide gel electrophoresis (SDS-PAGE)[1]. Following electroelution, the SDS and Coomassie Blue stain remaining from the purification process were removed by ion-pair extraction[10]. In this procedure, one part of sample is mixed with nineteen parts of freshly prepared extraction buffer (acetone–acetic acid–triethylamine, 90:5:5), placed on ice for 30 min, and centrifuged. This extraction is generally repeated and, finally, the sample is rinsed with acetone and dried. After extraction, the proteins were solubilized in 0.1% SDS at a concentration of 1 μg/μl and the purity of the sample was ascertained by amino-terminal sequence analysis.

The conditions used to screen the activity of endoglycosidases on this glycoprotein were as follows. For *endo-β*-N-acetylglucosaminidase H (*endo* H), 10 μg of glycoprotein in 10 μl of 0.1% SDS were mixed with 10 μl of 40 m*M* sodium phosphate (pH 6.0), 50 m*M* dithiothreitol. The sample was heated to 60°C for 5 min, cooled to 37°C, and then 0.1 μg of *endo* H, dissolved in 50 m*M* sodium phosphate, was added at a concentration of 0.05 μg/μl. The reaction was allowed to proceed at 37°C for 24 h. Aliquots were removed at 0, 4 and 24 h to study the time dependency of the reaction. The reactions were quenched by boiling the incubates in PAGE sample application buffer for 2 min to denature the enzyme and were then stored until electrophoresis. A sham digest (identical conditions except no enzyme was added), as well as positive (ovalbumin or ribonuclease B) and negative controls (bovine serum albumin) were routinely analyzed to ascertain the intrinsic degradation of the glycoprotein and the glycosidic and proteolytic activity of the endoglycosidases, respectively.

For *endo-β*-N-acetylglucosaminidase F (*endo* F), 10 μg of protein sample in 10 μl of 0.1% SDS were mixed with 10 μl of 200 m*M* sodium phosphate (pH 6.1), 100 m*M* EDTA, 2% NP-40, 0.1% SDS, and 2% 2-mercaptoethanol, and heated to 60°C

for 5 min. The sample was cooled to 37°C prior to the addition of 10–100 mU of *endo* F per μg of substrate. The reaction was carried out at 37°C for 24 h with the same time course and controls as described above for *endo* H. The samples, sham digests, and controls were then screened by SDS-PAGE[11]. Optimized conditions were used to digest 45 μg (640 pmol) of the test glycoprotein with *endo* F. For this reaction, the substrate was prepared as described for the 10 μg reaction, but 20 mU of *endo* F were used per μg of substrate. The reaction was carried out for 24 h at 37°C and the reaction products were monitored by SDS-PAGE. The reaction products were then extracted by the ion-pair method (described above) and lyophylized. Tryptic digestion was carried out by dissolving the sample in 5 μl of 8 *M* urea (ultrapure), 100 m*M* Tris (pH 8.1), sonicating the sample in a cup sonicator and then diluting it to 50 μl with 50 m*M* ammonium bicarbonate (pH 7.8). After checking the solubility of the protein by SDS-PAGE, tryptic digestion was carried out by adding 1% trypsin-TPCK by weight to the sample at both 0 and 2 h and allowing the reaction to proceed for a total of 4 h at 37°C[12]. The completeness of the reaction was determined by reversed-phase HPLC of 5% of each sample (35 pmol).

