SUMMARY

(In accordance with 40 CFR part 152, this summary is available for public release after registration)

STUDY TITLE

Characterization of the Aryloxyalkanoate Dioxygenase-1 (AAD-1) Protein Derived from Transgenic Maize Event DAS-40278-9

DATA REQUIREMENTS

Not Applicable

AUTHOR(S)

B. W. Schafer and S. K. Embrey

STUDY COMPLETED ON

20-May-2009

PERFORMING LABORATORY

Regulatory Sciences and Government Affairs — Indianapolis Lab
Dow AgroSciences LLC
9330 Zionsville Road
Indianapolis, Indiana 46268-1054

LABORATORY STUDY ID

080142

Characterization of the Aryloxyalkanoate Dioxygenase-1 (AAD-1) Protein Derived from Transgenic Maize Event DAS-40278-9

SUMMARY

Corn has been modified by the insertion of the *aad*-1 gene from *Sphingomonas herbicidivorans* which encodes the aryloxyalkanoate dioxygenase (AAD-1) protein. The trait confers tolerance to 2,4-dichlorophenoxyacetic acid and aryloxyphenoxypropionate (commonly referred to as "fop" herbicides such as quizalofop) herbicides which may be used as a selectable marker during plant transformation and in breeding nurseries. Whiskers transformation of corn with plasmid pDAS1740 was carried forward, through breeding, to produce plants containing event AAD-1 DAS-40278-9, which is the focus of this study. The AAD-1 protein is approximately 33 kDa in size.

To perform various toxicology, eco-toxicology, biochemical characterization, and enzymatic activity studies, large quantities of the AAD-1 protein are required. Because it is technically infeasible to extract and purify sufficient amounts of AAD-1 protein from transgenic plants, the protein was produced with a proprietary *Pseudomonas fluorescens* CellCap™ expression system.

The purpose of this study was to characterize the biochemical properties of the recombinant AAD-1 protein derived from the transgenic maize event DAS-40278-9 (event 278) and compare them with the properties of the previously characterized microbe-derived AAD-1 protein (Schafer, 2008 and Kuppannan, et al 2007). Sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS-PAGE, stained with Coomassie blue and glycoprotein detection methods), western blot, immunodiagnostic test strip assays, matrix assisted laser desorption/ionization time-of-flight mass spectrometry (MALDI-TOF MS) and protein sequencing analysis by tandem MS were used to characterize the biochemical properties of the protein. Utilizing these methods, the AAD-1 proteins from *P. fluorescens* and transgenic maize (event 278) were shown to be

biochemically equivalent. These data support the use of the microbial protein in studies supporting the registration of transgenic corn expressing the AAD-1 protein.

STUDY TITLE

Characterization of the Aryloxyalkanoate Dioxygenase-1 (AAD-1) Protein Derived from Transgenic Maize Event DAS-40278-9

DATA REQUIREMENTS

Not Applicable

AUTHOR(S)

B. W. Schafer and S. K. Embrey
317-337-3677
[bwschafer@dow.com]

STUDY COMPLETED ON

20-May-2009

PERFORMING LABORATORY

Regulatory Sciences and Government Affairs — Indianapolis Lab
Dow AgroSciences LLC
9330 Zionsville Road
Indianapolis, Indiana 46268-1054

LABORATORY STUDY ID

080142

STATEMENT OF NO DATA CONFIDENTIALITY CLAIMS

Compound: AAD-1

Title: Characterization of the Aryloxyalkanoate Dioxygenase-1 (AAD-1) Protein Derived from Transgenic Maize Event DAS-40278-9

No claim of confidentiality is made for any information contained in this study on the basis of its falling within the scope of FIFRA Section 10 (d)(1)(A)(B), or (C).*

Company: Dow AgroSciences LLC

Company Agent: P. L. Hunst

Title: Regulatory Manager

Signature: Penny L. Hunst

Date: 27-Feb-2009

*In the United States, the above statement supersedes all other statements of confidentiality that may occur elsewhere in this report.

THIS DATA MAY BE CONSIDERED CONFIDENTIAL IN COUNTRIES OUTSIDE THE UNITED STATES.

STATEMENT OF COMPLIANCE WITH GOOD LABORATORY PRACTICE STANDARDS

Title: Characterization of the Aryloxyalkanoate Dioxygenase-1 (AAD-1) Protein Derived from Transgenic Maize Event DAS-40278-9

Study Initiation Date: 09-Jul-2008

This report represents data generated after the effective date of the EPA FIFRA Good Laboratory Practice Standards.

United States Environmental Protection Agency
Title 40 Code of Federal Regulations Part 160
FEDERAL REGISTER, August 17, 1989

Organisation for Economic Co-Operation and Development
ENV/MC/CHEM(98)17, Paris January 26, 1998

All aspects of this study were conducted in accordance with the requirements for Good Laboratory Practice Standards, 40 CFR 160 with the following exceptions. The enzymatic peptide mass fingerprinting and MS/MS sequencing analysis were conducted in non-GLP laboratories. The GLP status of the commercial reference standards, such as horseradish peroxidase, soybean trypsin inhibitor, bovine serum albumin and protein molecular weight standards were unknown. The chain of custody of these standards was not monitored.

Penny K. Hunst — 27-Feb-2009

P. L. Hunst
Sponsor
Dow AgroSciences LLC

Date

Penny K. Hunst — 27-Feb-2009

P. L. Hunst
Submitter
Dow AgroSciences LLC

Date

B W Schafer — 20-MAY-2009

B. W. Schafer
Study Director/Author
Dow AgroSciences LLC

Study Completion Date

Dow AgroSciences Quality Assurance Unit
Good Laboratory Practice Statement Page

Study ID: 080142

Title: Characterization of Aryloxyalkanoate Dioxygenase-1 (AAD-1) Protein Derived from Transgenic Maize Event DAS-40278-9

Study Initiation Date: 09-July-2008 **Study Completion Date:** 20-May-2009

GLP Quality Assurance Inspections

Date of GLP Inspection(s)	Date Reported to the Study Director and to Management	Phases of the Study which received a GLP Inspection by the Quality Assurance Unit
27-June-2008	27-June-2008	Protocol Review
22-July-2008	22-July-2008	Immunodiagnostic Strip Test (Dip Stick Test)
05-Aug-2008	05-Aug-2008	Tissue Harvesting
24, 30, 31-Mar-2009	04-April-2009	Raw Data and Report Review; Test Substance Container and Sample Verification

QUALITY ASSURANCE STATEMENT:

The Quality Assurance Unit has reviewed the final study report and has determined that the report reflects the raw data generated during the conduct of this study.

Gaya Wickremsinhe — 20 May 2009

Gaya Wickremsinhe, RQAP-GLP
Dow AgroSciences, Quality Assurance

Date

SIGNATURE PAGE

B. W. Schafer
Author
Dow AgroSciences LLC

10-MAR-2009
Date

S. K. Embrey
Co-Author
Dow AgroSciences LLC

10 - Mar - 2009
Date

R. A. Herman
Peer Reviewer
Dow AgroSciences LLC

10-MARCH-2009
Date

G. Shan
Science Leader
Dow AgroSciences LLC

10-Mar-2009
Date

K. A. Clayton
Global Leader, Biotechnology Regulatory Sciences
Dow AgroSciences LLC

10 MAR 2009
Date

STUDY PERSONNEL

Title: Characterization of the Aryloxyalkanoate Dioxygenase-1 (AAD-1) Protein Derived from Transgenic Maize Event DAS-40278-9

Study Director: B. W. Schafer

Analysts: S. K. Embrey

STUDY PERSONNEL

Title: Characterization of the Aryloxyalkanoate Dioxygenase-1 (AAD-1) Protein Derived from Transgenic Maize Event DAS-40278-9

Study Director: B. W. Schafer

Analysts: S. K. Embrey

TABLE OF CONTENTS

TABLE OF CONTENTS (CONT.)

Characterization of the Aryloxyalkanoate Dioxygenase-1 (AAD-1) Protein Derived from Transgenic Maize Event DAS-40278-9

ABSTRACT

Corn has been modified by the insertion of the aad-1 gene from *Sphingomonas herbicidivorans* which encodes the aryloxyalkanoate dioxygenase (AAD-1) protein. The trait confers tolerance to 2,4-dichlorophenoxyacetic acid and aryloxyphenoxypropionate (commonly referred to as "fop" herbicides such as quizalofop) herbicides which may be used as a selectable marker during plant transformation and in breeding nurseries. Whiskers transformation of corn with plasmid pDAS1740 was carried forward, through breeding, to produce plants containing event AAD-1 DAS-40278-9 which is the focus of this study. The AAD-1 protein is approximately 33 kDa in size.

To perform various toxicology, eco-toxicology, biochemical characterization, and enzymatic activity studies, large quantities of the AAD-1 protein are required. Because it is technically infeasible to extract and purify sufficient amounts of AAD-1 protein from transgenic plants, the protein was produced with a proprietary *Pseudomonas fluorescens* CellCap™ expression system.

The purpose of this study was to characterize the biochemical properties of the recombinant AAD-1 protein derived from the transgenic maize event DAS-40278-9 (event 278) and compare them with the properties of the previously characterized microbe-derived AAD-1 protein (Schafer, 2008 and Kuppannan, et al 2007). Sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS-PAGE, stained with Coomassie blue and glycoprotein detection methods), western blot, immunodiagnostic test strip assays, matrix assisted laser desorption/ionization time-of-flight mass spectrometry (MALDI-TOF MS) and protein sequencing analysis by tandem MS were used to characterize the biochemical properties of the protein. Utilizing these methods, the AAD-1 proteins from P. fluorescens and transgenic maize (event 278) were shown to be biochemically equivalent. These data support the use of the microbial protein in studies supporting the registration of transgenic corn expressing the AAD-1 protein.

ABBREVIATIONS

EDTA	ethylenediaminetetraacetic acid
ESI-LC/MS	electrospray-ionization liquid chromatography/mass spectrometry
kDa	kiloDalton
MALDI-TOF MS	matrix assisted laser desorption/ionization time-of-flight mass spectrometry
µg	microgram
mA	milliAmps
µL	microliter
min	minutes
MS/MS	tandem mass spectrometry
MW	molecular weight
ng	nanogram
N-terminus	amino-terminus
C-terminus	carboxy-terminus
PBST	phosphate buffered saline + 0.05% Tween 20, pH 7.4
PMSF	phenylmethylsulphonyl fluoride
sec	seconds
SDS-PAGE	sodium dodecyl sulfate polyacrylamide gel electrophoresis
TSN	test substance number

INTRODUCTION

Corn has been modified by the insertion of the *aad*-1 gene from *Sphingomonas herbicidivorans* which encodes the aryloxyalkanoate dioxygenase (AAD-1) protein. The trait confers tolerance to 2,4-dichlorophenoxyacetic acid and aryloxyphenoxypropionate (commonly referred to as "fop" herbicides such as quizalofop) herbicides which may be used as a selectable marker during plant transformation and in breeding nurseries. Whiskers transformation of corn with plasmid pDAS1740 was carried forward, through breeding, to produce plants containing event AAD-1 DAS-40278-9, which is the focus of this study. The AAD-1 protein is approximately 33 kDa in size.

To perform various toxicology, eco-toxicology, biochemical characterization, and enzymatic activity studies, large quantities of the AAD-1 protein are required. Because it is technically infeasible to extract and purify sufficient amounts of AAD-1 protein from transgenic plants, the protein was produced with a proprietary *Pseudomonas fluorescens* CellCap™ expression system.

The purpose of this study was to characterize the biochemical properties of the recombinant AAD-1 protein derived from the transgenic maize event DAS-40278-9 and compare them with the properties of the previously characterized microbe-derived AAD-1 protein (Schafer, 2008 and Kuppannan, *et. al,* 2007). Sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS-PAGE, stained with Coomassie blue and glycoprotein detection methods), western blot, immunodiagnostic test strip assay, matrix assisted laser desorption/ionization time-of-flight mass spectrometry (MALDI-TOF MS) and peptide sequence analysis by tandem mass spectrometry were used to characterize the biochemical properties of the protein.

The biochemical and immunological methods employed in this study are among those that have been well established for protein analysis. SDS-PAGE separates proteins based on the apparent molecular weight (mass). Western blotting of proteins to a nitrocellulose membrane following SDS-PAGE, and immunodetection with a protein specific antibody (in addition to immunodiagnostic test strip assays) are widely used to identify the authenticity and quantity of a

molecule in a crude preparation. In addition, staining for carbohydrate moieties linked to polypeptides (following electrophoresis) is a standard test to detect post-translational glycosylation of proteins. Enzymatic peptide mass fingerprinting (by MALDI-TOF MS following trypsinolysis) and peptide sequencing by mass spectrometry are among the most powerful tools for comparing the amino acid sequence of related proteins.

MATERIALS AND METHODS

Test Substances

Immuno-purified, maize-derived AAD-1 protein (molecular weight: ~33 kDa) or crude aqueous extracts from corn stalk tissue. The seeds used were a F1 hybrid line that were hemizygous for the presence of the AAD-1 transgene (event DAS-40278-9; Source ID: ZQ07LQ570715—planted on June 12th and 20th, 2008 and harvested on August 5, 2008). Leaf punches from each individual plant were collected at the Dow AgroSciences greenhouse in Indianapolis, IN and were tested by an AAD-1 specific immunodiagnostic test strip (American Bionostica, Inc., Swedesboro, NJ) to confirm the presence of the AAD-1 protein (see following section for experimental details). After confirmation of the AAD-1 protein, all leaf and stalk tissues were harvested and transported to the laboratory as follows: The leaves were cut from the plant with scissors and placed in cloth bags and stored at approximately -20 °C for future use. Separately, the stalks were cut off just above the soil line, placed in cloth bags and immediately frozen at approximately -80 °C for ~6 hours. The stalks were then placed in a lyophilizer for 5 days to remove water. Once the tissues were completely dried they were ground to a fine powder with dry ice and stored at approximately -80 °C until needed.

Test System

The test system that AAD-1 and other reference or control substances were dissolved in, suspended in or applied to, include water, buffers/solvents, gels, nitrocellulose or PVDF membranes, and purification media depending on the test methods or procedure.

Purification of the AAD-1 Protein from Corn Stalk Tissue

The maize-derived AAD-1 protein was extracted from lyophilized stalk tissue in a phosphate-based buffer (see Table 1 for buffer components) by weighing out ~30 grams of lyophilized tissue into a chilled 1000 mL glass blender (Waring Commercial Laboratory Blender Model #: 51BL30) and adding 500 mL of extraction buffer. The tissue was blended on high for 60 seconds and the soluble proteins were harvested by centrifuging the sample for 20 minutes at 30,000 ×g. The pellet was re-extracted as described, and the supernatants were combined and filtered through a 0.45 µ filter. The filtered supernatants were loaded at approximately +4 °C onto an anti-AAD-1 immunoaffinity column that was conjugated with a monoclonal antibody prepared by Strategic Biosolution Inc. (MAb 473F185.1; Protein A purified; Lot #: 609.03C-2-4; 6.5 mg/mL (~35.2 mg total); Conjugated to CNBr-activated Sepharose 4B (GE Healthcare Cat #: 17-0430-01). The column preparation is documented in DAS Notebook F1257). The non-bound proteins were collected and the column was washed extensively with pre-chilled 20 mM ammonium bicarbonate buffer, pH 8.0. The bound proteins were eluted with 3.5 M NaSCN, (Sigma Cat #: S7757), 50 mM Tris (Sigma Cat #: T3038) pH 8.0 buffer. Seven 5-mL-fractions were collected and fraction numbers 2 → 7 were dialyzed overnight at approximately +4 °C against 10 mM Tris, pH 8.0 buffer. The fractions were examined by SDS-PAGE and western blot and the remaining samples were stored at approximately +4 °C until used for subsequent analyses.

Control Substance

The control substance used in this study was a non-transgenic maize-plant extract from pedigree XHH13. Seeds of the control corn line were planted, grown, and harvested under the same conditions as the transgenic plants described above. Maize leaf punches were harvested at the Dow AgroSciences greenhouse in Indianapolis, IN and each individual plant was tested using an immunodiagnostic test strip (American Bionostica, Inc., Swedesboro, NJ) to confirm the absence of the AAD-1 protein.

Reference Substances

1. Recombinant AAD-1 microbial protein, Batch #: 2 (Lot #: 480-15), molecular weight: ~33 kDa, 36.1% active ingredient (a.i.) AAD-1 to powder mass (Schafer, 2008). The microbial preparation was produced in *Pseudomonas fluorescens* strain DC485 at Dow AgroSciences Bioprocessing R &D facility in San Diego, CA and purified at the Dow AgroSciences Core Biotech R&D facility in Indianapolis, IN (Snodderley, *et al*., 2007). The lyophilized powder was sent to the Test Substance Coordinator at Dow AgroSciences located in Indianapolis. The material was designated TSN105930.

2. The commercially available reference substances used in this study are listed in the following table:

Reference Substance	Product Name	Lot Number	Assay	Reference
Soybean Trypsin Inhibitor	A component of the GelCode Glycoprotein Staining Kit	IA110577	Glycosylation	Pierce Cat #: 1856274
Horseradish Peroxidase	A component of the GelCode Glycoprotein Staining Kit	JG124509	Glycosylation	Pierce Cat #: 1856273
Bovine Serum Albumin Fraction V (BSA)	Pre-Diluted BSA Protein Assay Standard Set	FH71884A	Glycosylation, SDS-PAGE and Western Blot	Pierce Cat #: 23208

Reference Substance	Product Name	Lot Number	Assay	Reference
Prestained Molecular Weight Markers	Novex Sharp Prestained Protein Markers	469212 & 419493	Western Blot	Invitrogen Cat #: LC5800, Molecular Weight Markers of 260, 160, 110, 80, 60, 50, 40, 30, 20, 15, 10 and 3.5 kDa
Molecular Weight Markers	Invitrogen Mark12 Protein Marker Mix	39983 & 399895	SDS-PAGE	Invitrogen Cat #: LC5677, Molecular Weight Markers of 200, 116.3, 97.4, 66.3, 55.4, 36.5, 31.0, 21.5, 14.4, 6.0, 3.5 and 2.5 kDa

SDS-PAGE, Western Blot and Immunodiagnostic Test Strip Assay

To test for the expression of the AAD-1 protein, an immunodiagnostic strip analysis was performed. Four leaf punches were collected from each plant on July 22, 2008 for XHH13 and event DAS-40278-9 by pinching the tissue between the snap-cap lids of individually labeled 1.5-mL microfuge tubes. Upon receipt in the lab, 0.5 mL of AAD-1 extraction buffer (American Bionostica Cat #: 701B30) was added to each tube, and the tissue was homogenized using a disposable pestle followed by shaking the sample for ~10 seconds. After homogenization, the test strip (American Bionostica Cat #: 701V50) was placed in the tube and allowed to develop for ~ 5 minutes. The presence or absence of the AAD-1 protein in the plant extract was confirmed based on the appearance (or lack of appearance) of a test line on the immunodiagnostic strip. Once the expression of the AAD-1 protein was confirmed for the transgenic event, the maize stalk tissue was harvested (on August 5, 2008) and lyophilized and stored at approximately –80 °C until use.

Lyophilized tissue from event DAS-40278-9 and XHH13 stalk (~100 mg) were weighed out in 2-mL microfuge tubes and extracted with ~1 mL of PBST (Sigma Cat #: P3593) containing 10% plant protease inhibitor cocktail (Sigma Cat #: P9599). The extraction was facilitated by adding 4 small ball bearings and Geno-Grinding the sample for 1 minute. After grinding, the samples were centrifuged for 5 minutes at 20,000×g and the supernatants were mixed 4:1 with 5x Laemmli sample buffer (2% SDS, 50 mM Tris pH 6.8, 0.2 mg/mL bromophenol blue, 50% (w/w) glycerol containing 10% freshly added 2-mercaptoethanol) and heated for 5 minutes at

~100 °C. After a brief centrifugation, 45 µL of the supernatant was loaded directly onto a Bio-Rad Criterion SDS-PAGE gel (Bio-Rad Cat #:345-0032) fitted in a Criterion Cell gel module (Cat #: 165-6001). The positive reference standard, microbe-derived AAD-1 (TSN105930), was resuspended at 1 mg/mL in PBST pH 7.4 and further diluted with PBST. The sample was then mixed with Bio-Rad Laemmli buffer (Bio-Rad Cat #: 161-0737 with 5% 2-mercaptoethanol (Bio-Rad Cat #: 161-0710)) and processed as described earlier. The electrophoresis was conducted with Tris/glycine/SDS buffer (Bio-Rad, Cat #: 161-0772) at voltages of 150 - 200 V until the dye front approached the end of the gel. After separation, the gel was cut in half and one half was stained with Pierce GelCode Blue protein stain (Cat #: 24592) and the other half was electro-blotted to a nitrocellulose membrane (Bio-Rad, Cat #:162-0213) with a Mini trans-blot electrophoretic transfer cell (Bio-Rad Cat#: 170-3930) for 60 minutes under a constant voltage of 100 volts. The transfer buffer contained 20% methanol and Tris/glycine buffer from Bio-Rad (Cat #: 161-0771). For immunodetection, the membrane was probed with an AAD-1 specific polyclonal rabbit antibody (Strategic Biosolution Inc., Newark, DE, Protein A purified rabbit polyclonal antibody Lot #: DAS F1197-151, 1.6 mg/mL). A conjugate of goat anti-rabbit IgG (H+L) and alkaline phosphatase (Pierce Chemical, Cat #: 31340) was used as the secondary antibody. Sigma*Fast* BCIP/NBT substrate (Cat #: B5655) was used for development and visualization of the immunoreactive protein bands. The membrane was washed extensively with water to stop the reaction and a record of the results was captured with a digital scanner (Hewlett Packard, Model #: C7670A)

Detection of Post-translational Glycosylation

The immunoaffinity chromatography-purified, maize-derived AAD-1 protein (Fraction #3) was mixed 4:1 with 5x Laemmli buffer. The microbe-derived AAD-1, soybean trypsin inhibitor, bovine serum albumin and horseradish peroxidase were diluted with Milli-Q water to the approximate concentration of the plant-derived AAD-1 and mixed with Bio-Rad Laemmli buffer. The proteins were then heated at ~95 °C for 5 minutes and centrifuged at 20000×g for 2 minutes to obtain a clarified supernatant. The resulting supernatants were applied directly to a Bio-Rad

Criterion Gel (Cat #: 345-0111) and electrophoresed with XT MES running buffer (Bio-Rad, Cat#: 161-0789) essentially as described above except that the electrophoresis was run at 170 V for ~60 minutes. After electrophoresis, the gel was cut in half and one half was stained with GelCode Blue stain for total protein according to the manufacturers' protocol. After the staining was complete, the gel was scanned with a Molecular Dynamics densitometer to obtain a permanent visual record of the gel. The other half of the gel was stained with a GelCode Glycoprotein Staining Kit (Pierce, Cat #: 24562) according to the manufacturers' protocol to visualize glycoproteins. The procedure for glycoprotein staining is briefly described as follows: After electrophoresis, the gel was fixed in 50% methanol for 30 minutes and rinsed with 3% acetic acid. This was followed by an incubation period with the oxidation solution from the staining kit for 15 minutes. The gel was once again rinsed with 3% acetic acid and incubated with GelCode glycoprotein staining reagent for 15 minutes. Finally, the gel was immersed in the reduction solution for 5 minutes, and then rinsed with 3% acetic acid. The glycoproteins (with a detection limit as low as 0.625 ng per band) were visualized as magenta bands on a light pink background. After the glycoprotein staining was complete, the gel was scanned with a Hewlett Packard digital scanner to obtain a permanent visual record of the gel. After the image of the glycosylation staining was captured, the gel was stained with GelCode Blue to verify the presence of the non-glycosylated proteins.

Mass Spectrometry Peptide Mass Fingerprinting and Sequencing of Maize- and Microbe-Derived AAD-1

Mass Spectrometry analysis of the microbe- and maize-derived AAD-1 (TSN105930 and immunopurified Fraction #3, respectively) was conducted at the Analytical Sciences Laboratory of The Dow Chemical Company. The analytical summaries, which contain the methods and results, can be found in the Appendix starting on page 43.

Statistical Treatment of Data

No statistical treatments were used during this study.

RESULTS AND DISCUSSION

Immunodiagnostic Test Strip Assay

The presence of the AAD-1 protein in the leaf tissue of DAS-40278-9 was confirmed using commercially prepared immunodiagnostic test strips from American Bionostica. The strips were able to discriminate between transgenic and nontransgenic plants by testing crude leaf extracts (data not shown). The non-transgenic extracts (XHH13) did not contain detectable amounts of immunoreactive protein. This result was also confirmed by western blot analysis (Figure 1).

SDS-PAGE Analysis

In the *P. fluorescens*-produced AAD-1 (TSN105930), the major protein band, as visualized on Coomassie stained SDS-PAGE gels, was approximately 33 kDa (Figures 1 → 3). As expected, the corresponding maize-derived AAD-1 protein (event DAS-40278-9) was identical in size to the microbe-expressed proteins (Figures 2 and 3). Predictably, the plant purified fractions contained a minor amount of non-immunoreactive impurities in addition to the AAD-1 protein (Figure 2). The co-purified proteins were likely retained on the column by weak interactions with the column matrix or leaching of the monoclonal antibody off of the column under the harsh elution conditions. Other researchers have also reported the non-specific adsorption of peptides and amino acids on cyanogen-bromide activated Sepharose 4B immunoadsorbents (Kennedy and Barnes, 1983; Holroyde *et al*., 1976; Podlaski and Stern, 2008).

Western Blot Analysis

The microbe-derived AAD-1 protein showed a positive signal of the expected size by polyclonal antibody western blot analysis (Figures 1, Panel B). This was also observed in the DAS-40278-9 transgenic maize stalk extract. In the AAD-1 western blot analysis, no immunoreactive proteins

were observed in the control XHH13 extract and no alternate size proteins (aggregates or degradation products) were seen in the transgenic samples (Figure 1, Panel B).

Purification of the AAD-1 Protein from Transgenic Event DAS-40278-9 Extracts

Immunoaffinity chromatography was conducted on lyophilized DAS-40278-9 transgenic corn stalk extract. The protein that bound to the immunoaffinity column was examined by SDS-PAGE and the results showed that the eluted fractions contained the AAD-1 protein at an approximate molecular weight of 33 kDa (Figure 2). In addition, a western blot was also performed and was positive for the AAD-1 protein (data not shown). The maize-derived AAD-1 protein was isolated from ~30 g of lyophilized stalk material.

Detection of Glycosylation

Detection of carbohydrates, possibly covalently linked to maize-derived AAD-1 protein (event DAS-40278-9), was assessed by the GelCode Glycoprotein Staining Kit from Pierce. The immunoaffinity-purified AAD-1 protein was electrophoresed simultaneously with reference protein standards. A glycoprotein, horseradish peroxidase, was loaded as a positive indicator for glycosylation and non-glycoproteins, soybean trypsin inhibitor and bovine serum albumin, were employed as negative controls. The results showed that both the maize- and microbe-derived AAD-1 proteins had no detectable covalently linked carbohydrates (Figure 3). This result was also confirmed by peptide mass fingerprinting (see Appendices).

Enzymatic Peptide Mass Fingerprinting and MS/MS Sequence Analysis

The AAD-1 protein derived from transgenic corn stalk (event DAS-40278-9) was subjected to in-solution digestion by trypsin followed by MALDI-TOF MS and ESI-LC/MS (Karnoup and Kuppannan, 2009, see Appendices). The masses of the detected peptides were compared to those deduced based on potential protease cleavage sites in the sequence of maize-derived AAD-

1 protein. Figure 4 illustrates the theoretical cleavage which was generated *in silico* using Protein Analysis Worksheet (PAWS) freeware from Proteometrics LLC. The predicted amino acid digest (and molecular weights) of the maize-derived AAD-1 protein is also described in Table 2. The AAD-1 protein, once denatured, is readily digested by proteases and will generate numerous peptide peaks.

In the trypsin digest of the transgenic-maize-derived AAD-1 protein (event DAS-40278-9), the peptide coverage was excellent (96.6%). The detected peptide fragments covered nearly the entire protein sequence lacking only one small tryptic fragment at the C-terminal end of the protein, F^{248} to R^{253} and one short (2 amino acids) peptide fragment (Figure 5). This analysis confirmed the maize-derived protein amino acid sequence matched that of the microbe-derived AAD-1 protein. In the LC/MS chromatograms, there were several unidentified peptides detected in the tryptic digest preparations (data not shown). Many factors contribute to the formation of these unidentified peptides, such as over digestion (which results in non-specific cleavage), self-digestion products of trypsin, and random breakage of peptides during ionization. Unidentified peptides are expected and do not indicate the protein is different from the predicted sequence. Results of these analyses indicate that the amino acid sequence of the maize-derived AAD-1 protein was equivalent to the *P. fluorescens*-expressed protein characterized earlier (Kuppannan, *et al.,* 2007 and Schafer 2008).

Tryptic Peptide Fragment Sequencing

In addition to the peptide mass fingerprinting, the amino acid residues at the N- and C-termini of the maize-derived AAD-1 protein (immunoaffinity purified from maize event DAS-40278-9) were sequenced and compared to the sequence of the microbe-derived protein (Kuppannan *et al*., 2007, Appendices). The protein sequences were obtained, by tandem mass spectrometry, for the first 11 residues of the microbe- and maize-derived proteins (Tables 3 - 6). The amino acid sequences for both proteins were A^{1} H A A L S P L S Q R^{11} showing the N-terminal methionine had been removed by an aminopeptidase (Tables 3, 5 and 6). The N-terminal AAD-1 protein sequence was expected to be M^{1} A H A A L S P L S Q R^{12}. These results

suggest that during or after translation in maize and *P. fluorescens*, the N-terminal methionine is cleaved by a methionine aminopeptidase (MAP). MAPs cleave methionyl residues rapidly when the second residue on the protein is small, such as Gly, Ala, Ser, Cys, Thr, Pro, and Val (Walsh, 2006). In addition to the methionine being removed, a small portion of the N-terminal peptide of the AAD-1 protein was shown to have been acetylated after the N-terminal methionine was cleaved (Tables 4 and 6). This result is encountered frequently with eukaryotic (plant) expressed proteins since approximately 80-90% of the N-terminal residues are modified (Polevoda and Sherman, 2003). Also, it has been shown that proteins with serine and alanine at the N-termini are the most frequently acetylated (Polevoda and Sherman, 2002). The two cotranslational processes, cleavage of N-terminal methionine residue and N-terminal acetylation, are by far the most common modifications and occur on the vast majority (~85%) of eukaryotic proteins (Polevoda and Sherman, 2002). However, examples demonstrating biological significance associated with N-terminal acetylation are rare (Polevoda and Sherman, 2000).

In addition to N-acetylation, there was also slight N-terminal truncation that appeared during purification of the maize-derived AAD-1 protein (Table 6 and Figure 5). These "ragged-ends" resulted in the loss of amino acids A^2, H^3 and A^4 (in varying forms and amounts) from the maize-derived protein. This truncation is thought to have occurred during the purification of the AAD-1 protein as the western blot probe of the crude leaf extracts contained a single crisp band at the same MW as the microbe-derived AAD-1 protein. The extraction buffer for the western blotted samples contained an excess of a protease inhibitor cocktail which contains a mixture of protease inhibitors with broad specificity for the inhibition of serine, cysteine, aspartic, and metalloproteases, and aminopeptidases.

The C-terminal sequence of the maize- and microbe-derived AAD-1 proteins were determined as described above and compared to the expected amino acid sequences (Tables 7 – 9 and Figure 5). The results indicated the measured sequences were identical to the expected sequences, and both the maize- and microbe-derived AAD-1 proteins were identical and unaltered at the C-terminus.

CONCLUSIONS

It was demonstrated that the biochemical identity of *P. fluorescens*-produced AAD-1 protein was equivalent to the protein purified from stalk tissue of maize event DAS-40278-9 (except for the addition of a 42 Dalton acetyl group to the N-terminus of a small subset of the maize-derived protein and slight N-terminal truncation that appeared during purification). Both the maize- and microbe-derived AAD-1 proteins had an apparent molecular weight of ~33 kDa in crude leaf extracts and were immunoreactive to AAD-1 protein-specific antibodies in diagnostic test strip and western blot assays. The amino acid sequence of both proteins was confirmed by enzymatic peptide mass fingerprinting by MALDI-TOF MS and peptide sequence obtained from tandem mass spectrometry (>94% coverage). In addition, the lack of glycosylation of the maize-derived AAD-1 protein provided additional evidence that the AAD-1 protein produced by *P. fluorescens* and transgenic corn were biochemically equivalent molecules.

ARCHIVING

The original copy of the final report, protocol and all raw data records are filed in the Dow AgroSciences LLC archives at 9330 Zionsville Road in Indianapolis, IN 46268-1054. The raw data for the mass spectrometry characterization of the maize- and microbe-derived AAD-1 proteins is located in the Dow Chemical Company Analytical Sciences archives in the 1897 building, Midland, MI 48667.

REFERENCES

Holroyde, M. J., Chesher, J. M. E., Trayer, I. P., and Walker, D. G. Studies on the Use of Sepharose –*N*-(6-Aminohexanoyl)-2-Amino-2-Deoxy-D-Glucopyranose for the Large-Scale Purification of Hepatic Glucokinase. *Biochem. J.* **1976**. 153, 351-361.

Karnoup, A. S., and Kuppannan, K. Characterization of maize-derived AAD-1, Event # DAS-40278-9. **2009,** Laboratory Report Code: ML-AL MD-2008-006360, unpublished report of The Dow Chemical Company.

Kennedy J. F. and Barnes J. A. Immunochemical studies of the non-specific interactions of cyanogen bromide-activated macroporous agarose-based immunoadsorbents. *J Chromatogr*. **1983**. 281, 83-93.

Kuppannan, K., Hung, S. C., and Dielman, D. Characterization of AAD-1: Batch #1 480-14 and Batch #2 480-15. **2007**, Laboratory Report Code: ML-AL MD-2007-000127, unpublished report of The Dow Chemical Company.

Podlaski, F. J. and Stern, S. S. Site-Specific Immobilization of Antibodies to Protein G-Derivatized Solid Supports. **2008**, Methods in Molecular Biology, vol 147: Affinity Chromatography: Methods and Protocols. Humana Press., Totowa, NJ

Polevoda, B. and Sherman, F. N^{α}-terminal Acetylation of Eukaryotic Proteins. **2000**, *J. Biol. Chem*. 275:47, 36479–36482

Polevoda, B. and Sherman, F. The Diversity of Acetylated Proteins. **2002**, *Genome Biology*, 3(5):0006.1–0006

Polevoda, B. and Sherman, F. N-terminal Acetyltransferases and Sequence Requirements for N-terminal Acetylation of Eukaryotic Proteins, **2003**, *J. Mol. Biol.* 325, 595-622

REFERENCES (CONTINUED.)

Schafer, B. W. Certificate of Analysis for Test/Reference/Control Substances AAD-1 Batch #1 (TSN106003) and Batch #2 (TSN105930). Study Number: 08-162945, **2008**, unpublished report of Dow AgroSciences LLC.

Snodderley, E., Allen, J., Chew, L., Chiu, Y., Espinosa, O., Lin, G., Thomas, T., Mun, D., Patkar, A., and Shields, J. Production, Purification, and Analysis of AAD-1 Toxicology Lots TSN 106003 (Batch 1) and TSN 105930 (Batch 2). Study Number 070313, **2007**, unpublished report of Dow AgroSciences LLC.

Walsh, C. T. Posttranslational Modification of Proteins: Expanding Nature's Inventory, **2006,** Roberts & Company Publishers, ISBN: 0-9747077-3-2

Table 1. Maize-Derived AAD-1 Extraction Buffer Composition

Amount	Ingredient	Catalog Numbers
One foil packet	Phosphate Buffered Saline with 0.5% Tween[a] 20, pH 7.4	Sigma Cat #: P3563
2.0 mL	0.5 M EDTA	AccuGENE Cat #: 51234
1.0 mL	Triton X-100[b]	LabChem Cat #: LC26280-1
1.0 g	Octyl β-D-glucopyranoside	Sigma Cat #: O8001
To 1000 mL	Milli-Q H_2O	NA
Added to 500 mL of extraction buffer[c] immediately before extraction		
1.0 mL	0.5 M PMSF	Sigma Cat #: P7626
0.7 mL	β-mercaptoethanol	Bio-Rad Cat#: 161-0710

Notes:

a. Tween 20 = Polyoxyethylene (20) sorbitan monolaurate

b. Triton X-100 = 4-octylphenol polyethoxylate

c. The buffer was chilled to 4 ºC prior to extraction.

Table 2. A Summary of Peptide Mass Fingerprinting of Maize-Derived AAD-1(event DAS-40278-9) Based on MALDI MS and ESI/LC/MS

Fragment #	# of missed cleavages	Start	End	Sequence	Charge State	Theoretical monoisotopic m/z	Retention Time, min	Observed m/z LC-MS	Observed in-source peptide fragments	% ACN (elution from C18-zt; MALDI)	Observed m/z MALDI-MS	Modification	Comment
T1-Met	0	1	12	(-) MAHAALSPLSQR (F)	1	1281.67	ND	ND	ND	ND	ND	un-processed Met[1]	Met[1] is absent
	0	1	12		2	641.34							
	0	1	12		3	427.90							
T1	0	2	12	(-) AHAALSPLSQR (F)	1	1150.63	21.14	1150.64	942.54 (1+, y9), 871.50 (1+, y8), 848.41 (1+, b9), 800.46 (1+, y7), 761.42 (1+, b8), 687.38 (1+, y6), 648.34 (1+, b7), 630.36 (1+, b7-18), 600.34 (1+, y5), 551.30 (1+, b6), 533.28 (1+, b6-18), 503.30 (1+, y4), 464.26 (1+, b5), 390.21 (1+, y3), 351.18 (1+, b4)	5%, 20%, 25%	1150.69		sequence confirmed by MS/MS; **31% by UV, 32% by TIC**
	0	2	12		2	575.82		575.82					
T1 (N-Ac)	0	2	12	(-) Ac-AHAALSPLSQR (F)	1	1192.64	27.44	1192.64	942.51 (1+, y9), 871.49 (1+, y8), 800.46 (1+, y7), 687.38 (1+, y6), 600.34 (1+, y5), 593.31 (1+, b6), 575.30 (1+, b6-18), 540.30 (2+, y10), 506.27 (1+, b5), 393.20 (1+, b4), 390.22 (1+, y3), 365.17 (1+, a4)	5%, 20%	1192.75	N-Acetylated Ala	sequence confirmed by MS/MS; **3% by UV, 7% by TIC**
	0	2	12		2	596.83		596.83					
T1-cut1	0	5	12	(-?)(A?) ALSPLSQR (F)	1	871.50	22.56	871.50	687.38 (1+, y6), 600.34 (1+, y5), 503.29 (1+, y4), 390.21 (1+, y3)			3-aa truncation	possible: (a) truncation of N-terminus, (b) non-specific tryptic cleavage; **12% by UV, 12% by TIC**
	0	5	12		2	436.25		436.25					
T1-cut2	0	4	12	(-?)(H?) AALSPLSQR (F)	1	942.54	24.27	942.54	781.46 (1+, y8), 800.47 (1+, y7), 687.38 (1+, y6), 600.35 (1+, y5), 503.29 (1+, y4), 390.21 (1+, y3)			2-aa truncation	possible: (a) truncation of N-terminus, (b) non-specific tryptic cleavage; **6% by UV, 9% by TIC**
	0	4	12		2	471.77		471.77					
T1-cut3	0	3	12	(-?)(A?) HAALSPLSQR (F)	1	1079.60	20.71	1079.61	942.53 (1+, y9), 871.49 (1+, y8), 800.45 (1+, y7), 687.38 (1+, y6), 600.34 (1+, y5), 503.29 (1+, y4), 480.26 (1+, b5), 393.22 (1+, b4), 390.21 (1+, y3)	5%, 20%	1079.72	1-aa truncation	possible: (a) truncation of N-terminus, (b) non-specific tryptic cleavage; Coelutes with T16 peptide: if 1:1 amount in LC peak assumed -- **50% by UV, 40% by TIC**
	0	3	12		2	540.30		540.31					
T2	0	13	15	FER	1	451.23	5.50	451.22					
T3-nonspec	0	16	21	IAVQPL	1	640.39	32.74	640.39	509.30 (1+, b5), 412.26 (1+, b4), 357.22 (1+, y3)				non-specific tryptic cleavage
T3	0	16	35	IAVQPLTGVLGAEITGVDLR	1	2022.16	50.40		1362.79 (1+, b14), 1249.72 (1+, b13), 1143.63 (1+, y11), 1120.66 (1+, b12), 1049.60 (1+, b11), 1030.56 (1+, y10), 992.61 (1+, b10), 973.51 (1+, y9), 902.50 (1+, y8), 879.54 (1+, b9), 810.48 (2+, b17), 805.96 (2+, y16), 780.45 (1+, b8), 773.46 (1+, y7), 732.43 (2+, b15), 723.43 (1+, b7), 660.37 (1+, y6), 625.36 (1+, b13), 622.39 (1+, b6), 559.32 (1+, y5), 525.32 (2+, b11), 515.79 (2+, y10), 502.30 (1+, y4), 496.82 (2+, b10), 487.26 (2+, y9), 451.76 (2+, y8), 440.27 (2+, b9), 412.26 (1+, b4), 403.23 (1+, y3), 387.23 (2+, y7)	5%, 20%, 25%, 40%, 50%, 70%, 100%	2022.13		
	0	16	35		2	1011.58		1011.58					
	0	16	35		3	674.73		674.73					
T4-nonspec	0	36	53	EPLDDSTWNEILDAFHTY	1	2165.97	56.53		1300.56 (1+, b11), 1187.49 (1+, b10), 1058.47 (1+, b9), 992.96 (2+, b17), 979.50 (1+, y8), 944.41 (1+, b8), 942.44 (2+, b16), 873.91 (2+, b15), 866.41 (1+, y7), 800.38 (2+, b14), 753.33 (1+, y6), 707.34 (2+, b12), 650.80 (2+, b11), 638.30 (1+, y5), 594.25 (2+, b10), 567.26 (1+, y4), 420.20 (1+, y3)				non-specific tryptic cleavage
	0	36	53		2	1083.49		1083.49					
	0	36	53		3	722.66		722.67					
T4	0	36	73	EPLDDSTWNEILDAFHTYQVIYFPGQAIT NEQHIAFSR	1	4466.14	59.57		990.01 (2+, y17), 933.45 (3+, b23), 834.93 (2+, y15), 830.07 (3+, b21), 792.38 (3+, b20), 658.35 (2+, y11), 660.35 (3+, y17), 605.99 (3+, y16), 556.96 (3+, y15)	40%, 50%, 70%, 100%	4466.11		
	0	36	73		2	2233.57							
	0	36	73		3	1489.38		1489.42					
	0	36	73		4	1117.29		1117.32					
	0	36	73		5	894.03		894.06					
T6	0	75	85	FGPVDPVPLLK	1	1181.69	43.10	1181.70	1034.62 (1+, y10), 1035.59 (1+, b10), 977.59 (1+, y9), 922.49 (1+, b9), 880.54 (1+, y8), 781.49 (1+, y7), 712.37 (1+, b7), 666.45 (1+, y6), 613.32 (1+, b6), 569.35 (1+, y5), 516.25 (1+, b5), 489.3 (2+, y9), 470.33 (1+, y4), 401.22 (1+, b4), 373.23 (1+, y3)	20%, 25%, 40%, 100%	1181.70		
	0	75	85		2	591.35		591.34					
T7-nonspec	0	86	94	SIEGYPEVQ	1	1021.48	26.42	1021.49	875.43 (1+, b8), 776.36 (1+, b7), 550.26 (1+, b5), 472.24 (1+, y4)				non-specific tryptic cleavage
	0	86	94		2	511.25		ND					
T7	0	86	97	SIEGYPEVQMIR	1	1421.71	34.46	1421.70		5%, 20%, 25%, 40%, 50%	1421.75		
	0	86	97		2	711.36		711.36					
T9	0	99	105	EANESGR	1	762.34	4.21	762.34		50%	762.51		
T10	0	106	128	VIGDDWHTDSTFLDAPPAAVVMR	1	2513.21	43.86		1673.74 (1+, b15), 1602.69 (1+, b14), 1487.68 (1+, b13), 1374.60 (1+, b12), 1227.49 (1+, b11), 1151.03 (2+, y21), 1139.61 (1+, y11), 1104.50 (2+, b21), 1055.02 (2+, b20), 1026.54 (1+, y10), 1005.50 (2+, b19), 969.96 (2+, b18), 911.51 (1+, y9), 885.90 (2+, b16), 837.38 (2+, b15), 801.86 (2+, b14), 744.37 (2+, b13), 743.42 (1+, y7), 687.81 (2+, b12), 575.33 (1+, y5), 570.31 (2+, y11), 504.30 (1+, y4), 456.26 (2+, y9), 420.74 (2+, y8), 405.23 (1+. y3)	20%, 25%, 40%, 50%, 70%, 100%	2513.19		trace amount of Met127 oxydation also observed
	0	106	128		2	1257.11		1257.12					
	0	106	128		3	838.41		838.40					
	0	106	128		4	629.06		629.06					