The HPLC system used in this work consisted of a Waters (Milford, MA, U.S.A.) Model 730 system controller, two Model 6000A Waters solvent delivery systems, an Axiom (Cole Scientific, Calabasas, CA, U.S.A.) dynamic gradient mixer, a U-6K injector (Waters), and a Kratos (Ramsey, NJ, U.S.A.) Model 773 variable-wavelength UV detector. Peptide mapping was performed on a 250 × 4.6 mm I.D. Vydac C_{18} column (The Separations Group, Hesparia, CA, U.S.A.) by using either 10 m*M* potassium phosphate (pH 2.5) and acetonitrile (HPLC grade) or 0.1% trifluoroacetic acid (redistilled) in both the aqueous and acetonitrile phases. Gradient conditions and flow-rates are given in the figure legends where appropriate. Protein sequence analysis was performed with an Applied Biosystems (Foster City, CA, U.S.A.) Model 470A gas-phase protein sequencer operating under the 02NVAC program[18]. All sequencer chemicals were purchased from Applied Biosystems. PTH analysis was performed with an IBM (Danbury, CT, U.S.A.) cyano column (250 × 4.6 mm I.D.) at 38°C, with 80 m*M* sodium acetate (pH 5.25), containing 5% THF (HPLC grade) as the aqueous solvent and acetonitrile (HPLC grade) as the organic solvent.

RESULTS AND DISCUSSION

The glycoprotein of interest in this study, which was purified by micromethods of immunoaffinity chromatography[13] followed directly by SDS-PAGE and electroelution[1], was screened for its susceptibility to *endo* H and *endo* F in a time course study. The incubates were subjected to analysis by SDS-PAGE. As seen in Fig. 1, this glycoprotein showed no change in molecular weight following digestion with *endo* H but demonstrated a reduction by *ca.* 3000 dalton following digestion with *endo* F. Based upon these findings and the known specificity of *endo* H (it only cleaves high-mannose N-linked oligosaccharides) and *endo* F (cleaves both high-mannose and complex N-linked oligosaccharides), it was concluded that the observed decrease in molecular weight of the glycoprotein was likely the result of specific removal of complex oligosaccharide(s). It should be stressed that precautions were taken to ensure that the endoglycosidic reactions were performed in the presence of the appro-