Table 2 (Cont.) A Summary of Peptide Mass Fingerprinting of Maize-Derived AAD-1 (event DAS-40278-9) Based on MALDI MS and ESI/LC/MS

Fragment #	# of missed cleavages	Start	End	Sequence	Charge State	Theoretical monoisotopic m/z	Retention Time, min	Observed m/z LC-MS	Observed in-source peptide fragments	% ACN (elution from C18-zt; MALDI)	Observed m/z MALDI-MS	Modification	Comment
T11	0	129	170	AIDVPEHGGDTGFLSMYTAWETLSPTMQATIEGLNVVHSATR	1	4503.16	57.61		1290.11 (2+, b24), 1271.15 (2+, y23), 1246.57 (2+, b23), 1190.03 (2+, b22), 1178.11 (2+, y22), 1113.59 (2+, y21), 1063.07 (2+, y20), 1006.52 (2+, y19), 981.98 (2+, b19), 963.01 (2+, y18), 860.40 (3+, b24), 814.40 (2+, b16), 748.86 (2+, b15), 705.35 (2+, b14), 671.35 (3+, y19), 648.81 (2+, b13), 642.34 (3+, y18),	40%, 50%, 70%, 100%	4503.12		trace amount of Met oxydation also observed
	0	129	170		2	2252.08							
	0	129	170		3	1501.73		1501.80					
	0	129	170		4	1126.55		1126.55					
	0	129	170		5	901.44		901.42					
T12	0	171	181	VFGSLYQAQNR	1	1282.65	27.89	1282.66	1108.54 (1+, b10), 1036.52 (1+, y9), 994.49 (1+, b9), 979.49 (1+, y8), 892.47 (1+, y7), 795.40 (1+, b7), 779.38 (1+, y6), 616.32 (1+, y5), 592.30 (2+, y10), 518.77 (2+, y9), 504.28 (1+, b5), 488.26 (1+, y4), 417.22 (1+, y3), 391.20 (1+, b4)	5%, 20%, 25%, 40%, 70%	1282.62		
	0	171	181		2	641.83		641.83					
T14	0	183	189	FSNTSVK	1	782.40	11.87	782.41	635.34 (1+, y6), 548.31 (1+, y5), 434.26 (1+, y4)				
	0	183	189		2	391.71		391.71					
T15	0	190	198	VMDVDAGDR	1	977.44	17.67	977.44	878.37 (1+, y8), 747.33 (1+, y7), 688.33 (1+, b7), 632.30 (1+, y6), 560.29 (1+, b5), 533.23 (1+, y5), 445.22 (1+, b4), 418.21 (1+, y4)	5%	977.56		
	0	190	198		2	489.22		489.22					
T16	0	199	213	ETVHPLVVTHPGSGR	1	1585.85	20.80	1585.86	1119.59 (1+, y11), 1113.56 (1+, b10), 1022.56 (1+, y10), 976.54 (1+, b9), 958.53 (1+, b9-18), 909.49 (1+, y9), 758.42 (1+, b7-18), 711.35 (1+, y7), 659.35 (b6-18), 610.31 (1+, y6), 560.32 (2+, y11), 511.79 (2+, y10), 486.27 (3+, y14), 473.25 (1+, y5), 455.25 (2+, y9), 452.60 (3+, y13), 449.22 (1+, b4-18), 405.71 (2+, y8), 376.20 (1+, y4), 356.18 (2+, y7)	5%, 20%, 25%, 50%	1585.88		
	0	199	213		2	793.43		793.43					
	0	199	213		3	529.29		529.29					
T18	0	215	225	GLYVNQVYCQR	1	1399.68				5%, 20%	1399.82, 1400.56		possible Asn219 deamidation
T18-deamid	0	215	225	GLYV-isoD-QVYCQR	1	1400.66	38.72	1400.79	1067.59 (1+, y8)	5%, 20%	1399.82, 1400.56		Asn219 deamidation
	0	215	225		2	700.84		700.88					
T19	0	226	247	IEGMTDAESKPLLQFLYEHATR	1	2549.27	44.15		1036.51 (1+, y8), 952.01 (2+, y16), 916.50 (2+, y15), 812.75 (3+, y21), 769.73 (3+, y20), 776.37 (1+, y6), 744.40 (2+, y12)	20%, 25%, 40%, 50%, 70%, 100%	2549.24		
	0	226	247		2	1275.14		1275.15					
	0	226	247		3	850.43		850.43					
	0	226	247		4	638.07		638.07					
T19-KP	0	226	235	IEGMTDAESK	1	1080.49	16.60	1080.49	967.41 (1+, y9), 949.38 (1+, y9-18), 838.37 (1+, y8), 781.36 (1+, y7), 650.30 (1+, y6)				K-P cleavage
	0	226	235		2	540.75		**540.75**					
T20	0	248	253	FDFTCR	1	845.36	**ND**	**ND**	**ND**	**ND**	**ND**		
T24	0	259	272	DQVLVWDNLCTMHR	1	1786.84	**ND**	**ND**	**ND**	**ND**	**ND**		
T25	0	273	280	AVPDYAGK	1	820.42	15.75	820.43	650.31 (1+, y6), 553.27 (1+, y5), 438.24 (1+, y4)	5%, 50%	820.38		
T27	0	283	286	YLTR	1	552.31	10.72	552.31	389.25 (1+, y3)	5%	552.37		
T28	0	287	296	(R) TTVGGVRPAR (-)	1	1013.59	12.41	1013.58	695.40 (1+, y7-17), 482.27 (y4-17), 456.77 (2+, y9), 406.25 (2+, y8), 356.71 (2+, y7)	5%, 20%	1013.66		sequence confirmed by MS/MS
	0	287	296		2	507.30		**507.29**					
T3-4	1	16	73	IAVQPLTGVLGAEITGVDLREPLDDSTWNEILDAFHTYQVIYFPGQAITNEQHIAFSR	1	6473.24	63.24			50%, 70%, 100%	6473.24		
	1	16	73		2	3237.13							
	1	16	73		3	2158.42							
	1	16	73		4	1619.07		1619.18					
	1	16	73		5	1295.45		1295.37					
T5-6	1	74	85	RFGPVDPVPLLK	1	1337.79	38.69	1337.79	868.45 (1+, b8), 781.48 (1+, y7), 769.39 (1+, b7), 672.35 (1+, b6), 666.45 (1+, y6), 569.40 (1+, y5), 557.32 (1+, b5), 470.33 (1+, b4), 458.26 (1+, b4), 361.20 (1+, b3)	20%, 25%, 40%, 50%, 70%	1337.83		
	1	74	85		2	669.40		669.40					
	1	74	85		3	446.60		446.60					
T7-8	1	86	98	SIEGYPEVQMIRR	1	1577.81				40%, 50%, 70%	1577.84		
T8-9	1	98	105	REANESGR	1	918.44	3.87	918.38					
	1	98	105		2	459.72		**459.73**					
T13-14	1	182	189	RFSNTSVK	1	938.51	9.11	938.51	792.40 (1+, b7), 764.42 (1+, y7-18), 693.33 (1+, b6), 675.32 (1+, b6-18), 606.30 (1+, b5)	5%	938.61		
	1	182	189		2	469.76		**469.76**					
T17-18	1	214	225	KGLYVNQVYCQR	1	1527.77	34.32	1527.73					trace amount
	1	214	225		2	764.39		764.38					
T21-22	1	254	257	VRWK	1	588.36	22.27	588.31	489.24 (1+, y3)				
T24-25	1	259	280	DQVLVWDNLCTMHRAVPDYAGK	1	2588.24	44.11						trace amount
	1	259	280		2	1294.62		1294.62					
	1	259	280		3	863.42		863.42					
T27-28	1	283	296	(R) YLTRTTVGGVRPAR (-)	1	1546.88	45.92	1546.88					
	1	283	296		2	773.94		773.92					
	1	283	296		3	516.30		516.28					

Table 3. Amino acid sequence obtained for N-terminal peptide (m/z 575.78; retention time 21.19 min) of maize-derived AAD-1 (event DAS-40278-9).

Sequence: AHAALSPLSQR

Fragment ion masses: monoisotopic

Peptide mass $[M+2H]^{2+}$(monoisotopic): 575.82

Ion Table

	A	**H**	**A**	**A**	**L**	**S**	**P**	**L**	**S**	**Q**	**R**
a(+1)	44.05	181.11	252.15	323.18	436.27	523.30	620.35	733.44	820.47	948.53	
		181.12		**323.20**	**436.28**	**523.33**					
b(+1)	72.04	209.10	280.14	351.18	464.26	551.29	648.35	761.43	848.46	976.52	
		209.11	**280.15**	**351.19**	**464.27**	**551.30**					
y(+1)		1079.60	942.54	871.50	800.46	687.38	600.35	503.29	390.21	303.18	175.12
			942.55	**871.52**	**800.48**	**687.38**	**600.36**	**503.29**	**390.22**	**303.19**	**175.13**
y(+2)		540.30	471.77	436.25	400.74	344.19	300.68	252.15	195.61	152.09	88.06
		540.31									
y-NH_3 (+1)		1062.57	925.51	854.47	783.44	670.35	583.32	486.27	373.18	286.15	158.09
						670.36	**583.33**	**486.24**	**373.20**	**286.16**	**158.10**

Top: theoretical m/z of fragment ions
Bottom: observed m/z of fragment ions

Table 4. Amino acid sequence obtained for N-acetylated N-terminal peptide (m/z 596.83; retention time 27.47 min) of maize-derived AAD-1 (event DAS-40278-9).

Sequence: N-Acetyl-AHAALSPLSQR

Fragment ion masses: monoisotopic

Peptide mass $[M+2H]^{2+}$(monoisotopic): 596.83

Ion Table

	N-Ac-A	**H**	**A**	**A**	**L**	**S**	**P**	**L**	**S**	**Q**	**R**
a^{1+}	86.06	223.12	294.16	365.19	478.28	565.31	662.36	775.45	862.48	990.54	
		223.13	**294.16**	**365.21**	**478.26**	**565.31**					
b^{1+}	114.06	251.11	322.15	393.19	506.27	593.30	690.36	803.44	890.47	1018.53	
		251.13	**322.16**	**393.21**	**506.29**	**593.29**					
y^{1+}		1079.60	942.54	871.50	800.46	687.38	600.35	503.29	390.21	303.18	175.12
			942.55	**871.51**	**800.48**	**687.40**	**600.36**	**503.30**	**390.23**		**175.13**
y^{2+}		540.30	471.77	436.25	400.74	344.19	300.68	252.15	195.61	152.09	88.06
(y-NH		1062.57	925.51	854.47	783.44	670.35	583.32	486.27	373.18	286.15	158.09
						670.36	**583.34**	**486.30**	**373.20**	**286.16**	**158.10**

Top: theoretical m/z of fragment ions
Bottom: observed m/z of fragment ions

Table 5. Amino Acid Sequence Obtained for N-terminal Tryptic Peptide (1-11) m/z 575.82 of Microbe-Derived AAD-1 (sample Batch #2: 480-15).

Sequence: AHAALSPLSQR

Fragment ion masses: monoisotopic

Peptide mass $(MH)^{+2}$ (average): 576.16

Peptide mass $(MH)^{+2}$ (monoisotopic): 575.82

Ion Table

	A	**H**	**A**	**A**	**L**	**S**	**P**	**L**	**S**	**Q**	**R**
a(+1)	44.05	181.11	252.15	323.18	436.27	523.30	620.35	733.44	820.47	948.53	
		181.10		**323.19**	**436.21**	**523.25**					
b(+1)	72.04	209.10	280.14	351.18	464.26	551.29	648.35	761.43	848.46	976.52	
		209.09	**280.13**	**351.16**	**464.23**	**551.26**					
y(+1)		1079.60	942.54	871.50	800.46	687.38	600.35	503.29	390.21	303.18	175.12
			942.47	**871.44**	**800.42**	**687.33**	**600.35**	**503.25**	**390.18**	**303.16**	**175.11**
y-H_2O(+1)		1061.59	924.53	853.49	782.45	669.37	582.34	485.28	372.20		
						669.34					
y-NH_3(+1)		1062.57	925.51	854.47	783.44	670.35	583.32	486.27	373.18	286.15	158.09
						670.35	**583.28**		**373.15**		

Top mass: theoretical product fragment ions
Bottom mass: observed product fragment ions

Table 6. Summary of N-terminal Sequence Data of AAD-1 Maize- and Microbe-Derived Proteins

Source	Expected N-terminal Sequence[1]
P. fluorescens	M^{1} A H A A L S P L S Q R^{12}
Maize Event DAS-40278-9	M^{1} A H A A L S P L S Q R^{12}

Source	Detected N-terminal Sequence[2]	Relative[3] Abundance
P. fluorescens	A H A A L S P L S Q R^{12}	100%
Maize Event DAS-40278-9	A H A A L S P L S Q R^{12}	31%
Maize Event DAS-40278-9	$^{N\text{-}Ac}$A H A A L S P L S Q R^{12}	3%
Maize Event DAS-40278-9	H A A L S P L S Q R^{12}	50%
Maize Event DAS-40278-9	A A L S P L S Q R^{12}	6%
Maize Event DAS-40278-9	A L S P L S Q R^{12}	12%

[1]Expected N-terminal sequence of the first 12 amino acid residues of *P. fluorescens*- and maize-derived AAD-1.

[2]Detected N-terminal sequences of *P. fluorescens*- and maize-derived AAD-1.

[3]The tandem MS data for the N-terminal peptides revealed a mixture of AHAALSPLSQR (acetylated) and *N-Acetyl*-AHAALSPLSQR (acetylated). "Ragged N-terminal ends" were also detected (peptides corresponding to amino acid sequences HAALSPLSQR, AALSPLSQR, and ALSPLSQR). The relative abundance, an estimate of relative peptide fragment quantity, was made based on the corresponding LC peak areas measured at 214 nm.

Notes:

Numbers in superscript (R^{x}) indicate amino acid residue numbers in the sequence.

Amino acid residue abbreviations:

A:	alanine	H:	histidine
L:	leucine	M:	methionine
P:	proline	Q:	glutamine
R:	arginine	S:	serine
T:	threonine		

Table 7. Amino acid sequence obtained for C-terminal peptide (m/z 507.3; retention time 12.3 min) of maize-derived AAD-1 (event DAS-40278-9).

Sequence: TTVGGVRPAR

Fragment ion masses: monoisotopic

Peptide mass $[M+2H]^{2+}$(monoisotopic): 507.3

Ion Table

	T	**T**	**V**	**G**	**G**	**V**	**R**	**P**	**A**	**R**
a(+1)	74.06	175.11	274.18	331.2	388.22	487.29	643.39	740.44	811.48	
		175.12					**643.40**		**811.49**	
b-H_2O (+1)	84.04	185.09	284.16	341.18	398.2	497.27	653.37	750.43	821.46	
		185.10								
b(+1)	102.06	203.10	302.17	359.19	416.21	515.28	671.38	768.44	839.47	
		203.11							**839.47**	
y(+1)		912.54	811.49	712.42	655.4	598.38	499.31	343.21	246.16	175.12
			811.49	**712.42**	**655.40**		**499.32**	**343.22**		
y (+2)		456.77	406.25	356.71	328.20	299.69	250.16	172.11	123.58	88.06
		456.78	**406.26**							
y-H_2O(+1)		894.53	---	---	---	---	---	---	---	---
y-NH_3 (+1)		895.51	794.46	695.4	638.37	581.35	482.28	326.18	229.13	158.09
		895.58	**794.44**	**695.41**	**638.39**	**581.37**	**482.30**	**326.19**	**229.14**	

Top: theoretical m/z of fragment ions
Bottom: observed m/z of fragment ions

Table 8. Amino Acid Sequence Obtained for C-terminal Tryptic Peptide (286-295) m/z 507.3 of Microbe-Derived AAD-1 (sample Batch #2: 480-15).

Sequence: TTVGGVRPAR

Fragment ion masses: monoisotopic

Peptide mass $(MH)^{+2}$ (average): 507.59

Peptide mass $(MH)^{+2}$ (monoisotopic): 507.3

Ion Table

	T	**T**	**V**	**G**	**G**	**V**	**R**	**P**	**A**	**R**
a(+1)	74.06	175.11	274.18	331.2	388.22	487.29	643.39	740.44	811.48	
		175.12							**811.50**	
b(+1)	102.06	203.10	302.17	359.19	416.21	515.28	671.38	768.44	839.47	
		203.11					**671.36**		**839.47**	
y(+1)		912.54	811.49	712.42	655.4	598.38	499.31	343.21	246.16	175.12
			811.50	**712.44**	**655.40**		**499.33**	**343.21**		**175.12**
y-H_2O(+1)										
y-NH_3(+1)		895.51	794.46	695.4	638.37	581.35	482.28	326.18	229.13	158.09
			794.46	**695.40**	**638.38**	**581.36**	**482.30**	**326.18**		

Top mass: theoretical product fragment ions
Bottom mass: observed product fragment ions

Table 9. Summary of C-terminal Sequence Data of AAD-1 Maize- and Microbe-Derived Proteins

Source	**Expected C-terminal Sequence[1]**
P. fluorescens	^{287}T T V G G V R P A R^{296}
Maize Event DAS-40278-9	^{287}T T V G G V R P A R^{296}

Source	**Detected C-terminal Sequence[2]**
P. fluorescens	^{287}T T V G G V R P A R^{296}
Maize Event DAS-40278-9	^{287}T T V G G V R P A R^{296}

[1]Expected C-terminal sequence of the last 10 amino acid residues of *P. fluorescens*- and maize-derived AAD-1.

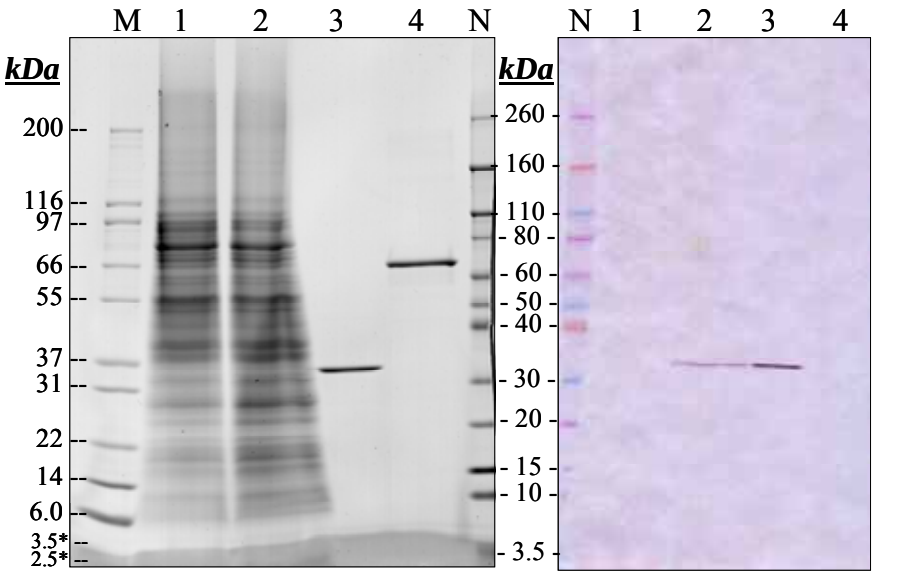
[2]Detected C-terminal sequences of *P. fluorescens*- and maize-derived AAD-1.

Notes:

Numbers in superscript (R^x) indicate amino acid residue numbers in the sequence.

Amino acid residue abbreviations:

A:	alanine	G:	glycine
P:	proline	R:	arginine
T:	threonine	V:	valine

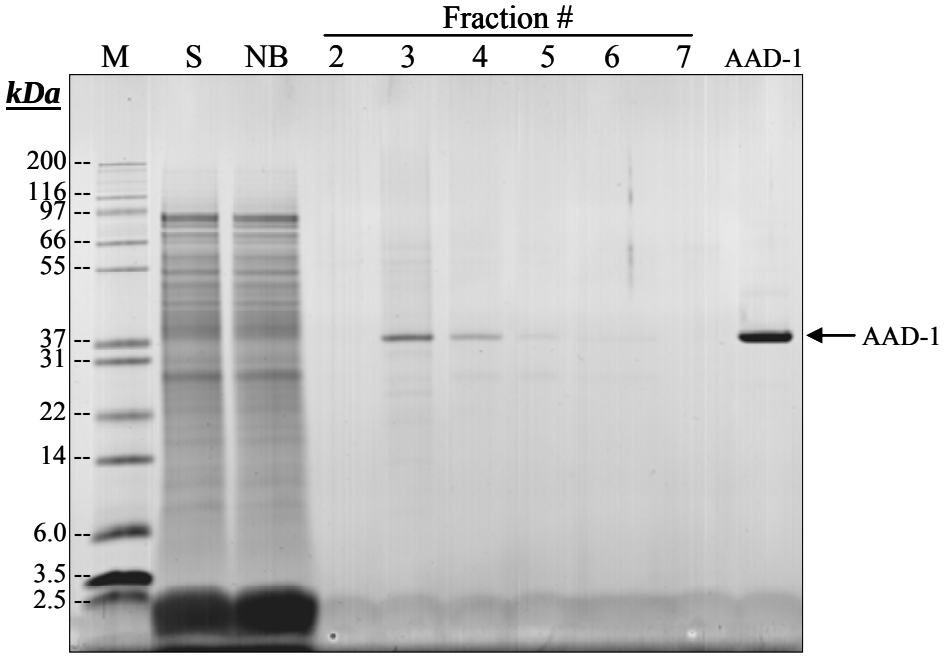


Panel A: SDS-PAGE gel Panel B: Western blot membrane

Lane	Sample	Loaded
M	Invitrogen Mark12 MW markers	10 µL
1	Nontransgenic stalk extract	36 µL
2	Event DAS-40278-9 stalk extract	36 µL
3_{gel}	Microbe-derived AAD-1	~ 1.0 µg
3_{blot}	Microbe-derived AAD-1	~ 10 ng
4	Bovine serum albumin (BSA)	1.0 µg
N	Novex (Invitrogen) prestained MW markers	5.0 µL

Lyophilized tissue from maize event DAS-40278-9 and XHH13 stalk (~100 mg) were weighed out in a 2-mL microfuge tubes and extracted with ~1 mL of PBST (Sigma Cat #: P3593) containing 10% plant protease inhibitor cocktail (Sigma Cat #: P9599). The extraction was facilitated by adding 4 small ball bearings and Geno-Grinding the sample for 1 minute. After grinding, the samples were centrifuged for 5 minutes at 20,000×g and the supernatants were mixed 4:1 with 5x Laemmli sample buffer (2% SDS, 50 mM Tris pH 6.8, 0.2 mg/mL bromophenol blue, 50% (w/w) glycerol containing 10% freshly added 2-mercaptoethanol) and heated for 5 minutes at ~100 °C. After a brief centrifugation, 45 µL of the supernatant was loaded directly on the gel. SDS-PAGE was performed with Bio-Rad Criterion gels (Bio-Rad Cat #:345-0032) fitted in a Criterion Cell gel module (Cat #: 165-6001). The positive reference standard, microbe-derived AAD-1 (TSN105930), was resuspended at 1 mg/mL in PBST pH 7.4 and further diluted with PBST and mixed with Bio-Rad Laemmli buffer (Bio-Rad Cat #: 161-0737 with 5% 2-mercaptoethanol (Bio-Rad Cat #: 161-0710)) and processed as described earlier. The electrophoresis was conducted with Tris/glycine/SDS buffer (Bio-Rad, Cat #: 161-0772) at voltages of 150 - 200 V until the dye front approached the end of the gel. After separation, the gel was cut in half and one half was stained with Pierce GelCode Blue protein stain (Cat #: 24592) and the remaining half was electro-blotted to a nitrocellulose membrane (Bio-Rad, Cat #:162-0213) with a Mini trans-blot electrophoretic transfer cell (Bio-Rad Cat#: 170-3930) for 60 minutes under a constant voltage of 100 volts. The transfer buffer contained 20% methanol and Tris/glycine buffer from Bio-Rad (Cat #: 161-0771). For immunodetection, the membrane was probed with an AAD-1 specific polyclonal rabbit antibody (Strategic Biosolution Inc., Newark, DE, Protein A purified rabbit polyclonal antibody Lot #: DAS F1197-151, 1.6 mg/mL). A conjugate of goat anti-rabbit IgG (H+L) and alkaline phosphatase (Pierce Chemical, Cat #: 31340) was used as the secondary antibody. Sigma*Fast* BCIP/NBT substrate (Cat #: B5655) was used for development and visualization of the immunoreactive protein bands.

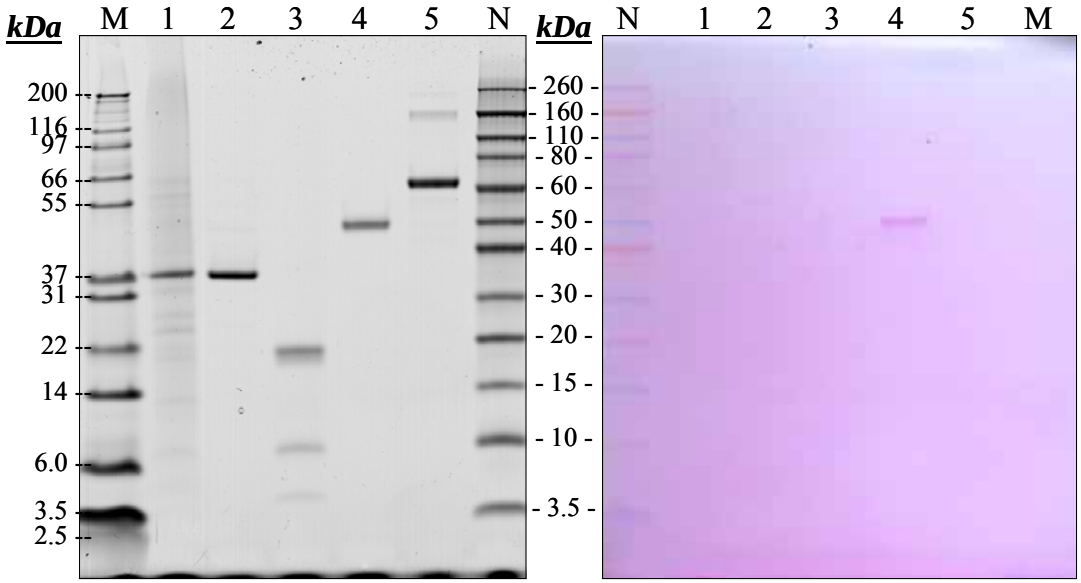
Figure 1. SDS-PAGE (Panel A) and Western Blot (Panel B) of Event DAS-40278-9 Extracts



Lane	Sample	Loaded
M	Invitrogen Mark12 MW markers	10 µL
S	Event DAS-40278-9 starting stalk extract	34 µL
NB	Immunoaffinity column (IAC) - non-bound	34 µL
2	IAC eluted fraction #2	12 µL
3	IAC eluted fraction #3	12 µL
4	IAC eluted fraction #4	12 µL
5	IAC eluted fraction #5	12 µL
6	IAC eluted fraction #6	12 µL
7	IAC eluted fraction #7	12 µL
AAD-1	Microbe-derived AAD-1	~ 1.0 µg

The AAD-1 protein was extracted from lyophilized stalk tissue (event DAS-40278-9) in a phosphate based buffer (see Table 1 for buffer components) by weighing out ~30 grams of lyophilized tissue into a chilled 1000 mL glass blender (Waring Commercial Laboratory Blender Model #: 51BL30) and adding 500 mL of buffer. The tissue was blended on high for 60 seconds and the soluble proteins were harvested by centrifuging the sample for 20 minutes at 30,000 ×g. The pellet was re-extracted as described and the supernatants were combined and filtered through a 0.45 µ filter. The filtered supernatants were loaded at 4 °C onto an anti-AAD-1 immunoaffinity column that was conjugated with a monoclonal antibody prepared by Strategic Biosolution Inc. (MAb 473F185.1; Protein A purified; Lot #: 609.03C-2-4; 6.5 mg/mL (~35.2 mg total); Isotype: IgG1, kappa - conjugated to CNBr-activated Sepharose 4B (GE Healthcare Cat #: 17-0430-01). The column preparation is documented in DAS Notebook F1257). The non-bound proteins were collected and the column was washed extensively with pre-chilled 20 mM ammonium bicarbonate buffer, pH 8.0. The bound proteins were eluted with 3.5 M NaSCN, (Sigma Cat #: S7757), 50 mM Tris (Sigma Cat #: T3038) pH 8.0 buffer. Seven 5 mL-fractions were collected and fraction numbers 2 – 7 were dialyzed overnight at 4 °C against 10 mM Tris pH 8.0 buffer. The fractions were examined by SDS-PAGE and western blot (data not shown) and the remaining samples were stored at 4 °C until used for subsequent analyses.

Figure 2. SDS-PAGE of DAS Event 40278-9 Crude Extracts and Immunopurified Maize-Derived AAD-1



Panel A: SDS-PAGE gel stained for total protein

Panel B: SDS-PAGE gel stained for glycoproteins

Lane	***Sample***	***Loaded***
M	Invitrogen Mark12 MW markers	10 µL
1	Maize-derived AAD-1 (Event DAS-40278-9)	32 µL
2	Microbe-derived AAD-1 (TSN105930)	1.0 µg
3	Soybean Trypsin Inhibitor (negative control)	1.0 µg
4	Horseradish Peroxidase (positive control)	1.0 µg
5	Bovine serum albumin (negative control)	1.0 µg
N	Novex (Invitrogen) prestained MW markers	5.0 µL

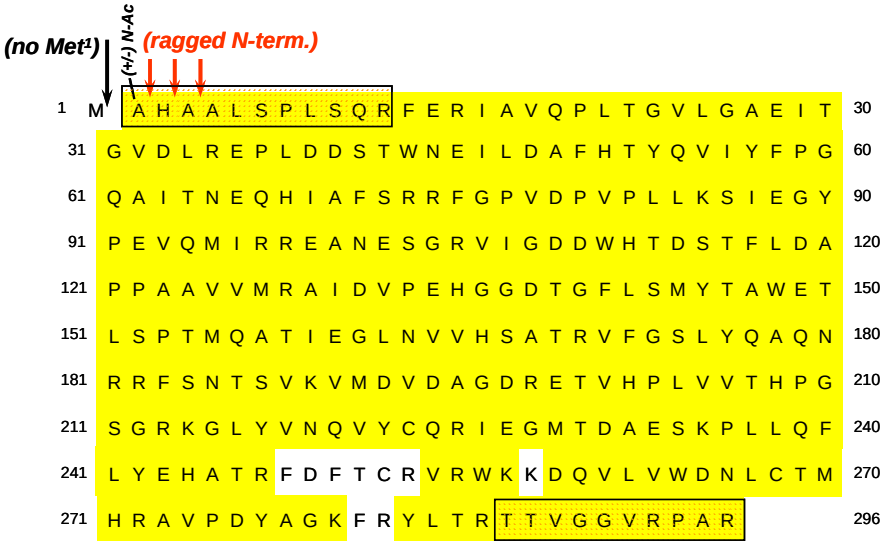
The immunoaffinity chromatography-purified, maize-derived AAD-1 protein (Fraction #3) was mixed 4:1 with 5x Laemmli buffer. The microbe-derived AAD-1, soybean trypsin inhibitor, bovine serum albumin and horseradish peroxidase were diluted with Milli-Q water to the approximate concentration of the plant-derived AAD-1 and mixed with Bio-Rad Laemmli buffer. The proteins were then heated at ~95 °C for 5 minutes and centrifuged at 20000×g for 2 minutes to obtain a clarified supernatant. The resulting supernatants were applied directly to a Bio-Rad Criterion Gel (Cat #: 345-0111) and electrophoresed with XT MES running buffer (Bio-Rad, Cat#: 161-0789) at 170 V for ~60 minutes. After electrophoresis, the gel was cut in half and one half was stained with GelCode Blue stain (Pierce, Cat #: 24592) for total protein according to the manufacturers' protocol. After the staining was complete, the gel was scanned with a Molecular Dynamics densitometer to obtain a permanent visual record of the gel. The remaining half of the gel was stained with a GelCode Glycoprotein Staining Kit (Pierce, Cat #: 24562) according to the manufacturers' protocol to visualize the glycoproteins. The procedure for glycoprotein staining is briefly described as follows: After electrophoresis, the gel was fixed in 50% methanol for 30 minutes and rinsed with 3% acetic acid. This was followed by an incubation period with the oxidation solution from the staining kit for 15 minutes. The gel was once again rinsed with 3% acetic acid and incubated with GelCode glycoprotein staining reagent for 15 minutes. Finally, the gel was immersed in the reduction solution for 5 minutes, and then rinsed with 3% acetic acid. The glycoproteins (with a detection limit as low as 0.625 ng per band) were visualized as magenta bands on a light pink background. After the glycoprotein staining was complete, the gel was scanned with a Hewlett Packard digital scanner to obtain a permanent visual record of the gel.

Figure 3. SDS-PAGE gel stained with GelCode Blue Total Protein Stain (Panel A) and GelCode Glycoprotein Stain (Panel B)

```
  1  M A H A A L S P L S Q R f e r I A V Q P L T G V L G A E I T  30
 31  G V D L R e p l d d s t w n e i l d a f h t y q v i y f p g  60
 61  q a i t n e q h i a f s r R f g p v d p v p l l k S I E G Y  90
 91  P E V Q M I R r E A N E S G R v i g d d w h t d s t f l d a 120
121  p p a a v v m r A I D V P E H G G D T G F L S M Y T A W E T 150
151  L S P T M Q A T I E G L N V V H S A T R v f g s l y q a q n 180
181  r R f s n t s v k V M D V D A G D R e t v h p l v v t h p g 210
211  s g r K g l y v n q v y c q r I E G M T D A E S K P L L Q F 240
241  L Y E H A T R f d f t c r V R w k K d q v l v w d n l c t m 270
271  h r A V P D Y A G K f r Y L T R t t v g g v r p a r       296
```

Theoretical trypsin cleavage (cut at lysine (K) and arginine (R)) of the AAD-1 protein. Alternating blocks of upper (**black**) and lower case (**red**) letters within the amino acid sequence are used to differentiate the potential peptides after trypsin digestion. The numbers on the left and right sides indicate the amino acid residue numbers.

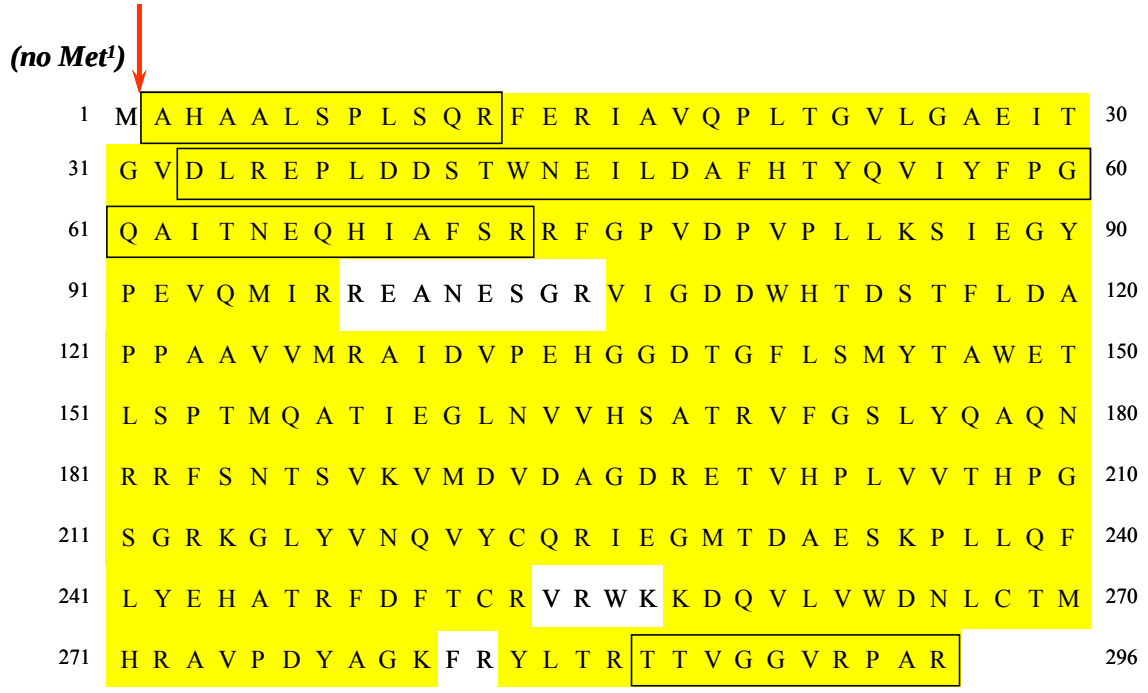
Figure 4. Expected Amino Acid Sequence and Trypsin Peptide Fragments of AAD-1 (MW ~ 33 kDa).



Sequence coverage = 96.6%

The numbers on the left and right sides of the protein sequence indicate the amino acid residue numbers. Letters in yellow represent peptide fragments detected by enzymatic peptide mass fingerprinting. The letters in **orange** blocks indicate the peptide sequence was confirmed by tandem MS sequencing. The overall sequence coverage was 96.6%. The "**(±) N-Ac"** on the N-terminal residue indicates the protein was partially acetylated in plants. The **black** arrow indicates the N-terminal methionine was removed by an aminopeptidase for all detected peptides. The **red** arrows indicate that trace amounts of the N-terminal peptide were found to have various additional truncations.

Figure 5. Sequence coverage of immunopurified, maize-derived AAD-1 (event DAS-40278-9) based on enzymatic peptide mass fingerprinting and MS/MS sequencing



Sequence coverage = 94.9%

The numbers on the left and right sides of the protein sequence indicate the amino acid residue numbers. Letters in yellow represent peptide fragments detected by enzymatic peptide mass fingerprinting. The letters in **orange** blocks indicate the peptide sequence confirmed by tandem MS sequencing. The overall sequence coverage was 94.9%. The **red** arrow indicates the N-terminal methionine was removed by an aminopeptidase.

Figure 6. Sequence coverage of microbe-derived AAD-1 based on enzymatic peptide mass fingerprinting and MS/MS sequencing

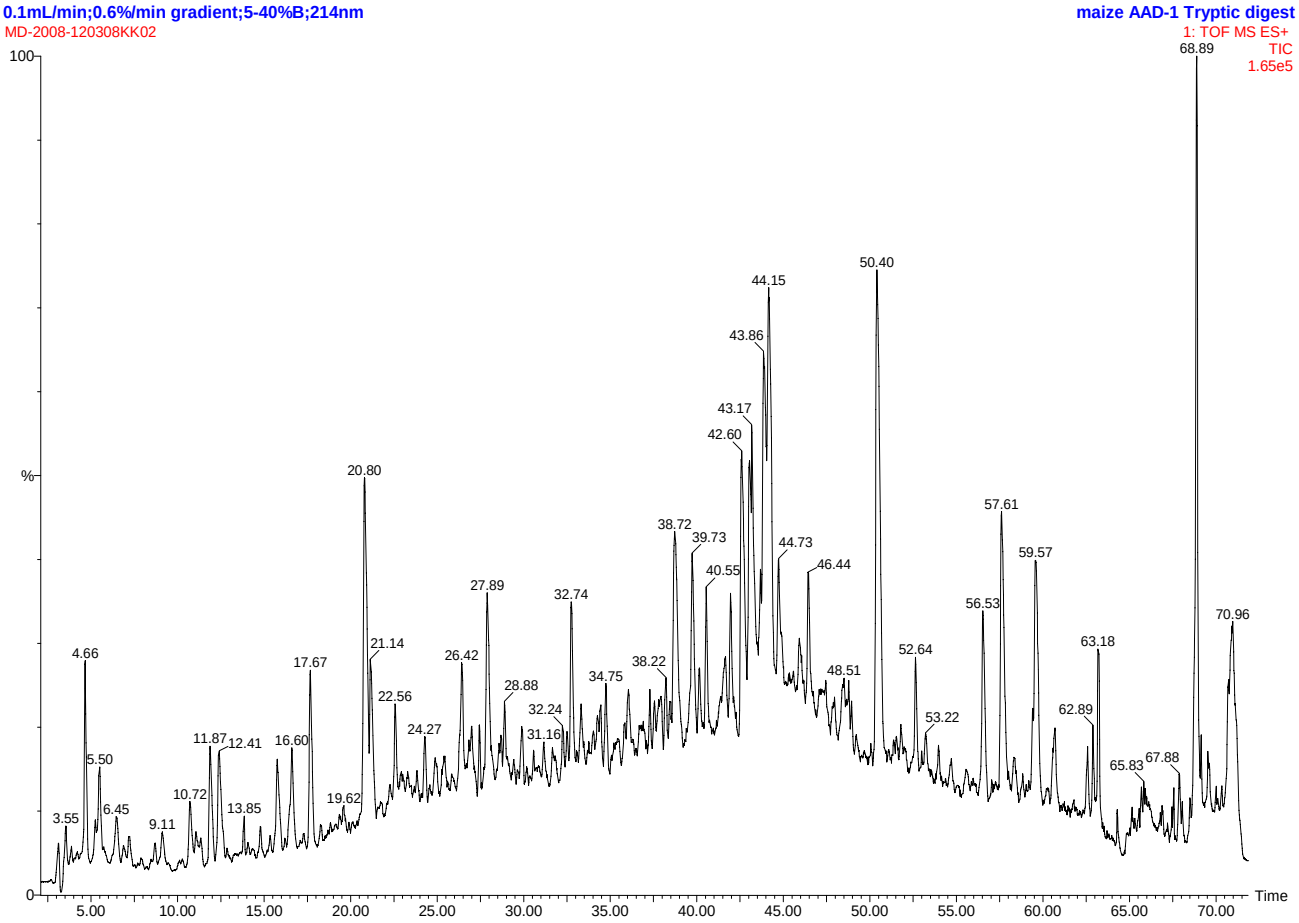
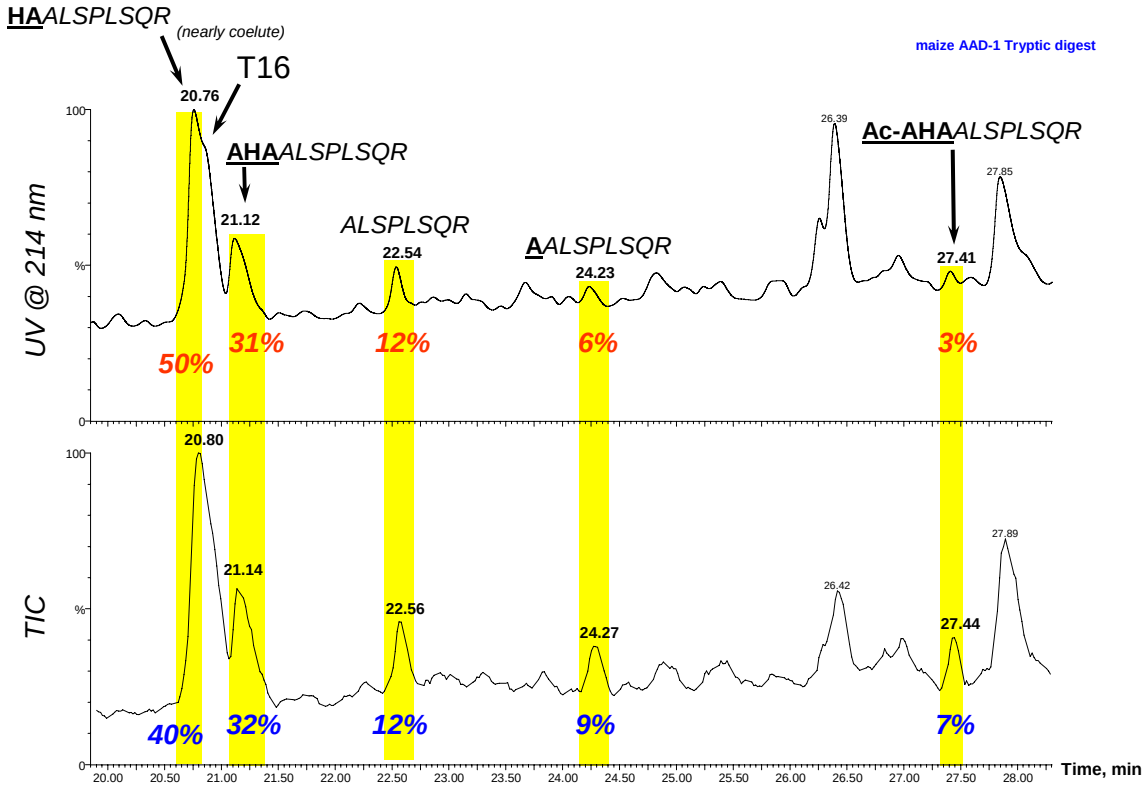


Figure 7. LC/MS Chromatogram for Maize-Derived AAD-1 (event DAS-40278-9).



Note: Both UV-absorption at 214 nm (upper panel) and total ion current (TIC; lower panel) are shown. Rough estimates of relative fragment abundances were made based on the corresponding LC peak areas.

Figure 8. Zoom of the chromatogram of Figure 7, focusing on the eluting N-terminal fragment variants of Maize-Derived AAD-1 (event DAS-40278-9).

APPENDIX

Biochem. J. (1976) **153**, 351–361
Printed in Great Britain

Studies on the Use of Sepharose–*N*-(6-Aminohexanoyl)-2-Amino-2-Deoxy-D-Glucopyranose for the Large-Scale Purification of Hepatic Glucokinase

By MICHAEL J. HOLROYDE, JANICE M. E. CHESHER, IAN P. TRAYER and DERYCK G. WALKER
Department of Biochemistry, University of Birmingham, P.O. Box 363, Birmingham B15 2TT, U.K.

(*Received* 25 *July* 1975)

The synthesis of *N*-(6-aminohexanoyl)-2-amino-2-deoxy-D-glucose is described and it was shown to be a competitive inhibitor (K_I, 0.75mM) with respect to glucose of rat hepatic glucokinase (EC 2.7.1.2). After attachment to CNBr-activated Sepharose 4B, this derivative was able to remove glucokinase quantitatively from crude liver extracts and release it when the columns were developed with glucose, glucosamine, *N*-acetylglucosamine or KCl. Repeated exposure of the columns to liver extracts led to rapid loss in their effectiveness as affinity matrices because proteins other than glucokinase are bound to the columns. The nature of such protein binding and methods for the rejuvenation of 'used' columns are discussed along with the effect of the mode of preparation of the Sepharose–ligand conjugate and the concentration of bound ligand on the purification of glucokinase. Glucose 6-phosphate dehydrogenase is cited as an example of both non-specific protein binding to the affinity column and of the importance of the control of ligand concentration in removing such non-specifically bound proteins. Some guidelines emerged that should be generally applicable to other systems, particularly those which involve affinity chromatography of enzymes that are present in tissue extracts in very low amounts and possess only a relatively low association constant for the immobilized ligand.

Despite the impact of affinity chromatography on protein-separation technology [see review by Cuatrecasas (1972) and references therein] most of the current literature on this subject has concerned itself with small-scale purifications, often using commercially available enzymes in model systems (e.g. Mosbach *et al.*, 1972; Brodelius & Mosbach, 1973; Lowe *et al.*, 1974*a*; Lowe & Mosbach, 1975). Only in a relatively few instances has the technology been applied to the large-scale purification of enzymes to homogeneity (e.g. Trayer & Hill, 1971; Barker *et al.*, 1972; Nicolas *et al.*, 1972). This is particularly true when either the ligand to be immobilized has only a relatively high dissociation constant (about 10^{-3}M) for the enzyme under study or the enzyme is present in the tissue source in only very small amounts.

In the large-scale purification of rat hepatic glucokinase (EC 2.7.1.2) (the following paper, Holroyde *et al.*, 1976), the key step was the affinity chromatography of the enzyme on a Sepharose–*N*-(6-aminohexanoyl)-2-amino-2-deoxy-D-glucopyranose matrix. Our initial attempts (Chesher *et al.*, 1973) to utilize this immobilized glucosamine derivative, which in free solution is a competitive inhibitor with respect to glucose of the glucokinase reaction (K_I, 0.75mM), in a chromatographic procedure gave variable results. A more detailed appraisal now permits us to report the various parameters, such as the control of ligand concentration on the Sepharose matrix and the elution conditions used that can be varied to obtain optimum results. Correct washing procedures were found to be essential when processing crude extracts from up to 100 rat livers to maintain the effectiveness of the affinity matrix.

Although the results presented here are devoted to the affinity chromatography of glucokinase, some guidelines emerge that should be generally applicable to other systems. Some of this work has been reported in preliminary form (Holroyde & Trayer, 1974).

Materials and Methods

Materials

Chemicals. Coenzymes, nucleotides, Dowex ion-exchange resins and dithiothreitol were purchased from Sigma (London) Chemical Co., Kingston-upon-Thames, Surrey, U.K. Ethyl trifluorothiol acetate was from Pierce Chemical Co., Rockford, Ill., U.S.A., *N*-ethoxycarbonylethoxy-1,2-dihydroquinoline was obtained from Calbiochem, San Diego, Calif., U.S.A., and CNBr was from R. Emanuel, Wembley, Middx., U.K. Glucosamine hydrochloride, *N*-glycylglycine, *N*-(*N*-glycylglycyl)-glycine and 6-aminohexanoic acid were supplied by

BDH Chemicals, Poole, Dorset, U.K. Sepharose 4B was from Pharmacia (G.B.) Ltd., London W.5, U.K., and DEAE-cellulose (DE 52) was from Whatman Biochemicals, Maidstone, Kent, U.K. All other chemicals were AnalaR grade and used as supplied.

Proteins. Yeast hexokinase [ATP–D-hexose 6-phosphotransferase, EC 2.7.1.1; type C-302 (300units/mg)] and glucose 6-phosphate dehydrogenase [D-glucose 6-phosphate–$NADP^+$ oxidoreductase, EC 1.1.1.49, from yeast, type VII (345units/mg)] were purchased from Sigma Chemical Co. Pronase (B grade) was from Calbiochem.

Glucokinase preparations. The crude extracts of rat liver were prepared as described by Chesher *et al.* (1973). The partially purified enzyme preparations (post-DEAE-cellulose treatment) were obtained as described by Holroyde *et al.* (1975).

Methods

Preparation of N-(6-aminohexanoyl)-2-amino-2-deoxy-D-glucopyranose. The synthesis of an amide linkage between the amino group of glucosamine and the carboxyl group of 6-aminohexanoic acid is described below. This involves prior blocking of the amino group of the amino acid before the coupling reaction and deblocking afterwards. This scheme has also been carried out with galactosamine (R. Barker, personal communication) and with a variety of other amino acids (A. S. Warsy & I. P. Trayer, unpublished work).

N-Trifluoroacetyl-6-aminohexanoic acid. To 0.1mol (13.2g) of 6-aminohexanoic acid, dissolved in 100ml of water and maintained at 4°C in an ice-bath, were added 18ml of ethyl trifluorothiol acetate in 1ml batches during a 3–4h period with vigorous stirring. The emulsion was maintained at pH9.5 by the occasional addition of 1M-NaOH. The reaction was judged complete (3–4h) when only a pale-pink (and not blue) colour was obtained after spotting a sample on paper, spraying with ninhydrin and heating at 100°C for 5min. The reaction mixture was adjusted to pH5 with trifluoroacetic acid and concentrated to dryness at 40°C. The residue was redissolved in ethanol and reconcentrated from this (three times in all) and then three times from aq. 50% (v/v) ethanol to remove ethyl mercaptan. The residue was redissolved in 100ml of 50% (v/v) ethanol, adjusted to pH1.5 with trifluoroacetic acid, and passed through a column (30cm×3cm) of Dowex 50 (X8; H^+ form; 20–50 mesh) equilibrated to 50% (v/v) ethanol to remove any unchanged amine. The acidic eluate was concentrated to dryness under vacuum at 45°C and then dried *in vacuo* over anhydrous $CaSO_4$ for 24h. This material (in aqueous solution) migrated as a single anionic component on high-voltage paper electrophoresis at pH6.5, was negative to ninhydrin but gave a yellow acidic spot when stained with Bromophenol Blue. It was stored in a desiccator at room temperature and was stable over several years. The overall yield of pure compound was 20.3g (90%) and was used in the subsequent stages without further characterization.

N-(N-Trifluoroacetyl-6-aminohexanoyl)-2-amino-2-deoxy-D-glucopyranose. Glucosamine hydrochloride (2.15g, 10mmol) and *N*-trifluoroacetyl-6-aminohexanoic acid (2.27g, 10mmol) were dissolved in 400ml of water and adjusted to pH7.5 with 1M-NaOH. To this solution was added 15g (60mmol) of *N*-ethoxycarbonylethoxy-1,2-dihydroquinoline (Belleau & Malek, 1968) and the suspension was shaken vigorously at 37°C for 16h. During the first 2–4h of the reaction, three additions of 100ml of ethanol were made to bring the *N*-ethoxycarbonylethoxy-1,2-dihydroquinoline into solution slowly so that activation of the carboxyl groups occurred throughout the reaction. After 16h, the reaction mixture was evaporated to dryness at 40–50°C and the residue extracted twice with 200ml of water. The water-soluble extracts were combined, adjusted to pH1.5 with trifluoroacetic acid, and passed through a column (10cm×3cm) of Dowex 50 (X8; H^+ form; 20–50 mesh) to remove unchanged glucosamine. The column was washed with 200ml of water. The eluate and washings were combined, adjusted to pH7 with 1M-NaOH, and evaporated to dryness. The product contained some unchanged *N*-trifluoroacetyl-6-aminohexanoic acid which was removed at a later stage.

The reaction was monitored by high-voltage paper electrophoresis at pH6.5. The disappearance of glucosamine (which is positive to both ninhydrin and the silver nitrate test for reducing sugars) correlates with the appearance of the product (ninhydrin-negative and silver nitrate-positive) at the origin. The yield was 70–80%. This material was chemically characterized in our early preparations (see below), but as a routine in later preparations was immediately deblocked as follows.

N-(6-Aminohexanoyl)-2-amino-2-deoxy-D-glucopyranose. The above product (about 3.5g) was dissolved in 12–13ml of 1.0M-piperidine at room temperature (20°C) and the solution was immediately chilled in an ice bath. After standing at 0°C for 2h, the reaction mixture was adjusted to pH6.0 with 0.5M-acetic acid and applied directly to a column (140cm×3.0cm) of Sephadex G-10 equilibrated in water. The product was located in the eluate by testing samples for reducing sugar with ferricyanide (Park & Johnson, 1949). Two ferricyanide-positive peaks emerged from the column. The first contained the final product which was further purified as described below. The second peak, which emerged together with the salt peak, located by conductivity measurements, contained some neutral material positive to silver nitrate together with some 6-aminohexanoate and piperidine.

Further deblocking of this unidentified material did not yield any product and so this second peak was discarded. After concentrating the appropriate fractions to dryness, the product was dissolved in 1mM-HCl and applied to a column (25cm×0.9cm) of Dowex 50 (X8; H^+ form; 200–400 mesh) previously equilibrated with 1mM-HCl. The column was washed with about 100ml of 1mM-HCl and a linear gradient (400ml of 1mM-HCl and 400ml of 0.3M-LiCl in 1mM-HCl) was applied. The initial breakthrough peak was examined by high-voltage paper electrophoresis at pH6.5 and contained 6-aminohexanoic acid and neutral silver nitrate-positive material. This latter material could not be further deblocked to form additional product and appeared to be derived from the coupling agent *N*-ethoxycarbonyl-ethoxy-1,2-dihydroquinoline and was discarded. The single peak of material that emerged under the gradient and gave a positive reaction for reducing sugar was concentrated to dryness. It was desalted, if necessary, with Sephadex G-10 as described above, but it can be used for coupling to CNBr-activated agarose without further purification.

As judged by the ferricyanide procedure, the overall yield was 1.9g (65%) and the product was stored, after adjusting it to pH7.0 with 1M-NaOH, in portions at −12°C.

The final product was not crystallized but was shown to be pure because it migrated as a single cationic species on high-voltage paper electrophoresis (mobility relative to glucosamine: glucosamine, 1.0; product 0.81) and on ascending paper chromatography in solvents 1 and 2 (see below) ($R_F1 = 0.65$; $R_F2 = 0.56$). The single component gave positive tests with both ninhydrin and silver nitrate. The product was further characterized by chromatographic analysis on an amino acid analyser (see below) after acid hydrolysis (6M-HCl, 20h at 105°C). The hydrolysate was shown to contain only 6-aminohexanoic acid and glucosamine in a molar ratio of 1:1. On this basis, and the elementary analyses found for the *N*-trifluoroacetyl derivative (elementary analyses; found: C, 43.5, H, 6.1; N, 7.8; $C_{13}H_{23}N_2O_6F_3$ requires C, 43.3; H, 6.4; N, 7.8%), the structure shown in Fig. 1 was assigned to this product.

The equivalent *N*-glycylglycine and *N*-(*N*-glycylglycyl)glycine derivatives of glucosamine were prepared in a manner identical with that described above. These derivatives were shown to be pure by high-voltage paper electrophoresis at pH6.5 and were characterized by chromatography on the amino acid analyser (see below) after acid hydrolysis. The expected molar ratios of glycine/glucosamine were found.

Preparation and characterization of Sepharose adsorbents. The ligands were attached to CNBr-activated Sepharose 4B (Axen *et al.*, 1967) under the conditions described by Trayer *et al.* (1974). These procedures generally gave a coupling efficiency of 80–95% as judged by the analytical procedures described below.

As a routine, the ligands were left to couple to the activated gel at a concentration of 10–12μmol/g of gel on a rotary mixer overnight at 4°C. For the 'direct-linked' gels described in Fig. 6, the ligands were coupled at different concentrations, ranging from 2 to 10μmol/g of gel. (N.B. 1g of gel in these experiments is defined as the weight of Sepharose 4B taken after packing it down in a Buchner funnel connected to a water suction pump until the Sepharose 4B began to 'crack'. This is not entirely satisfactory but does give reproducible results in our experience.)

After coupling, unchanged groups on the Sepharose 4B were treated by washing the gels with 1M-ethanolamine adjusted to pH8 with HCl. The gels (up to 100g) were then washed on a Buchner funnel with 0.1M-Na_2CO_3, pH9.5 (2litres), water (1litre), 2M-KCl containing 1mM-HCl (2litres) and water (1litre). The gels were characterized by taking a known weight of gel and washing by centrifugation in a bench centrifuge three times with acetone. The dried residue was hydrolysed *in vacuo* in a sealed tube with 2ml of 6M-HCl at 105°C for 24h. The black precipitate that formed was removed by centrifugation, washed with 3×5ml of water and the combined supernatants were concentrated to dryness from water. The residue was dissolved in a known volume of 0.2M-sodium citrate adjusted to pH2.2 with 1M-HCl and an appropriate portion of it subjected to chromatography on a Beckman Spinco model 120B amino acid analyser. Losses due to mechanical transfer were eliminated by using norleucine as an internal standard. Known concentrations of glucosamine and 6-aminohexanoic acid were similarly hydrolysed before their use as standards on the analyser to correct for any destruction occurring during hydrolysis.

Dilution of substituted gels to a required ligand concentration was performed by weighing out the substituted gels and fresh Sepharose 4B as described above and mixing in the appropriate proportions. When packing columns of diluted gels, the mixture was stirred continuously and small samples were withdrawn for introducing to the glass columns; this procedure decreased the possibility of local regions of either unsubstituted or substituted gel in the resulting column.

Operation of the affinity columns. These were always run at 4°C and were unpacked and washed by suction on a Buchner funnel with 2M-KCl/6M-urea after every use. Pronase treatment, when required, was carried out as follows: the gels (200–300g) were first washed with the KCl/urea solution and then suspended in 20mM-triethanolamine/HCl, pH7.0, at room temperature. Pronase (20mg) was added

and the gel suspension gently stirred. Two further additions of Pronase (20mg each) were made after 2h and 4h and the suspension was left stirring gently overnight. The treated gels were then washed carefully on a Buchner funnel with water (1 litre), the KCl/urea solution (3 litres), water (10 litres) followed by the appropriate operating buffer (1 litre). The gels were stored in this buffer containing 0.2% sodium azide at 4°C.

The progress of glucose gradients was monitored by subjecting samples to the phenol/H_2SO_4 reaction for neutral sugars (Ashwell, 1966). Recoveries of glucokinase from these columns was always greater than 80% (usually nearer 100%) unless specifically stated to the contrary.

Enzyme assays. Glucokinase activity was measured at 30°C in a total volume of 0.75ml by the coupled assay of Parry & Walker (1966) as described by Storer & Cornish-Bowden (1974). Low-K_m hexokinase activities were estimated in the presence or absence of glucokinase as described by Holroyde *et al.* (1975). Any contaminating activities which may interfere with these assays and the method of correction for them are also described in that paper. Glucose 6-phosphate dehydrogenase activity was also measured by the increase in E_{340} as $NADP^+$ was reduced in a total volume of 0.75ml containing the same reagents as for the glucokinase assay except that the glucose, ATP and glucose 6-phosphate dehydrogenase were replaced by 20mM-glucose 6-phosphate.

One unit of enzyme activity is defined as that which catalyses the formation of 1 μmol of either glucose 6-phosphate/min (glucokinase) or 6-phosphogluconate (glucose 6-phosphate dehydrogenase) at 30°C.

Analytical procedures. High-voltage paper electrophoresis and paper chromatography were carried out as described by Trayer *et al.* (1974). The solvents (referred to above) are: solvent 1, isobutyric acid/NH_3 (sp.gr. 0.88)/water (66:1:33, by vol.); solvent 2, propan-2-ol/aq. 0.25M-NH_4HCO_3 (13:7, v/v).

Chromatograms and electrophoretograms were developed with the ninhydrin/cadmium reagent for amines (Barrolier, 1957) and with Bromophenol Blue [a 0.1% solution in aq. 95% (v/v) ethanol titrated to its end point with 5M-NH_3] for acidic and basic groups. Reducing sugars on chromatograms were revealed by treatment with the silver nitrate reagent. A saturated solution of silver nitrate in water (0.1ml) was added to 20ml of acetone and water added until the white precipitate disappeared. The dried chromatography paper was dipped in this solution and then immediately sprayed with 0.5M-NaOH in aq. 95% (v/v) ethanol. Reducing sugars appeared as dark-brown spots.

Salt peaks were detected in fractions after gel filtration by monitoring on a conductivity meter, type CDM (Radiometer, Copenhagen, Denmark).

Elementary analyses were performed by the Department of Chemistry (Analytical Section) of this University.

Results

Initial studies on the affinity chromatography of glucokinase on immobilized glucosamine derivatives

The structure of the glucosamine derivative prepared by synthesizing an amide link between the amino group of glucosamine and the carboxyl group of 6-aminohexanoic acid is shown in Fig. 1(*a*). The free ligand was a competitive inhibitor with respect to glucose in the glucokinase reaction, with a K_i (0.75mM) similar to that of free glucosamine (Fig. 1*b*), and was not phosphorylated. Another glucosamine derivative prepared by synthesizing a glycosidic bond between position 1 of *N*-acetylglucosamine and 6-aminohexan-1-ol (Barker *et al.*, 1972, 1974) did not inhibit the glucokinase reaction nor did this enzyme bind to this derivative linked to Sepharose. This result is consistent with the view that the position 1 of glucose is essential for the binding of glucose to a variety of hexokinase enzymes during their reaction (Crane, 1962).

When rat liver extracts were applied to conjugates of *N*-(6-aminohexanoyl)-2-amino-2-deoxy-D-glucopyranose with Sepharose then all of the glucokinase activity was retained on these columns and could be released by the introduction of glucose to the

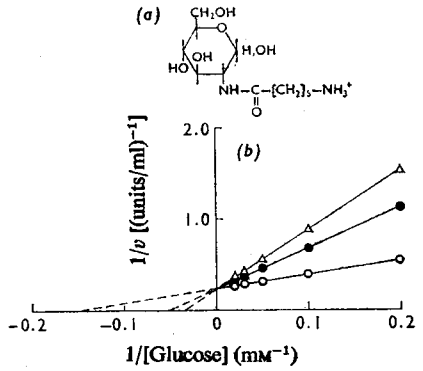


Fig. 1. *Inhibition of glucokinase activity by N-(6-aminohexanoyl)-2-amino-2-deoxy-D-glucopyranose*

(*a*) Structure of *N*-(6-aminohexanoyl)-2-amino-2-deoxy-D-glucopyranose. (*b*) Full details of the reaction mixture are given in the Materials and Methods section. The enzyme (0.02unit/ml) was incubated at 30°C with 5, 10, 20, 30 and 50mM-glucose in the absence and presence of the glucosamine derivative. ○, Glucose only; ●, glucose +0.66mM-glucosamine derivative; △, glucose+1.33mM-glucosamine derivative.

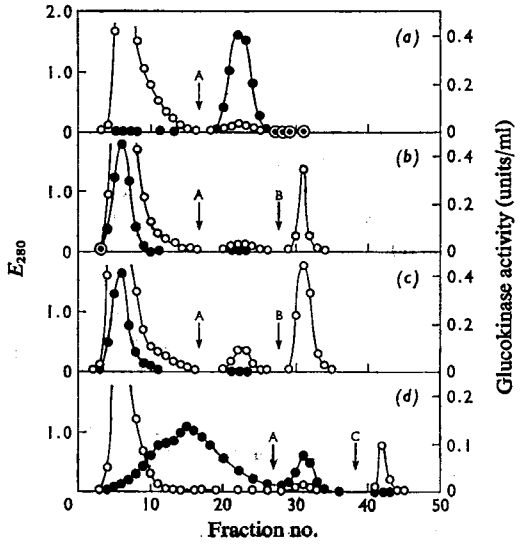


Fig. 2. *Chromatography of rat liver extracts on various Sepharose derivatives*

To all columns (15cm×0.8cm) 2.5ml of liver extract (containing 0.8unit of glucokinase activity/ml) was applied. The columns were operated in 20mM-triethanolamine/HCl, pH7.0, containing 10mM-KCl, 4mM-EDTA, 7.5mM-$MgCl_2$, 1mM-dithiothreitol and 5% (v/v) glycerol at 20ml/h; 1.5ml fractions were collected. At A, 0.5M-glucose and at C, 0.5M-KCl, were included in the developing buffers. At B, the buffer was changed to 2M-KCl/6M-urea. (*a*) Sepharose–*N*-(6-aminohexanoyl-2-amino-2-deoxy-D-glucopyranose (ligand concentration: 2μmol/g wet weight of packed gel); (*b*) Sepharose–glucosamine (6μmol/g wet weight of packed gel); (*c*) Sepharose–6-aminohexanoate (6μmol/g wet weight of packed gel); (*d*) Sepharose–*N*-(*N*-glycylglycyl)-2-amino-2-deoxy-D-glucopyranose (8μmol/g wet weight of packed gel). ○, E_{280}; ●, glucokinase activity.

developing buffers (Fig. 2*a*). Under these conditions the liver low-K_m hexokinase activity appeared not to be retained on the columns but to pass through unretarded. Subsequent work has shown that mammalian hexokinases can be bound to these glucosamine derivatives under certain conditions (A. S. Warsy, M. J. Holroyde & I. P. Trayer, unpublished work). In separate experiments, yeast hexokinase did not exhibit any interaction with these glucosamine-containing matrices.

Glucokinase eluted from these derivatives (Fig. 2*a*) was found to be purified some 500- to 1000-fold. Rechromatography of the glucokinase fraction on the affinity columns under similar conditions resulted in additional purification but the final product was still heavily contaminated by other proteins and in our best preparations the glucokinase only represented some 5% (w/w) of the total protein present. These columns lost much of their effectiveness as affinity matrices when repeatedly exposed to liver extracts. The apparent 'capacity' of the columns to bind glucokinase was gradually diminished with each subsequent usage. Their working life could be prolonged somewhat by exhaustive washing between runs with 6M-urea/2M-KCl but after four or five chromatography runs their effectiveness was much decreased. There was no evidence of any enzymic degradation of the ligand attached to the columns and subsequent experimentation suggested that the problem was a gradual build-up of apparently non-specific protein on the affinity matrix. As a consequence the interaction of proteins from the liver extracts with the component parts of the glucosamine derivative attached to Sepharose was investigated.

Although it was not possible to show any interaction between glucokinase and glucosamine attached directly to CNBr-activated Sepharose-4B without the intervening 6-aminohexanoate 'spacer', molecule (Fig. 2*b*), these derivatives did remove a considerable amount of protein when liver extracts were applied. A small proportion of this protein could be eluted by 0.5M-glucose and a large amount of the remaining protein could be released by eluting the columns with 6M-urea/2M-KCl (Fig. 2*b*). Even more protein was removed from liver extracts when these were applied to a Sepharose–6-aminohexanoate conjugate but all the glucokinase activity passed through unretarded. Again glucose caused some protein to be eluted and most of the remaining bound protein was released by the combined action of 6M-urea/2M-KCl (Fig. 2*c*).

Clearly, a series of combined interactions between the glucosamine derivative and the proteins found in the liver extract was occurring. Although the results in Fig. 2 support the view that the interaction between glucokinase and the Sepharose–glucosamine derivatives are truly specific, interactions of an ion-exchange and/or hydrophobic nature also occur between the other proteins in the crude extract and the CNBr-activated Sepharose matrix and the polymethylene backbone of the 'spacer' molecule. O'Carra *et al.* (1973) have also described hydrophobic interactions between the polymethylene 'spacer' groups most often used in affinity chromatographic matrices and proteins. In an effort to minimize this effect glucosamine derivatives containing more hydrophilic 'spacer' arms were synthesized by attaching either *N*-glycylglycine or *N*-(*N*-glycylglycyl)glycine in amide linkage to the glucosamine in place of the 6-aminohexanoate. Both the free diglycyl- and triglycyl- derivatives of glucosamine were competitive inhibitors of the glucokinase reaction with respect to glucose, having K_i values of 0.75 and 1.4mM respectively. When

attached to Sepharose 4B at ligand concentrations of up to 8μmol/g, however, neither of these immobilized derivatives exhibited any strong affinity for the enzyme (Fig. 2*d*). Some interaction was observed since the glucokinase emerged from these columns after the main protein peak, but no conditions could be found which caused the glucokinase to bind without considerable leakage. The use of these immobilized glycyl glucosamine derivatives was thus discontinued.

The problems encountered by applying crude liver extracts directly to the affinity columns thus appeared too great to be overcome easily and the subsequent experiments described below were therefore performed on extracts first purified some 200-fold by batchwise chromatography on DEAE-cellulose (Holroyde *et al.*, 1975) to increase the relative amounts of glucokinase to other proteins in the preparation.

Importance of adequate washing procedures between affinity-chromatography operations

Despite such partial purification of glucokinase before the affinity-chromatography step, the Sepharose–glucosamine derivatives still increasingly lost much of their effectiveness to bind the enzyme after several operations. This lowering in their apparent 'capacity' to bind the enzyme was now less than when liver extracts were applied directly but the problem remained even after exhaustive washing with the urea/KCl solution between operations. The physical appearance of the affinity matrix after six or so operations indicated clumping of the gel particles and the matrix had a pale-yellow colour. This contrasted with the clean white appearance of an unused matrix and suggested that the washing procedures were not preventing a gradual build-up of protein on the matrix. When the 'aged' matrices were left overnight with a non-specific proteinase preparation, Pronase, then the original 'capacity' of the columns could be virtually fully recovered (Fig. 3). Consequently, the working life of the columns was greatly prolonged for use in large-scale purification procedures by introducing the Pronase treatment after every second use of the columns. The urea/KCl wash was performed as a routine after every use and after the Pronase treatment.

Effect of ligand concentration

The concentration of ligand attached to the Sepharose matrix affects both the binding of glucokinase and non-specific protein–column interactions and hence the purification of the enzyme. Fig. 4 describes a series of column profiles where the same amount of a glucokinase preparation was

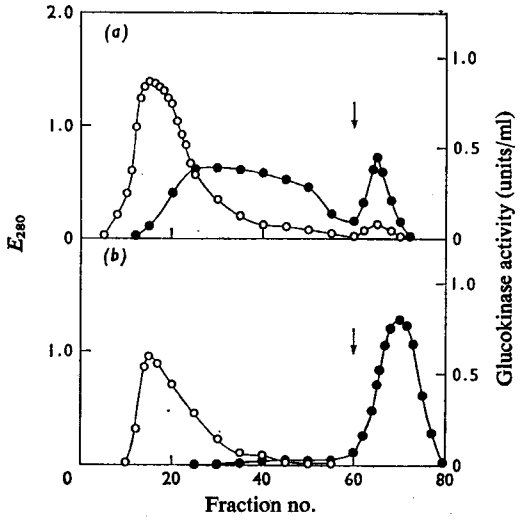


Fig. 3. *Effect of Pronase treatment of the glucosamine derivative affinity columns*

In each case glucokinase (70units in 70ml), purified by DEAE-cellulose chromatography, was applied to the columns (10cm×3cm) operated in the buffer described in Fig. 2. The arrows indicate where 0.5M-glucose was included in the developing buffer. The columns were operated at 30ml/h, and 7ml fractions were collected. (*a*) Sepharose–*N*-(6-aminohexanoyl)-2-amino-2-deoxy-D-glucopyranose that had been used in previous experiments using liver extracts at least six times. Between operations this matrix had been washed with 2M-KCl/6M-urea. (*b*) The same column treated with Pronase overnight (full details in the Materials and Methods section). ○, E_{280}; ●, glucokinase activity.

applied to columns of the Sepharose matrix covalently substituted with different amounts of *N*-(6-aminohexanoyl)-2-amino-2-deoxy-D-glucopyranose. After washing the columns, a linear gradient from 0 to 1.0M-glucose was applied. At low concentrations on the column (Fig. 4*a*), the enzyme activity appeared before the inclusion of glucose in the developing buffer. Increasing the ligand concentration caused the glucokinase to be bound quantitatively and necessitated increasingly higher concentrations of glucose to effect its elution (Figs. 4*b* and 4*c*). At very high ligand concentrations, no enzyme was eluted by up to 1M-glucose and 0.5M-KCl was needed to release the bound enzyme (Fig. 4*d*).

The size of the protein breakthrough peak before application of the glucose gradient also varied. This became noticeably smaller as the ligand concentration was increased and indicated that binding of non-specific protein was also increasing. A comparison of the specific enzymic activities

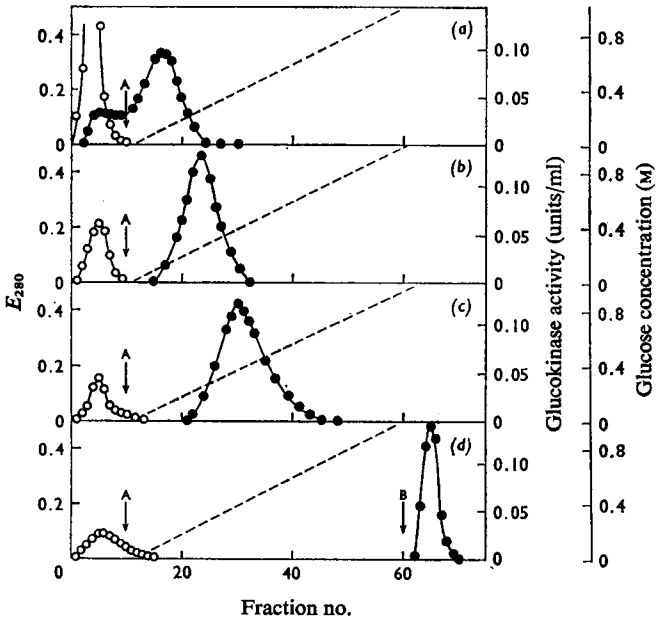


Fig. 4. *Effect of ligand concentrations on the elution of glucokinase by glucose from Sepharose–N-(6-aminohexanoyl)-2-amino-2-deoxy-D-glucopyranose*

Each column (10cm×0.8cm) was equilibrated with the buffer described in Fig. 2 and operated at 20ml/h; 3.0ml fractions were collected. Glucokinase (2ml, 2units/ml) purified by DEAE-cellulose chromatography was applied to each column followed by 25ml of the equilibration buffer. At A, the columns were developed with a linear gradient formed from 75ml of the column buffer and 75ml of 1M-glucose dissolved in this buffer. At B, 0.5M-KCl was included in the buffer. (*a*) 1.2μmol of glucosamine derivative coupled per g wet weight of packed gel; (*b*) 3.75μmol/g; (*c*) 6.0μmol/g; (*d*) 10μmol/g. In this experiment the final ligand concentrations were achieved by diluting the 10μmol/g gel material with unsubstituted Sepharose (see the Materials and Methods section). ○, E_{280}; ●, glucokinase activity; ----, glucose concentration.

of the eluting enzyme confirmed this; the enzyme fraction eluted with glucose shown in Fig. 4(*a*) had a higher specific activity (10units/mg) than the enzyme eluted in Fig. 4(*c*) (1.5units/mg). Thus for optimal use it was necessary to control the ligand concentration coupled to the Sepharose matrix to balance the degree of purification obtained against efficient binding of the enzyme. In practice, this meant working at as low a ligand concentration as would quantitatively remove the glucokinase from the protein applied to the column.

The above experiments were conducted with affinity columns that were linked initially at a concentration of about 10μmol of glucosamine ligand/g of Sepharose and then 'diluted' to the required concentration with unsubstituted Sepharose 4B (see the Materials and Methods section). Dilution would give rise to columns that contained local regions of high ligand concentration interspersed with unsubstituted Sepharose. The experiments were repeated with affinity columns prepared at different ligand concentrations by direct coupling and the results compared (Fig. 5). The two lines converge at higher ligand concentrations, i.e. where dilution with unsubstituted Sepharose of the 'diluted' series is minimal. At lower ligand concentrations a higher concentration of glucose is required to elute the enzyme from a 'diluted' column containing a given ligand concentration than from the corresponding direct-linked gel. More non-specific protein bound to the direct-linked gels as judged by the amount of protein passing unretarded through the columns whereas the enzyme eluted from the 'diluted' columns, was of a consistently higher specific activity. This was particularly apparent when large amounts of enzyme were being processed. The presence of the non-specific protein may also have influenced the tightness of binding of glucokinase.

The results of these studies have been incorporated into a successful purification scheme for

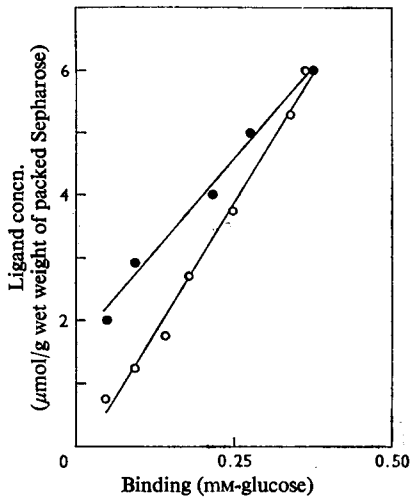


Fig. 5. *Comparison of the elution of glucokinase from affinity columns containing different concentrations of N-(6-aminohexanoyl)-2-amino-2-deoxy-D-glucopyranose prepared by either dilution or direct-linking*

The experiments were conducted exactly as described in Fig. 4. 'Binding' refers to the concentration of glucose in the fraction where the maximum glucokinase activity eluted from the column was found. Columns containing different ligand concentrations were prepared either by diluting a stock ligand–Sepharose conjugate (10μmol/g) with unsubstituted Sepharose (○) or by coupling the ligand to the CNBr-activated Sepharose directly at different ligand/gel ratios (●). Full details are given in the Materials and Methods section.

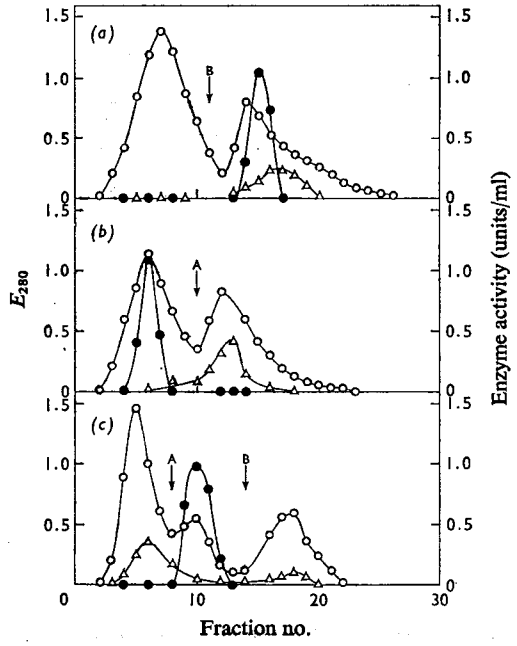


Fig. 6. *Non-specific binding of glucose 6-phosphate dehydrogenase to various Sepharose derivatives*

Liver extract (3ml) purified by DEAE-cellulose chromatography and containing 5 units of glucokinase and 2.5 units of glucose 6-phosphate dehydrogenase activity was applied to each column (5cm×0.8cm). The column and samples were equilibrated with the buffer described in Fig. 2 and the various additions were made where indicated: A, 1.0M-glucose; B, 0.5M-KCl. The columns were operated at 25ml/h, and 2ml fractions were collected. (*a*) and (*c*) Sepharose-*N*-(6-aminohexanoyl)-2-amino-2-deoxy-D-glucopyranose at coupled ligand concentrations of 4μmol/g wet weight of packed gel (*a*) and 2μmol/g of gel (*c*). (*b*) Sepharose-6-aminohexanoate (6μmol/g of gel). ○, E_{280}; ●, glucokinase activity; △, glucose 6-phosphate dehydrogenase activity.

glucokinase where, by using a ligand concentration of 2μmol/g obtained by dilution after covalently coupling the ligand to the CNBr-activated Sepharose at about 8μmol/g, the binding of non-specific proteins was lowered, thus facilitating a 20- to 50-fold purification step with a high yield in addition to that obtained on DEAE-cellulose (Holroyde *et al.*, 1975).

Other considerations

Glucokinase could also be displaced from the affinity column by including *N*-acetylglucosamine in the developing buffer. Elution by this compound, which is a competitive inhibitor with respect to glucose of the glucokinase reaction, did not offer any advantages over glucose elution since both enzyme preparations were of similar specific activity (results not shown in detail), but did substantiate the true affinity nature of the interactions between the glucokinase and the immobilized glucosamine ligand. Bound glucokinase could also be eluted by raising the KCl concentration, although in this case the eluted enzyme activity peak was of a much lower specific activity, because additional non-specific protein material had been removed from the column (Fig. 6*a*). For large-scale purifications, however, it seemed better to elute the enzyme from the column as a sharp peak by developing it with a high concentration of glucose (1M) rather than eluting by a glucose gradient and spreading the enzyme activity over a greater number of fractions with a concomitant decrease in specific activity of the final preparation.

Glucose was, nevertheless, eluting proteins other than glucokinase from these columns. Whether or not these proteins were binding specifically to the immobilized ligand was not investigated. Glucose 6-phosphate dehydrogenase (EC 1.1.1.49) has been identified (Storer, 1975) as one of the impurities and in this instance, at least, the interaction with the affinity matrix appeared to be non-specific. Although this dehydrogenase bound to the immobilized glucosamine derivative and could be displaced by either glucose or KCl (Fig. 6*a*), neither glucosamine, 6-aminohexanoic acid nor *N*-(6-aminohexanoyl)-2-amino-2-deoxy-D-glucopyranose inhibited the enzymic activity in free solution.

In Fig. 6(*a*), 0.5M-KCl was used as the eluent; a similar elution profile could be obtained with glucose except that less E_{280}-absorbing material was then released. Further, glucose 6-phosphate dehydrogenase bound to a Sepharose–6-aminohexanoate derivative and could be eluted with glucose from this matrix (Fig. 6*b*). Control of ligand concentrations was found to be effective in separating the two enzymic activities, for the binding of the dehydrogenase to the immobilized glucosamine derivative was only noticed at ligand concentrations of 4μmol/g or higher. If the columns were operated at 2μmol/g (the value chosen for large-scale purification work from the earlier studies) then most of the dehydrogenase passed through the column before the application of the glucose step, albeit in a position slightly behind the main protein peak (Fig. 6*c*).