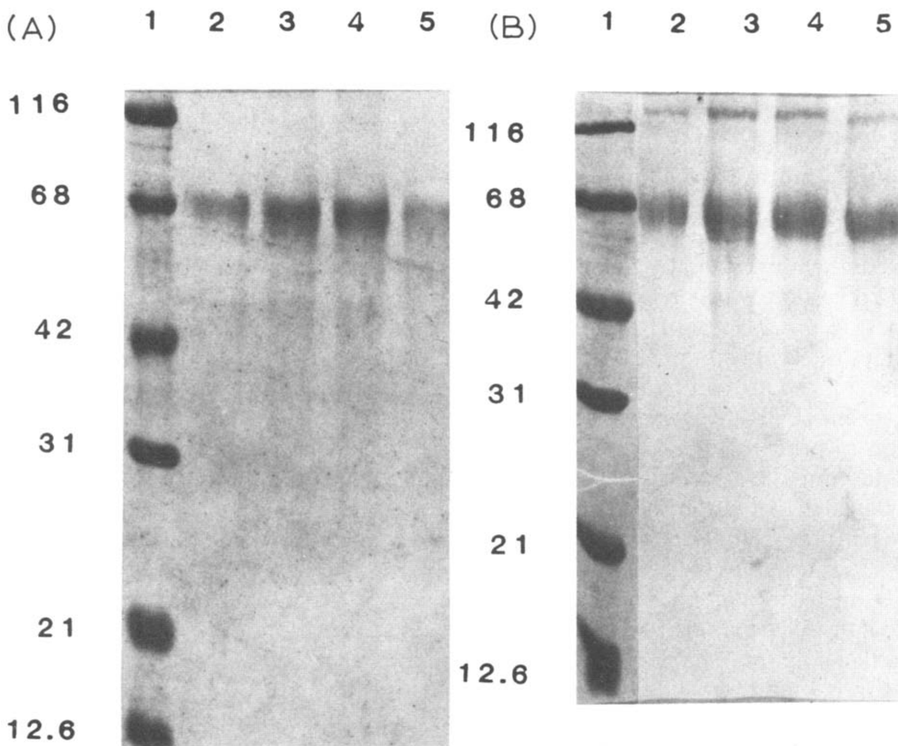

Fig. 1. (A) Analysis of analytical-scale enzymatic deglycosylation reaction by SDS-PAGE. In this experiment, the herpesvirus glycoprotein which was the subject of this study was treated with *endo* H as described in the text. Aliquots of the reaction mixture were removed at 0, 4 and 24 h and quenched by boiling in sample application buffer. These samples were then subjected to analysis by SDS-PAGE to look for molecular weight changes resulting from carbohydrate removal. Sham reactions, as well as positive and negative controls were run to demonstrate the specificity of the reaction (controls not illustrated). Lane 1 represents molecular weight markers whose molecular weight in kilodaltons is depicted on the left. Lane 2 represents an *endo* H sham reaction. Lanes 3, 4 and 5 represent the 0, 4 and 24 h time points of the *endo* H reaction, respectively. No changes in the molecular weight of the glycoprotein of interest was observed following treatment with *endo* H. (B) *Endo* F was used to treat the same herpesvirus glycoprotein, the reaction time course and controls were run as previously described in Fig. 1A. Lane 1 represents molecular weight markers whose molecular weight is depicted in kilodaltons on the left. Lane 2 represents an *endo* F sham reaction. Lanes 3, 4 and 5 represent the 0, 4 and 24 h time points of the *endo* F reaction, respectively. As noted from this figure, treatment of this glycoprotein with *endo* F results in an endoglycosidase specific reduction of *ca.* 2000–3000 dalton indicative of removal of *ca.* one site of N-linked glycosylation as described in the text.

priate inhibitors, where necessary. (EDTA has been shown to inhibit any proteolytic activity still associated with *endo* F[7], but PMSF, which has been used with *endo* F[14], was not needed in this study.) Furthermore, proteolytic activity from the endoglycosidases (measured by incubation with non-glycosylated substrates) and from possible contamination of buffers or proteins (measured by sham digests) could not be detected. The activity of each endoglycosidase was also assessed by including glycoproteins, known to be susceptible, under conditions identical with those for the glycoprotein of interest as positive controls.

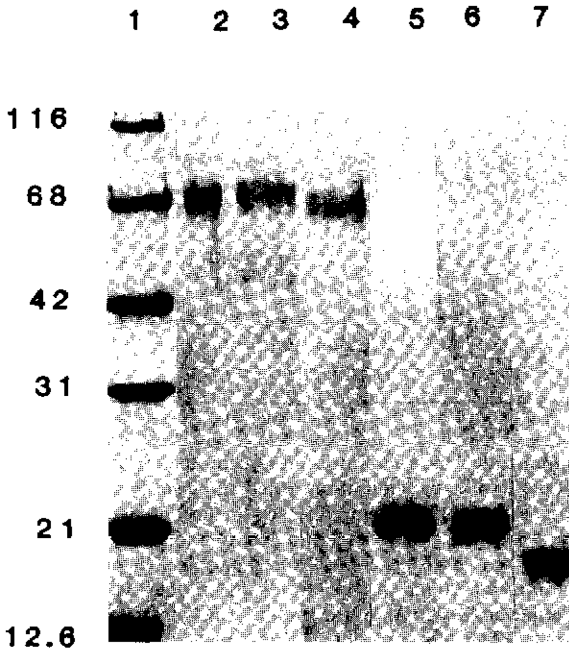

Fig. 2. Analysis of preparative scale (640 pmol) enzymatic deglycosylation of the herpesvirus glycoprotein by *endo* F. Again, sham reactions, as well as both positive and negative controls were run to demonstrate the specificity of the reaction. Lane 1 represents molecular weight markers whose molecular weight is given in kilodaltons on the left. Lanes 2, 3 and 4 represent the sham reaction of the glycoprotein treated with *endo* F, a 0 h time point and a 24 h time point, respectively. Lanes 5, 6 and 7 represent the reaction products from *endo* F treatment of ribonuclease B (a positive control) of a sham reaction, a 0 h time point and a 24 h time point, respectively. As seen in this figure, the results were identical with the analytical-scale reaction showing an endoglycosidase specific decrease of *ca.* 2000–3000 dalton following treatment with *endo* F.