Discussion

The results presented here indicate the necessity for careful control of the various parameters involved in the operation and positioning of an affinity-chromatography step in a purification scheme. This is particularly important if the enzyme to be purified is only present as a minor component in the starting material. In our case glucokinase represents only some 0.005% (w/w) of the total protein in the rat liver extracts after centrifugation. A series of different interactions must occur between the applied proteins and the gel matrix that, unless controlled, can severely limit the working life of the columns. Fortunately, these problems can be virtually eliminated by (*a*) extensive column-washing procedures between successive utilizations, (*b*) the introduction of other purification step(s) before the affinity-column step to increase the relative concentration of required enzyme in the applied protein and (*c*) careful control of immobilized ligand concentration.

The yield of glucokinase in the glucose eluates from these columns was always in excess of 80% even in large-scale preparations suggesting that its interaction with the affinity column was essentially of a specific nature. It was not determined whether the proteins other than glucokinase present in the glucose eluates from these affinity columns operated in the purification scheme were other glucosamine-binding proteins found in rat liver or whether they were the results of non-specific interactions. The latter seems most likely in view of the results obtained with glucose 6-phosphate dehydrogenase. The elution of this enzyme by glucose from immobilized ligands that do not affect its activity in free solution is difficult to explain. It only bound weakly to all the columns tested and may have been interacting, possibly through hydrogen bonds, with groups produced by the immobilization procedure (Ahrgren *et al.*, 1972) rather than any introduced ligands. Elution with glucose may have resulted from an effect on the hydration shell around the protein.

The results shown in Fig. 2 clearly indicate that most of the non-specific interactions are occurring either by an ion-exchange process [presumably due to the charged groups introduced by the CNBr-activation procedure (Ahrgren *et al.*, 1972)] or by hydrophobic interactions (presumably due to the presence of the pentamethylene 'spacer' group). These interactions are particularly troublesome and we have estimated that up to 25% (w/w) of the protein in a crude liver extract can be absorbed on a Sepharose–aminohexanoate matrix at a ligand concentration of 6μmol/g. The use of immobilized polymethylene ligands, often containing charged groups, in hydrophobic chromatography is well documented (Er-el *et al.*, 1972; Yon, 1972; Shaltiel & Er-el, 1973). The 'disappearance' of protein from the breakthrough peak of the direct-linked gels compared with the 'diluted' gels (Fig. 4) points also to the effects of the modified gel matrix (by the CNBr-activation procedure) in contributing to this non-specific binding. We were surprised, however, that the combined action of urea and KCl could not completely remove all protein from these columns. Additional protein could be removed after the washing procedure if the gels were left overnight in 8M-urea, or by washing with an organic solvent [aq. 80% (v/v) dioxan] (Lowe & Mosbach, 1975). These treatments, however, were still not sufficient to completely renew an 'aged' column and it was more convenient and efficient to incubate with Pronase (Fig. 3).

All of the columns used in this study were prepared by linking the pre-synthesized and characterized glucosamine ligands directly to the CNBr-activated Sepharose. The columns could have been prepared more simply by the methods outlined by Cuatrecasas (1970), i.e. preparation of a Sepharose–aminohexanoate derivative and coupling the glucosamine directly to this. Since the coupling is unlikely to be 100% efficient and is chemically ambiguous, our experience suggests that the non-specific binding

encountered with the defined ligands would have been considerably enhanced in this situation. In addition careful control of ligand concentration coupled to the gel would not have been possible. This was essential in order to help minimize non-specific binding and to maximize the purification of the enzyme.

The relationship between immobilized ligand concentration and the concentration of glucose required for elution is linear up to about 6 μmol/g. Above this concentration increasingly higher concentrations of glucose are required to achieve elution, suggesting that additional factors, such as hydrophobic binding, may well play an increasing role in binding the enzyme to the column. Up to 6 μmol/g it is most convenient to express the binding in terms of two competing equilibria, i.e., the competition of the immobilized ligand and the eluting molecule, glucose, for the same site on the enzyme. This is an oversimplified treatment since the presence of non-specifically bound protein may also indirectly influence these equilibria. The effect of ligand concentration on the binding of enzymes was first demonstrated by Harvey *et al.* (1974) in their small-scale model system studies. These workers (Lowe *et al.*, 1974*b*) have also shown that the total amount of immobilized ligand and even the column geometry must also be considered.

The dramatic effect of change in ligand concentration on the affinity chromatography of enzymes is probably only applicable in a situation as here where there is a relatively low association constant between the immobilized ligand and the enzyme. If the association constant is very high then it may not be possible to control the elution of the enzyme from the column so effectively by adjusting the immobilized ligand concentration.

The results obtained with the gels in which hydrophilic glycine arms were used were at first sight surprising. Both the diglycine and the 6-aminohexanoate derivatives of glucosamine were equally effective at inhibiting the glucokinase reaction in free solution ($K_i = 0.75$ mM). When attached to an agarose matrix, however, the glucosamine derivative with the hydrophobic 'spacer' moiety was much more effective as an affinity matrix. Barry & O'Carra (1973) and O'Carra *et al.* (1974) have suggested that the stronger binding at similar ligand concentrations found with the more hydrophobic derivative can be best explained by a hydrophobic interaction between the enzyme and the 'spacer' molecule reinforcing the specific enzyme–ligand interaction. This may well be the case in many systems (Lowe & Mosbach, 1975) but an alternative explanation is possible. The results with the glycyl derivative (Fig. 2) were exactly those found when the hexanoate–glucosamine derivative was operated at a ligand concentration of about 1 μmol/g. The hydrophilic glucosamine derivative may not have been as physically available for interaction with the enzyme, perhaps owing to hydrogen-bonding along its whole length to the polysaccharide matrix, thus decreasing the effective ligand concentration.

Although our objective was to determine as precisely as possible the best operating conditions for this affinity step in the large-scale purification of hepatic glucokinase (Holroyde *et al.*, 1975), the results obtained should be applicable to many enzyme purification systems, particularly where the enzyme under study is present in the tissue extract in very low amounts and only possesses a relatively low association constant for the immobilized ligand. It is noteworthy that if our preliminary experiments with this system had been conducted under different conditions, e.g. at a lower ligand concentration, then the lack of binding that would have been observed might well have discouraged us from further use of this method.

The early phase of this investigation, was supported in part by a grant to D. G. W. from the Wellcome Trust. We thank that body and the Medical Research Council for a grant to D. G. W., I. P. T. and Dr. A. Cornish-Bowden for studies on glucokinase. M. J. H. holds a Science Research Council Training Award.

References

Ahrgren, L., Kagedal, L. & Akerstrom, S. (1972) *Acta Chem. Scand.* **26**, 285–288
Ashwell, G. (1966) *Methods Enzymol* **8**, 93–95
Axen, R., Porath, J. & Ernbäck, S. (1967) *Nature (London)* **214**, 1302–1304
Barker, R., Olsen, K. W., Shaper, J. H. & Hill, R. L. (1972) *J. Biol. Chem.* **247**, 7135–7147
Barker, R., Chiang, C.-K., Trayer, I. P. & Hill, R. L. (1974) *Methods Enzymol.* **34**, 317–328
Barrolier, J. (1957) *Hoppe-Seyler's Z. Physiol. Chem.* **309**, 219–223
Barry, S. & O'Carra, P. (1973) *Biochem. J.* **135**, 595–607
Belleau, B. & Malek, G. (1968) *J. Am. Chem. Soc.* **90**, 1651–1652
Brodelius, P. & Mosbach, K. (1973) *FEBS Lett.* **35**, 223–226
Chesher, J. M. E., Trayer, I. P. & Walker, D. G. (1973) *Biochem. Soc. Trans.* **1**, 876
Crane, R. K. (1962) *Enzymes* **6**, 47–66
Cuatrecasas, P. (1970) *J. Biol. Chem.* **245**, 3059–3065
Cuatrecasas, P. (1972) *Adv. Enzymol. Relat. Areas Mol. Biol.* **36**, 28–89
Er-el, Z., Zaidenzaig, Y. & Shaltiel, S. (1972) *Biochem. Biophys. Res. Commun.* **49**, 383–390
Harvey, M. J., Lowe, C. R., Craven, D. B. & Dean, P. D. G. (1974) *Eur. J. Biochem.* **41**, 335–340
Holroyde, M. J. & Trayer, I. P. (1974) *Biochem. Soc. Trans.* **2**, 1310–1311
Holroyde, M. J., Allen, M. B., Storer, A. C., Warsy, A. S., Chesher, J. M. E., Trayer, I. P., Cornish-Bowden, A. & Walker, D. G. (1976) *Biochem. J.* **153**, 363–373

Lowe, C. R. & Mosbach, K. (1975) *Eur. J. Biochem.* **52**, 99–105

Lowe, C. R., Harvey, M. J. & Dean, P. D. G. (1974*a*) *Eur. J. Biochem.* **41**, 347–351

Lowe, C. R., Harvey, M. J. & Dean, P. D. G. (1974*b*) *Eur. J. Biochem.* **41**, 341–345

Mosbach, K., Guildford, H., Ohlsson, R. & Scott, M. (1972) *Biochem. J.* **127**, 625–631

Nicolas, J. C., Pons, M., Descomps, B. & Crastes de Paulet, A. (1972) *FEBS Lett.* **23**, 175–179

O'Carra, P., Barry, S. & Griffin, T. (1973) *Biochem. Soc. Trans.* **1**, 289–290

O'Carra, P., Barry, S. & Griffin, T. (1974) *FEBS Lett.* **43**, 169–175

Park, J. T. & Johnson, M. J. (1949) *J. Biol. Chem.* **181**, 149–151

Parry, M. J. & Walker, D. G. (1966) *Biochem. J.* **99**, 266–274

Shaltiel, S. & Er-el, Z. (1973) *Proc. Natl. Acad. Sci. U.S.A.* **70**, 778–781

Storer, A. C. (1975) Ph.D. Thesis, University of Birmingham

Storer, A. C. & Cornish-Bowden, A. (1974) *Biochem. J.* **141**, 205–209

Trayer, I. P. & Hill, R. L. (1971) *J. Biol. Chem.* **246**, 6666–6675

Trayer, I. P., Trayer, H. R., Small, D. A. P. & Bottomley, R. C. (1974) *Biochem. J.* **139**, 609–623

Yon, R. J. (1972) *Biochem. J.* **126**, 765–767

DOW CONFIDENTIAL - Do not share without permission

Technology Report The Dow Chemical Company		CRI Number Laboratory Report Code ML-AL MD-2008-006360
Department	Geographic Location	Date Issued
Analytical Sciences	Midland	1/28/2009
Page Count	Protocol Study Number	Report Status
19		Final

Title

Characterization of AAD-1 from transgenic maize

Author(s): Last Name and Initials (Master Numbers)	Author(s) Signature / Date
Karnoup, Anton (AS) (u369292)	01/28/09
Kuppannan, Krishna (K) (u386368)	01-28-2009
Reviewer Name(s)	**Reviewer and/or Supervisor Signature(s)/Date**
Young, Scott (SA) (U289561)	01-28-2009

Patent Status

Disclosure Submitted	Case Filed	No Action Required
[ ]	[ ]	[x]

Descriptive Summary and Conclusions

A sample of maize-derived aryloxyalkanoate dioxygenase (AAD-1) was submitted by Barry Schafer of Dow AgroSciences for characterization. In conjunction with Dow AgroSciences characterization, Analytical Sciences Laboratory was requested to provide analytical data on peptide mass fingerprinting and N-terminal and C-terminal sequencing. Peptide mass fingerprinting/mapping was accomplished by in-solution tryptic digest followed by ESI LC-MS and MALDI-TOF MS. The peptide mass fingerprinting resulted in 96.6% mass coverage for the maize-expressed AAD-1 recombinant protein sample. The N-terminal and C-terminal sequences for the AAD-1 sample were determined by a combination of in-solution digestion with trypsin, followed by LC-MS/MS. The tandem MS data for the N-terminal peptides revealed the following sequences: a mixture of *N-Acetyl*-AHAALSPLSQR and AHAALSPLSQR (un-acetylated). "Ragged N-terminal ends" were also detected (peptides corresponding to amino acid sequences HAALSPLSQR, AALSPLSQR, and ALSPLSQR). The tandem MS data for the C-terminal peptide revealed the following sequence: TTVGGVRPAR.

DISTRIBUTION LIST

Schafer, Barry - U097380 Biotech Reg Sciences, 306/A2/768, Indianapolis
Young, Scott (SA) GL Analytical Mol Spec, E89/1897, Midland
Clayton, Kathryn (KA) Regulatory Labs, 306/A2/775, INDIANAPOLIS
Deshmukh, Balasaheb (BK) Analytical Sciences, L08/1897, Midland

INTRODUCTION

A sample of recombinant aryloxyalkanoate dioxygenase (AAD-1) purified from transgenic maize was submitted by Barry Schafer of Dow AgroSciences for characterization. In conjunction with Dow AgroSciences characterization, Analytical Sciences Laboratory was requested to provide analytical data on peptide mass fingerprinting and N-terminal and C-terminal sequencing. Original experimental data are stored in the raw data packet ML-AL MD-2008-006360 [1].

EXPERIMENTAL

Sample Preparation:

A sample of recombinant affinity-purified AAD-1 (approximately 2.4 mL), was submitted by Barry Schafer (Dow AgroSciences, Indianapolis, IN). The protein isolation is documented in the study file of Dow AgroSciences report #:080142.

Prior to analysis by mass-spectrometry, the sample was dried in a centrifugal evaporator.

In-solution Protein Reduction, Alkylation and Processing with Trypsin:

Equipment:

a) Mettler AE168 analytical balance serial no. F00518
b) Eppendorf Centrifuge, Model 5415D, serial no. 5425 17645
c) Eppendorf, Thermomixer R, serial no. 5355 20846
d) Centrifugal evaporator (Centrivap), Labconco, cat. no. 7812013, S/N 051146935 A
e) Eppendorf adjustable pipettes: 2.5 µL serial no. 296447, 2-20 µL serial no. 286820, 10-100µL serial no. 289560, and 1000µL serial no. 33165
f) Fisher Vortex Genie 2, serial no. 2-156856
g) Siliconized microcentrifuge tubes, 1.5 mL, Fisher, cat no. 02-681-320
h) Parafilm
i) Eppendorf pipette tips (epTips) 10µL
j) Fisher brand Reditip General Purpose, 200µL and 1000µL

Reagents and Standards:

1. Fisher, acetonitrile, cat no. A998-1
2. Sigma, ammonium bicarbonate, cat no. A-6141
3. Pierce, Dithiothreitol (DTT), cat no. 20290
4. Sigma, Iodoacetamide (IAA), Sigma, cat no. I-1149
5. Roche, Trypsin, cat no. 1-418-025
6. Fluka, 98% Formic Acid (FA), Lot no. 1255194
7. Fisher, Trifluoroacetic acid (TFA), cat no. 04902-100
8. Deionized water

Reagent Solution Preparation:

a. 50 mM Ammonium Bicarbonate buffer: dissolved 197.66 mg NH_4HCO_3 in 50 mL of Milli-Q water, adjust pH to 7.5 with HCl.
b. DTT solution (100 mM; prepared fresh): dissolved 15.4 mg DTT in 1 mL of water.
c. Alkylating reagent (IAA) (200 mM; prepared fresh): dissolved 37 mg IAA in 1 mL of water.
d. Trypsin solution: Dissolved 100 µg of dried trypsin in 1 mL of 25 mM ammonium bicarbonate buffer immediately prior to digestion procedure.

In-solution Protein Processing (Reduction/ alkylation/ digestion with trypsin):

a. The dried sample was dissolved in 360 µL of protein dissolution buffer (6M guanidine hydrochloride/ 0.4M ammonium bicarbonate, pH 7.8), 40 µL of 100 mM DTT (reducing reagent) solution was added, the tube was sealed and the sample was incubated at 65°C for 30 minutes (in a thermomixer).
b. Sample was cooled to room temperature and 80 µL of 200 mM IAA (Iodoacetamide, alkylating reagent) solution was added to the tube. The sample was incubated in the dark at room temperature for 2 hours.
c. The sample was centrifuged briefly and desalting of the reduced/alkylated protein sample was performed using a NAP-5 gravity cartridge (Sephadex G-25) (Pharmacia Amersham, cat no. 52-2074-00), as per the manufacturer's procedure. NAP-5 cartridge was pre-equilibrated with 50 mM Ammonium Bicarbonate buffer, pH 7.5, and protein elution was performed with the same buffer (1 mL).

d. Approximately 100 μL of Trypsin solution was added to the tube with the desalted reduced/alkylated protein, and the sample was incubated for 1 hour at 37 °C. An additional 100 μL of Trypsin solution was added to the sample tube, and the digest was incubated for 16 hours (overnight) at 37 °C.
e. The digest was concentrated by drying in a centrifugal evaporator and re-dissolving in 150 μL of deionized water (to a final protein digest concentration of approximately 0.1 mg/mL).

MALDI MS of Tryptic digest

Reagents, Materials, and Standards:

1. siliconized 0.6-ml microcentrifuge tubes, Fisher, cat no. 02-681-330
2. Pipette tips 0.2-10 μL, Eppendorf
3. Zip Tips C18, Millipore, cat no. ZTC18S096
4. Milli-Q water (18 Mohm cm-1, TOC 30-20 ppb)
5. Acetonitrile (Baker analyzed HPLC solvent, JT Baker), Lot no. B51822
6. α-cyano-4-hydroxycinnamic acid (CHCA), Fluka, cat no. 28480
7. Trifluoroacetic acid (TFA), Fisher, cat no. 04902-100
8. ProteoMass MALDI-MS calibration kit, Sigma, cat. no. MS-CAL2

Equipment:

a) Applied Biosystems Voyager DE STR MALDI-TOF mass spectrometer, serial no. 4260
b) Applied Biosystems, Voyager Biospectrometry Workstation v5.1 software
c) Data Explorer Software v 4.0, part no. 4317717
d) Micromass, MassLynx v3.5 software
e) Applied Biosystems, Voyager Sample Plate, catalog no. v700666

Analytical Procedure:

a. Fifty μL of the reconstituted sample (concentrated tryptic digest) was aliquoted in a fresh 0.6-mL siliconized microfuge tube and purified using a Millipore C18 reversed-phase ZipTip. The ZipTip resin was activated with three 20-μL repetitions of 50% acetonitrile. Next, the resin was rinsed with three 20-μL repetitions of 0.1% aqueous trifluoroacetic acid. The peptide sample was

passed slowly through the fresh activated/equilibrated resin 10 times (used solution was returned back to vial with reconstituted peptide sample), and the bound sample was washed with three 20-µL repetitions of 0.1% aqueous trifluoroacetic acid. Finally, the peptide sample was eluted onto a MALDI plate step-wise with 3 µL each of 5%, 20%, 40%, 50%, 70%, and 100% acetonitrile containing 0.1% trifluoroacetic acid.

b. Each eluted ZipTip fraction on MALDI plate was covered with 1 µL of matrix solution (matrix solution was prepared in the following manner: Approximately 20 mg of α-cyano-4-hydroxycinnamic acid (CHCA) was transferred to a 1.5-mL microfuge tube. One mL of 50% acetonitrile containing 0.1% trifluoroacetic acid was added to the dry matrix. The matrix solution was mixed thoroughly by vortexing for 1 minute, then centrifuging for 1 min to sediment undissolved substance).

c. After allowing the samples to air dry, the peptide fragments were ready for analysis by MALDI-TOF MS. All mass spectra were acquired on an Applied Biosystems Voyager DE-STR MALDI-TOF mass-spectrometer equipped with a nitrogen laser (337 nm, 3-nsec pulse width, 20-Hz repetition rate). The following mass spectrometer settings were used:

Mode of operation: reflector
Extraction mode: delayed
Polarity: positive
Acquisition control: manual
Accelerating voltage: 20000 V
Grid voltage: 64%
Mirror voltage ratio: 1.12
Extraction delay time: 215 nsec
Acquisition mass range: 500-5500 Da
Number of laser shots: 200-250/spectrum
Laser intensity: 1300 - 2000
Low mass gate: 500-1900 Da
Timed ion selector: off

External mass calibration was performed with peptide standards utilizing a Sigma mass calibration kit (cat. no. MS-CAL2), consisting of the calibration mixture (monoisotopic $(M+H)^+$ m/z values used): Bradykinin (fragment 1-7), m/z 757.3997; Angiotensin II, m/z 1046.5423; $P_{14}R$ synthetic peptide, m/z 1533.8582; ACTH (fragment 18-39), m/z 2465.1989; Insulin oxidized B chain (bovine), m/z 3494.6513.

ESI/LC-MS and MS/MS of Tryptic digest

Reagents and Materials:

1. 0.1% formic acid in Acetonitrile (LC/MS reagent, JT Baker), Lot no. G15E12
2. 0.1% formic acid in water (J.T. Baker), Lot no. G16503
3. Milli-Q water
4. 98% Formic Acid (Fluka), Lot no. 1255194
5. Poly-DL-Alanine, Sigma, cat. no. P9003, Lot no. 97H5912
6. Leucine Enkephalin acetate salt, Sigma, cat. no. L-9133, Lot no. 095K5109
7. Waters total recovery HPLC vials, P/N 186000384c, lot no. 0384661180

Analytical Procedure:

ESI/LC/MS: The sample (tryptic digest, 150 μL) was centrifuged and transferred to a Waters HPLC vial, and analyzed by LC/MS. All mass spectra were acquired on a Waters Q-Tof micro MS system (S/N YA137). The following liquid chromatography and mass spectrometer settings were used:

LC :	Acquity UPLC system
Mobile Phase A :	0.1% formic acid (FA) in water
Mobile Phase B :	0.1% formic acid (FA) in acetonitrile
Column :	2.1x150mm Acquity BEH C18 1.7μm 135Å; S/N: 01245523640B05 Part No: 186002353
Flow rate :	100 μL/min
Column temperature :	50 °C
Injection volume :	20μL (full loop mode)
Injection loop :	20μL
UV detection :	214nm

Time, min	Flow rate, mL/min	% MPA	%MPB	Curve
Initial	0.100	95	5	Initial
5.00	0.100	95	5	6
63.00	0.100	60	40	6
64.50	0.300	60	40	6
69.00	0.300	10	90	6
70.00	0.300	10	90	6
71.00	0.300	95	5	6
79.00	0.300	95	5	6
80.00	0.100	95	5	6
85.00	0.100	95	5	6
90.00	0.000	95	5	6

MS :	Q-TOF micro mass spectrometer (S/N YA137)
ESI :	Micromass lock-spray electrospray interface
Mode :	+TOFMS
Scan :	350 to 1900 amu in MS mode; 90 to 1600 amu in MS/MS mode
Capillary :	2.85 kV
Sample Cone :	20 V
Extraction Cone :	1 V
Source Block :	90°C
Desolvation Temperature :	300°C
Desolvation Gas :	300 L/hr
MCP :	2350 V

Methods:

The sample was injected using the full loop configuration. After sample injection, the column was held at 5%MPB for 5 minutes. The gradient from 5%MPB to 40%MPB was then employed. At the end of the gradient, the MPB concentration was increased to 90% to allow removal of any hydrophobic components. The column was then re-equilibrated to the initial conditions.

The Time of Flight (ToF) analyzer was calibrated daily (for MS) using a 0.1 mg/mL (100 ppm) solution (in acetonitrile) of Poly-Alanine at 20 µL/min flow rate. For tandem MS (MS/MS) data acquisition, the fragments of Glu-Fibrinopeptide was used. For acquisition of lock-mass data, 2.5 µM Leucine-Enkephalin peptide solution (0.1% formic acid in 50% acetonitrile was used as the solvent) flowing at 3µL/min was used during LC-MS and LC-MS/MS experiments. Data acquisition was performed with a cycle time of 1 scan/sec (scan acquisition time: 0.88 sec; interscan delay: 0.1sec) in the MS and MS/MS modes. The lock mass channel was sampled every 7 sec during MS analysis and 10 sec during tandem MS analysis. The reference ion used was the singly charged Leucine-Enkephalin ion at m/z 556.2771.

In MS/MS mode, mass scan range set to 90-1600 amu. Tandem MS data acquisition used user-specified m/z analysis mode (where the precursor ion to be chosen was specified in a file). Charge state dependent collision energy profiles were used to apply the appropriate collision energies for the selected precursor ions.

RESULTS AND DISCUSSION

In this study, the numbering of the amino acid residues is in accordance with the theoretical sequence of the recombinant AAD-1 protein sample starting with Met^1 and containing a total of 296 amino acid residues (**Figure 1**).

Peptide Mass Fingerprinting:

MALDI MS and ESI/LC-MS analysis was used to generate peptide coverage map, N-terminal and C-terminal sequences, and to determine process sites using in-solution digestion with trypsin. **Figure 2** shows the LC-MS chromatograms (total ion current (TIC) *vs* elution time) of the tryptic digest preparation of the AAD-1 protein. **Figure 3** shows part of the same chromatogram: a zoom-in on the eluting N-terminal variant fragments (both TIC and UV-absorption at 214 nm *vs* elution time). A summary of mass spectral data plus assignments from MALDI MS and ESI/LC-MS analyses is presented in **Table I**. The overall sequence coverage was approximately 96.6% (**Figure 1**).

N-Terminal and C-Terminal Sequencing by LC Tandem MS:

N- and C-terminal peptides observed by LC-MS analyses were further analyzed by tandem MS to confirm their amino acid sequences. The results obtained for the N-terminal and C-terminal peptides are shown in **Tables II-IV**. Sequence tags were generated from the trypsin peptides with m/z 596.83 (N-terminal N-acetylated "full" peptide, $[M+2H]^{2+}$), m/z 575.82 (N-terminal un-acetylated "full" peptide, $[M+2H]^{2+}$), and m/z 507.30 (C-terminal peptide, $[M+2H]^{2+}$). LC tandem MS ion spectra were acquired for each individual peptide at specific retention time obtained in the peptide mass fingerprint study. The fragment mass ions for N-terminal peptide were consistent with the N-terminal peptide sequences: *N-Acetyl*-AHAALSPLSQR (**Table II**) and AHAALSPLSQR (un-acetylated) **(Table III**). "Ragged N-terminal ends" were also detected (peptides corresponding to amino acid sequences HAALSPLSQR, AALSPLSQR, and ALSPLSQR (**Figure 1**)).

The fragment mass ion assignments for C-terminal peptide were consistent with the C-terminal peptide sequence TTVGGVRPAR (**Table IV**).

Sequences of many of the internal tryptic peptides were confirmed by the observed in-source fragments (**Table I**).

REFERENCES

1. Raw data packet ML-AL MD-2008-006360.

Table I. A summary of peptide mapping of maize-derived AAD-1 based on MALDI MS and ESI/LC/MS.

Fragment #	# of missed cleavages	Start	End	Sequence	Charge State	Theoretical monoisotopic m/z	Retention Time, min	Observed m/z LC-MS	Observed in-source peptide fragments	% ACN (elution from C18-zt; MALDI)	Observed m/z MALDI MS	Modification	Comment
T1-Met	0	1	12	(-) MAHAALSPLSQR (F)	1	1281.67	ND	ND	ND	ND	ND	un-processed Met[1]	Met[1] is absent
	0	1	12		2	641.34							
	0	1	12		3	427.90							
T1	0	2	12	(-) AHAALSPLSQR (F)	1	1150.63	21.14	1150.64	942.54 (1+, y9), 871.50 (1+, y8), 848.41 (1+, b9), 800.46 (1+, y7), 761.42 (1+, b8), 687.38 (1+, y6), 648.34 (1+, b7), 630.36 (1+, b7-18), 600.34 (1+, y5), 551.30 (1+, b6), 533.28 (1+, b6-18), 503.30 (1+, y4), 464.26 (1+, b5), 390.21 (1+, y3), 351.18 (1+, b4)	5%, 20%, 25%	1150.69		sequence confirmed by MS/MS; **31% by UV, 32% by TIC**
	0	2	12		2	575.82		575.82					
T1 (N-Ac)	0	2	12	(-) Ac-AHAALSPLSQR (F)	1	1192.64	27.44	1192.64	942.51 (1+, y9), 871.49 (1+, y8), 800.46 (1+, y7), 687.38 (1+, y6), 600.34 (1+, y5), 593.31 (1+, b6), 575.30 (1+, b6-18), 540.30 (2+, y10), 506.27 (1+, b5), 393.20 (1+, b4), 390.22 (1+, y3), 365.17 (1+, a4)	5%, 20%	1192.75	N-Acetylated Ala	sequence confirmed by MS/MS; **3% by UV, 7% by TIC**
	0	2	12		2	596.83		596.83					
T1-cut1	0	5	12	(-?)(A?) ALSPLSQR (F)	1	871.50	22.56	871.50	687.38 (1+, y6), 600.34 (1+, y5), 503.29 (1+, y4), 390.21 (1+, y3)			3-aa truncation	possible: (a) truncation of N-terminus, (b) non-specific tryptic cleavage; **12% by UV, 12% by TIC**
	0	5	12		2	436.25		436.25					
T1-cut2	0	4	12	(-?)(H?) AALSPLSQR (F)	1	942.54	24.27	942.54	781.46 (1+, y8), 800.47 (1+, y7), 687.38 (1+, y6), 600.35 (1+, y5), 503.29 (1+, y4), 390.21 (1+, y3)			2-aa truncation	possible: (a) truncation of N-terminus, (b) non-specific tryptic cleavage; **6% by UV, 9% by TIC**
	0	4	12		2	471.77		471.77					
T1-cut3	0	3	12	(-?)(A?) HAALSPLSQR (F)	1	1079.60	20.71	1079.61	942.53 (1+, y9), 871.49 (1+, y8), 800.45 (1+, y7), 687.38 (1+, y6), 600.34 (1+, y5), 503.29 (1+, y4), 480.26 (1+, b5), 393.22 (1+, b4), 390.21 (1+, y3)	5%, 20%	1079.72	1-aa truncation	possible: (a) truncation of N-terminus, (b) non-specific tryptic cleavage; Coelutes with T16 peptide: if 1:1 amount in LC peak assumed -- **50% by UV, 40% by TIC**
	0	3	12		2	540.30		540.31					
T2	0	13	15	FER	1	451.23	5.50	451.22					
T3-nonspec	0	16	21	IAVQPL	1	640.39	32.74	640.39	509.30 (1+, b5), 412.26 (1+, b4), 357.22 (1+, y3)				non-specific tryptic cleavage
T3	0	16	35	IAVQPLTGVLGAEITGVDLR	1	2022.16	50.40		1362.79 (1+, b14), 1249.72 (1+, b13), 1143.63 (1+, y11), 1120.66 (1+, b12), 1049.60 (1+, b11), 1030.56 (1+, y10), 992.61 (1+, b10), 973.51 (1+, y9), 902.50 (1+, y8), 879.54 (1+, b9), 810.48 (2+, b17), 805.96 (2+, y16), 780.45 (1+, b8), 773.46 (1+, y7), 732.43 (2+, b15), 723.43 (1+, b7), 660.37 (1+, y6), 625.36 (1+, b13), 622.39 (1+, b6), 559.32 (1+, y5), 525.32 (2+, b11), 515.79 (2+, y10), 502.30 (1+, y4), 496.82 (2+, b10), 487.26 (2+, y9), 451.76 (2+, y8), 440.27 (2+, b9), 412.26 (1+, b4), 403.23 (1+, y3), 387.23 (2+, y7)	5%, 20%, 25%, 40%, 50%, 70%, 100%	2022.13		
	0	16	35		2	1011.58		1011.58					
	0	16	35		3	674.73		674.73					
T4-nonspec	0	36	53	EPLDDSTWNEILDAFHTY	1	2165.97	56.53		1300.56 (1+, b11), 1187.49 (1+, b10), 1058.47 (1+, b9), 992.96 (2+, b17), 979.50 (1+, y8), 944.41 (1+, b8), 942.44 (2+, b16), 873.91 (2+, b15), 866.41 (1+, y7), 800.38 (2+, b14), 753.33 (1+, y6), 707.34 (2+, b12), 650.80 (2+, b11), 638.30 (1+, y5), 594.25 (2+, b10), 567.26 (1+, y4), 420.20 (1+, y3)				non-specific tryptic cleavage
	0	36	53		2	1083.49		1083.49					
	0	36	53		3	722.66		722.67					
T4	0	36	73	EPLDDSTWNEILDAFHTYQVIYFPGQAIT NEQHIAFSR	1	4466.14	59.57		990.01 (2+, y17), 933.45 (3+, b23), 834.93 (2+, y15), 830.07 (3+, b21), 792.38 (3+, b20), 658.35 (2+, y11), 660.35 (3+, y17), 605.99 (3+, y16), 556.96 (3+, y15)	40%, 50%, 70%, 100%	4466.11		
	0	36	73		2	2233.57							
	0	36	73		3	1489.38		1489.42					
	0	36	73		4	1117.29		1117.32					
	0	36	73		5	894.03		894.06					
T6	0	75	85	FGPVDPVPLLK	1	1181.69	43.10	1181.70	1034.62 (1+, y10), 1035.59 (1+, b10), 977.59 (1+, y9), 922.49 (1+, b9), 880.54 (1+, y8), 781.49 (1+, y7), 712.37 (1+, b7), 666.45 (1+, y6), 613.32 (1+, b6), 569.35 (1+, y5), 516.25 (1+, b5), 489.3 (2+, y9), 470.33 (1+, y4), 401.22 (1+, b4), 373.23 (1+, y3)	20%, 25%, 40%, 100%	1181.70		
	0	75	85		2	591.35		591.34					
T7-nonspec	0	86	94	SIEGYPEVQ	1	1021.48	26.42	1021.49	875.43 (1+, b8), 776.36 (1+, b7), 550.26 (1+, b5), 472.24 (1+, y4)				non-specific tryptic cleavage
	0	86	94		2	511.25		ND					
T7	0	86	97	SIEGYPEVQMIR	1	1421.71	34.46	1421.70		5%, 20%, 25%, 40%, 50%	1421.75		
	0	86	97		2	711.36		711.36					
T9	0	99	105	EANESGR	1	762.34	4.21	762.34		50%	762.51		
T10	0	106	128	VIGDDWHTDSTFLDAPPAAVVMR	1	2513.21	43.86		1673.74 (1+, b15), 1602.69 (1+, b14), 1487.68 (1+, b13), 1374.60 (1+, b12), 1227.49 (1+, b11), 1151.03 (2+, y21), 1139.61 (1+, y11), 1104.50 (2+, b21), 1055.02 (2+, b20), 1026.54 (1+, y10), 1005.50 (2+, b19), 969.96 (2+, b18), 911.51 (1+, y9), 885.90 (2+, b16), 837.38 (2+, b15), 801.86 (2+, b14), 744.37 (2+, b13), 743.42 (1+, y7), 687.81 (2+, b12), 575.33 (1+, y5), 570.31 (2+, y11), 504.30 (1+, y4), 456.26 (2+, y9), 420.74 (2+, y8), 405.23 (1+. y3)	20%, 25%, 40%, 50%, 70%, 100%	2513.19		trace amount of Met127 oxydation also observed
	0	106	128		2	1257.11		1257.12					
	0	106	128		3	838.41		838.40					
	0	106	128		4	629.06		629.06					

Table I. (continued)

Fragment #	# of missed cleavages	Start	End	Sequence	Charge State	Theoretical monoisotopic m/z	Retention Time, min	Observed m/z LC-MS	Observed in-source peptide fragments	% ACN (elution from C18-zt; MALDI)	Observed m/z MALDI-MS	Modification	Comment
T11	0	129	170	AIDVPEHGGDTGFLSMYTAWETLSPTMQATIEGLNVVHSATR	1	4503.16	57.61		1290.11 (2+, b24), 1271.15 (2+, y23), 1246.57 (2+, b23), 1190.03 (2+, b22), 1178.11 (2+, y22), 1113.59 (2+, y21), 1063.07 (2+, y20), 1006.52 (2+, y19), 981.98 (2+, b19), 963.01 (2+, y18), 860.40 (3+, b24), 814.40 (2+, b16), 748.86 (2+, b15), 705.35 (2+, b14), 671.35 (3+, y19), 648.81 (2+, b13), 642.34 (3+, y18),	40%, 50%, 70%, 100%	4503.12		trace amount of Met oxydation also observed
	0	129	170		2	2252.08							
	0	129	170		3	1501.73		1501.80					
	0	129	170		4	1126.55		1126.55					
	0	129	170		5	901.44		901.42					
T12	0	171	181	VFGSLYQAQNR	1	1282.65	27.89	1282.66	1108.54 (1+, b10), 1036.52 (1+, y9), 994.49 (1+, b9), 979.49 (1+, y8), 892.47 (1+, y7), 795.40 (1+, b7), 779.38 (1+, y6), 616.32 (1+, y5), 592.30 (2+, y10), 518.77 (2+, y9), 504.28 (1+, b5), 488.26 (1+, y4), 417.22 (1+, y3), 391.20 (1+, b4)	5%, 20%, 25%, 40%, 70%	1282.62		
	0	171	181		2	641.83		641.83					
T14	0	183	189	FSNTSVK	1	782.40	11.87	782.41	635.34 (1+, y6), 548.31 (1+, y5), 434.26 (1+, y4)				
	0	183	189		2	391.71		391.71					
T15	0	190	198	VMDVDAGDR	1	977.44	17.67	977.44	878.37 (1+, y8), 747.33 (1+, y7), 688.33 (1+, b7), 632.30 (1+, y6), 560.29 (1+, b5), 533.23 (1+, y5), 445.22 (1+, b4), 418.21 (1+, y4)	5%	977.56		
	0	190	198		2	489.22		489.22					
T16	0	199	213	ETVHPLVVTHPGSGR	1	1585.85	20.80	1585.86	1119.59 (1+, y11), 1113.56 (1+, b10), 1022.56 (1+, y10), 976.54 (1+, b9), 958.53 (1+, b9-18), 909.49 (1+, y9), 758.42 (1+, b7-18), 711.35 (1+, y7), 659.35 (b6-18), 610.31 (1+, y6), 560.32 (2+, y11), 511.79 (2+, y10), 486.27 (3+, y14), 473.25 (1+, y5), 455.25 (2+, y9), 452.60 (3+, y13), 449.22 (1+, b4-18), 405.71 (2+, y8), 376.20 (1+, y4), 356.18 (2+, y7)	5%, 20%, 25%, 50%	1585.88		
	0	199	213		2	793.43		793.43					
	0	199	213		3	529.29		529.29					
T18	0	215	225	GLYVNQVYCQR	1	1399.68				5%, 20%	1399.82, 1400.56		possible Asn219 deamidation
T18-deamid	0	215	225	GLYV-isoD-QVYCQR	1	1400.66	38.72	1400.79	1067.59 (1+, y8)	5%, 20%	1399.82, 1400.56		Asn219 deamidation
	0	215	225		2	700.84		700.88					
T19	0	226	247	IEGMTDAESKPLLQFLYEHATR	1	2549.27	44.15		1036.51 (1+, y8), 952.01 (2+, y16), 916.50 (2+, y15), 812.75 (3+, y21), 769.73 (3+, y20), 776.37 (1+, y6), 744.40 (2+, y12)	20%, 25%, 40%, 50%, 70%, 100%	2549.24		
	0	226	247		2	1275.14		1275.15					
	0	226	247		3	850.43		850.43					
	0	226	247		4	638.07		638.07					
T19-KP	0	226	235	IEGMTDAESK	1	1080.49	16.60	1080.49	967.41 (1+, y9), 949.38 (1+, y9-18), 838.37 (1+, y8), 781.36 (1+, y7), 650.30 (1+, y6)				K-P cleavage
	0	226	235		2	540.75		**540.75**					
T20	0	248	253	FDFTCR	1	845.36	**ND**	**ND**	**ND**	**ND**	**ND**		
T24	0	259	272	DQVLVWDNLCTMHR	1	1786.84	**ND**	**ND**	**ND**	**ND**	**ND**		
T25	0	273	280	AVPDYAGK	1	820.42	15.75	820.43	650.31 (1+, y6), 553.27 (1+, y5), 438.24 (1+, y4)	5%, 50%	820.38		
T27	0	283	286	YLTR	1	552.31	10.72	552.31	389.25 (1+, y3)	5%	552.37		
T28	0	287	296	(R) TTVGGVRPAR (-)	1	1013.59	12.41	1013.58	695.40 (1+, y7-17), 482.27 (y4-17), 456.77 (2+, y9), 406.25 (2+, y8), 356.71 (2+, y7)	5%, 20%	1013.66		sequence confirmed by MS/MS
	0	287	296		2	507.30		**507.29**					
T3-4	1	16	73	IAVQPLTGVLGAEITGVDLREPLDDSTWNEILDAFHTYQVIYFPGQAITNEQHIAFSR	1	6473.24	63.24			50%, 70%, 100%	6473.24		
	1	16	73		2	3237.13							
	1	16	73		3	2158.42							
	1	16	73		4	1619.07		1619.18					
	1	16	73		5	1295.45		1295.37					
T5-6	1	74	85	RFGPVDPVPLLK	1	1337.79	38.69	1337.79	868.45 (1+, b8), 781.48 (1+, y7), 769.39 (1+, b7), 672.35 (1+, b6), 666.45 (1+, y6), 569.40 (1+, y5), 557.32 (1+, b5), 470.33 (1+, b4), 458.26 (1+, b4), 361.20 (1+, b3)	20%, 25%, 40%, 50%, 70%	1337.83		
	1	74	85		2	669.40		669.40					
	1	74	85		3	446.60		446.60					
T7-8	1	86	98	SIEGYPEVQMIRR	1	1577.81				40%, 50%, 70%	1577.84		
T8-9	1	98	105	REANESGR	1	918.44	3.87	918.38					
	1	98	105		2	459.72		**459.73**					
T13-14	1	182	189	RFSNTSVK	1	938.51	9.11	938.51	792.40 (1+, b7), 764.42 (1+, y7-18), 693.33 (1+, b6), 675.32 (1+, b6-18), 606.30 (1+, b5)	5%	938.61		
	1	182	189		2	469.76		**469.76**					
T17-18	1	214	225	KGLYVNQVYCQR	1	1527.77	34.32	1527.73					trace amount
	1	214	225		2	764.39		764.38					
T21-22	1	254	257	VRWK	1	588.36	22.27	588.31	489.24 (1+, y3)				
T24-25	1	259	280	DQVLVWDNLCTMHRAVPDYAGK	1	2588.24	44.11						trace amount
	1	259	280		2	1294.62		1294.62					
	1	259	280		3	863.42		863.42					
T27-28	1	283	296	(R) YLTRTTVGGVRPAR (-)	1	1546.88	45.92	1546.88					
	1	283	296		2	773.94		773.92					
	1	283	296		3	516.30		516.28					

Table II. Amino acid sequence obtained for *N-acetylated* N-terminal peptide (m/z 596.83; retention time 27.47 min) of maize-derived AAD-1.