To identify the specific site(s) of N-linked glycosylation, a preparative reaction was performed on 45 μg of this 70 000-dalton glycoprotein (640 pmol) by using *endo* F under the same conditions as in the analytical reaction. A parallel sham reaction was also performed with 45 μg of glycoprotein. The reaction products from both the *endo* F digest and the sham reaction were then evaluated by SDS-PAGE (Fig. 2). The sham digest showed no decrease in molecular weight, while the *endo* F digest resulted in a molecular weight decrease of *ca.* 3000 dalton. These results were in agreement with our analytical-scale experiments. Following ion-pair extraction of both the *endo* F and sham reaction products, the resulting polypeptides were resuspended and subjected to proteolytic digestion with trypsin. The completeness of the tryptic digestion on the *endo* F and sham-treated glycoprotein was ascertained by direct analysis of 5% of each sample (35 pmol) by reversed-phase HPLC, as seen in Fig. 3. Individually, each of these tryptic profiles illustrates that the proteolytic digestion proceeded well. Furthermore, when these profiles are compared, it is apparent that a major peak at 44 min in the sham (without *endo* F) tryptic profile is shifted by *ca.* 1 min in the *endo* F-treated profile (Fig. 3). This peak shift is significant since no other peptide shifts of this magnitude were observed, and it is consistent with previous

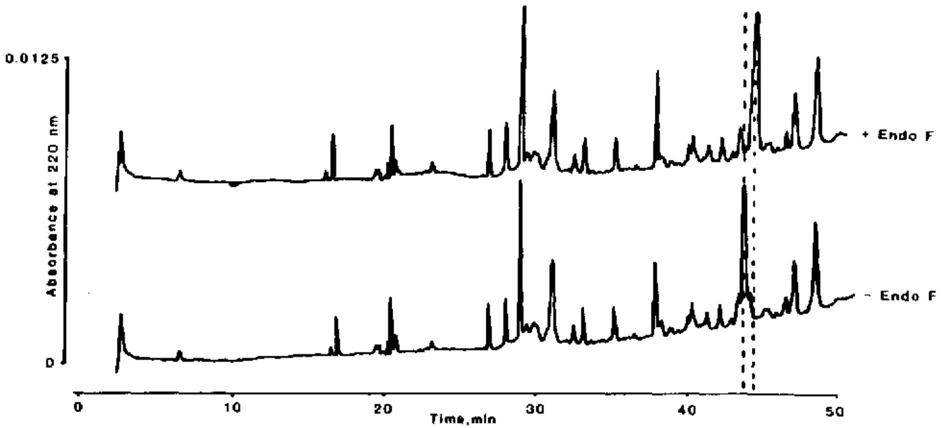

Fig. 3. Comparative HPLC peptide maps of untreated and *endo* F-treated glycoprotein. In this figure, 35 pmol of the peptide mixture resulting from the tryptic digestion of untreated and *endo* F-treated glycoprotein were applied in the digestion buffer to a Vydac C_{18} column (250 × 4.6 mm I.D.) equilibrated in 20 m*M* sodium phosphate (pH 2.5). The chromatograms were developed with a gradient of 0–45% acetonitrile over 45 min, followed directly with a 5-min gradient of 45–70% acetonitrile. The flow-rate was 1 ml/min and the column effluent was monitored at 220 nm. As seen in this figure, there is one peak in the chromatogram (highlighted by a dashed line at *ca.* 45 min of the profile) which exhibits an increase in its retention time following *endo* F treatment of the glycoprotein. This observation, that the deglycosylated peptide (*endo* F treated) had an increased retention time compared to the glycosylated peptide (untreated) is consistent with previous observations[13,14] and became the focus of our search to identify the site of *endo* F-sensitive glycosylation.

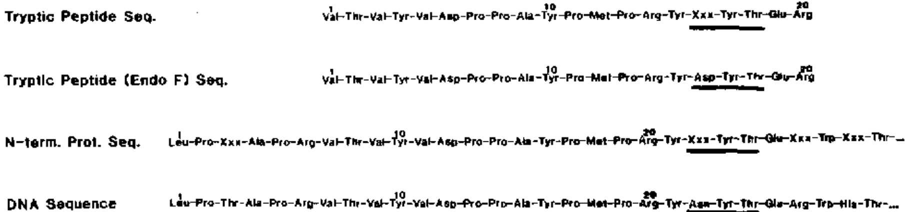

Fig. 4. This figure summarizes the results from direct protein sequence analysis of the peptides identified in Fig. 3, as well as the amino terminal sequence of the protein from protein and DNA sequence analysis. The sequence of the untreated glycoprotein tryptic peptide and the *endo* F-treated glycoprotein tryptic peptide were identical, with one exception at position 16 of the sequence. As described in the text, carbohydrate modification of amino acids precludes their identification by the commonly used methods in direct protein sequence analysis. Thus, a gap was observed in the sequence of the untreated (glycosylated) tryptic peptide, while an Asp was observed in this position of the *endo* F-treated (deglycosylated) tryptic peptide (details in text). Together, the gap in the untreated glycoprotein tryptic peptide sequence and the Asp in the same position of the *endo* F-treated glycoprotein tryptic peptide strongly suggest that this is a site (underlined) of N-linked glycosylation. Subsequently, the DNA generated protein sequence showed that Asn is coded for in this position (underlined), creating the required Asn-Xxx-Thr site needed for N-linked glycosylation.

observations that the removal of oligosaccharides from a glycopeptide increases its retention time[13,14].

In a subsequent experiment, the peptide thus identified was collected from both the sham and *endo* F glycoprotein tryptic peptide maps obtained with the remaining 90% of each sample (*ca.* 570 pmol) and its sequence was determined by gas-phase protein sequence analysis using *ca.* 20% of the collected peptide. The results of this experiment (details will be published elsewhere) are given in Fig. 4. This figure shows that the sequence obtained from the two isolated peptides is consistent with the results obtained from amino-terminal sequence analysis of this glycoprotein[15] and from the corresponding DNA sequence[16]. Upon closer inspection of the peptide sequence it was observed that there was a gap at position 16 of the peptide from the sham digest but that there was Asp (in lower than expected yield) at the same position of the peptide from the *endo* F digest. A gap in the corresponding position of the amino-terminal protein sequence was also noted. Finally, the DNA sequence of this region shows that an Asn residue is coded for in this position, and this creates an Asn-Xxx-Thr site required for N-linked glycosylation. The observation of blanks in the protein sequence of the glycosylated polypeptides is consistent with the observation that the polarity of either Asn-oligosaccharide complex or Asn-N-acetylglucosamine (the normal endpoint of endoglycosidase cleavage) generally does not permit the extraction of their corresponding anilinothiazolinone (ATZ) derivatives from the reaction vessel of a gas-phase sequencer (J. J. L'Italien, unpublished observations). Furthermore, even when the carbohydrate-modified ATZ derivatives of ASN can be recovered from the reaction vessel (in solid-phase sequencers they are recovered from the reaction vessel in trifluoroacetic acid) and subsequently converted into their corresponding phenylthiohydantoin (PTH) derivatives, they are too hydrophilic to be detected in most PTH identification systems (J. J. L'Italien, unpublished observations).

The identification of Asp in position 16 of the *endo* F-treated peptide results from the fact that commercial preparations of *endo* F contain a mixture of two endoglycosidases[17]. One of these is the true *endo* F, which cleaves the di-N-acetylchitobiosyl bond so that one acetylglucosamine residue remains on the released oligosaccharide and the other remains on the asparagine residue of polypeptide. The second glycosidase, termed peptide: N-glycosidase F, cleaves at the glucosamine junction, releasing the intact oligosaccharide and converting the Asn of the polypeptide into an Asp residue at the site of cleavage[17]. These enzymes have different pH optima, the *endo* F activity being optimal at pH 4 and the peptide:N-glycosidase F activity being optimal at pH 9.3[17]. Additional studies have shown that at pH 6 there is a *ca.* 50:50 mixture of activity from each enzyme[17]. Therefore, the observation of Asp at the former site of glycosylation is also consistent with the known enzymatic activity. In the future, all screening of glycopeptides will be performed at both pH 6 and 9. If there is degradation of the substrate at both pH values, then the final reaction will be performed at pH 9. This gives a uniform product (only the peptide:N-glycosidase F is active at this pH), which can be directly identified by the usual methods of PTH analysis from protein microsequence analysis.