Sequence: N-Acetyl-AHAALSPLSQR
Fragment ion masses: monoisotopic
Peptide mass $[M+2H]^{2+}$(monoisotopic): 596.83

Ion Table

	N-Ac-A	***H***	***A***	***A***	***L***	***S***	***P***	***L***	***S***	***Q***	***R***
a^{1+}	86.06	223.12	294.16	365.19	478.28	565.31	662.36	775.45	862.48	990.54	
		223.13	**294.16**	**365.21**	**478.26**	**565.31**					
b^{1+}	114.06	251.11	322.15	393.19	506.27	593.30	690.36	803.44	890.47	1018.53	
		251.13	**322.16**	**393.21**	**506.29**	**593.29**					
y^{1+}		1079.60	942.54	871.50	800.46	687.38	600.35	503.29	390.21	303.18	175.12
			942.55	**871.51**	**800.48**	**687.40**	**600.36**	**503.30**	**390.23**		**175.13**
y^{2+}		540.30	471.77	436.25	400.74	344.19	300.68	252.15	195.61	152.09	88.06
$(y\text{-}NH_3)^{1+}$		1062.57	925.51	854.47	783.44	670.35	583.32	486.27	373.18	286.15	158.09
						670.36	**583.34**	**486.30**	**373.20**	**286.16**	**158.10**

Top: theoretical m/z of fragment ions
Bottom: observed m/z of fragment ions

Table III. Amino acid sequence obtained for un-acetylated N-terminal peptide (m/z 575.78; retention time 21.19 min) of maize-derived AAD-1.

Sequence: AHAALSPLSQR
Fragment ion masses: monoisotopic
Peptide mass $[M+2H]^{2+}$(monoisotopic): 575.82

Ion Table

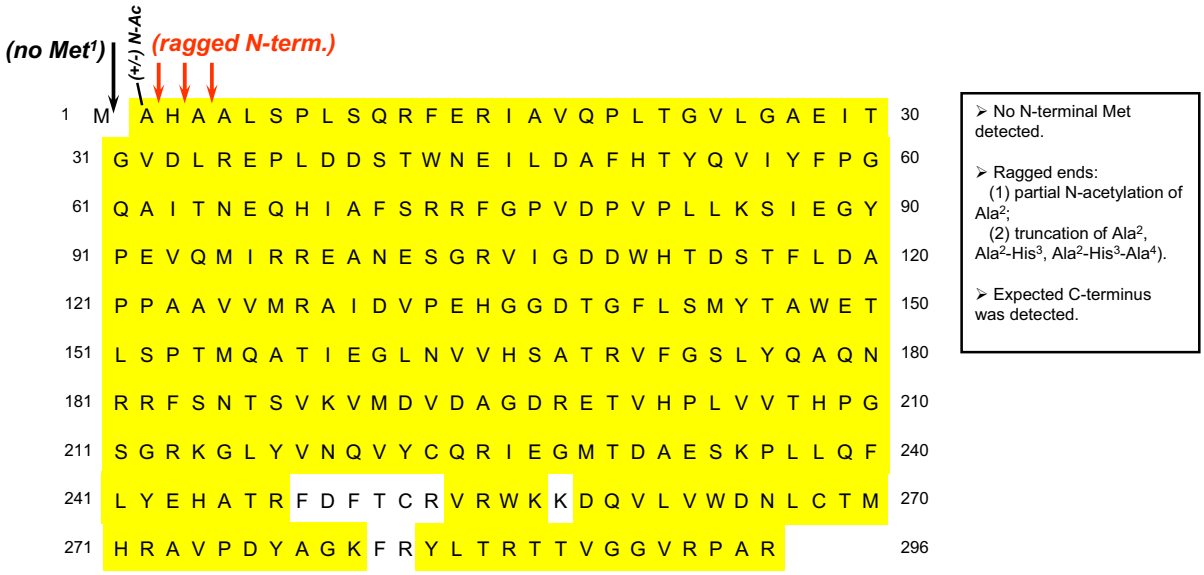
	A	***H***	***A***	***A***	***L***	***S***	***P***	***L***	***S***	***Q***	***R***
a(+1)	44.05	181.11	252.15	323.18	436.27	523.30	620.35	733.44	820.47	948.53	
		181.12		**323.20**	**436.28**	**523.33**					
b(+1)	72.04	209.10	280.14	351.18	464.26	551.29	648.35	761.43	848.46	976.52	
		209.11	**280.15**	**351.19**	**464.27**	**551.30**					
y(+1)		1079.60	942.54	871.50	800.46	687.38	600.35	503.29	390.21	303.18	175.12
			942.55	**871.52**	**800.48**	**687.38**	**600.36**	**503.29**	**390.22**	**303.19**	**175.13**
y(+2)		540.30	471.77	436.25	400.74	344.19	300.68	252.15	195.61	152.09	88.06
		540.31									
y-NH_3(+1		1062.57	925.51	854.47	783.44	670.35	583.32	486.27	373.18	286.15	158.09
						670.36	**583.33**	**486.24**	**373.20**	**286.16**	**158.10**

Top: theoretical m/z of fragment ions
Bottom: observed m/z of fragment ions

Figure 1. Theoretical amino acid sequence, sequence coverage (highlighted) and modifications observed in this study for maize-expressed AAD-1. Removal of N-terminal Met[1] (cleavage indicated by an arrow) and partial acetylation of the resulting N-terminal Ala was observed. N-terminal "ragged ends" were also observed. Highlighting indicates observed AAD-1 sequence fragments. Overall sequence coverage is 96.6% of theoretical value.

Maize AAD-1 sequence coverage from tryptic digest

Small polar:	**D(19)**	**E(16)**	**N(8)**	**Q(13)**			
Large polar:	**K(7)**	**R(23)**	**H(10)**				
Small non-polar:	**S(15)**	**T(23)**	**A(23)**	**G(21)**			
Large non-polar:	**L(22)**	**I(12)**	**V(28)**	**M(8)**	**F(13)**	**Y(10)**	**W(5)**
Special:	**C(3)**	**P(17)**					



Sequence coverage = 96.6%

Figure 2. LC/MS chromatogram for maize-derived AAD-1 (tryptic digest).

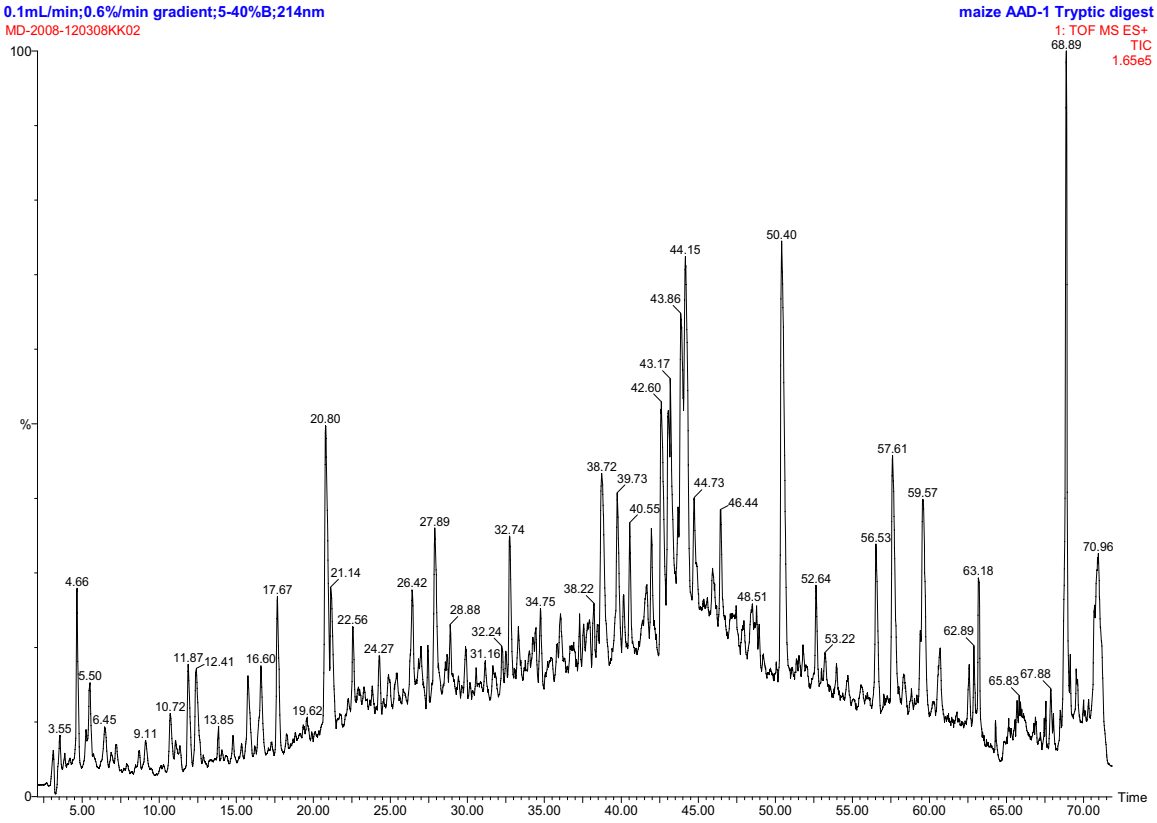
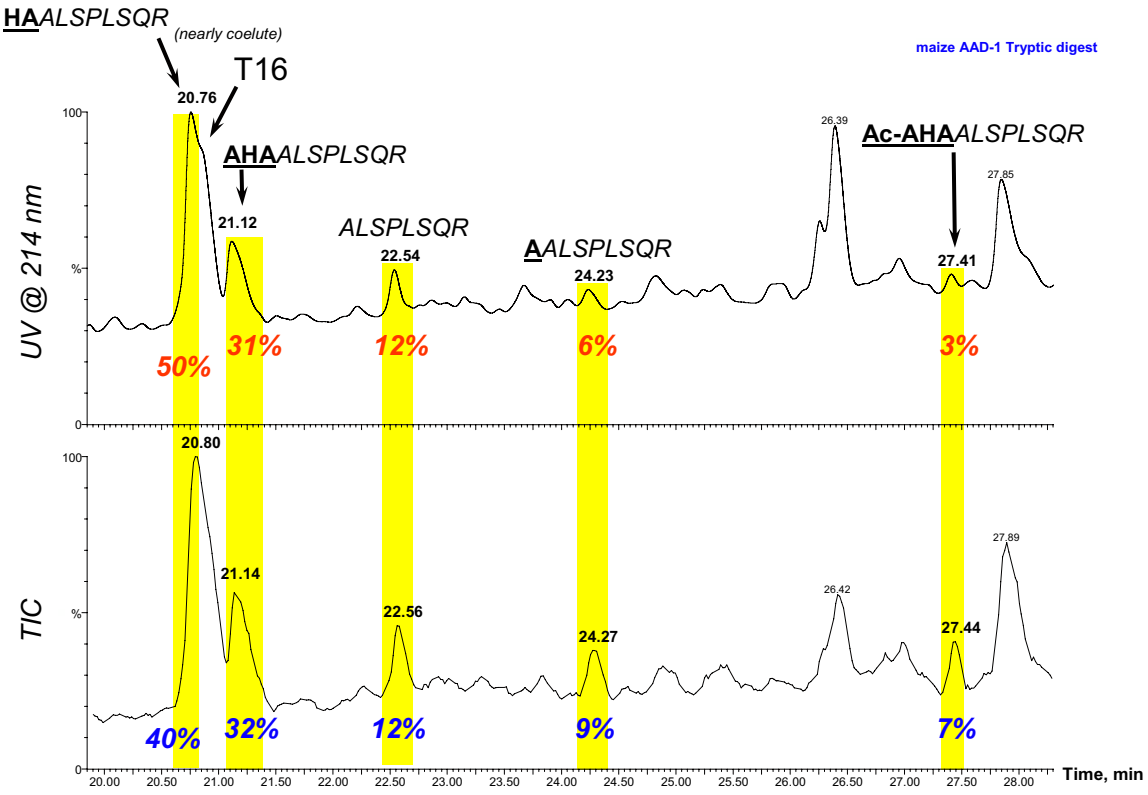


Figure 3. Part of the chromatogram of Figure 2, focusing on the eluting N-terminal fragment variants. Both UV-absorption at 214 nm (upper panel) and total ion current (TIC; lower panel) are shown. Rough estimates of relative fragment abundances were made based on the corresponding LC peak areas.



Journal of Chromatography, 281 (1983) 83–93
Elsevier Science Publishers B.V., Amsterdam — Printed in The Netherlands

CHROM. 16,190

IMMUNOCHEMICAL STUDIES OF THE NON-SPECIFIC INTERACTIONS OF CYANOGEN BROMIDE-ACTIVATED MACROPOROUS AGAROSE-BASED IMMUNOADSORBENTS

JOHN F. KENNEDY* and JUNOR A. BARNES
Research Laboratory for Bioactive Carbohydrates and Proteins, Department of Chemistry, University of Birmingham, P.O. Box 363, Birmingham B15 2TT (U.K.)

(First received November 22nd, 1982; revised manuscript received August 4th, 1983)

SUMMARY

This paper reports studies of the origin of the undesirable non-specific adsorption in immunoadsorption chromatography. The non-specificity of cyanogen bromide (CNBr)-activated macroporous agarose (Sepharose 4B)-based immunoadsorbents has been assessed from a comparative study of the following parameters: (1) The non-specific adsorption of protein (IgG) on unsubstituted cyanogen bromide-activated Sepharose 4B. (2) The non-specific adsorption of proteins (human serum IgG and sheep IgG) on cyanogen bromide-activated Sepharose 4B–sheep IgG immunoadsorbent. (3) The non-specific adsorption of proteins, immunoglobulin G(IgG) and human serum albumin (HSA) on cyanogen bromide-activated Sepharose 4B–sheep anti-human IgG immunoadsorbent and (4) the non-specific adsorption of peptides and amino acids on cyanogen-bromide activated Sepharose 4B–sheep anti-human IgG immunoadsorbent.

INTRODUCTION

The currently high popularity of solid phase techniques in chemical, biochemical and immunochemical procedures is an index of the intrinsic advantages afforded by immobilized reagents in analytical and preparative applications. Despite the invaluable contribution of immunoadsorption chromatography[1] for isolating antigens and antibodies, its operational efficiency has been severely compromised by various non-biospecific interactions resulting in rather poor small-scale separations and purifications. Furthermore, few successful large-scale separations have been reported in the literature.

The adsorption, onto an adsorbent, of a specific protein from a loaded heterogeneous sample is determined by the optimum conditions favouring complex formation between the immobilized ligand and the compound to be isolated. Such conditions are determined by the following parameters: the solvent[2]; the polarity of the eluent, its ionic strength and pH[3]; temperature[4,5]; charge of the activated matrix and ligand[6,7]; the ratio of the amount of covalently coupled ligand:activated matrix[7,8];

0021-9673/83/$03.00

the ratio of the amount of protein loaded:amount of ligand immobilized[8]; and the accessibility of the reactive groups and the degree of hydrophilicity of the matrix and ligand[5,7,9]. The successful application of immunoadsorption chromatography is largely attributed to the wide use of cyanogen bromide (CNBr)-derivatized macroporous agarose, introduced by Porath and co-workers[10]. The 4% (weight per expanded volume) beaded macroporous agarose available commercially from Pharmacia as Sepharose 4B is the most popular matrix for the synthesis of immunoadsorbents. The popularity of Sepharose has been attributed to its relative inertness[11] (*i.e.* the matrix does not act as an adsorbent), porosity[5], the apparent stability of the derivatized products[12] and their commercial availability.

Eveleigh and Levy[7] identified two types of non-specific interactions: (a) a non-reversible primary reaction that occurs in the first exposure of an immunoadsorbent to a complex biological sample, attributable to inadequate blocking of matrix active sites, remaining from coupling of the ligand to the matrix, with a low-molecular-weight reagent; (b) a reversible non-specific adsorption of certain components that results in a contamination of the eluted fractions and a partial blockage of otherwise specific binding sites. This latter proces presents a tedious problem in terms of obtaining highly pure products by the single step procedure of passing a heterogeneous sample such as serum through an immunoadsorbent column.

Non-specific interactions have been largely attributed to hydrophobic ligands or spacer molecules[13–16], charged groups[6,17–19], steric hindrance arising from the mode of attachment of the ligand[5,11] and occlusion of the ligand by the matrix backbone[5,11]. Although the salt sensitivity of the adsorption of some proteins suggests the possibility of some electrostatic interaction, some types of non-specific binding cannot be reversed by salt, polarity-reducing agents or denaturing agents[5,7,20]. In addition, some proteins such as albumin and ovalbumin[9,20] show irregular adsorption/elution features. O'Carra suggested that when biospecific adsorption and non-biospecific adsorption are balanced, they could act synergistically[5]. The obscure mechanisms of certain non-specific interactions and the severe disparity between the amount of adsorbed and eluted substances[5,7,8] prompted this investigation of the cyanogen bromide-activated matrix, the covalently coupled protein ligand and the loaded protein as possible sources of non-specific interference.

MATERIALS

Cyanogen bromide-activated macroporous agarose, CNBr-Sepharose 4B, was purchased from Pharmacia (Uppsala, Sweden). Sheep anti-human IgG (batch Z511G, 24.7 mg/ml), sheep anti-human human serum albumin (HSA) (batch Z464), donkey anti-sheep IgG (batch Z592A) and normal human serum standard (NIRDL, normal human serum standard code No. BR 99) were obtained from Seward Laboratories (London, U.K.). Agarose for immunochemical studies was purchased from Fisons (Loughborough, U.K.). The amino acid standard solution (2.5 μmole/ml) was purchased from Pierce (Rockford, IL, U.S.A.) and the peptides were obtained from Sigma (St. Louis, MO, U.S.A.). Chromatography columns GA 10 × 15 (15 × 1.0 cm) were bought from Wright Scientific (Kenley, U.K.). All materials used for preparing buffers and salt solutions were of AnalaR grade.

METHODS

Treatment of unsubstituted cyanogen bromide-activated Sepharose 4B

Cyanogen bromide-activated Sepharose 4B was swollen in and washed with 10^{-3} *M* hydrochloric acid to remove dextran and lactose stabilisers, filtered, blocked with aqueous ethanolamine solution (1 *M*) by gently stirring magnetically for one hour and then equilibrated in sodium citrate buffer (0.2 *M*, pH 6.5). After degassing, the gel (1 g dry gel, approximately 3.5 ml swollen) was packed in the column, washed with ammonia solution (0.5 *M*, pH 11.5) and then equilibrated in NaCl–Tris buffer [0.5 *M* NaCl, 0.1 *M* tris(hydroxymethyl)aminomethane–HCl, 0.5% (w/v) sodium azide] pH 8.0.

Aliquots (0.1 ml, 0.78 mg IgG) of pooled whole human serum were loaded in three serial adsorption–desorption cycles followed by loadings of 0.2-ml aliquots of the serum (1.56 mg IgG) for twelve cycles and 1-ml aliquots of the serum (7.8 mg IgG) for two subsequent cycles. After each loading the column was washed with NaCl–Tris buffer (pH 8.0) until the $A^{1\,cm}_{280\,nm}$ readings corresponded to that of the buffer alone. The column was next eluted with ammonia solution (0.5 *M*, pH 11.5) until the $A^{1\,cm}_{280\,nm}$ absorbances of the 4-ml fractions were negligible (≤ 0.005). After elution with ammonia solution the column was re-equilibrated in NaCl–Tris buffer (pH 8.0) before another cycle was commenced.

Preparation of immunoadsorbent columns

Cyanogen bromide-activated Sepharose–sheep anti-human IgG immunoadsorbent columns (1 g dry gel, ≈ 3 ml packed bed volume) for assessment of the non-specific adsorption of IgG and HSA from whole human serum, and of amino acids and peptides from test solutions, were prepared as described previously[19].

Prior to loading a column with serum it was washed with ammonia solution (0.5 *M*, pH 11.5) and then equilibrated with NaCl–Tris buffer (pH 8.0). Two 3-ml control immunoadsorbent (cyanogen bromide-activated Sepharose 4B–sheep anti-human IgG) columns for the adsorption of aminoacids and peptides contained 2.61 and 9.9 mg, respectively, sheep anti-human IgG per ml of swollen gel. The columns were washed with ammonia solution (0.5 *M*, pH 11.5) and equilibrated in phosphate buffered saline (0.01 *M* Na_2HPO_4/NaH_2PO_4, 0.15 *M* NaCl, pH 7.2), before being loaded with mixed amino acids and mixed peptides. Each column was loaded serially with 250 μl (625 nmole) amino acid standard solution or 0.5 ml (90–123 nmole) mixed peptide solution. The columns were then washed with phosphate buffered saline until the $A^{1\,cm}_{250\,nm}$ of the amino acid eluate was negligible (< 0.005) and the absorbance of the peptide eluate was also small (< 0.005). The columns were then eluted with ammonia solution (0.5 *M*, pH 11.5) and the ammonia eluates from the columns loaded with amino acids were monitored for absorbance at 280 nm while the eluates from the columns loaded with peptides were monitored at 220 nm. The pooled ammonia eluates from each adsorption–desorption cycle were analysed on a Locarte automatic amino acid analyser of the University of Birmingham Macromolecular Analysis Service.

Isolation of sheep IgG

Sheep IgG was isolated from 4 ml normal sheep serum (NSS) by precipitation

with 1.3 ml saturated ammonium sulphate (to give a final concentration of 33.3% (w/v) with respect to saturated ammonium sulphate). The precipitate was dissolved in phosphate buffer (0.03 *M*, pH 7.2) and dialysed extensively against saline (0.9% (w/v) sodium chloride). The dialysed fraction was next loaded onto an anion-exchange diethylaminoethyl (DEAE)-cellulose column (10 × 1 cm) equilibrated in phosphate buffer (0.03 *M*, pH 7.2) and subsequently eluted with the same buffer and collected in 4.0 ml fractions on an LKB Uvicord II automatic fraction collector at a flow-rate of 20 ml/h controlled by an LKB Perpex peristaltic pump. The column effluents were monitored at 280 nm and the IgG eluted in the first peak was pooled, dialysed extensively against saline and concentrated by negative-pressure dialysis[21]. The IgG purity was determined by Ouchterlony double immunodiffusion[22] and immunoelectrophoresis[23] using donkey anti-sheep serum. Sheep IgG (1 ml, 30 mg) equilibrated in sodium citrate buffer (0.2 *M*, pH 6.5) was added to cyanogen-activated Sepharose 4B (1 g dry gel) swollen in and washed with 10^{-3} *M* hydrochloric acid and equilibrated in sodium citrate buffer (6.0 ml). A 73-% coupling efficiency was achieved resulting in 27.9 mg IgG/3 ml packed gel volume.

Immunochemical assays

The pooled unadsorbed and pooled desorbed fractions from four 3-ml immunoadsorbent columns (CNBr-Sepharose 4B–sheep anti-human IgG) of identical binding capacity, were quantitated by the radial immunodiffusion method of Mancini[24], using the sheep anti-human albumin and donkey anti-sheep IgG to quantitate the unadsorbed and absorbed fractions, and the purity of the fractions was determined by immunoelectrophoresis[23] and Ouchterlony double immunodiffusion[22]. The potential antigen binding capacity (C) of an antibody immunoadsorbent column was derived from the equation: $C = (P - p)T$ where P is the amount of antibody added to the actual gel, p is the amount of uncoupled antibody and T is the titre of the antibody. The titre is defined as the amount of antigen in mg which reacts with a known amount of antiserum in mg at the equivalence point[25]. The equivalence point was determined by precipitin titration followed by manual nephelometric quantitation[26].

RESULTS

Non-specific adsorption of protein (IgG) on ethanolamine-substituted cyanogen bromide-activated Sepharose 4B

The contribution of the ethanolamine-substituted cyanogen bromide-activated Sepharose 4B to the non-specific adsorption of human serum IgG was demonstrated by a plot of the IgG adsorbed by and eluted from the ethanolamine-substituted cyanogen bromide-activated gel as a function of serial adsorption–desorption cycles (Fig. 1). When a 3-ml cyanogen bromide-activated Sepharose 4B column was loaded with whole human serum (0.1 ml, 0.78 mg IgG) for the first four adsorption–desorption cycles followed by a loading of 0.2 ml serum (1.56 mg IgG) for the succeeding ten cycles, there was no detectable adsorption of IgG on the cyanogen bromide-activated Sepharose matrix. However, after the fifteenth cycle when the IgG load (1 ml, 7.8 mg) was increased ten-fold with respect to the initial load, there was a retention of 1.9 mg (24.4% of the loaded IgG) and 1.6 mg (20.5%) in the sixteenth and seventeenth cycles respectively with the corresponding IgG elution of 0.46 mg (24.2% of the IgG retained) and 0.36 mg (22.5%).

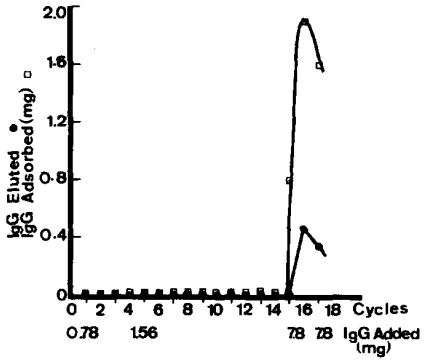


Fig. 1. Non-specific adsorption and elution profiles of human serum IgG on ethanolamine-substituted CNBr-activated Sepharose 4B.

The purity of the eluted IgG was confirmed by immunoelectrophoresis and two-dimensional electrophoresis with the aid of sheep anti-whole human serum. These results demonstrate that although the IgG in serum was not significantly adsorbed on the ethanolamine substituted cyanogen bromide-activated Sepharose 4B matrix after loadings of small volumes of serum in serial adsorption–desorption cycles, there was a pronounced retention of IgG when the volume of serum (1 ml) and hence the amount of IgG (7.8 mg) loaded, was increased ten-fold in the fifteenth cycle. The marked decrease in the amount of IgG retained on the matrix after the peak adsorption in the sixteenth cycle was indicative of the saturation of the column non-specific binding sites. Once a saturation level is attained as a result of specific or non-specific interaction, the amount of material adsorbed decreases asymptotically.

Non-specific adsorption of human serum IgG on a control adsorbent (CNBr-Sepharose 4B–sheep IgG) column

In order to assess the contribution of a macromolecular protein ligand to the

TABLE I

QUANTITATIVE RESULTS OF THE NON-SPECIFIC ADSORPTION OF HUMAN IgG ON A CONTROL ADSORBENT (CNBr-SEPHAROSE 4B–SHEEP IgG) COLUMN

Description	*Cycle*		
	1	*2*	*3*
Application			
IgG applied (mg)	2.32	2.32	2.32
Washing			
IgG unadsorbed (mg)	1.6	1.6	1.92
% IgG unabsorbed	69.0	69.0	83.0
Retention			
IgG adsorbed (mg)	0.72	0.72	0.40
% IgG adsorbed	31.0	31.0	17.2
Elution			
IgG eluted (mg)	Nil	Nil	Nil

TABLE II

QUANTITATIVE RESULTS OF THE NON-SPECIFIC ADSORPTION OF SHEEP SERUM IgG ON A CONTROL ADSORBENT (CNBr-SEPHAROSE 4B–SHEEP IgG) COLUMN

Description	*Cycle*		
	1	*2*	*3*
Application			
Sheep IgG loaded (mg)	2.7	2.7	5.4
Washing			
Sheep IgG unadsorbed (mg)	2.51	2.75	5.16
% sheep IgG unadsorbed	93.0	101.9	95.4
Retention			
Sheep IgG adsorbed (mg)	0.19	0	0.25
% sheep IgG adsorbed	7.0	0.0	4.6
Elution			
Sheep IgG eluted (mg)	Nil	Nil	Nil

phenomenon of non-specific adsorption, sheep IgG was employed as a model non-specific ligand instead of the specific sheep anti-human IgG ligand. When pooled whole human serum (0.2 ml, 2.32 mg IgG) was loaded onto a 3-ml (10.9 mg sheep IgG/ml gel) control adsorbent column (cyanogen bromide-activated Sepharose 4B–sheep IgG) in three successive adsorption–elution cycles and the unadsorbed protein was washed off with NaCl–Tris buffer (pH 8.0) followed by elution with ammonia solution (0.5 *M*, pH 11.5), the results of the column performance were as illustrated in Table I.

Non-specific adsorption of sheep serum IgG on a control adsorbent (cyanogen bromide-Sepharose 4B–sheep IgG) column

When a 3-ml control adsorbent column (cyanogen bromide-Sepharose 4B–sheep IgG) was loaded with normal sheep serum (0.1 ml, 2.7 mg IgG) in two successive adsorption elution cycles, followed by a loading of 0.2 ml serum (5.4 mg sheep IgG) in the third cycle and the unadsorbed protein in each cycle was washed off with NaCl–Tris buffer (pH 8.0) followed by elution with ammonia (0.5 *M*, pH 11.5); the results of the column performance were as illustrated in Table II.

Non-specific adsorption of proteins (IgG and HSA) on cyanogen bromide-activated Sepharose 4B–sheep anti-human IgG immunoadsorbent

The non-specific adsorption of protein associated with the interaction of a macromolecular antibody ligand (sheep anti-human IgG), covalently coupled to CNBr-activated Sepharose 4B, with the complementary antigen (human serum IgG) was assessed by loading immunoadsorbent columns of constant binding capacities with varying volumes of whole human serum and monitoring the adsorption and elution patterns of endogenous IgG (Fig. 2) and HSA (Fig. 3). The composite plots show the percentage IgG eluted as a function of the adsorption–desorption cycles resulting from the loading of four immunoadsorbent columns (3 ml) of identical binding capacities (1.6 mg IgG) with pooled whole human serum (0.1–0.7 ml). Although the antigen yield increased with increasing antigen load, the maximum yield

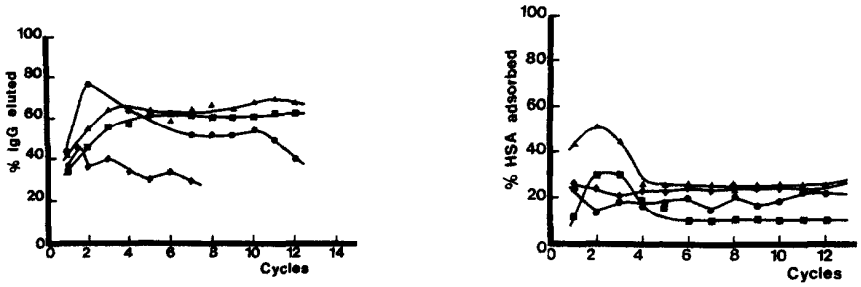


Fig. 2. Composite plots of the percentage human serum IgG eluted on serial adsorption–desorption cycles when increasing amounts of antigen in the form of whole human serum (WHS) were loaded on (CNBr-activated Sepharose 4B–sheep anti-human IgG) immunoadsorbent columns (3 ml gel) of the same potential binding capacity (1.6 mg IgG). Amounts of WHS loaded: ■, 0.1 ml (1.16 mg IgG); ▲, 0.15 ml (1.74 mg); ●, 0.4 ml (4.64 mg); ◆, 0.7 ml (8.12 mg).

Fig. 3. Composite plots of the percentage HSA adsorbed in serial adsorption–desorption cycles when increasing amounts of antigen in the form of whole human serum was loaded on (CNBr-activated Sepharose 4B–sheep anti-human IgG) immunoadsorbent columns (3 ml) of the same potential binding capacity. Amounts of HSA applied: ■, 2.85 mg; ▲, 4.275 mg; ●, 11.4 mg; ◆, 19.74 mg.

was obtained when the antigen load approximated to three-fold the potential binding capacity of the column. Composite plots of HSA adsorbed against cycles indicated the adsorption of 10–40% of the loaded HSA in the first cycles when the HSA loaded was 2.85 and 4.28 mg respectively. No significant amount of HSA was eluted with 0.5 *M* ammonia. Quantitation of the eluates by Mancini radial immunodiffusion revealed the eluted HSA to be less than 9 μg/ml (0.05–0.35% of the loaded HSA).

TABLE III

NON-SPECIFIC ADSORPTION OF AMINO ACIDS ON CYANOGEN BROMIDE-ACTIVATED SEPHAROSE 4B-SHEEP ANTI-HUMAN IgG IMMUNOADSORBENT

Amino acid applied	*Percent of applied amino acid adsorbed per cycle*		
	Cycle 1	*Cycle 2*	*Cycle 3*
L-Aspartic acid	55.6	0	0
L-Threonine	6.8	4.6	2.1
L-Serine	71.6	0	0
L-Glutamic acid	42.0	0	0
L-Proline	58.0	0	0
Glycine	8.8	1.9	0
L-Cystine	9.6	5.9	2.6
L-Valine	5.2	0	0
L-Methionine	5.2	1.0	0
L-Isoleucine	5.6	0	0
L-Leucine	9.6	0	0
L-Tyrosine	3.6	0	0
L-Phenylalanine	10.0	2.4	0
L-Histidine	9.0	1.4	0
L-Lysine	10.4	2.8	Not resolved
L-Arginine	20.8	7.3	Not resolved

TABLE IV

NON-SPECIFIC ADSORPTION OF PEPTIDES ON CYANOGEN BROMIDE-ACTIVATED SEPHAROSE 4B–SHEEP ANTI-HUMAN IgG IMMUNOADSORBENT

Peptide applied	*Percent of applied peptide adsorbed per cycle*					
	Cycle 1	*Cycle 2*	*Cycle 3*	*Cycle 4*	*Cycle 5*	*Cycle 6*
Glycyl-L-aspartic acid	0	0	0	0	0	0
Glycyl-L-serine	39.0	20.0	18.3	0	0	0
Glycyl-glycyl-glycine	0	0	0	0	0	0
Glycyl-L-phenylalanine	10	3	0	0	0	0
L-Phenylalanyl-glycine	6.0	0	0	0	0	0

Non-specific adsorption of peptides and amino acids on immunoadsorbent (cyanogen bromide-activated Sepharose 4B–sheep anti-human IgG) column

The adsorption characteristics of amino acids (Table III) and peptides (Table IV) on the test immunoadsorbent (cyanogen bromide-activated Sepharose 4B–sheep anti-human IgG) column proved useful models for exploring the cationic charge distribution on CNBr-activated polysaccharide matrix. A more distinct adsorption profile was apparent for the mixed amino acids loaded than for the loaded mixed peptides.

DISCUSSION

The chromatographic inertness of unsubstituted agarose matrix is commonly assumed by a majority of investigators, but in many cases the controls demonstrating negative adsorption on the matrix were inadequate. Small molecules such as ethanolamine or lysine are commonly employed to block residual active sites in the matrix after ligand substitution of the cyanogen bromide derivatized Sepharose matrix. Furthermore, the contribution of such molecules to non-specific binding has largely been assumed to be negligible although the non-specific adsorption of poly(A) in the nucleic acid field on ethanolamine-substituted Sepharose 2B has been described[27]. In the present work, although only the adsorption and elution of human serum IgG on ethanolamine-substituted cyanogen bromide-activated Sepharose 4B was monitored (Fig. 1), non-specific adsorption was dominant when the serum load was significantly increased. A negligible amount of protein was adsorbed when 0.2 ml of human serum (about 7% of the column volume) was loaded in ten serial adsorption–desorption cycles. The marked retention of IgG in the fifteenth cycle accompanied the ten-fold increase in the initial serum load. The decrease in the amount of IgG adsorbed in the sixteenth cycle indicates the saturation of the adsorptive sites in the column. The results illustrated in Fig. 1 demonstrate that overloading the Sepharose-based immunoadsorbent not only increases the non-specific adsorption of extraneous proteins from a heterogeneous sample but also severely limits the successful re-use of the column.

Although some of the protein could have been trapped in the interstices of the matrix or coupled to unblocked residual active groups, the simultaneous effects of electrostatic and non-ionic interactions appear to be implicated in the adsorption of

protein on cyanogen bromide-activated gel. As the cationic charges on the isourea derivatives[6,17] became neutralised, other non-covalent interactions would tend to become predominant. Jencks[28] suggested that hydrophobic bonding may be the most single important factor in non-covalent interactions in aqueous solutions where the strengths of electrostatic, charge transfer and hydrogen bonds are reduced by the charge solvating hydrogen bonding ability of water.

The results illustrated in Fig. 1 and in Table I suggest that the sheep IgG ligand contributed positively to non-biospecific adsorption. The 31% and 33% human IgG adsorbed on the adsorbent (cyanogen bromide-activated Sepharose 4B–sheep IgG) column in the first and second adsorption–elution cycles (Table I) respectively cannot be attributed to biospecific adsorption. In addition, comparable loadings of human serum on the ethanolamine-blocked cyanogen bromide-activated Sepharose 4B column (Fig. 1) showed no significant adsorption of human IgG after twelve serial adsorption–elution cycles. The adsorption of human serum IgG must be attributed mainly to the interaction of the loaded antigen with the sheep IgG ligand. The 50% decrease in the column binding capacity in the third adsorption–elution cycle (Table I) is another demonstration of the column saturation effect illustrated in Fig. 1.

In contrast, the loading of sheep serum on the control adsorbent (cyanogen bromide-activated Sepharose 4B–sheep IgG) column resulted in the adsorption of 7% and 0% IgG in the first and second adsorption–elution cycles respectively (Table II). However, a two-fold increase in the amount of sheep serum loaded resulted in only 4.6% adsorption of the loaded IgG. Although some of the IgG adsorbed in cycle 1 was washed off in the unadsorbed fraction in cycle 2, these results would suggest that there was no significant adverse non-specific interaction of sheep serum IgG with the sheep IgG ligand covalently coupled to the Sepharose matrix. This finding would justify the addition of sheep IgG to a coupling mixture in order to minimise the amount of specific ligand substituted. Nevertheless, the cummulative non-specific retention of small amounts of protein on adsorbent columns during serial adsorption–elution cycles would limit the useful re-use of the column by occlusion of active sites and contribute to contamination of eluates with the non-specifically adsorbed proteins.

It is noticeable that in cycles 1 and 3 (Table II) only 7% and 5% respectively of the loaded IgG were retained on the cyanogen bromide-activated Sepharose 4B–sheep IgG adsorbent column loaded with sheep serum. In contrast, 21% and 17.2% of the loaded IgG were retained in cycles 1 and 3 respectively on a similar adsorbent column when human serum was loaded (Table I). In both cases there was no detectable desorption of IgG with ammonia solution (0.5 *M*, pH 11.5). These results demonstrate the retention of protein on the columns, and indicate a stronger interaction of the cyanogen bromide-activated Sepharose 4B–sheep IgG adsorbent with human serum IgG than with sheep serum IgG.

The difference between the amount of protein (from whole human serum) adsorbed on and eluted from an immunoadsorbent (cyanogen bromide-activated Sepharose 4B–sheep anti-human IgG) column (true biospecific) is highlighted in Fig. 2 which shows the percentage of the adsorbed IgG eluted as a function of the adsorption–elution cycles. Although the yield of antigen increased with increasing antigen load, the highest antigen yield resulted when the antigen load approximated to three-fold the potential binding capacity of the column. This corresponded to a

mean percentage IgG elution of 55.0% of the IgG adsorbed in twelve serial adsorption–elution cycles. The non-specific adsorption of contaminated protein is illustrated by the adsorption of HSA from the whole human serum applied to the column as shown in Fig. 3. The HSA loading of 11.4 mg coincided with the IgG load of approximately three-fold the column potential binding capacity and resulted in comparatively low HSA adsorption. Despite the loading of the heterogeneous serum, the immunoadsorbent columns, which had the smallest volumes of serum loaded, showed no significant reduction in biospecific activity after twelve serial cycles (Fig. 2). The column which had the largest volume of serum loaded showed the greatest decrease in biospecific activity after six cycles (Fig. 2).

Many studies have been reported in the literature describing the interaction of enzymes with substituted cyanogen bromide-activated Sepharose 4B[14,15]. However, no report could be found describing the molecular interaction of amino acids and peptides with immunoadsorbents. The adsorption characteristics of amino acids (Table III) and peptides (Table IV) on the test immunoadsorbent (cyanogen bromide-activated Sepharose 4B–sheep anti-human IgG) gave some insights into the nature of protein–protein interactions with charged absorbents[19]. The significant adsorption of certain amino acids (L-serine, L-proline, L-aspartic acid and L-glutamic acid) and the dipeptide (L-glycyl-L-serine) indicate that contributions from ionic and some non-ionic effects could be concomitantly involved in the adsorption process. This is highlighted by the adsorption of the hydroxymonoamino-monocarboxylic α-amino acids L-serine (72%) and L-threonine (7%) and the cyclic α-amino acid L-proline (58%). The major difference between L-serine and L-threonine is the presence in L-threonine of a methyl group adjacent to the hydroxyl group. The strongly hydrophobic methyl group of threonine could have inhibited adsorption on the hydrophilic Sepharose beads as a result of steric hindrance and possibly hydrogen bonding with the lone pair electrons on the hydroxy group oxygen atom. In contrast, the adsorption of L-proline appears to be determined partly by the hydrophobic methylene groups of the pyrolidine ring. The adsorption of the mono-aminodicarboxylic α-amino acids, L-aspartic acid (56%) and L-glutamic acid (42%), also reflects the contribution of the charged polar groups at pH 6.7 to the adsorption process.

The comparative analysis of the adsorptive characteristics of four dipeptides and one tripeptide (Table IV) showed that only glycyl-L-serine was adsorbed significantly under the experimental conditions. The high adsorption in the first cycle of L-glycyl-L-serine was consistent with the high adsorption of serine (72%) in the first cycle. The high adsorption of aminoacids and peptides in the first cycle could also reflect some binding by unblocked active groups in the adsorbent. In addition, the comparatively small adsorption of peptides could be a reflection of the strong pH dependence observed by Joustra and Axen (1975) for the coupling of glycyl-L-leucine to cyanogen bromide-activated Sepharose 4B[29].

Particularly for immunoadsorption, the attractive features of Sepharose and other forms of macroporous agarose compared to other matrix materials are: their commercial availability; porosity of the beads which allows access to molecules in the million dalton range; their hydrophilic nature; relative chromatographic inertness[11]; and the apparent stability of the derivatized products[12]. Notable limitations of the agarose matrix include: the presence of ester sulphate groups and carboxy groups derived from actual formation with endogenous pyruvic acid; the cationic

Table IV. Amino acid sequence obtained for C-terminal peptide (m/z 507.3; retention time 12.3 min) of maize-derived AAD-1.