CONCLUSIONS

The use of specific endoglycosidases and comparative proteolytic peptide map-

ping provides an excellent method for the direct microscale identification of sites of N-linked glycosylation. Endoglycosidase F (which includes peptide:N-glycosidase F, as commercially available) may be the optimal preparation to use in this type of study because of its ability to cleave both high-mannose and complex oligosaccharides from glycopolypeptides. In addition, the specific use of peptide:N-glycosidase through the control of pH can completely remove the oligosaccharide of the glycopolypeptide and give an unmodified Asp at the former site of glycosylation. This presents an advantage for identification of sites by direct polypeptide microsequence analysis, because the Asp residue can be identified directly without modification of a PTH analysis system. Thus, by comparison of glycosidase-treated and untreated glycopolypeptides, sites of N-linked oligosaccharide can be directly identified. Finally, as illustrated here, endoglycosidases can be used to identify sites of N-linked glycosylation on proteins that have been purified by SDS-PAGE followed by electroelution, which presents a convenient purification route for glycoproteins for this type of study.

REFERENCES

1 J. J. L'Italien, in J. Shively (Editor), *Methods of Protein Microcharacterization*, Humana Press, Clifton, NJ, 1986, Ch. 10, pp. 279–314.
2 F. Wold, *Ann. Rev. Biochem.*, 50 (1981) 783–814.
3 R. Kornfeld and S. Kornfeld, *Ann. Rev. Biochem.*, 45 (1976) 217–237.
4 E. L. Chambers, B. C. Pressman and B. Rose, *Biochem. Biophys. Res. Commun.*, 60 (1974) 126–132.
5 R. Kornfeld and S. Kornfeld, *Ann. Rev. Biochem.*, 54 (1985) 631–664.
6 A. L. Tarentino and F. Maley, *J. Biol. Chem.*, 249 (1974) 811–817.
7 J. H. Elder and S. Alexander, *Proc. Natl. Acad. Sci., U.S.A.*, 79 (1982) 4540–4544.
8 J. J. L'Italien and J. E. Strickler, *Anal. Biochem.*, 127 (1982) 198–212.
9 J. ,W. Chase, J. J. L'Italien, J. B. Murphy, E. K. Spicer and K. R. Williams, *J. Biol. Chem.*, 259 (1984) 805–814.
10 W. H. Koningsberg and L. Henderson, *Methods Enzymol.*, 91 (1983) 254–259.
11 U. K. Laemmli, *Nature (London)*, 227 (1970) 680–685.
12 J. J. L'Italien and R. A. Laursen, *J. Biol. Chem.*, 256 (1981) 8092–8101.
13 J. J. L'Italien (Editor), *Proteins: Structure and Function*, Plenum Press, New York, 1986, in press.
14 T. H. Plummer, Jr., J. H. Elder, S. Alexander, A. W. Phelan and A. L. Tarentino, *J. Biol. Chem.*, 259 (1984) 10 700–10 704.
15 J. J. L'Italien, R. Hansen, J. P. Phelps and T. E. Zamb, in preparation.
16 T. E. Zamb, in preparation.
17 R. B. Trimble and F. Maley, *Anal. Biochem.*, 141 (1984) 515–522.
18 R. M. Hewick, M. W. Hunkapiller, L. E. Hood and W. J. Dreyer, *J. Biol. Chem.*, 256 (1981) 7990–7997.