Sequence: TTVGGVRPAR
Fragment ion masses: monoisotopic
Peptide mass $[M+2H]^{2+}$(monoisotopic): 507.3

Ion Table

	T	***T***	***V***	***G***	***G***	***V***	***R***	***P***	***A***	***R***
a(+1)	74.06	175.11	274.18	331.2	388.22	487.29	643.39	740.44	811.48	
		175.12					**643.40**		**811.49**	
b-H_2O (+1)	84.04	185.09	284.16	341.18	398.2	497.27	653.37	750.43	821.46	
		185.10								
b(+1)	102.06	203.10	302.17	359.19	416.21	515.28	671.38	768.44	839.47	
		203.11							**839.47**	
y(+1)		912.54	811.49	712.42	655.4	598.38	499.31	343.21	246.16	175.12
			811.49	**712.42**	**655.40**		**499.32**	**343.22**		
y (+2)		456.77	406.25	356.71	328.20	299.69	250.16	172.11	123.58	88.06
		456.78	**406.26**							
y-H_2O(+1)		894.53	---	---	---	---	---	---	---	---
y-NH_3 (+1)		895.51	794.46	695.4	638.37	581.35	482.28	326.18	229.13	158.09
		895.58	**794.44**	**695.41**	**638.39**	**581.37**	**482.30**	**326.19**	**229.14**	

Top: theoretical m/z of fragment ions
Bottom: observed m/z of fragment ions

charge associated with cyanogen bromide-activated agarose; low mechanical and chemical rigidity; low effective ligand concentration within the gel matrix; solubility in hot water and non-aqueous solvents and the partial retention of biospecifically and non-biospecifically adsorbed molecules resulting in inefficiency and restrictions in re-use[8].

Despite the advantages afforded by cyanogen bromide-activated agarose for synthesizing adsorbents, the inherent nature of the support imposes certain restrictions in its efficient utilization[6–8,19]. The most effective chromatographic systems will best be achieved by the careful control of the matrix activation in order to limit the number of active sites generated and the cross-linking of the matrix[7] and careful attention to coupling and column operation protocols in order to obtain the best balance between the desired biospecific and the adventitious non-biospecific binding resulting from the gross electrostatic, hydrophobic and other non-covalent interactions. For achievement of this optimum an understanding at the planning stage of both the non-specific adsorption likely, as discussed in this paper, and the steps which may be taken (judicious choice of the gel: protein ratio, the composition of eluents and the amount of protein loaded onto the column), as discussed previously[8], are necessary.

REFERENCES

1 D. H. Campbell, E. Leuscher and L. S. Lerman, *Proc. Nat. Acad. Sci. U.S.*, 37 (1951) 575.
2 G. L. Amidon, R. S. Pearlman and S. T. Anik, *J. Theor. Biol.*, 77 (1979) 161.
3 P. Cuatrecasas, *Adv. in Enzymol.*, 36 (1972) 29.
4 T. K. Korpela and E. Mäkinen, *J. Chromatogr.*, 174 (1979) 361.
5 P. O'Carra, in B. Spencer (Editor), *Industrial Aspects of Biochemistry*, North-Holland Publishing Company, 1974, p. 107.
6 A. H. Nishikawa and P. Bailon, *Arch. Biochem. Biophys.*, 168 (1975) 576.
7 J. W. Eveleigh and D. E. Levy, *J. Solid Phase Biochem.*, 2 (1977) 45.
8 J. F. Kennedy and J. A. Barnes, *Int. J. Biol. Macromol.*, 2 (1980) 289.
9 B. H. J. Hofstee and N. F. Otillio, *J. Chromatogr.*, 159 (1978) 57.
10 J. Porath, R. Axén and S. Ernback, *Nature (London)*, 215 (1967) 1491.
11 J. Porath, *Methods Enzymol.*, 34 (1974) 13.
12 S. Hjertén, J. Rosengren and S. Påhlman, *J. Chromatogr.*, 101 (1974) 281.
13 Z. Er-El, Y. Zaidenzaig and S. Shaltiel, *Biochem. Biophys. Res. Commun.*, 49 (1972) 383.
14 B. H. J. Hofstee, *Anal. Biochem.*, 52 (1973) 430.
15 B. H. J. Hofstee, *Biochem. Biophys. Res. Commun.*, 50 (1973) 751.
16 A. H. Nishikawa and P. Bailon, *Anal. Biochem.*, 68 (1975) 274.
17 R. Jost, T. Miron and M. Wilchek, *Biochim. Biophys. Acta*, 362 (1974) 75.
18 R. L. Schnaar, T. F. Sparks and S. Roseman, *Anal. Biochem.*, 79 (1977) 513.
19 J. F. Kennedy, J. A. Barnes and J. B. Matthews, *J. Chromatogr.*, 196 (1980) 379.
20 T. Gribnau, Th. van Lith, A. van Sommeren, F. Roeles, H. van Hell and A. Schuurs, *Synthesis and application of immunosorbents based on halopyrimidine of reactive azo-dye activated polysaccharides —influence of ligand modification on the performance of immunosorbents, Third International Symposium on Affinity Chromatography and Molecular Interaction, Strasbourg, 1979.*
21 L. Hudson and F. C. Hay, *Practical Immunology*, Blackwell, Oxford, 1976, p. 196.
22 O. Ouchterlony, *Progr. Allergy*, 5 (1958) 1.
23 P. Grabar and C. A. Williams, *Biochim. Biophys. Acta*, 30 (1953) 365.
24 G. Mancini, A. D. Carbonara and J. F. Heremans, *Immunochemistry*, 2 (1965) 235.
25 J. Feger, in H. Peters (Editor), *Protides of the Biological Fluids*, Pergamon, Oxford, 1975, p. 495.
26 L. M. Killingsworth and J. Savory, *Clin. Chem.*, 18 (1972) 335.
27 J.-L. Ochoa, J. Kempf and J. M. Egly, *Int. J. Biol. Macromol.*, 2 (1980) 33.
28 W. P. Jencks, *Catalysis in Chemistry and Enzymology*. McGraw-Hill, New York, 1969.
29 M. Joustra and R. Axén, in H. Peters (Editor), *Protides of the Biological Fluids, XXII Colloquium, Brugge*, Pergamon, Oxford, 1975, p. 525.

DOW CONFIDENTIAL - Do not share without permission

Technology Report The Dow Chemical Company		CRI Number Laboratory Report Code ML-AL MD-2007-000127
Department	Geographic Location	Date Issued
Analytical Sciences	Midland	3/12/2007
Page Count	Protocol Study Number	Report Status
35		Final

Title

Characterization of AAD-1: Batch #1 480-14 and Batch #2 480-15

Author(s): Last Name and Initials (Master Numbers)	Author(s) Signature / Date
Kuppannan, Krishna (K) (U386368) Hung, Shao-Ching (SC) (U382160) Dielman, Demetrius (D) (U391474)	3.12.2007 3/12/2007 03-12-07
Reviewer Name(s)	Reviewer and/or Supervisor Signature(s)/Date
Young, Scott (SA) (U289561)	03-12-07

Patent Status

Disclosure Submitted	Case Filed	No Action Required
[ ]	[ ]	[ ]

Descriptive Summary and Conclusions

Two batches of purified recombinant aryloxyalkanoate dioxygenase (AAD-1) (batch #1: 480-14 and batch #2: 480-15) were submitted by Barry Schafer of Dow AgroSciences for characterization. In conjunction with Dow AgroSciences characterization, Analytical Sciences Laboratory was requested to provide analytical data on the native state of the intact protein, intact molecular weight, peptide mass fingerprinting, and N-terminal and C-terminal sequencing. One-dimensional SDS-PAGE analysis of both batches 480-14 and 480-15 resulted in an apparent molecular weight of 33kDa. Native state of AAD-1 for both batches were accomplished by size exclusion chromatographic (SEC) analyses. The molecular mass was determined to be approximately 148 kDa, which suggests either a tetrameric or a pentameric structure in the native state of AAD-1. Intact molecular weight analyses were accomplished by electrospray ionization/liquid chromatography/mass spectrometry (ESI/LC/MS). The mass spectrum revealed the presence of a principal mass component at m/z 33151 (des-Met[1]; batch #1: 480-14) and m/z 33150 (des-Met[1]; batch #2: 480-15). The experimentally observed mass des-Met[1] is within 0.01% of the calculated average mass of AAD-1 lacking a methionine based on the amino acid sequence. Peptide mass fingerprinting was accomplished by in-solution trypsin digests followed by ESI/LC/MS. The peptide mass fingerprinting resulted in 94.3% mass coverage for both AAD-1 recombinant protein samples. The N-terminal and C-terminal sequences for both batches were determined by a combination of in-solution digestion with trypsin, followed by tandem MS. The tandem MS data for both the N-terminal and C-terminal peptides revealed the following sequences AHAALSPLSQR and TTVGGVRPAR, respectively. In addition, an internal peptide was sequenced by a combination of in-solution trypsin and endoproteinase Asp-N.

DISTRIBUTION LIST

CRI, 566, Midland, MI
Schafer, Barry BIOTECH REG SCIENCES, 306/A2/768, INDIANAPOLIS
Clayton, Kathryn (KA) Regulatory Labs, 306/A2/775, INDIANAPOLIS
Young, Scott (SA) GL Analytical Mol Spec, E89/1897, Midland
Frawley, Nile (NN) Analytical Sciences, 1897B/B51, MIDLAND
O'Connor, Paul (PJ) Analytical Sciences, 1897 Bldg., Office E21, MIDLAND

TABLE OF CONTENTS

TABLE OF TABLES

TABLE OF FIGURES

INTRODUCTION

Two batches of purified recombinant aryloxyalkanoate dioxygenase (AAD-1) (Batch #1: 480-14 and Batch #2: 480-15) were submitted by Barry Schafer of Dow AgroSciences for characterization. In conjunction with Dow AgroSciences characterization, Analytical Sciences Laboratory was requested to provide analytical data on the native state of the intact protein, intact molecular weight, peptide mass fingerprinting, and N-terminal and C-terminal sequencing. Original experimental data are stored in the raw data packet ML-AL MD-2007-000127.

EXPERIMENTAL

Sample Preparation:

Two samples of recombinant purified AAD-1 (Batch #1: 480-14 and Batch #2: 480-15) approximately 25 mg of each was submitted by Barry Schafer (Dow AgroSciences, Indianapolis, IN). The sample was prepared as follows for the individual analysis:

1) SDS-PAGE, SEC, Intact MW –ESI/LC/MS: The lyophilized AAD-1 two batches were resuspended in PBS buffer at a final concentration of 1 mg/mL.
2) All other analysis: were prepared as described below.

SDS-PAGE Analysis:

Equipment:

a) Bio-Rad Criterion Cell cat. No. 165-6001
b) Bio-Rad PowerPac 1000 cat. # 165-5054
c) Traceable Digital Thermometer model # NEW 15-078J
d) Fisher brand Heating Block
e) Eppendorf Centrifuge, Model 5415D
f) Eppendorf pipette's 2-20 μL, and 10-100 μL adjustable pipette
g) Aros 160 Orbital Shaker
h) Fisher Vortex Genie 2
i) Eppendorf safe-lock microfuge tubes 1.5 mL, cat # 22 36 332-8
j) Bio-Rad gel loading tips, cat # 223-9917
k) Parafilm
l) Graduated cylinder, 1000 mL
m) Fluor-S MultiImager cat. # 170-7700, Quantity One Version 4.2 software

Reagents and Standards:

1. Laemmli Sample Buffer, Bio-Rad, lot no. 88934
2. β-mercaptoethanol, Fisher, Certified lot no. 004508
3. 4-20% Tris-HCl Criterion Precast Gel, Bio-Rad, lot no. C053101C1

4. Tris/Glycine/SDS Running Buffer, Bio-Rad, lot no. 68199A
5. Coomassie Stain Solution, Bio-Rad, lot no. 68198A
6. Destaining solution, Bio-Rad, lot no. 71169A
7. Certified Precision Unstained Protein Standards, Bio-Rad, lot no. 90310

Analytical Procedure:

The molecular weight of AAD-1 [Batch #1: 480-14 and Batch #2: 480-15] was determined by high-resolution SDS-PAGE gel electrophoresis analysis using an internal standard. The preparation of reagents, samples, and internal standards are shown below:

a. Laemmli sample buffer was prepared by adding ~50-μL of β-mercaptoethanol to ~950-μL of sample buffer. The sample buffer was thoroughly mixed by a vortex.
b. The test substances AAD-1 [Batch #1: 480-14 and Batch #2: 480-15] were diluted 1:1 in Laemmli sample buffer by transferring 10-μL of sample and 10-μL of sample buffer to a 1.5-mL microfuge tube. After briefly mixing the test substance in the Laemmli buffer, the microfuge tube was sealed with Parafilm, and placed in a pre-heated heat block set between 95-100 °C for ~1.5 minutes. The microfuge tube was removed from the heating block, water was wiped off, and the microfuge tube was briefly centrifuged.

Instrumental Conditions:

A 4-20%, Tris-HCl Criterion precast gel was removed from the storage container, the comb removed from the gel, the wells thoroughly rinsed with deionized water, and the tape removed from the bottom of the cassette. The Criterion gel was inserted into one of the slots in the Criterion tank. The upper buffer chamber of the Criterion gel was filled with approximately 1X Tris/Glycine/SDS (100 mL of 10X Tris/Glycine/SDS added to 900 mL of deionized water) premixed running buffer. The remaining running buffer was added to the lower buffer chamber. Approximately 14-μL of certified unstained precision protein standards, Bio-Rad, was loaded into 1 well with a pipette using gel-loading tips. Approximately 20-μL of each test substance was loaded into a well with a pipette using gel-loading tips (equivalent to ~10 μg). After applying the samples to the gel, the lid was placed on the tank, the electrical leads were plugged into the power supply, and the power was turned on. Fifty milliamps was applied to the Criterion cell at constant current for 1.25 hours until the dye reached the bottom of the gel. After the electrophoresis was complete, the power supply was turned off and the electrical leads were disconnected. The gel was removed from the Criterion gel cassette, transferred to Coomassie Stain Solution, Bio-Rad, in the gel cassette tray to cover gel ~40-mL, and placed on an orbital shaker at 35 rpm for approximately 30 minutes. The Coomassie Stain Solution was discarded and replaced with ~40-mL of Bio-Rad Destaining Solution, and placed on an orbital shaker at 35 rpm for ~30 minutes. The Destaining Solution was replaced with fresh Destaining Solution and the gel was destained for approximately 16 hours. Gel was stored in destain solution.

Methods for determining molecular weight of AAD-1: After destaining was complete, gel was removed, and an image was acquired by utilizing the Bio-Rad Fluor-S MultiImager, as specified by the manufacturer. The captured image was then analyzed using Quantity One-Version 4.2 software utilizing software tools for determining molecular weight. The molecular weight value was determined relative to the certified protein standards defined for the gel, the band's position in the lane.

Size Exclusion Liquid Chromatographic (SEC) Analysis of Intact Protein:

Reagents and Standards:

1. Phosphate buffered saline (PBS) solution provided by Barry Schafer.
2. Molecular weight standards: thyroglobulin (670 kDa), bovine gamma-globulin (158 kDa), chicken ovalbumin (44 kDa), equine myoglobin (17 kDa), vitamin B12 (1350 Da), Bio-Rad, cat. no. 151-1901; aprotinin (6500 Da), Sigma-Aldrich, cat. no. A3886.
3. Blue dextran 2000 (Amersham Pharmacia), cat no.17-0360-01, Lot no. 283873.
4. Mobile phase A – 10 mM PBS buffer, pH 7.4. (diluted from 10X PBS buffer pH 7.4, Ambion, cat. No. 9624, lot no. 045P20A).

Analytical Procedure:

AAD-1 batches [Batch #1: 480-14 and Batch #2: 480-15] were prepared in PBS buffer at 1 mg/mL concentration, and these sample preparations were analyzed directly by SEC. All SEC analyses were acquired on an Agilent 1100 system with the following conditions.

LC:	Agilent 1100 system
Columns:	Superdex 200 PC 3.2/30, GE Healthcare, cat no. 17-1089-01, Lot no. 10005918
Mobile phase:	100% A, isocratic
Flow (mL/min):	0.05
Column temp (°C):	Room temperature
UV detection:	260 and 280 nm
Injection volume:	10-μL

SEC analysis of molecular weight standards was performed before and after the analysis of AAD-1 batches for accurate determination of molecular weight. Blue dextran 2000 (1 mg/mL) was used for determination of column void volume.

ESI/LC-MS for Intact Protein:

Reagents and Standards:

1. Acetonitrile (Baker analyzed HPLC solvent, JT Baker), Lot no. B13814

2. Trifluoroacetic Acid (Fisher), Lot no. 032234
3. Deionized water from MilliQ
4. Poly-DL-Alanine, Sigma, Catalog no. P9003, Lot no. 97H5912
5. [Glu^1]-fibrino peptide B, Sigma, Catalog no. F3261.

Analytical Procedure:

ESI/LC/MS: The resulting sample preparations were analyzed directly by mass spectrometry. All mass spectra were acquired on a Waters Q-Tof Micro MS system (S/N YA137). The following mass spectrometer settings were used.

LC	: Acquity UPLC system
Mobile Phase A	: 0.01% trifluoroacetic acid in water
Mobile Phase B	: 0.008% trifluoroacetic acid in acetonitrile
Column	: 2.1x150mm Symmetry 300C18 3.5µm 300Å; S/N: 01283608610502 Part No: 186000188
Flow rate	: 100µL/min
Column temperature	: 50 °C
Injection volume	: 10µL
Injection loop	: 20µL
UV detection	: 214 nm

Time, min	Flow rate, mL/min	% MPA	%MPB	Curve
Initial	0.100	90.0	10.0	Initial
3.00	0.100	76.0	24.0	6
19.00	0.100	44.0	56.0	6
21.00	0.300	10.0	90.0	6
25.00	0.300	10.0	90.0	6
26.00	0.300	90.0	10.0	6
32.00	0.300	90.0	10.0	6
33.00	0.100	90.0	10.0	6
35.00	0.100	90.0	10.0	6

MS	: QTOFmicro mass spectrometer (S/N YA137)
ESI	: Micromass lock-spray electrospray interface
Mode	: +TOFMS
Scan	: 500 to 2400 amu (+)
Capillary	: 2.8 kV
Sample Cone	: 25 V
Extraction Cone	: 0.7V
Source Block	: 100°C
Desolvation Temperature	: 300°C
Desolvation Gas	: 350L/hr
MCP	: 2350 V

The Micromass supplied electrospray maximum entropy algorithm software (MAXENT 1) was used to transform the spectra to a mass axis and to resolution enhance the transformed spectra. The maximum entropy algorithm was set to optimize the spectra at peak widths of 0.35 Da with a resolution of

1Da/channel. The resulting resolution-enhanced spectral peaks were then integrated to display the correct ion abundance for distribution analysis.

In-solution Protein Processing with Trypsin and with Trypsin/AspN:

Equipment:

a) Mettler AE168 analytical balance serial no. F00518
b) Eppendorf Centrifuge, Model 5415D, serial no. 5425 17645
c) Eppendorf, Thermomixer R, serial no. 5355 20846
d) Eppendorf, Vacufuge, serial no. 5301 01600
e) Eppendorf pipette's, 2.5 μL serial no. 296447, 2-20 μL serial no. 286820, 10-100μL serial no. 289560, and 1000μL serial no. 33165 adjustable
f) Fisher Vortex Genie 2, serial no. 2-156856
g) Eppendorf safe-lock microfuge tubes 1.5 mL
h) Siliconized microcentrifuge tubes, 0.6 mL, Fisher
i) Syringe filters, sterile, 0.22 μm
j) Parafilm
k) Graduated cylinders 100, 250, and 1000-mL
l) Labonco Centrivap, serial no. 051146935 A
m) Eppendorf pipette tips (epTips) 10μL
n) Fisher brand Reditip General Purpose, 200μL and 1000μL
o) Aros 160 Orbital Shaker, serial no. 660980711950
p) VWR disposable/conical microcentrifuge tubes with attached caps, 0.65mL and 1.7mL
q) NAP-5 gravity cartridges (Sephadex G-25), Pharmacia Amersham, cat no. 52-2074-00
r) Bio-Rad gel loading tips
s) Graduated cylinders 100, 250, and 1000-mL

Reagents and Standards:

1. Bio-Rad, Tris, cat no. 161-0715
2. Fisher, acetonitrile, cat no. A998-1
3. Sigma, ammonium bicarbonate, cat no. A-6141
4. Pierce, Guanidine hydrochloride, (Gu:HCl), 8M solution, cat. No. 24115
5. Pierce, Dithiothreitol (DTT), cat no. 20290
6. Sigma, Iodoacetamide (IAA), Sigma, cat no. I-1149
7. Roche, Trypsin, cat no. 1-418-025
8. Roche, Endoproteinase Asp-N, cat. No. 11-054-589-001
9. Deionized water

Reagent Solution Preparation:

a. Protein dissolution buffer (6M Gu:HCl/ 400 mM ammonium bicarbonate, pH 7.8): to 316 mg of ammonium bicarbonate, add 7.5 mL 8 M Gu:HCl solution. Add 2.5 mL water. Adjust pH to 7.8 with NaOH. Filter through 0.22 μm sterile syringe filter.
b. 25 mM Ammonium Bicarbonate buffer: dissolve 98.83 mg NH_4HCO_3 in 50 mL of Milli-Q water. pH should be ~ 8 without adjustment. Filter through 0.22 μm sterile syringe filter.
c. DTT solution (100 mM; prepare fresh): dissolve 15.4 mg DTT in 1 mL of water.
d. Alkylating reagent (IAA) (200 mM; prepare fresh): dissolve 37 mg IAA in 1 mL of water.
e. Trypsin solution: Once prepared, solution can be used within 2 days. Dissolve contents of 1 vial (25 μg) of dried trypsin with 500 μL of 25 mM ammonium bicarbonate buffer.
f. 100 mM Tris buffer (digestion buffer): dissolve 121.1 mg Tris in 10-mL of Milli-Q water. Adjust pH to 8-8.5 with HCl. Filter through 0.22 μm sterile syringe filter.

In-solution Protein Processing (Reduced/alkylated):

a. Reduction and carboxymethylation of proteins: approximately 720-µL of protein dissolution buffer, 6M guanidine hydrochloride/0.4M ammonium bicarbonate, pH 7.8, was added to 1 mg of AAD-1 [Batch #1: 480-14 and Batch #2: 480-15] and samples were mixed by pipette action. Eighty microliters of 100 mM DTT (reducing reagent) solution was added to each tube. Tubes were sealed, vortexed, and incubated at 65 ºC for 40 min in a thermomixer at 1100 rpm. Tubes were then cooled to room temperature, centrifuged for 30 sec. and 160 µL of 200 mM IAA (alkylating reagent) solution was added to each tube. Tubes were incubated in the dark at room temperature for 1 hour. Three hundred and twenty microliters of DTT solution was added to consume unreacted IAA and the tubes were allowed to stand for 20 min at room temperature. The total reaction volume is approximately 1,280 µL.
b. Desalting of the reduced/alkylated protein samples were performed using NAP-5 gravity cartridges (Sephadex G-25) as per the manufacturer's procedure. NAP-5 cartridges were pre-equilibrated with 100 mM Tris buffer, pH 8.4, and protein elution was performed with the same buffer (final volume 1-mL).
c. Tryptic digestions of reduced/alkylated proteins: in parallel, 100-µL of trypsin solution was added to the 1-mL of reduced/alkylated protein samples. The digests were incubated for 16 hrs at 37 ºC in a thermomixer at 900 rpm. Samples were evaporated to dryness in a Speedvac and resuspended in dionized Milli-Q water to bring the final concentration to approximately 2.5 mg/mL.
d. In solution protein processing with Asp-N after digestion: approximately 500 µg of dried trypsin digests was reconstituted with 100 μL of MilliQ water prior to Asp-N digestion. Asp-N (2 μg) was reconstituted with 2 μL of MilliQ water, and the 100 μL of trypsin digests was added to the tube

containing Asp-N. The whole solution was then transferred to a microfuge tube and sealed with parafilm followed by incubation at 37 °C for 2 hours.

ESI/LC-MS and MS/MS of Trypsin digest

Reagents:

1. Acetonitrile (Baker analyzed HPLC solvent, JT Baker), Lot no. B51822
2. MilliQ water
3. 98% Formic Acid (Fluka), Lot no. 1255194
4. Poly-DL-Alanine, Sigma, Catalog no. P9003, Lot no. 97H5912
5. [Glu^1]-fibrino peptide B, Sigma, Catalog no. F3261.

Analytical Procedure:

ESI/LC/MS: The resulting sample preparations were reconstituted by dissolving in water to a concentration of approximately 2 µg/µL and analyzed directly by mass spectrometry. All mass spectra were acquired on a Waters Q-Tof Micro MS system (S/N YA137). The following liquid chromatography and mass spectrometer settings were used:

LC : Acquity UPLC system
Mobile Phase A : 0.1% formic acid in water
Mobile Phase B : 0.1% formic acid in acetonitrile
Column : 2.1x150mm Acquity BEH C18 1.7µm 135Å; S/N: 01245523640B05 Part No: 186002353
Flow rate : 100µL/min
Column temperature : 50 °C
Injection volume : 10µL
Injection loop : 20µL
UV detection : 214nm

Time, min	Flow rate, mL/min	% MPA	%MPB	Curve
Initial	0.100	95.0	5.0	Initial
5.00	0.100	95.0	5.0	6
63.00	0.100	60.0	40.0	6
64.00	0.300	60.0	40.0	6
65.00	0.300	10.0	90.0	6
70.00	0.300	10.0	90.0	6
71.00	0.300	95.0	5.0	6
79.00	0.300	95.0	5.0	6
80.00	0.100	95.0	5.0	6
85.00	0.100	95.0	5.0	6

MS : QTOFmicro mass spectrometer (S/N YA137)
ESI : Micromass lock-spray electrospray interface
Mode : +TOFMS
Scan : 350 to 1900 amu (+)
Capillary : 2.8 kV
Sample Cone : 27 V
Extraction Cone : 1V

Source Block	: 100°C
Desolvation Temperature	: 250°C
Desolvation Gas	: 500L/hr
MCP	: 2350 V

Methods:

Both batches [Batch #1: 480-14 and Batch #2: 480-15] were injected using the partial loop configuration. For most of the analyses, 20µg of AAD-1 trypsin digest was used.

After sample injection, the column was held at 5%MPB for 5 minutes. The gradient from 5%MPB to 40%MPB was then employed. At the end of the gradient, the MPB concentration was increased to 90% to allow removal of any hydrophobic components. The column was then re-equilibrated to the initial conditions.

Electrospray ionization (ESI) source with a lockspray interface was used. The capillary was held at 2800V and sample cone was set at 27V. The collision energy was held at 6V. The MCP detector was held at 2350V.

The Time of Flight (ToF) analyzer was calibrated daily using a 20µM solution (0.1% formic acid in 98% acetonitrile was used as the solvent) of Poly Alanine at 10µL/min flow rate. The same instrument parameter file (with the calibration parameters) was used for both MS and MS/MS data acquisitions. The calibration was checked with 1µM [Glu^1]-fibrino peptide solution (0.1% formic acid in 50% acetonitrile was used as the solvent) flowing at 3µL/min; the same solution was used for acquiring lock-mass data during LC-MS and LC-MS/MS experiments. Data acquisition was performed with a cycle time of 1 scan/sec (scan acquisition time: 0.88sec; interscan delay: 0.1sec) in the MS and MS/MS modes. The lock mass channel was sampled every 7 sec during MS analysis and 10 sec during tandem MS analysis. The reference ion used was the doubly charged [Glu^1]-fibrino peptide ion at m/z 785.8426.

In MS/MS mode, default collision energy was set to 23V and the mass scan range set to 100-1900 amu. Tandem MS data acquisition used user-specified m/z analysis mode (where the precursor ion to be chosen was specified in a file). In the user-specified m/z analysis mode, the precursor ion with its elution time (with a 60sec time window) was specified. Collision energies were also specified for each of the precursor ions.

RESULTS AND DISCUSSION

In this study, the numbering of the amino acid residues is in accordance with the theoretical sequences of the recombinant AAD-1 protein sample starting with Met^1 and containing a total of 296 residues (**Figure 1**).

SDS-PAGE:

The enriched AAD-1 [Batch #1: 480-14 and Batch #2: 480-15] were resolved by high resolution SDS-PAGE (**Figure 2**). One predominant band with an apparent molecular weight of 33 kDa was resolved in both batches. The gel image was acquired by a Fluoro-S MultiImager.

Size Exclusion Chromatography:

The molecular weight of the native AAD-1 was determined by size exclusion chromatography (**Figure 3**). A molecular mass of 148 kDa was obtained for both batches [Batch #1: 480-14 and Batch #2: 480-15] based on the molecular weight calibration curve at the fractionation range for Superdex 200 column in PBS buffer at room temperature (**Figure 3d**). The observed molecular weight is between a tetrameric and a pentameric structure based on the calculated subunit molecular mass, 33.15 kDa.

ESI Intact Mass Spectral Characterizations:

The purified AAD-1 [Batch #1: 480-14 and Batch #2: 480-15] were processed in two different methods. The batches were either directly solubilized in PBS or solubilized in protein dissolution buffer followed by reduction with DTT and alkylation with iodoacetamide. The solubilized proteins were then analyzed by ESI/LC/MS using a Symmetry C18 column for separation. The chromatography for AAD-1 [Batch #1: 480-14 and Batch #2: 480-15] solubilized in PBS revealed the presence of one major peak at retention times of 13.36 and 13.3 min, respectively (**Figures 4** and **5**). The chromatography for reduced and alkylated AAD-1 [Batch #1: 480-14 and Batch #2: 480-15] revealed the presence of one major peak at retention times of 13.42 and 13.36 min, respectively (**Figures 6** and **7**). Unique features of the spectra were the broad charge distribution and partial resolution of the peaks. Ions related to the monomer form of AAD-1 were observed in each sample lot. For the batches dissolved in PBS, the transformed and integrated maximum entropy spectra revealed the presence of a principal mass component at m/z 33151 (Batch #1: 480-14), and m/z 33150 (Batch #2: 480-15) (**Figures 4** and **5**). For the reduced and alkylated batches, the transformed and integrated maximum entropy spectra revealed the presence of a principal mass component at m/z 33322 (Batch #1: 480-14), and m/z 33322 (Batch #2: 480-15) (**Figures 6** and **7**). The calculated molecular weight for des-Met[1]AAD-1 and reduced and alkylated des-Met[1]AAD-1 are 33153 Da (m/z 33154^{+1}) and 33324Da (m/z 33325^{+1}), respectively. Ions related to the monomer form of AAD-1 (des-Met[1]) were observed for both batches under non-reduced and reduced conditions. Thus the observed masses are within 0.01% of the theoretical molecular weight of AAD-1 (des-Met[1]) (**Table I**).

Peptide Mass Fingerprinting:

ESI/LC/MS analysis was used to generate peptide coverage maps, N-terminal and C-terminal sequence, and determination of process sites using in-solution digest with trypsin for both batches. As shown in **Figure 8,** the LC-MS chromatograms of the trypsin digests of both batches are very similar. A summary

of mass spectral data plus assignments from ESI/LC/MS analysis is presented in **Table II**. The sequence coverage's are approximately 94.3% for both sample batches [Batch #1: 480-14 and Batch #2: 480-15]. All peptides were further analyzed by LC tandem MS to confirm their sequences.

LC Tandem MS:

All peptides observed by LC-MS analyses were further analyzed by tandem MS to confirm their amino acid sequences. The results obtained for the N-terminal, C-terminal, and T4-0 peptide from the two different lots [Batch #1: 480-14 and Batch #2: 480-15] are shown in **Tables III-XIV**. Sequence tags were generated from the trypsin peptides with m/z 575.8 (N-terminal peptide, $[M+2H]^{2+}$), m/z 507.3 (C-terminal peptide, $[M+2H]^{2+}$), and m/z 1117.65 (internal peptide T4-0, $[M+4H]^{4+}$). LC tandem MS ion spectra were acquired for each individual peptide at specific retention time obtained in the peptide mass fingerprint study. Multiple tandem MS experiments were performed for each sample, with different collision energy parameters. During manual processing of the tandem MS data, for most of the peptides, the scans from more than one experiment was summed to improve the S/N ratio and sequencing performed on the resulting spectrum. Due to the fact that T4-0 peptide is large and highly charged, tandem MS spectra of this peptide does not have sufficient b, and y ions to completely sequence the peptide. Hence, the trypsin digests from both batches [Batch #1: 480-14 and Batch #2: 480-15] were further subjected to endoproteinase Asp-N digest. Selected peptides that cover critical residues (aa 32-37 with m/z 742.6^{+1}; aa 38-46 with m/z 1092.7^{+1}; aa 39-46 with m/z 977.7^{+1}) in the T4-0 tryptic peptide were subjected to tandem MS analysis. The fragment mass ions for N-terminal peptide for all sample lots were consistent with the N-terminal peptide sequence H-AHAALSPLSQR-OH (**Tables III** and **IX**). The fragment mass ion assignments for C-terminal peptide from both batches were consistent with the C-terminal peptide sequence H-TTVGGVRPAR-OH (**Tables IV** and **X**). The fragment mass ion assignments for the T4-0 peptide from sample 17 were consistent with the peptide sequence H-EPLDDSTWNEILDAFHTYQVIYFPGQAITNEQHIAFSR-OH (**Table V**). In addition, the Asp-N fragments from both batches [Batch #1: 480-14 and Batch #2: 480-15] with m/z 742.6, m/z 1092.7, and m/z 977.7 were sequenced by LC tandem MS; and the corresponding fragment mass ion assignments were consistent with the sequences (32-37) H-DLREPL-OH, (38-46) H-DDSTWNEIL-OH, and (39-46) H-DSTWNEIL-OH, respectively (**Tables VI, VII, VIII, XI, XII, XIII**). The tandem MS data for the rest of the peptides are included in the raw data packet.

REFERENCE

1. Raw data packet ML-AL MD-2007-000127.

Table I: Molecular weight of intact AAD-1 by ESI/LC/MS

Sample Lot #	Residues	Processing	Mass Charge	Theoretical	Observed
Batch #1: 480-14	2-296 (des-Met[1])	Reduced and alkylated	M^{+1}	33325	33322
Batch #2: 480-15	2-296 (des-Met[1])	Reduced and alkylated	M^{+1}	33325	33322
Batch #1: 480-14	2-296 (des-Met[1])	in PBS	M^{+1}	33154	33151
Batch #2: 480-15	2-296 (des-Met[1])	in PBS	M^{+1}	33154	33150

Table II: Trypsin digest peptide mass fingerprinting of AAD-1 [Batch #1: 480-14 and Batch #2: 480-15].

Frag #	Residue #	Sequence	Charge State	Theoretical m/z	Observed m/z (Batch#1: 480-14)	Observed m/z (Batch#2: 480-15)
T1-0	1--11	(-) AHAALSPLSQR(F)	1	1150.63	1150.63	1150.62
T1-0	1--11	(-) AHAALSPLSQR(F)	2	575.82	**575.82***	**575.82***
T2-0	13-15	(R) FER(I)	1	451.23	451.23	451.23
T3-0	15-34	(R) IAVQPLTGVLGAEITGVDLR(E)	2	1011.58	1011.58	1011.58
T3-0	15-34	(R) IAVQPLTGVLGAEITGVDLR(E)	3	674.73	674.72	674.73
T3-1	15-72	(R) IAVQPLTGVLGAEITGVDLREPLDDSTWNEILDAFHTYQVIYFPGQAITNEQHIAFSR(R)	5	*1295.45*	*1295.54*	*1295.41*
T3-1	15-72	(R) IAVQPLTGVLGAEITGVDLREPLDDSTWNEILDAFHTYQVIYFPGQAITNEQHIAFSR(R)	6	*1079.71*	*1079.76*	*1079.70*
T3-1	15-72	(R) IAVQPLTGVLGAEITGVDLREPLDDSTWNEILDAFHTYQVIYFPGQAITNEQHIAFSR(R)	4	*1619.07*	*1619.15*	*1619.09*
T3-1	15-72	(R) IAVQPLTGVLGAEITGVDLREPLDDSTWNEILDAFHTYQVIYFPGQAITNEQHIAFSR(R)	5	*1295.45*	*1295.73*	*1295.68*
T3-1	15-72	(R) IAVQPLTGVLGAEITGVDLREPLDDSTWNEILDAFHTYQVIYFPGQAITNEQHIAFSR(R)	6	*1079.71*	*1079.86*	*1079.83*
T3-1	15-72	(R) IAVQPLTGVLGAEITGVDLREPLDDSTWNEILDAFHTYQVIYFPGQAITNEQHIAFSR(R)	4	*1619.07*		*1619.18*
T3-3	15-84	(R) IAVQPLTGVLGAEITGVDLREPLDDSTWNEILDAFHTYQVIYFPGQAITNEQHIAFSRRFGPVDPVPLLK(S)	7	*1114.13*		*1114.14*
T3-3	15-84	(R) IAVQPLTGVLGAEITGVDLREPLDDSTWNEILDAFHTYQVIYFPGQAITNEQHIAFSRRFGPVDPVPLLK(S)	6	*1299.65*		*1299.62*
T3-3	15-84	(R) IAVQPLTGVLGAEITGVDLREPLDDSTWNEILDAFHTYQVIYFPGQAITNEQHIAFSRRFGPVDPVPLLK(S)	5	*1559.38*		*1559.31*
T4-0	35-72	(R) EPLDDSTWNEILDAFHTYQVIYFPGQAITNEQHIAFSR(R)	3	1489.38	1489.28	1489.34
T4-0	35-72	(R) EPLDDSTWNEILDAFHTYQVIYFPGQAITNEQHIAFSR(R)	4	1117.29	**1117.27***	**1117.28***
T5-1	73-84	(R) RFGPVDPVPLLK(S)	1	1337.79	1337.76	1337.77
T5-1	73-84	(R) RFGPVDPVPLLK(S)	2	669.40	669.40	669.40
T6-0	74-84	(R) FGPVDPVPLLK(S)	1	1181.69	1181.68	1181.68
T6-0	74-84	(R) FGPVDPVPLLK(S)	2	591.35	591.34	591.34
T7-0	85-96	(K) SIEGYPEVQMIR(R)	1	1421.71	1421.70	1421.67
T7-0	85-96	(K) SIEGYPEVQMIR(R)	2	711.36	711.35	711.36
T8-1	97-104	(R) REANESGR(V)	1	918.44	ND	ND
T9-0	98-104	(R) EANESGR(V)	1	762.34	ND	ND
T10-0	105-127	(R) VIGDDWHTDSTFLDAPPAAVVMR(A)	2	1257.11	1257.11	1257.10
T10-0	105-127	(R) VIGDDWHTDSTFLDAPPAAVVMR(A)	3	838.41	838.41	838.40
T11-0	128-169	(R) AIDVPEHGGDTGFLSMYTAWETLSPTMQATIEGLNVVHSATR(V)	3	1501.73	1501.70	1501.72
T11-0	128-169	(R) AIDVPEHGGDTGFLSMYTAWETLSPTMQATIEGLNVVHSATR(V)	4	1126.55	1126.54	1126.55
T11-0	128-169	(R) AIDVPEHGGDTGFLSMYTAWETLSPTMQATIEGLNVVHSATR(V)	5	901.44	901.41	901.42
T12-0	170-180	(R) VFGSLYQAQNR(R)	1	1282.65	1282.64	1282.64
T12-0	170-180	(R) VFGSLYQAQNR(R)	2	641.83	641.83	641.83
T13-1	181-188	(R) RFSNTSVK(V)	1	938.51	938.49	938.50
T13-1	181-188	(R) RFSNTSVK(V)	2	469.76	469.75	469.76
T14-0	182-188	(R) FSNTSVK(V)	1	782.40	782.40	782.40
T14-0	182-188	(R) FSNTSVK(V)	2	391.71	391.71	391.71
T15-0	189-197	(K) VMDVDAGDR(E)	1	977.44	977.43	977.43
T15-0	189-197	(K) VMDVDAGDR(E)	2	489.22	489.22	489.22
T15-1	189-212	(K) VMDVDAGDRETVHPLVVTHPGSGR(K)	2	1272.64	1272.61	1272.61
T15-1	189-212	(K) VMDVDAGDRETVHPLVVTHPGSGR(K)	3	848.76	848.76	848.76
T15-1	189-212	(K) VMDVDAGDRETVHPLVVTHPGSGR(K)	4	636.82	636.82	636.82
T15-1	189-212	(K) VMDVDAGDRETVHPLVVTHPGSGR(K)	5	509.66	509.65	509.65
T16-0	198-212	(R) ETVHPLVVTHPGSGR(K)	2	793.43	793.43	793.43
T16-0	198-212	(R) ETVHPLVVTHPGSGR(K)	3	529.29	529.28	529.28
T17-1	213-224	(R) KGLYVNQVYCQR(I)	2	764.39	764.41	764.39
T18-0	214-224	(K) GLYVNQVYCQR(I)	1	1399.68	1399.64	1399.66
T18-0	214-224	(K) GLYVNQVYCQR(I)	2	700.34	700.34	700.34
T19-0	225-246	(R) IEGMTDAESKPLLQFLYEHATR(F)	2	1275.14	1275.14	1275.13
T19-0	225-246	(R) IEGMTDAESKPLLQFLYEHATR(F)	3	850.43	850.43	850.43
T19-0	225-246	(R) IEGMTDAESKPLLQFLYEHATR(F)	4	638.07	638.07	638.07
T20-0	247-252	(R) FDFTCR(V)	1	845.36	845.36	845.36
T20-0	247-252	(R) FDFTCR(V)	2	423.18	423.18	423.18
T21-1	253-256	(R) VRWK(K)	1	588.36	ND	ND
T22-1	255-257	(R) WKK(D)	1	461.29	ND	ND
T23-1	257-271	(K) KDQVLVWDNLCTMHR(A)	2	957.97	957.96	957.97
T23-1	257-271	(K) KDQVLVWDNLCTMHR(A)	3	638.98	638.98	638.98
T24-0	258-271	(K) DQVLVWDNLCTMHR(A)	2	893.92	893.93	893.92
T24-0	258-271	(K) DQVLVWDNLCTMHR(A)	3	596.28	596.28	596.29
T25-0	272-279	(R) AVPDYAGK(F)	1	820.42	820.42	820.42
T26-1	280-285	(K) FRYLTR(T)	1	855.48	ND	ND
T27-0	282-285	(R) YLTR(T)	1	552.31	552.31	552.31
T28-0	286-295	(R) TTVGGVRPAR(-)	1	1013.59	1013.57	1013.57
T28-0	286-295	(R) TTVGGVRPAR(-)	2	507.30	**507.29***	**507.29***
monoisotopic mass shown; m/z in italics are average masses						
*Further analyzed by MS/MS						
*Further digested using endoproteinase AspN followed by MS/MS						

Fragments nomenclature: *e.g.* T1-0 represents first tryptic peptide with zero missed cleavage site. T3-1 represents third tryptic peptide with one missed cleavage site.

Table III: Amino acid sequence obtained for N-terminal tryptic peptide (1-11) m/z 575.82 of AAD-1 sample Batch #1: 480-14.

Sequence: AHAALSPLSQR
Fragment ion masses: monoisotopic
Peptide mass $(MH)^{+2}$ (average): 576.16
Peptide mass $(MH)^{+2}$ (monoisotopic): 575.82

Ion Table

	A	*H*	*A*	*A*	*L*	*S*	*P*	*L*	*S*	*Q*	*R*
a(+1)	**44.05**	**181.11**	**252.15**	**323.18**	**436.27**	**523.30**	**620.35**	**733.44**	**820.47**	**948.53**	
		181.11									
b(+1)	**72.04**	**209.10**	**280.14**	**351.18**	**464.26**	**551.29**	**648.35**	**761.43**	**848.46**	**976.52**	
		209.11	280.15	351.18	464.25	551.29					
y(+1)		**1079.60**	**942.54**	**871.50**	**800.46**	**687.38**	**600.35**	**503.29**	**390.21**	**303.18**	**175.12**
			942.55	871.48	800.48	687.38	600.35		390.2		175.13
y(+2)		**540.30**	**471.77**	**436.25**	**400.74**	**344.19**	**300.68**	**252.15**	**195.61**	**152.09**	**88.06**
		540.31									
y-NH_3 (+1		**1062.57**	**925.51**	**854.47**	**783.44**	**670.35**	**583.32**	**486.27**	**373.18**	**286.15**	**158.09**
							583.34		373.17		

Top mass: theoretical product fragment ions
Bottom mass: observed product fragment ions

Table IV: Amino acid sequence obtained for C-terminal tryptic peptide (286-295) m/z 507.3 of AAD-1 sample Batch #1: 480-14.

Sequence: TTVGGVRPAR
Fragment ion masses: monoisotopic
Peptide mass $(MH)^{+2}$ (average): 507.59
Peptide mass $(MH)^{+2}$ (monoisotopic): 507.3

Ion Table

	T	*T*	*V*	*G*	*G*	*V*	*R*	*P*	*A*	*R*
a(+1)	**74.06**	**175.11**	**274.18**	**331.2**	**388.22**	**487.29**	**643.39**	**740.44**	**811.48**	
		175.12							811.43	
b-H_2O (+1)	**84.04**	**185.09**	**284.16**	**341.18**	**398.2**	**497.27**	**653.37**	**750.43**	**821.46**	
		185.1			398.2					
b(+1)	**102.06**	**203.10**	**302.17**	**359.19**	**416.21**	**515.28**	**671.38**	**768.44**	**839.47**	
		203.11							839.48	
y(+1)		**912.54**	**811.49**	**712.42**	**655.4**	**598.38**	**499.31**	**343.21**	**246.16**	**175.12**
				712.39				343.22		
y (+2)		**456.77**	**406.25**	**356.71**	**328.20**	**299.69**	**250.16**	**172.11**	**123.58**	**88.06**
		456.75	406.23							
y-H_2O(+1)		**894.53**	**---**	**---**	**---**	**---**	**---**	**---**	**---**	**---**
y-NH_3 (+1)		**895.51**	**794.46**	**695.4**	**638.37**	**581.35**	**482.28**	**326.18**	**229.13**	**158.09**
			794.42	695.41	638.38			326.18	229.14	

Top mass: theoretical product fragment ions
Bottom mass: observed product fragment ions

Table V: Amino acid sequence obtained for tryptic peptide T4-0 (35-72) m/z 1117.3 of AAD-1 sample Batch #1: 480-14.

Sequence: EPLDDSTWNEILDAFHTYQVIYFPGQAITNEQHIAFSR
Fragment ion masses: monoisotopic
Peptide mass $(MH)^{+4}$ (average): 1117.98
Peptide mass $(MH)^{+4}$ (monoisotopic): 1117.29
Ion Table

	E	***P***	***L***	***D***	***D***	***S***	***T***	***W***	***N***	***E***
a(+1)	102.06	199.11	312.19	427.22	542.25	629.28	730.33	916.41	1030.45	1159.49
		199.12								
b-H_2O (+1)	112.04	209.09	322.18	437.2	552.23	639.26	740.31	926.39	1040.43	1169.48
		209.11	**322.19**	**437.21**	**552.25**	**639.28**	**740.37**			**1169.51**
y(+1)										
y (+2)								1854.92	1761.88	1704.85
y (+3)	1446.37	1414.02	1376.32	1337.98	1299.64	1270.63	1236.95	1174.92	1136.91	1446.37
							1236.97	**1174.96**		

	I	***L***	***D***	***A***	***F***	***H***	***T***	***Y***	***Q***	***V***
a(+1)	1272.57	1385.66	1500.69	1571.72	1718.79	1855.85				
		1385.65								
b-H_2O (+1)	1282.56	1395.64	1510.67	1581.71	1728.78	1865.83				
y(+1)										
y (+2)	1640.33	1583.79	1527.25	1469.74	1434.22	1360.68	1292.15	1241.63	1160.10	1096.07
y (+3)	1093.89	1056.20	1018.50	980.16	956.48	907.46	861.77	828.09	773.73	731.05
		1056.21	**1018.57**	**980.15**						

	I	***Y***	***F***	***P***	***G***	***Q***	***A***	***I***	***T***	***N***
a(+1)										
b-H_2O (+1)										
y(+1)			1815.91	1668.85	1571.79	1514.77	1386.71	1315.68	1202.59	1101.54
				1668.88	**1571.8**				**1202.63**	
y (+2)	1046.53	989.99	908.46	834.93	786.40	757.89	693.86	658.34	601.80	551.28
	1046.55	**990.02**	**908.42**	**834.94**				**658.37**	**601.82**	
y (+3)	698.03	660.33	605.98	556.95	524.60	505.60	462.91	439.23	401.54	367.85

	E	***Q***	***H***	***I***	***A***	***F***	***S***	***R***
a(+1)								
b-H_2O (+1)								
y(+1)	987.50	858.46	730.40	593.34	480.26	409.22	262.15	175.12
		858.44	**730.43**	**593.34**	**480.27**	**409.22**	**262.16**	
y (+2)	494.25	429.73	365.70	297.17	240.63	205.11	131.58	88.06
y (+3)	329.84	286.82	244.14	---	---	---	---	---

Top mass: theoretical product fragment ions; **Bottom mass**: observed product fragment ions

Table VI: Amino acid sequence obtained for endoproteinase Asp-N digest peptide (39-46) m/z 997.7 of AAD-1 sample Batch #1: 480-14.

Sequence: DSTWNEIL
Fragment ion masses: monoisotopic
Peptide mass $(MH)^{+1}$ (average): 978.0499
Peptide mass $(MH)^{+1}$ (monoisotopic): 977.4580

Ion Table

	D	*S*	*T*	*W*	*N*	*E*	*I*	*L*
b-H_2O (+1)	**98.02**	**185.06**	**286.1**	**472.18**	**586.23**	**715.27**	**828.35**	
			286.11	472.19				
b-NH_3 (+1)	**---**	**---**	**---**	**---**	**587.21**	**716.25**	**829.34**	
					587.22	716.28		
b(+1)	**116.03**	**203.07**	**304.11**	**490.19**	**604.24**	**733.28**	**846.36**	
			304.13	490.19	604.25	733.3	846.37	
y(+1)		**862.43**	**775.4**	**674.35**	**488.27**	**374.23**	**245.19**	**132.1**
								132.12

Top mass: theoretical product fragment ions
Bottom mass: observed product fragment ions

Table VII: Amino acid sequence obtained for endoproteinase Asp-N digest peptide (38-46) m/z 1092.7 of AAD-1 sample Batch #1: 480-14.

Sequence: DDSTWNEIL
Fragment ion masses: monoisotopic
Peptide mass $(MH)^{+1}$ (average): 1093.1386
Peptide mass $(MH)^{+1}$ (monoisotopic): 1092.4849

	D	*D*	*S*	*T*	*W*	*N*	*E*	*I*	*L*
a (+1)	**88.04**	**203.07**	**290.1**	**391.15**	**577.23**	**691.27**	**820.31**	**933.4**	
		203.08			577.22				
b-H_2O (+1)	**98.02**	**213.05**	**300.08**	**401.13**	**587.21**	**701.25**	**830.3**	**943.38**	
				401.14	587.23		830.32	943.4	
b-NH_3 (+1)	**---**	**---**	**---**	**---**	**---**	**702.24**	**831.28**	**944.36**	
						702.27	831.31	944.4	
b(+1)	**116.03**	**231.06**	**318.09**	**419.14**	**605.22**	**719.26**	**848.31**	**961.39**	
		231.06	318.11	419.16	605.23	719.26	848.33	961.42	
y(+1)		**977.46**	**862.43**	**775.4**	**674.35**	**488.27**	**374.23**	**245.19**	**132.1**
					674.35	488.28			132.12

Top mass: theoretical product fragment ions
Bottom mass: observed product fragment ions

Table VIII: Amino acid sequence obtained for endoproteinase Asp-N digest peptide (32-37) m/z 742.6 of AAD-1 sample Batch #1:480-14.

Sequence: DLREPL
Fragment ion masses: monoisotopic
Peptide mass $(MH)^{+1}$ (average): 742.8533
Peptide mass $(MH)^{+1}$ (monoisotopic): 742.4099

Ion Table

	D	*L*	*R*	*E*	*P*	*L*
a (+1)	**88.04**	**201.12**	**357.23**	**486.27**	**583.32**	
		201.12		486.29		
b(+1)	**116.03**	**229.12**	**385.22**	**514.26**	**611.32**	
			385.23	514.26		
y(+1)		**627.38**	**514.3**	**358.2**	**229.16**	**132.1**
		627.39			229.16	

Top mass: theoretical product fragment ions
Bottom mass: observed product fragment ions

Table IX: Amino acid sequence obtained for N-terminal tryptic peptide (1-11) m/z 575.82 of AAD-1 sample Batch #2: 480-15.

Sequence: AHAALSPLSQR
Fragment ion masses: monoisotopic
Peptide mass $(MH)^{+2}$ (average): 576.16
Peptide mass $(MH)^{+2}$ (monoisotopic): 575.82

Ion Table

	A	H	A	A	L	S	P	L	S	Q	R
a(+1)	**44.05**	**181.11**	**252.15**	**323.18**	**436.27**	**523.30**	**620.35**	**733.44**	**820.47**	**948.53**	
		181.10		323.19	436.21	523.25					
b(+1)	**72.04**	**209.10**	**280.14**	**351.18**	**464.26**	**551.29**	**648.35**	**761.43**	**848.46**	**976.52**	
		209.09	280.13	351.16	464.23	551.26					
y(+1)		**1079.60**	**942.54**	**871.50**	**800.46**	**687.38**	**600.35**	**503.29**	**390.21**	**303.18**	**175.12**
			942.47	871.44	800.42	687.33	600.35	503.25	390.18	303.16	175.11
y-H_2O(+1)		**1061.59**	**924.53**	**853.49**	**782.45**	**669.37**	**582.34**	**485.28**	**372.20**		
						669.34					
y-NH_3 (+1)		**1062.57**	**925.51**	**854.47**	**783.44**	**670.35**	**583.32**	**486.27**	**373.18**	**286.15**	**158.09**
						670.35	583.28		373.15		

Top mass: theoretical product fragment ions
Bottom mass: observed product fragment ions

Table X: Amino acid sequence obtained for C-terminal tryptic peptide (286-295) m/z 507.3 of AAD-1 sample Batch #2: 480-15.

Sequence: TTVGGVRPAR
Fragment ion masses: monoisotopic
Peptide mass $(MH)^{+2}$ (average): 507.59
Peptide mass $(MH)^{+2}$ (monoisotopic): 507.3

Ion Table

	T	T	V	G	G	V	R	P	A	R
a(+1)	**74.06**	**175.11**	**274.18**	**331.2**	**388.22**	**487.29**	**643.39**	**740.44**	**811.48**	
		175.12							811.50	
b(+1)	**102.06**	**203.10**	**302.17**	**359.19**	**416.21**	**515.28**	**671.38**	**768.44**	**839.47**	
		203.11					671.36		839.47	
y(+1)		**912.54**	**811.49**	**712.42**	**655.4**	**598.38**	**499.31**	**343.21**	**246.16**	**175.12**
			811.50	712.44	655.40		499.33	343.21		175.12
y-H_2O(+1)										
y-NH_3 (+1)		**895.51**	**794.46**	**695.4**	**638.37**	**581.35**	**482.28**	**326.18**	**229.13**	**158.09**
			794.46	695.40	638.38	581.36	482.30	326.18		

Top mass: theoretical product fragment ions
Bottom mass: observed product fragment ions

Table XI: Amino acid sequence obtained for endoproteinase Asp-N digest peptide (39-46) m/z 997.7 of AAD-1 sample Batch #2: 480-15.

Sequence: DSTWNEIL
Fragment ion masses: monoisotopic
Peptide mass $(MH)^{+1}$ (average): 978.0499
Peptide mass $(MH)^{+1}$ (monoisotopic): 977.4580

Ion Table

	D	*S*	*T*	*W*	*N*	*E*	*I*	*L*
b-H_2O (+1)	**98.02**	**185.06**	**286.1**	**472.18**	**586.23**	**715.27**	**828.35**	
			286.11	475.2				
b-NH_3 (+1)	**---**	**---**	**---**	**---**	**587.21**	**716.25**	**829.34**	
					587.23	716.29		
b(+1)	**116.03**	**203.07**	**304.11**	**490.19**	**604.24**	**733.28**	**846.36**	
			304.12	490.2	604.21	733.3	846.38	
y(+1)		**862.43**	**775.4**	**674.35**	**488.27**	**374.23**	**245.19**	**132.1**
								132.1

Top mass: theoretical product fragment ions
Bottom mass: observed product fragment ions

Table XII: Amino acid sequence obtained for endoproteinase Asp-N digest peptide (38-46) m/z 1092.7 of AAD-1 sample Batch #2: 480-15.

Sequence: DDSTWNEIL
Fragment ion masses: monoisotopic
Peptide mass $(MH)^{+1}$ (average): 1093.1386
Peptide mass $(MH)^{+1}$ (monoisotopic): 1092.4849

Ion Table

	D	*D*	*S*	*T*	*W*	*N*	*E*	*I*	*L*
a (+1)	**88.04**	**203.07**	**290.1**	**391.15**	**577.23**	**691.27**	**820.31**	**933.4**	
		203.08							
b-H_2O (+1	**98.02**	**213.05**	**300.08**	**401.13**	**587.21**	**701.25**	**830.3**	**943.38**	
				401.15	587.23		830.36	943.4	
b-NH_3 (+1	**---**	**---**	**---**	**---**	**---**	**702.24**	**831.28**	**944.36**	
						702.26	831.32	944.41	
b(+1)	**116.03**	**231.06**	**318.09**	**419.14**	**605.22**	**719.26**	**848.31**	**961.39**	
		231.06	318.1	419.16	605.23	719.28	848.3	961.42	
y(+1)		**977.46**	**862.43**	**775.4**	**674.35**	**488.27**	**374.23**	**245.19**	**132.1**
						488.28			132.12

Top mass: theoretical product fragment ions
Bottom mass: observed product fragment ions

Table XIII: Amino acid sequence obtained for endoproteinase Asp-N digest peptide (32-37) m/z 742.6 of AAD-1 sample Batch #2: 480-15.

Sequence: DLREPL
Fragment ion masses: monoisotopic
Peptide mass $(MH)^{+1}$ (average): 742.8533
Peptide mass $(MH)^{+1}$ (monoisotopic): 742.4099

Ion Table

	D	*L*	*R*	*E*	*P*	*L*
a (+1)	**88.04**	**201.12**	**357.23**	**486.27**	**583.32**	
		201.15		486.28		
b(+1)	**116.03**	**229.12**	**385.22**	**514.26**	**611.32**	
			385.22	514.26		
y(+1)		**627.38**	**514.3**	**358.2**	**229.16**	**132.1**
		627.38			229.15	132.11

Top mass: theoretical product fragment ions
Bottom mass: observed product fragment ions

Figure 1: Amino acid sequence of AAD-1 with theoretical mass. a) Unmodified AAD-1 b) Cysteines alkylated with Iodoacetamide

a)

AAD-1

[1-296] mass = 33283.7

Small polar:	**D(19)**	**E(16)**	**N(8)**	**Q(13)**			
Large polar:	**K(7)**	**R(23)**	**H(10)**				
Small non-polar:	**S(15)**	**T(23)**	**A(23)**	**G(21)**			
Large non-polar:	**L(22)**	**I(12)**	**V(28)**	**M(8)**	**F(13)**	**Y(10)**	**W(5)**
Special:	**C(3)**	**P(17)**					

```
  1 M A H A A L S P L S Q R F E R I A V Q P L T G V L G A E I T  30
 31 G V D L R E P L D D S T W N E I L D A F H T Y Q V I Y F P G  60
 61 Q A I T N E Q H I A F S R R F G P V D P V P L L K S I E G Y  90
 91 P E V Q M I R R E A N E S G R V I G D D W H T D S T F L D A 120
121 P P A A V V M R A I D V P E H G G D T G F L S M Y T A W E T 150
151 L S P T M Q A T I E G L N V V H S A T R V F G S L Y Q A Q N 180
181 R R F S N T S V K V M D V D A G D R E T V H P L V V T H P G 210
211 S G R K G L Y V N Q V Y C Q R I E G M T D A E S K P L L Q F 240
241 L Y E H A T R F D F T C R V R W K K D Q V L V W D N L C T M 270
271 H R A V P D Y A G K F R Y L T R T T V G G V R P A R         296
```

b)

AAD-1

[1-296] mass = 33454.8

Small polar:	**D(19)**	**E(16)**	**N(8)**	**Q(13)**			
Large polar:	**K(7)**	**R(23)**	**H(10)**				
Small non-polar:	**S(15)**	**T(23)**	**A(23)**	**G(21)**			
Large non-polar:	**L(22)**	**I(12)**	**V(28)**	**M(8)**	**F(13)**	**Y(10)**	**W(5)**
Special:	**C(3)**	**P(17)**					

C[223] + 57.1 **C[252] + 57.1** **C[268] + 57.1**

```
  1 M A H A A L S P L S Q R F E R I A V Q P L T G V L G A E I T  30
 31 G V D L R E P L D D S T W N E I L D A F H T Y Q V I Y F P G  60
 61 Q A I T N E Q H I A F S R R F G P V D P V P L L K S I E G Y  90
 91 P E V Q M I R R E A N E S G R V I G D D W H T D S T F L D A 120
121 P P A A V V M R A I D V P E H G G D T G F L S M Y T A W E T 150
151 L S P T M Q A T I E G L N V V H S A T R V F G S L Y Q A Q N 180
181 R R F S N T S V K V M D V D A G D R E T V H P L V V T H P G 210
211 S G R K G L Y V N Q V Y C Q R I E G M T D A E S K P L L Q F 240
241 L Y E H A T R F D F T C R V R W K K D Q V L V W D N L C T M 270
271 H R A V P D Y A G K F R Y L T R T T V G G V R P A R         296
```

Figure 2: A captured image of a Coomassie-stained one-dimensional 4-20% SDS-PAGE analysis of AAD-1 [Batch #1: 480-14 and Batch #2: 480-15]. Lane "S" Molecular weight standards.

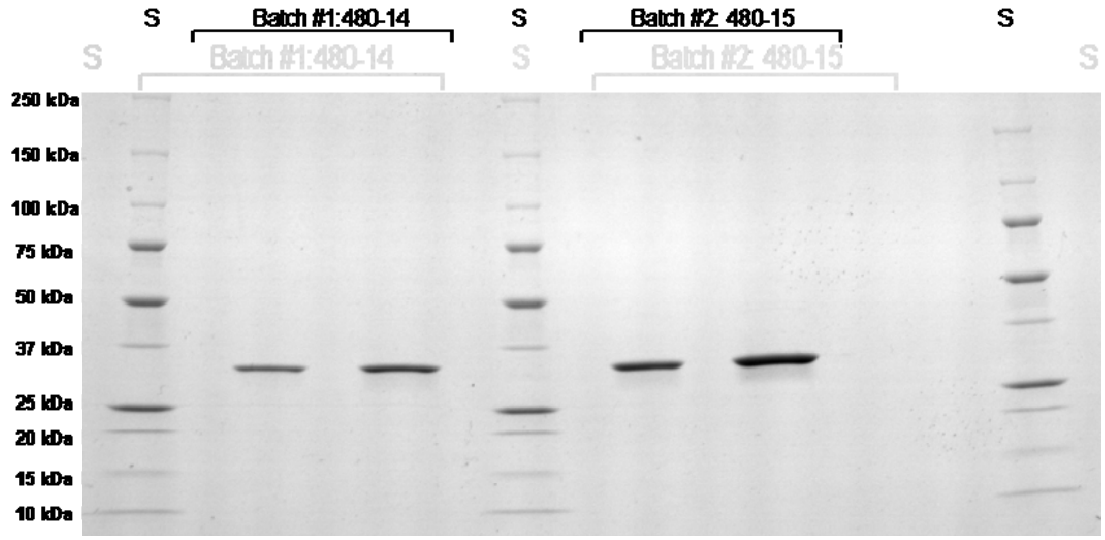


Figure 3: Size Exclusion Chromatographic Analysis of AAD-1. (a) elution profile of molecular weight standards; (b) elution profile of AAD-1 Batch #1: 480-14; (c) elution profile of AAD-1 Batch #2: 480-15; (c) protein molecular weight calibration curve for the Superdex 200 column in 10 mM PBS buffer, pH 7.4, at room temperature (♦ and solid line), and the calculated native AAD-1 MW (■).

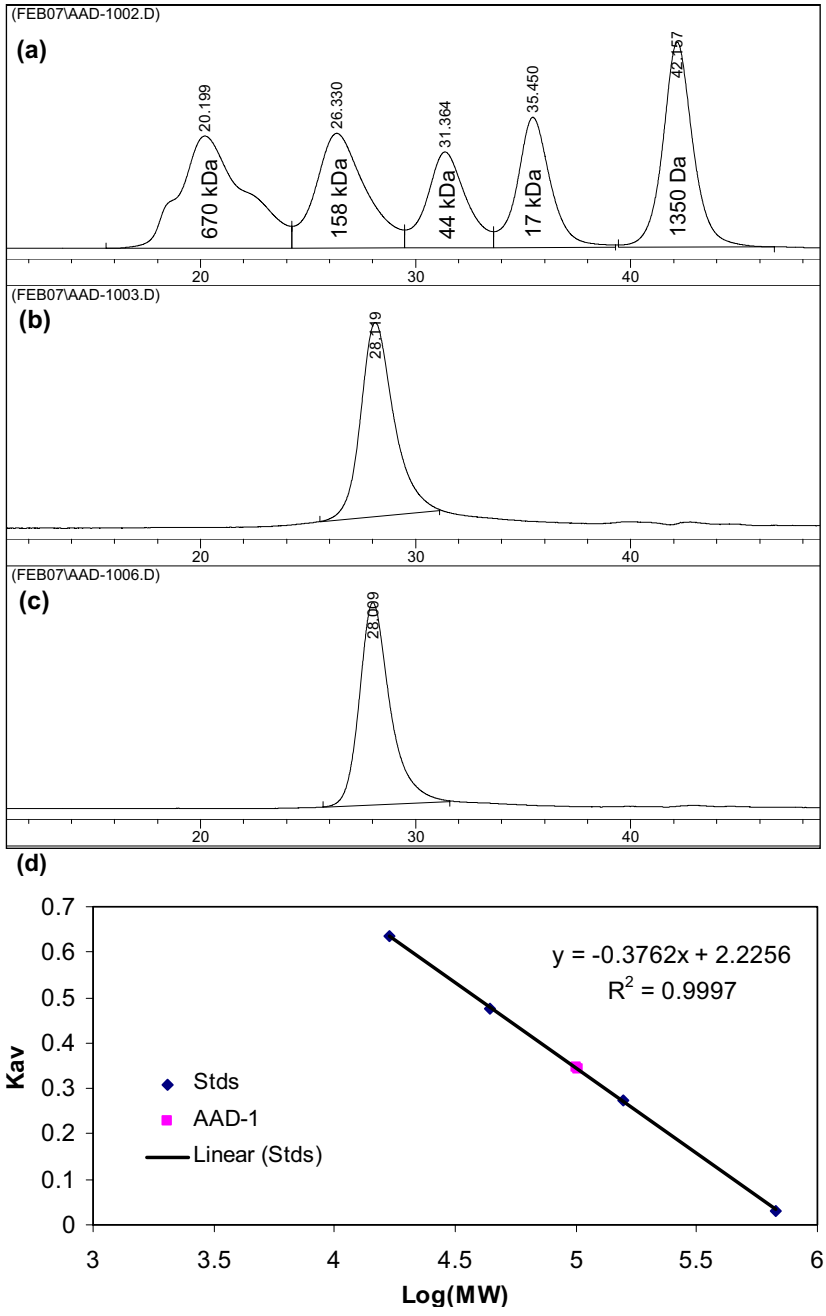
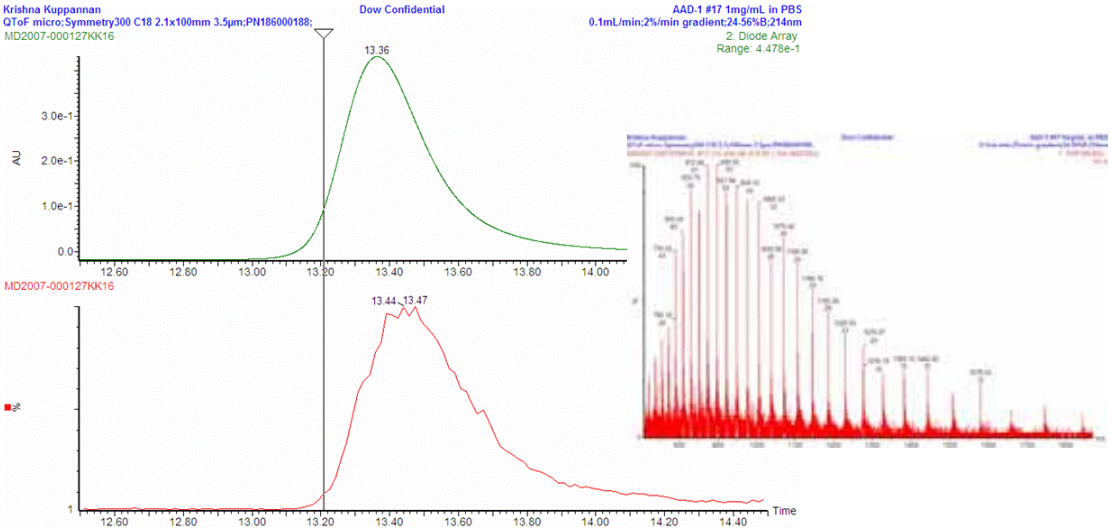


Figure 4: AAD-1 sample Batch #1: 480-14 in PBS analyzed by UPLC-MS. (a) Chromatograph (top: UV trace; bottom: MS TIC); inset: MS spectrum (b) deconvoluted mass spectrum.

(a)



(b)

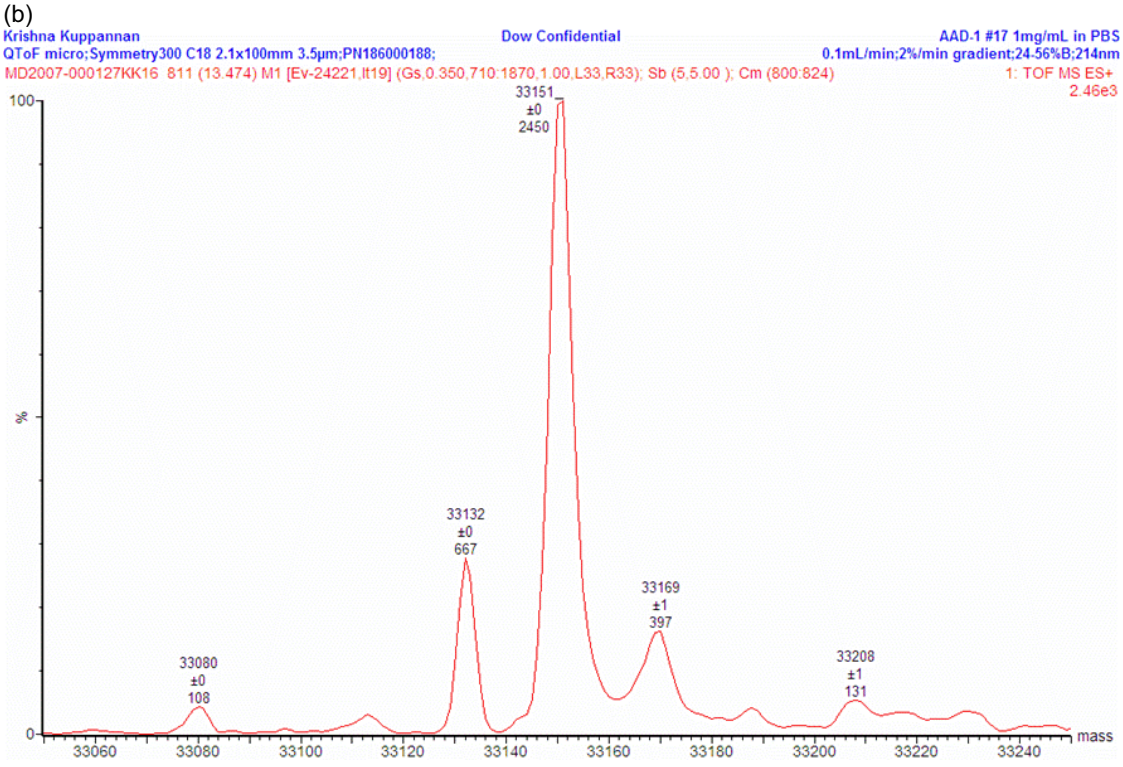
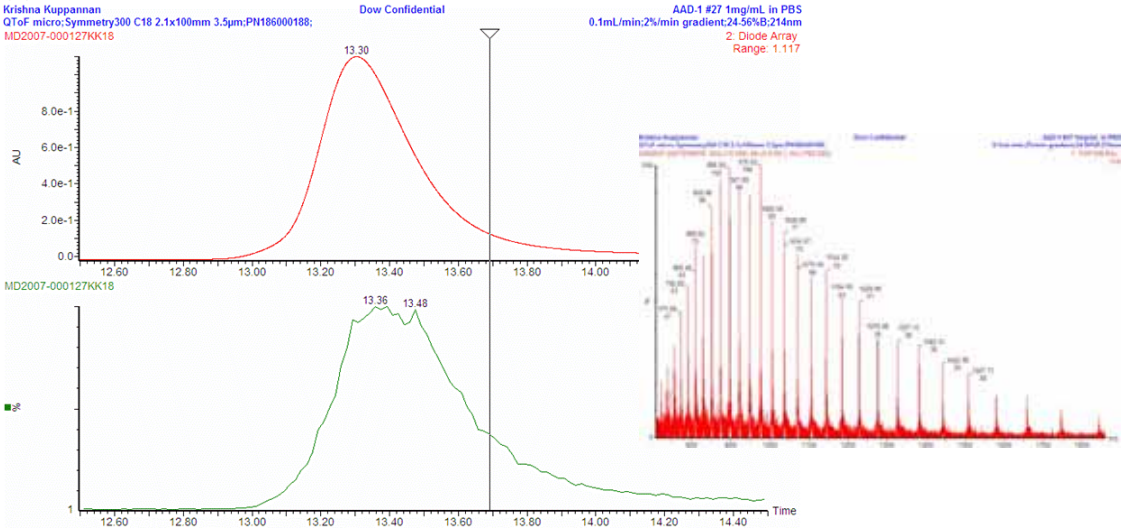


Figure 5: AAD-1 sample Batch #2: 480-15 in PBS analyzed by UPLC-MS. (a) Chromatograph (top: UV trace; bottom: MS TIC); inset: MS spectrum (b) deconvoluted mass spectrum.

(a)



(b)

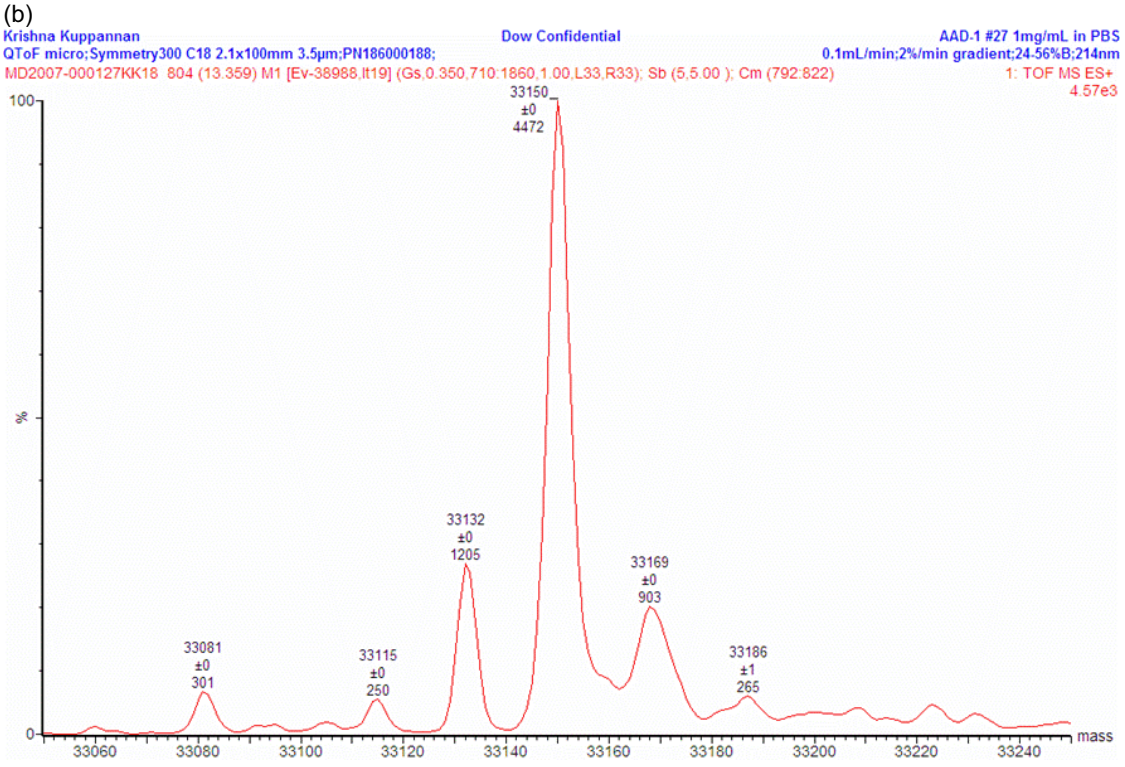
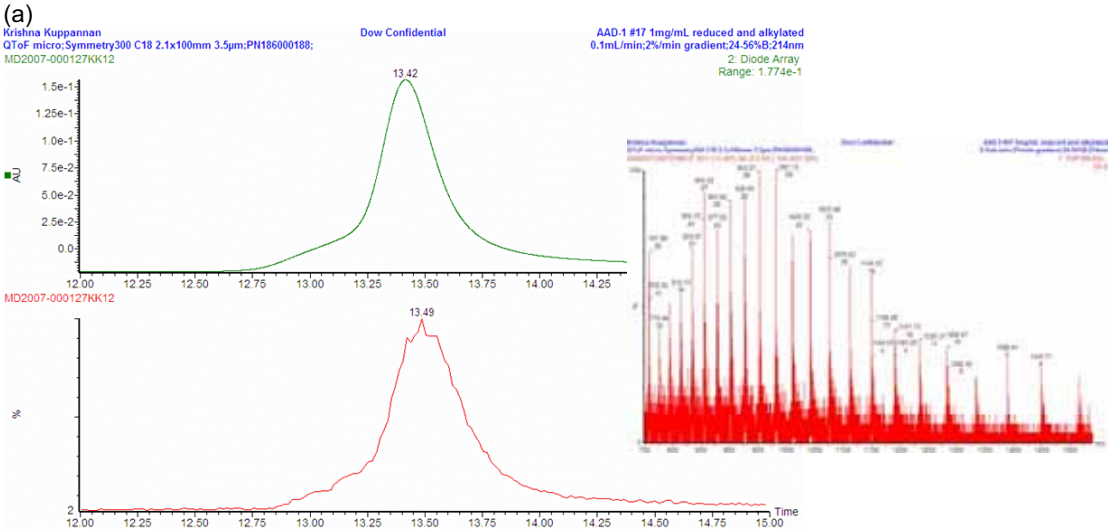


Figure 6: AAD-1 sample Batch #1: 480-14 reduced and alkylated prior to analysis by UPLC-MS. (a) Chromatograph (top: UV trace; bottom: MS TIC); inset: MS spectrum (b) deconvoluted mass spectrum.

(a)



(b)

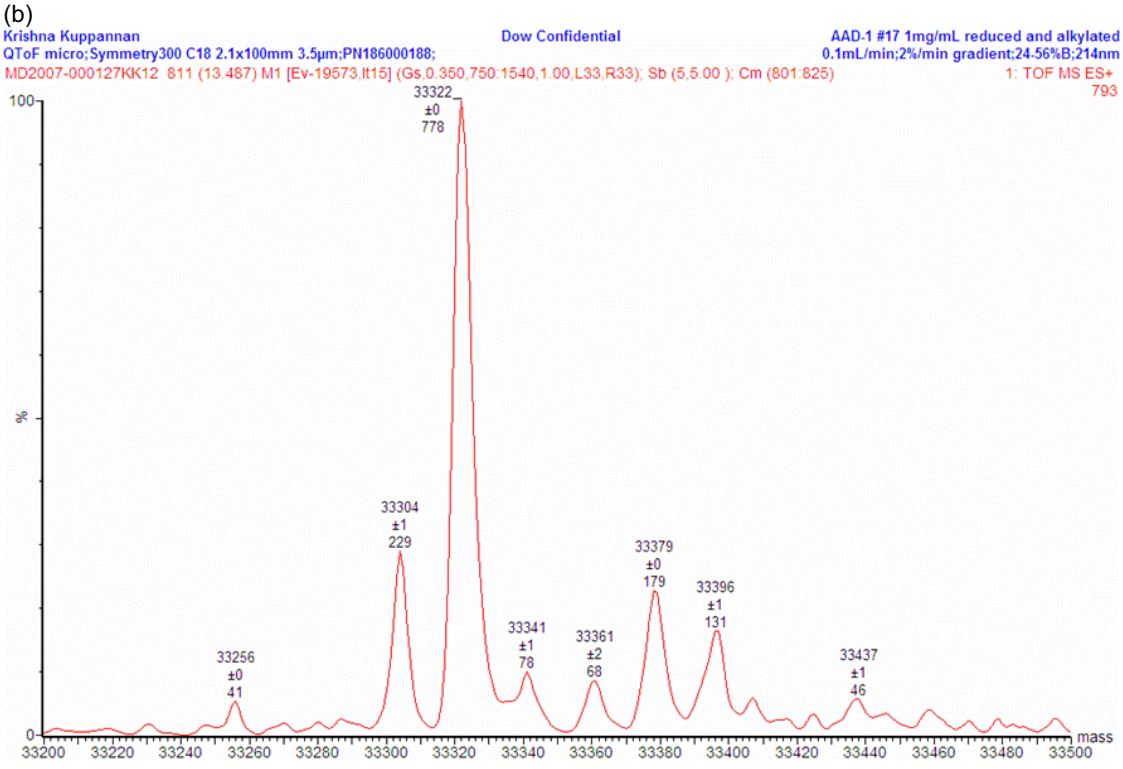
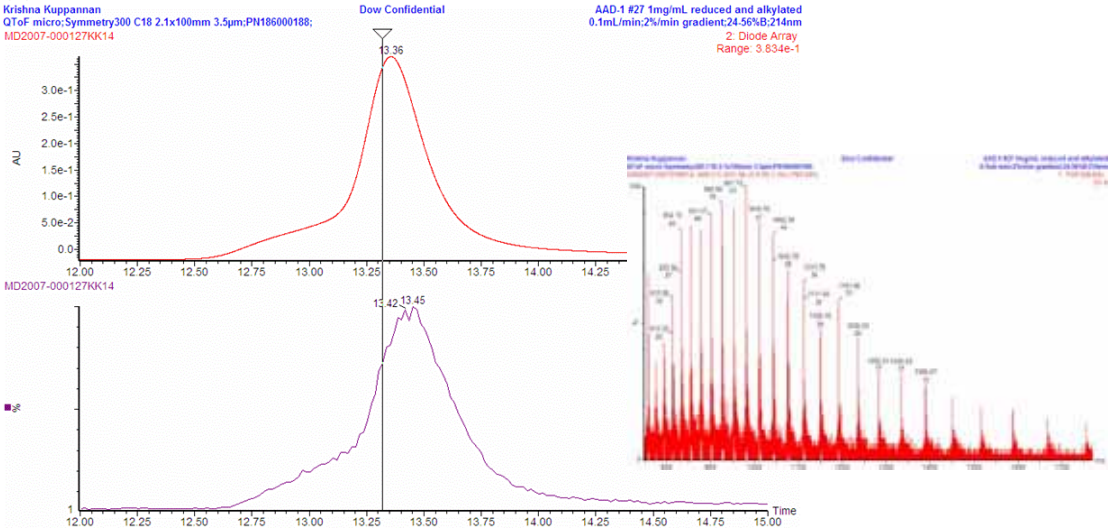


Figure 7: AAD-1 sample Batch #2: 480-15 reduced and alkylated prior to analysis by UPLC-MS. (a) Chromatograph (top: UV trace; bottom: MS TIC); inset: MS spectrum (b) deconvoluted mass spectrum.

(a)



(b)

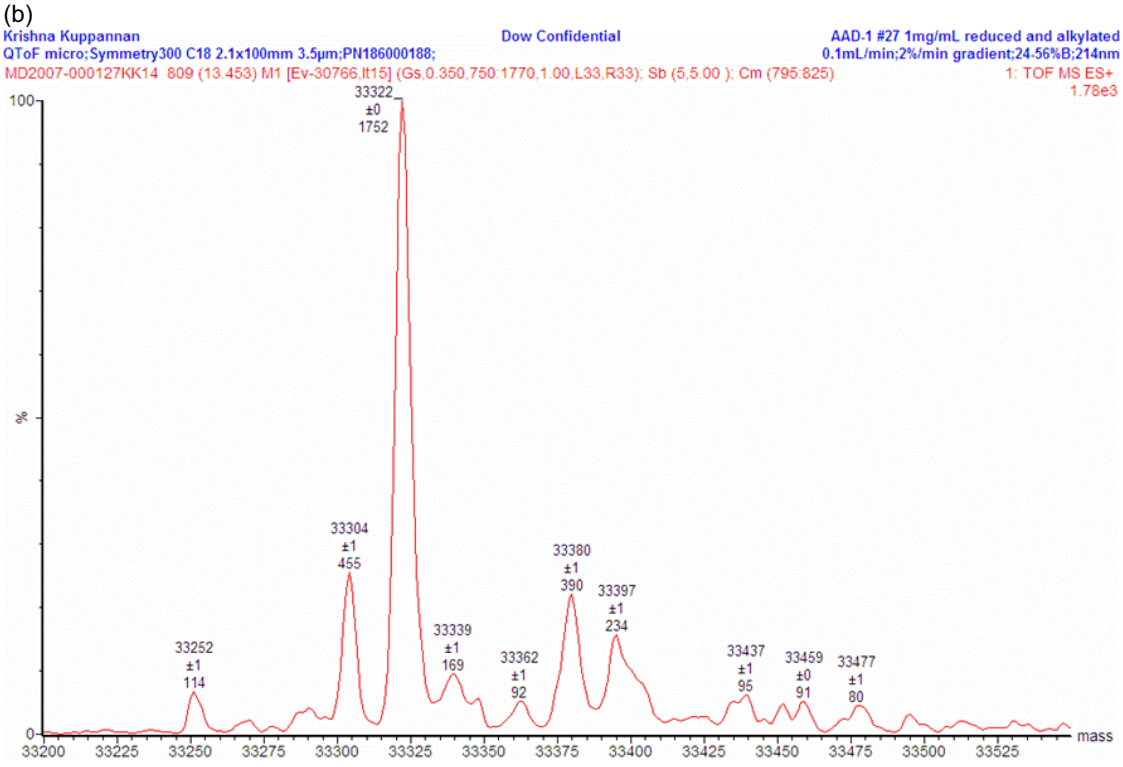
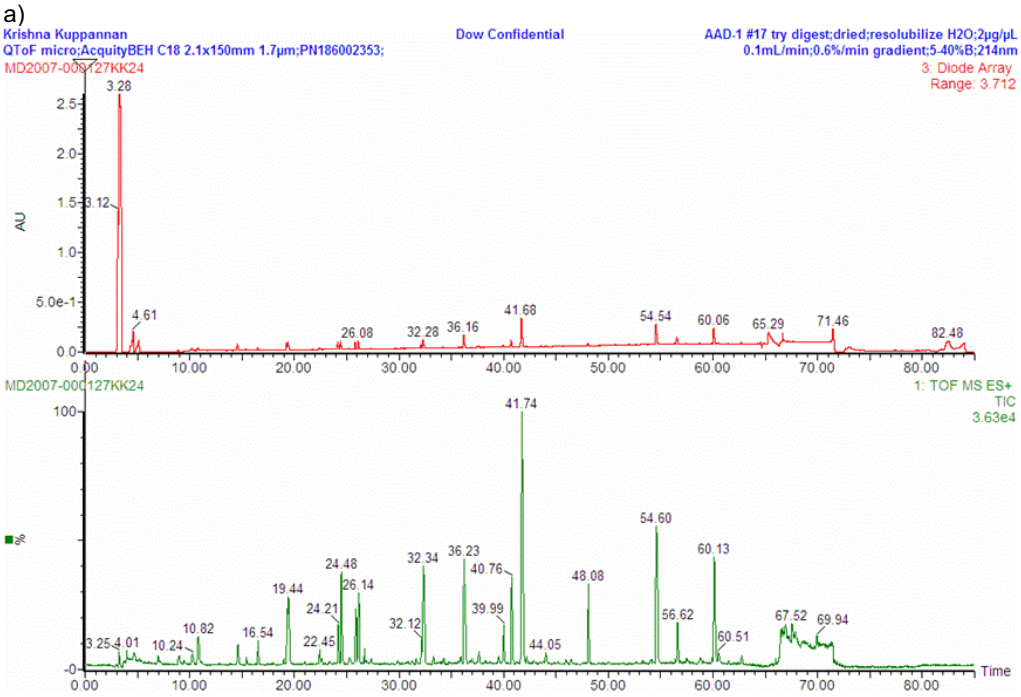
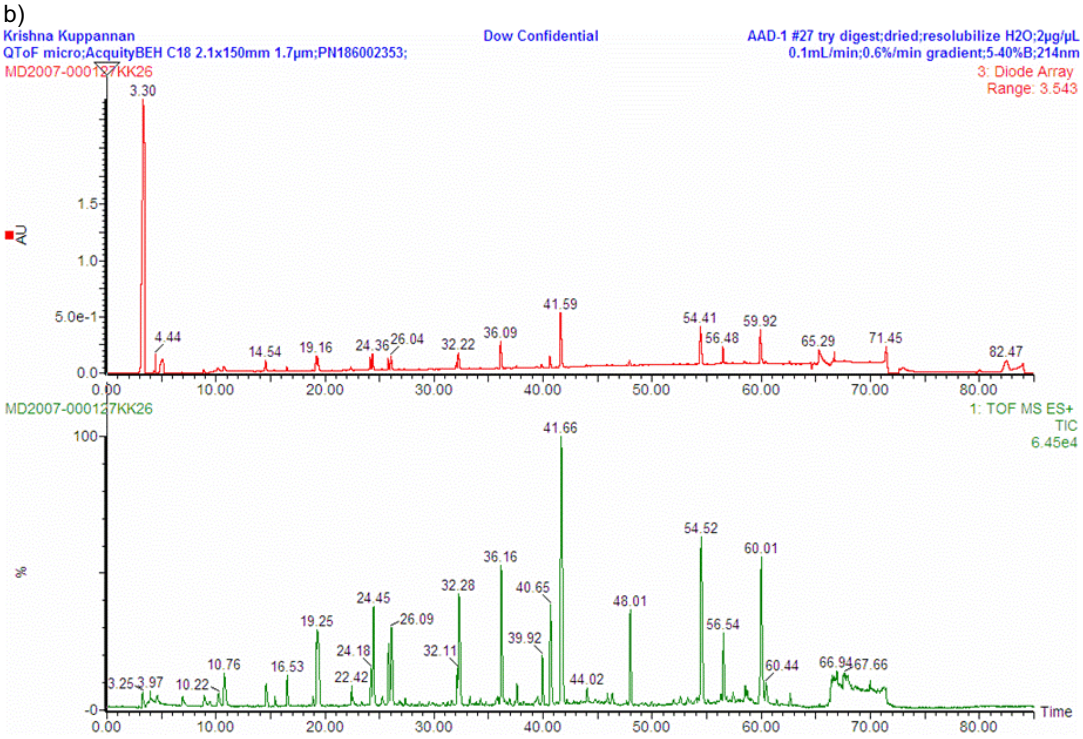


Figure 8: ESI-LC/MS chromatograms of AAD-1 tryptic digest for batches (a) Batch #1: 480-14 and (b) Batch #2: 480-15 (top: UV trace; bottom: MS TIC).

a)



b)



THE JOURNAL OF BIOLOGICAL CHEMISTRY
Vol. 275, No. 47, Issue of November 24, pp. 36479–36482, 2000

Minireview

N^{α}-terminal Acetylation of Eukaryotic Proteins*

Published, JBC Papers in Press, September 29, 2000,
DOI 10.1074/jbc.R000023200

Bogdan Polevoda and Fred Sherman‡

From the Department of Biochemistry and Biophysics, University of Rochester School of Medicine and Dentistry, Rochester, New York 14642

The two cotranslational processes, cleavage of N-terminal methionine residues and N-terminal[1] acetylation, are by far the most common modifications, occurring on the vast majority of eukaryotic proteins. Studies with the yeast *Saccharomyces cerevisiae* revealed three N-terminal acetyltransferases, NatA, NatB, and NatC, that acted on groups of substrates, each containing degenerate motifs. Orthologous genes encoding the three N-terminal acetyltransferases and the patterns of N-terminal acetylation suggest that eukaryotes generally use the same systems for N-terminal acetylation. The biological significance of this N-terminal modification varies with the particular protein, with some proteins requiring acetylation for function, whereas others do not.

Methionine Cleavage

Cleavage of N-terminal methionine residues is by far the most common modification, occurring on the vast majority of proteins. Proteins from prokaryotes, mitochondria, and chloroplasts initiate with formylmethionine, whereas proteins from the cytosol of eukaryotes initiate with methionine. The formyl group is usually removed from prokaryotic proteins by a deformylase, resulting in methionine at N termini. The methionine at N termini is cleaved from nascent chains of most prokaryotic and eukaryotic proteins. Results with altered iso-1-cytochromes *c* from yeast (1) were the basis for the hypothesis that methionine is cleaved from penultimate residues having radii of gyration of 1.29 Å or less (glycine, alanine, serine, cysteine, threonine, proline, and valine) (2), a hypothesis that was confirmed from the results of a complete set of altered iso-1 having all possible amino acids at the penultimate position (3). A similar pattern of cleavage also was observed with prokaryotic systems *in vivo* (4, 5) and *in vitro* (6, 7), and other eukaryotic systems *in vivo* (8, 9) and *in vitro* (10). Only minor differences were observed between the quantitative results obtained *in vivo* with yeast iso-1-cytochrome *c* and with the two proteins, TimJp and TimLp, from *Escherichia coli* (3–5). The lack of action of methionine aminopeptidase on proteins with large penultimate residues can now be explained by steric hindrance, as deduced from the high resolution crystal structure of the inhibitor complex of methionine aminopeptidase from *E. coli* (11).

However, N-terminal methionine is completely or partially retained on certain exceptional proteins having penultimate residues with intermediate sizes of side chains. The mature forms of the S27A, S27B, and L42B ribosomal proteins have Met-Val-Leu or Met-Val-Asn terminal regions (12). Furthermore, Moerschell *et al*. (3) demonstrated antepenultimate (the third residue) proline residues can inhibit methionine cleavage from certain residues with intermediate sizes of side chains. Also, methionine cleavage was completely inhibited from the Met-Val-Pro sequence of a mutant human hemoglobin (13). In other studies with *E. coli*, antepenultimate proline residues partially inhibited cleavage from Met-Ala-Pro (4, 14) and Met-Thr-Pro (15).

Methionine excision occurs before completion of the nascent chain and before other N-terminal processing events, such as N-terminal acetylation (16, 17). *S. cerevisiae* contains two types of methionine aminopeptidases, Map1p and Map2p (18). Mutants containing either *map1* or *map2* null mutations are viable, but the *map1 map2* double mutants are nonviable (18). Thus, removal of N-terminal methionine is an essential function in yeast, as in prokaryotes, but the process can be carried out by either of two enzymes. There are probably a number of reasons why N-terminal methionine removal is required for viability. For example, *N*-myristoylation is essential for growth, and *N*-myristoyltransferase requires a free N-terminal glycine (19). Also, N-terminal residues can be important for the activity of a variety of diverse proteins, such as actin (20, 21) and proteasome subunits (22, 23).

Cotranslational N-terminal Acetylation

N-terminal acetylation of proteins is catalyzed by NATs that transfer acetyl groups from acetyl-CoA to termini of α-amino groups. Similar to N-terminal methionine cleavage, N-terminal acetylation is one of the most common protein modifications in eukaryotes, occurring on approximately 85% of the different varieties of eukaryotic proteins but rarely on prokaryotic proteins (16). *In vitro* studies indicated that N-terminal acetylation of eukaryotic proteins occurs cotranslationally when there are between 20 and 50 residues extruding from the ribosome (16, 17).

N-terminal acetylation also can occur at internal sites after specific proteolytic processing of the completely translated protein, as in the cases of peptide hormones (24). Posttranslational acetylation of each of the different proteins occurs with different NATs having different specificities, and these differ from the set of NATs carrying out cotranslational acetylation.

Interestingly, internal threonine residues were acetylated when the propeptide region of several proteasome subunits was replaced by a ubiquitin sequence (22). When such artificial protein fusions are expressed in yeast, ubiquitin is rapidly cleaved by deubiquitinating enzymes, presumably on the growing nascent chain, and N termini can serve as substrates for co-translational acetylation. On the other hand, after translation of normal mRNA containing the propeptide region and after near completion of proteasome assembly, subunits are processed at a conserved Gly-Thr motif, exposing the catalytic N-terminal Thr residues, which are not acetylated. Thus, in this instance N-terminal acetylation did not occur when the appropriate N-terminus was formed posttranslationally. Furthermore, artificial constructs with a Met-Thr terminus have the expected cleavage of methionine and N-terminal acetylation of the penultimate threonine residue, resulting in the lack of function.

Eukaryotic proteins susceptible to N-terminal acetylation have a variety of different N-terminal sequences with no simple consensus motifs and with no dependence on a single type of residue. Eukaryotic proteins with serine and alanine termini are the most frequently acetylated, and these residues along with methionine, glycine, and threonine account for over 95% of the N-terminal acetylated residues (16, 17, 25, 26). However, only subsets of proteins with any of these N-terminal residues are acetylated, and none of these N-terminal residues guarantee acetylation.

Three Different NATs

Ard1p, Nat3p, and Mak3p are related to each other by amino acid sequence and are believed to be the catalytic subunits of three different NATs, designated NatA, NatB, and NatC, respectively, each acting on different sets of proteins having different N-terminal regions (27–29) (Fig. 1). Ard1p activity requires two subunits, Ard1p itself and Nat1p. A *nat1*$^{-}$ mutant was originally uncovered by screening a collection of heavily mutagenized strains for protein acetyltransferase activity *in vitro* (30). The previously identified

* This minireview will be reprinted in the 2000 Minireview Compendium, which will be available in December, 2000. This research was supported by National Institutes of Health Grant GM12702.
‡ To whom correspondence should be addressed: Dept. of Biochemistry and Biophysics, Box 712, University of Rochester Medical School, Rochester, NY 14642. Tel.: 716-275-6647; Fax: 716-275-6007; E-mail: Fred_Sherman@urmc.rochester.edu.
[1] The abbreviations used are: N-terminal, NH_2-terminal, N^{α}-terminal; NAT, N-terminal acetyltransferase.

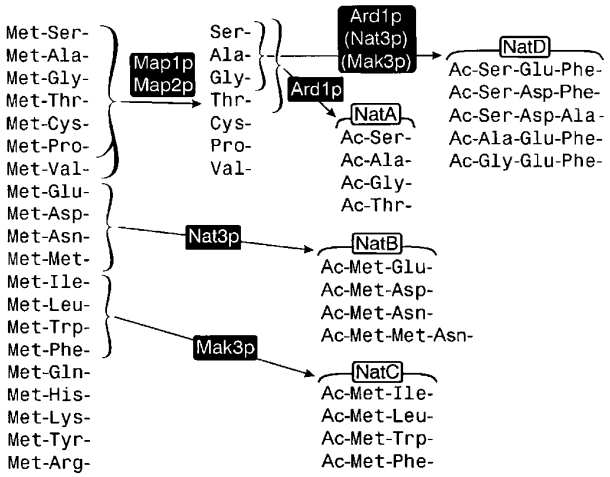


FIG. 1. **The pathways of N-terminal processing.** The two methionine aminopeptidases, Map1p and Map2, cleave N-terminal methionine residues that have small side chains (glycine, alanine, serine, cysteine, threonine, proline, and valine), although methionine is retained on some proteins having penultimate residues of valine. Subsequently, N termini of NatA, NatB, and NatC substrates are acetylated by the Ard1p, Nat3p, and Mak3p acetyltransferases, respectively. In addition, acetylation of the NatD substrates requires all three NATs. Except for possibly the Met-Glu and Met-Asp NatB substrates, not all proteins with the designated N-termini are acetylated.

ard1$^-$ mutant was first suspected to be related to *nat1*$^-$ because of certain similar phenotypes (30). Overexpression of both Ard1p and Nat1p subunits is required for increased NAT activity *in vivo* (30), and both interact with each other to form an active complex *in vitro* (31). The *MAK3* gene encodes a NAT that is required for the N-terminal acetylation of the killer viral major coat protein, *gag*, with an Ac-Met-Leu-Arg-Phe terminus. *MAK3* was first identified from *mak3*$^-$-deficient mutants that did not assemble or maintain the L-A double-stranded RNA viral particle (28, 32). The co-purification of Mak3p, Mak10p, and Mak31p suggests that these three subunits form a complex that is required for N-terminal acetylation (33). Nat3p was originally identified on the basis of similarities of its amino acid sequence to those of Ard1p and Mak3p (27).

Sequences Required for N-terminal Acetylation

Previous attempts to predict N-terminal acetylation based on the properties of amino acid residues distributed along the N-terminal region were unsuccessful (16, 26, 34, 35). However, new insights on this problem have been provided by using yeast mutants deleted in one or another of these NAT genes. The substrate specificities for each of the Ard1p, Nat3p, and Mak3p enzymes were deduced from considering the lack of acetylation of the following groups of protein in mutants containing one or another of the *ard1*-Δ, *nat1*-Δ, *nat3*-Δ, or *mak3*-Δ deletions: mutationally altered iso-1-cytochromes *c* (27); mutationally altered β-galactosidases (32); abundant proteins (27, 36, 37); ribosomal proteins (12); and 20 S proteasome subunits (38).

As summarized in Fig. 1, subclasses of proteins with Ser, Ala, Gly, or Thr termini are not acetylated in *ard1*-Δ mutants (NatA substrates) (27); proteins with Met-Glu or Met-Asp termini and subclasses of proteins with Met-Asn termini are not acetylated in *nat3*-Δ mutants (NatB substrates) (27); and subclasses of proteins with Met-Ile, Met-Leu, Met-Trp, or Met-Phe termini are not acetylated in *mak3*-Δ mutants (NatC substrates) (27, 32, 38). In addition, a special subclass of NatA substrates with Ser-Glu, Ser-Asp, Ala-Glu, or Gly-Glu termini, designated NatD substrates, is also only partially acetylated in *nat3*-Δ and *mak3*-Δ mutants (Table I, Fig. 1) (12, 27).

One possible interpretation of the NatD substrates is that they are a subclass of NatA substrates requiring, in addition to Ard1p and Nat1p, other factors whose full active form requires acetylation by Mak3p and Nat3p. The lack of acetylation of the auxiliary factors may result in no or only partial acetylation, depending on the particular NatD substrate, as summarized in Table I.

Generally, acetylation cannot be definitively predicted from the primary amino acid sequence. Only the NatB substrates have common sequences that can be easily deciphered, and these are composed of Ac-Met-Glu, Ac-Met-Asp, Ac-Met-Asn-Asn, and probably Ac-Met-Met-Asn sequences. As emphasized by Moerschell *et al*. (3), all seven eukaryotic Met-Glu and Met-Asp proteins uncovered in literature and data base searches were N-terminally acetylated, but not any of the six prokaryotic proteins with the same N-terminal residues. Furthermore, all 11 normal yeast proteins having Met-Glu and Met-Asp were acetylated (3, 12, 27, 32). However, there are Met-Glu and Met-Asp iso-1-cytochromes *c* with reduced efficiency of acetylation, including *CYC1–838* (55%) and *CYC1–878* (67%) (Table II). We suggest that all the NatB substrates contain any one of these required sequences, but acetylation is diminished by inhibitory residues. For example, from the result with the Ac-Met-Asp-Pro iso-1 (*CYC1–878*) having only 67% acetylation, one can suggest that adjacent proline residues diminish the action of Nat3p. Similarly, Moerschell *et al*. (3) demonstrated that antepenultimate proline residues can inhibit methionine cleavage from certain residues.

An alignment of the N-terminal region of NatD substrates and related sequences reveals an obvious requirement, but not sufficiency, for acidic residues, Glu or Asp, at the antepenultimate position (Table I).

We suggest that NATs act on substrates with specific but degenerate sequences and that the activities can be diminished by suboptimal residues. We further suggest that acetylation can be diminished by the inhibitory residues situated anywhere on the nascent chain at the time of this addition. Thus, the degree of acetylation is the net effect of positive optimal or suboptimal residues and negative inhibitory residues. Furthermore, this lack of acetylation could be because of the absence of required residues or the presence of inhibitory residues. For example, the antepenultimate residue, Asn, in the *CYC1–872* sequence Ac-Met-Asn-Asn can be considered optimal, allowing complete acetylation, whereas the antepenultimate residue, Phe, in the *CYC1–849* sequence Ac-Met-Asn-Phe can be considered suboptimal, resulting in only 79% acetylation (Table II). On the other hand, the antepenultimate residues, Gln, in the *CYC1–9-CB* sequence Met-Asn-Gln prevents acetylation because it is not part of a required sequence. From the other point of view, the Lys residue in the *CYC1–838* sequence Ac-Met-Glu-Phe-Lys and the Pro residue in the *CYC1–878* Ac-Met-Asp-Pro-Leu can be considered inhibitory residues. Because the identities of required and inhibitory residues are unknown, the ability of a protein to be acetylated cannot be definitively predicted from the primary sequence. Also, it is unclear if an amino acid position is occupied by required or inhibitory residues. For example, in the NatC series Met-Leu-Arg-*Any*, represented by the *CYC1–1201*, L-A *gag*, JC33B, and Ilv5p proteins (Table II), it is unknown whether the "*Any*" residue is part of the required sequence or if it depicts residues inhibiting the action of the Met-Leu-Arg sequence. Other examples of NatC substrates and related nonacetylated proteins are also presented in Table II.

The NatA substrates appear to be the most degenerate, encompassing a wide range of sequences, especially those with N-terminal residues of serine or alanine. In fact, approximately 90 and 50% of the Ser and Ala proteins, respectively, are acetylated (16). Examples of related acetylated and nonacetylated proteins having serine, alanine, or threonine termini are presented in Table II. Whereas the reason for the lack of acetylation of most of these proteins is unclear, the N-terminal region of many of the nonacetylated proteins related to both NatA and NatB substrates contains basic residues, lysine, arginine, and histidine, as well as proline residues, whereas some nonacetylated proteins related to NatC substrates contain acidic residues, such as glutamic acid. Because Erg7p but not the *CYC1–1371* iso-1-cytochrome *c* is acetylated (Table II), one can suggest that inhibitory residues could occupy sites further than five amino acid residues from the terminus. Because the required and inhibitory residues may affect acetylation to various degrees and because inhibitory residues may possibly occupy various sites in the nascent chain, predicting acetylated and nonacetylated sequences is still unreliable.

TABLE I
Examples of N-terminal acetylation of NatD and other substrates

Type	Gene	Sequence	Approximate acetylation: Normal	NatA *ard1, nat1*	NatB *nat3*	NatC *mak3*	Ref.
			%	%	%	%	
	CYC1–795	Thr-Glu-Phe-Leu-Ala-	0				27
NatA	*CYC1–1383*	Ac-Ser-Glu-Ile-Thr-Ala-	100	0	100	100	[a]
NatB	*CYC1–853*	Ac-Met-Glu-Phe-Leu-Ala-	100	100	0	100	27
NatC	*CYC1–1162*	Ac-Met-Ile-Arg-Leu-Lys-	94	100	100	0	27
NatD	*CYC1–963*	Ac-Gly-Glu-Phe-Leu-Ala-	100	0	50	60	[b]
	CYC1–962	Ac-Ala-Glu-Phe-Leu-Ala-	100	0	10	40	[b]
	RPS20	Ac-Ser-Asp-Phe-Gln-Lys-	100	0	90	60	12
	RPS24A	Ac-Ser-Asp-Ala-Val-Thr-	100	0	100	40	12

[a] B. Polevoda and F. Sherman, unpublished result.
[b] The percent N-terminal acetylation was determined by mass spectrometry (B. Polevoda and F. Sherman, unpublished result), correcting the values determined earlier by HPLC (27).

TABLE II
Examples of similar sequences that are completely, partially or not acetylated

The percentage acetylation values are estimates. Residues that appear to be interfering with acetylation are underlined; certain residues that appear to be required for acetylation are highlighted in black.

	Proteins or genes	Sequence	Approx. % acetylation	Ref.
NatB	*CYC1-853*	Ac-**Met**-**Glu**-Phe-Leu-Ala-	100%	(3)
	CYC1-838	Ac-**Met**-**Glu**-Phe-Lys-Ala-	55%	(3)
	CYC1-878	Ac-**Met**-**Asp**-Pro-Leu-Ala-	67%	(3)
	CYC1-872	Ac-**Met**-**Asn**-**Asn**-Leu-Ala-	100%	(3)
	CYC1-849	Ac-**Met**-**Asn**-**Phe**-Leu-Ala-	79%	(3)
	CYC1-9-CB	**Met**-**Asn**-**Gln**-Phe-Lys-	0%	(3)
NatC	L-A *gag*	Ac-**Met**-**Leu**-Arg-Phe-Val-	100%	(28)
	JC33B	**Met**-**Leu**-Arg-Glu-Val-	0%	(28)
	JC32A	Ac-**Met**-**Leu**-Ala-Phe-Val-	100%	(28)
	JC32B	**Met**-**Leu**-Glu-Phe-Val-	0%	(28)
NatA	Cct2p	Ac-**Ser**-Val-Gln-Ile-Phe-	(100%)	(36)
	Mef1p	**Ser**-Val-Gln-Lys-Met-	(0%)	(36)
	Pyc1p	Ac-**Ser**-Gln-Arg-Lys-Phe-	(100%)	(36, 37)
	RPL3p	**Ser**-His-Arg-Lys-Tyr-	0%	(12)
	Grs1p	Ac-**Ser**-Val-Glu-Asp-Ile-	(100%)	(37)
	Yrb1p	**Ser**-Ser-Glu-Asp-Lys-	(0%)	(36)
	Efb1p	Ac-**Ala**-Ser-Thr-Asp-Phe-	(100%)	(37)
	Tpi1p	**Ala**-Arg-Thr-Phe-Phe-	0%	(27, 36, 37)
	Erg7p	Ac-**Thr**-Glu-Phe-Tyr-Ser-	100%	(see 27)
	CYC1-1371	**Thr**-Glu-Phe-Tyr-Ser-	0%	(27)
	Cyc1p	**Thr**-Glu-Phe-Lys-Ala-	0%	(1)

Biological Importance of N-terminal Acetylation

The finding that N-terminal acetylation, occurring posttranslationally, causes increased melanotropic effects of α-melanocyte-stimulating hormone while it reduces the analgesic action of β-endorphin is the clearest example of the biological importance of this modification (39, 40). However, there are surprisingly few examples demonstrating the biological importance of N-terminal acetylation occurring cotranslationally.

Alterations at N-termini, including loss of acetylation, decreased thermal stabilities of NADP-specific glutamate dehydrogenase from *Neurospora crassa* (41) and the *E. coli* ribosomal 5 S protein (42). Nonacetylated cytoplasmic actin from cultured *Drosophila* cells is less efficient in the assembly of microfilaments than the acetylated form (43). Hershko *et al*. (44) observed that N-terminal-acetylated cytochrome *c* and enolase from mammalian sources were not degraded *in vitro*, in contrast to the nonacetylated correspondents from yeast, which were good substrates. Also, Matsuura *et al*. (45) suggested that N-terminal acetylation protected apocytochrome *c* from degradation *in vitro*. It should be emphasized that in these and other examples, the proteins lacking acetylated termini also had other differences in amino acid sequences. In contrast, R. E. Cohen *et al*.[2] used *NAT1*$^+$ and *nat1*$^-$ yeast strains to prepare acetylated and unacetylated pairs of rat and yeast cytochrome *c*, respectively, and observed equal extents of ubiquitin conjugation within each pair, although both yeast forms were more highly ubiquitinated than both of the rat forms. Thus, the difference in ubiquitination of mammalian and yeast cytochrome *c* is because of differences other than N-terminal acetylation. Furthermore, Mayer *et al*. (46) observed ubiquitin-dependent degradation of N-terminal acetylated proteins in a crude reticulocyte lysate.

A significant means for assessing the general importance of N-terminal acetylation comes from the phenotypic defects in the *ard1*-Δ (or *nat1*-Δ), *mak3*-Δ, and *nat3*-Δ mutants. As described above, the silent mating loci, particularly *HMLα*, are partially derepressed in *nat1*$^-$ and *ard1*$^-$ mutants, leading to a partial mating defect of *MAT***a** strains. In addition *nat1*$^-$ and *ard1*$^-$ mutants exhibit defects of slow growth; inability of homozygous diploid strains to sporulate; and the failure to enter G_0 when limited for nutrients (30). Presumably, these multiple defects are because of the lack of N-terminal acetylation of one or more specific proteins requiring acetylation for function. Diminished function by the lack of acetylation of the *SIR3* protein, for example, can explain the partial derepression of *HML* (47), whereas diminished function of any one of a number of proteins in the cAMP pathway can explain the failure to enter G_0 and the inability of homozygous diploids to sporulate (30). Also, as described above, the lack of N-terminal acetylation of the viral major coat protein, *gag,* in *mak3*$^-$ strains prevents assembly or maintenance of the viral particle (28). Also *mak3*$^-$ strains do not utilize nonfermentable carbon sources at 37 °C, probably because of the lack of acetylation of a still unidentified protein (27, 28). Similar to the other mutants, *nat3*-Δ mutants exhibit multiple defective phenotypes, including lack of growth on YPG medium at 37 °C, reduced growth on medium containing NaCl, and reduced mating of the *MATα* cells. Such defects could arise from the lack of acetylation of any of a number of proteins essential for different processes. Whereas the unacetylated proteins responsible for these defects are not easily identified, the temperature and NaCl sensitivity could be attributed to lack of acetylation of actin (Act1p), which contains a normal N-terminal sequence, Ac-Met-Asp-Ser-Glu. Many actin mutants are temperature- and NaCl-sensitive, including *act1–136*, which has D2A replacement (48). Furthermore, nonacetylated actin of *Dictyostelium discoideum* prepared *in vitro* weakened the interaction with actinomyosin (20).

Not only can the lack of acetylation result in various defects, but abnormal acetylation also can prevent normal functions. The acetylation of the N-terminal catalytic threonine residue of various 20 S proteasome subunits causes the loss of specific peptidase activities (22). Obviously, both N-terminal acetylation and the lack of N-terminal acetylation have evolved to meet the individual requirements of specific enzymes.

[2] R. E. Cohen and C. W. Sokolik, unpublished result.

Interestingly, the Met-Glu-Ile N-terminal sequence of Mak3p is characteristic of a NatB substrate. The acetylation of NatC substrates in the NatB mutant, *nat3*-Δ, indicates that acetylation of Mak3p is not required for its function.

The viability of *ard1*-Δ, *nat1*-Δ, *mak3*-Δ, and *nat3*-Δ mutants lacking NATs suggests that the role of acetylation may be subtle and not absolute for most proteins. Possibly only a subset of proteins actually requires this modification for activity or stability, whereas the remainder is acetylated only because their termini fortuitously correspond to consensus sequences. Clearly, N-terminal acetylation does not necessarily protect proteins from degradation, as often supposed, nor does it play any obvious role in protection of proteins from degradation by the "N-end rule" pathway. When asked, "what is the function of N-terminal acetylation?" we reply, "it varies, just like the function, for example, of a histidine residue."

Generality of N-terminal Acetylation of Eukaryotic Proteins

The similarity in the pattern of N-terminal acetylation of proteins from higher eukaryotes and *S. cerevisiae* (3) suggests that the same systems may operating in all eukaryotes. Also, orthologous genes encoding the three N-terminal acetyltransferases indicate that the same or similar N-terminal acetyltransferases may be operating in higher eukaryotes. Species containing orthologs of the yeast Ard1p include *Caenorhabditis elegans*, *Drosophila melanogaster*, *Mus musculus*, *Arabidopsis thaliana,* and *Homo sapiens*; of the yeast Nat3p include *C. elegans*, *D. melanogaster*, and *H. sapiens*; and of the yeast Mak3p include *A. thaliana.*[3] The presence of one or another ortholog in *Methanobacterium thermoautotrophicum*, *Aeropyrum pernix,* and *Pyrococcus abyssi* suggests that archaeobacteria employ the eukaryotic systems for N-terminal acetylation.

Acknowledgments—We thank Drs. Reed Wickner and Rolf Sternglanz for critically reading the manuscript.

REFERENCES

1. Tsunasawa, S., Stewart, J. W., and Sherman, F. (1985) *J. Biol. Chem.* **260,** 5382–5391
2. Sherman, F., Stewart, J. W., and Tsunasawa, S. (1985) *Bioessays* **3,** 27–31
3. Moerschell, R. P., Hosokawa, Y., Tsunasawa, S., and Sherman, F. (1990) *J. Biol. Chem.* **265,** 19638–19643
4. Hirel, H-P., Schmitter, J-M., Dessen, P., Fayat, G., and Blanquet, S. (1989) *Proc. Natl. Acad. Sci. U. S. A.* **86,** 8247–8251
5. Dalbïge, H., Bayen, S., and Pedersen, J. (1990) *FEBS Lett.* **266,** 1–3
6. Ben-Bassat, A., Bauer, K., Chang, S. Y., Myambo, K., Boosman, A., and Chang, S. (1987) *J. Bacteriol.* **169,** 751–757
7. Wingfield, P., Graber, P., Turcatti, G., Movva, N. R., Pelletier, M., Craig, S., Rose, K., and Miller, C. G. (1989) *Eur. J. Biochem.* **180,** 23–32
8. Huang, S., Elliott, R. C., Liu, P. S., Koduri, R. K., Weickmann, J. L., Lee, J. H., Blair, L. C., Gosh-Dastidar, P., Bradshaw, R. A., Bryan, K. M., Einarson, B., Kendall, R. L., Kolacs, K. H., and Saito, K. (1987) *Biochemistry* **26,** 8242–8246
9. Boissel, J. P., Kasper, T. J., and Bunn, H. F. (1988) *J. Biol. Chem.* **263,** 8443–8449
10. Chang, Y.-H., Teichert, U., and Smith, J. A. (1990) *J. Biol. Chem.* **265,** 19892–19897
11. Lowther, W. T., Orville, A. M., Madden, D. T., Lim, S., Rich, D. H., and Matthews, B. W. (1999) *Biochemistry* **38,** 7678–7688
12. Arnold, R., Polevoda, B., Reilly, J. P., and Sherman, F. (1999) *J. Biol. Chem.* **274,** 37035–37040
13. Prchal, J. T., Cashman, D. P., and Kan, Y. W. (1986) *Proc. Natl. Acad. Sci. U. S. A.* **83,** 24–27
14. Yamada, T., Kato, K., Kawahara, K., and Nishimura, O. (1986) *Biochem. Biophys. Res. Commun.* **135,** 837–843
15. Devlin, P. E., Drummond, R. J., Toy, P., Mark, D. F., Watt, K. W., and Devlin, J. J. (1988) *Gene* (*Amst.*) **65,** 13–22
16. Driessen, H. P., de Jong, W. W., Tesser, G. I., and Bloemendal, H. (1985) *CRC Crit. Rev. Biochem.* **18,** 281–325
17. Kendall, R. L., Yamada, R., and Bradshaw, R. A. (1990) *Methods Enzymol.* **185,** 398–407
18. Li, X., and Chang, Y.-H. (1995) *Proc. Natl. Acad. Sci. U. S. A.* **92,** 12357–12361
19. Towler, D. A., Eubanks, S. R., Towery, D. S., Adams, S. P., and Glaser, L. (1987) *J. Biol. Chem.* **262,** 1030–1036
20. Abe, A., Saeki, K., Yasunaga, T., and Wakabayashi, T. (2000) *Biochem. Biophys. Res. Commun.* **268,** 14–19
21. Hansen, J. E., Marner, J., Pavlov, D., Rubenstein, P. A., and Reisler, E. (2000) *Biochemistry* **39,** 1792–1799
22. Arendt, C. S., and Hochstrasser, M. (1999) *EMBO J.* **18,** 3575–3585
23. Groll, M., Heinemeyer, W., Jager, S., Ullrich, T., Bochtler, M., Wolf, D. H., and Huber, R. (1999) *Proc. Natl. Acad. Sci. U. S. A.* **96,** 10976–10983
24. Dores, R. M., McDonald, L. K., Steveson, T. C., and Sei, C. A. (1990) *Brain Behav. Evol.* **36,** 80–99
25. Bradshaw, R. A., Brickey, W. W., and Walker, K. W. (1998) *Trends Biochem. Sci.* **23,** 263–267
26. Persson, B., Flinta, C., von Heijne, G., and Jörnvall, H. (1985) *Eur. J. Biochem.* **152,** 523–527
27. Polevoda, B., Norbeck, J., Takakura, H., Blomberg, A., and Sherman, F. (1999) *EMBO J.* **18,** 6155–6168
28. Tercero, J. C., and Wickner, R. B. (1992) *J. Biol. Chem.* **267,** 20277–20281
29. Tercero, J. C., Riles, L. E., and Wickner, R. B. (1992) *J. Biol. Chem.* **267,** 20270–20276
30. Mullen, J. R., Kayne, P. S., Moerschell, R. P., Tsunasawa, S., Gribskov, M., Colavito-Shepanski, M., Grunstein, M., Sherman, F., and Sternglanz, R. (1989) *EMBO J.* **8,** 2067–2075
31. Park, E.-C., and Szostak, J. W. (1992) *EMBO J.* **11,** 2087–2093
32. Tercero, J. C., Dinman, J. D., and Wickner, R. B. (1993) *J. Bacteriol.* **175,** 3192–3194
33. Rigaut, G., Shevchenko, A., Rutz, B., Wilm, M., Mann, M., and Seraphin, B. (1999) *Nature Biotechnol.* **17,** 1030–1032
34. Flinta, C., Persson, B., Jörnvall, H., and von Heijne, G. (1986) *Eur. J. Biochem.* **154,** 193–196
35. Augen, J., and Wold, F. (1986) *Trends Biochem. Sci.* **11,** 494–497
36. Perrot, M., Sagliocco, F., Mini, T., Monbirot, C., Schneider, U., Shevchenko, A., Mann, M., Jeno, P., and Boucherie, H. (1999) *Electrophoresis* **20,** 2280–2298
37. Garrels, J. I., McLaughlin, C. S., Warner, J. R., Futcher, B., Latter, G. I., Kobayashi, R., Schwender, B., Volpe, T., Anderson, D. S., Mesquita-Fuentes, R., and Payne, W. E. (1997) *Electrophoresis* **18,** 1347–1360
38. Kimura, Y., Takaoka, M., Tanaka, S., Sassa, H., Tanaka, K., Polevoda, B., Sherman, F., and Hirano, H. (2000) *J. Biol. Chem.* **275,** 4635–4639
39. Symth, D. G., Massey, D. E., Zakarian, S., and Finnie, D. A. (1979) *Nature* **279,** 252–254
40. Symth, D. G., and Zakarian, S. (1980) *Nature* **288,** 613–615
41. Siddig, M. A. M., Kinsey, J. A., Fincham, J. R. S., and Keighren, M. (1980) *J. Mol. Biol.* **137,** 125–135
42. Cumberlidge, A. G., and Isono, K. (1979) *J. Mol. Biol.* **131,** 169–189
43. Berger, E. M., Cox, G., Weber, L., and Kenney, J. L. (1981) *Biochem. Genet.* **19,** 321–331
44. Hershko, A., Heller, H., Eytan, E., Kaklij, G., and Rose, I. A. (1984) *Proc. Natl. Acad. Sci. U. S. A.* **81,** 7051–7025
45. Matsuura, S., Arpin, M., Hannum, C., Margoliash, E., Sabatini, D. D., and Morimoto, T. (1981) *Proc. Natl. Acad. Sci. U. S. A.* **78,** 4368–4372
46. Mayer, A., Siegel, N. R., Schwartz, A. L., and Ciechanover, A. (1989) *Science* **244,** 1480–1483
47. Stone, E. M., Swanson, M. J., Romeo, A. M., Hicks, J. B., and Sternglanz, R. (1991) *Mol. Cell. Biol.* **11,** 2253–2262
48. Wertman, K. F., Drubin, D. G., and Bostein, D. (1992) *Genetics* **132,** 337–350

[3] B. Polevoda and F. Sherman, unpublished results.

http://genomebiology.com/2002/3/5/reviews/0006.1

Review

The diversity of acetylated proteins

Bogdan Polevoda and Fred Sherman

Address: Department of Biochemistry and Biophysics, University of Rochester School of Medicine and Dentistry, Rochester, New York 14642, USA.

Correspondence: Fred Sherman. E-mail: Fred_Sherman@urmc.rochester.edu

Published: 30 April 2002

Genome ***Biology*** 2002, **3(5)**:reviews0006.1–0006

The electronic version of this article is the complete one and can be found online at http://genomebiology.com/2002/3/5/reviews/0006

Abstract

Acetylation of proteins, either on various amino-terminal residues or on the ε-amino group of lysine residues, is catalyzed by a wide range of acetyltransferases. Amino-terminal acetylation occurs on the bulk of eukaryotic proteins and on regulatory peptides, whereas lysine acetylation occurs at different positions on a variety of proteins, including histones, transcription factors, nuclear import factors, and α-tubulin.

Modification of proteins extends the range of possible molecular structures beyond the limits imposed by the 20 encoded amino acids and, if reversible, gives a means of control and signaling. Many proteins are acetylated, both co- and post-translationally, and at least for eukaryotic proteins, acetylation is the most common covalent modification out of over 200 types that have been reported. Acetylation of proteins is catalyzed by a wide range of acetyltransferases that transfer acetyl groups from acetyl-coenzyme A to either the α-amino group of amino-terminal residues or to the ε-amino group of lysine residues at various positions. (The α-amino group designates the position of the central carbon atom of amino acids, whereas the ε-amino group of lysine residues designates the position of a carbon atom in the side chain.) As shown in Table 1, amino-terminal acetylation occurs co-translationally on the bulk of acetylated eukaryotic proteins [1-3] and post-translationally on prokaryotic ribosomal proteins [4,5] and on processed eukaryotic regulatory peptides [6]. Amino-terminal acetylation is one of the most common protein modifications in eukaryotes, occurring on approximately 85% of eukaryotic proteins, but is rare for prokaryotic proteins [1-3]. Furthermore, ε-lysine acetylation occurs post-translationally on histones, high mobility group (HMG) proteins, transcription factors, nuclear receptors [7-9], and α-tubulin [10]. Acetylation affects many protein functions, including enzymatic activity, stability, DNA binding, protein-protein interaction, and peptide-receptor recognition, and occurs on numerous and diverse proteins.

Amino-terminal acetylation

Cotranslational amino-terminal acetylation of eukaryotic proteins

Studies *in vitro* indicate that amino-terminal acetylation of eukaryotic proteins takes place when there are between 20 and 50 residues protruding from the ribosome [1,11]. Proteins susceptible to amino-terminal acetylation have a variety of different amino-terminal sequences, with no simple consensus motifs and no dependence on a single type of residue [1,3,12]. Proteins with serine and alanine termini are the most frequently acetylated, and these residues, along with methionine, glycine, and threonine, account for over 95% of the amino-terminal acetylated residues [1,2]. Only subsets of proteins with any of these amino-terminal residues are acetylated, however, and none of them guarantees acetylation [3]. The complexity of the termini that are acetylated is due to the presence of multiple N-acetyltransferases (NATs; Tables 1,2), each acting on different groups of amino-acid sequences and whose specificity is determined by two or more residues at the amino-terminal positions [13]. Unlike the situation for histones and other proteins with acetylated ε-lysine residues, amino-terminal modifications are irreversible.

Studies with the yeast *Saccharomyces cerevisiae* have revealed three amino-terminal acetyltransferases, NatA, NatB, and NatC, that act on different groups of substrates; each group of substrates has a different degenerate motif recognized by the NAT [3]. As shown in Table 2, all

Table 1

Acetylated proteins and the corresponding acetyltransferases that act either cotranslationally (Co) or post-translationally (Post)

Acetylated proteins	Residues	Process	Acetyltransferases	References
Majority of eukaryotic proteins	α-Ser, α-Ala, and so on	Co	NatA, NatB, and NatC	Reviewed in [3]
E. coli ribosomal proteins (S18, S5, and L12)	α-Ser, α-Ala	Post	RimI, RimJ, and RimL	[4,5]
Regulatory peptides (β-endorphin, α-MSH, enkephalin, GHRF)	α-Tyr, α-Ser, and α-Ala	Post	Unknown	[6,17]
Histones (H2A, H2B, H3, H4)	ε-Lys	Co and Post	GNAT group: Gcn5, PCAF, Hat1, Elp3, and Hpa2	Reviewed in [7,21,25]
			MYST group: Esa1, MOF, Sas2, Sas3, Tip60, and MORF	
			p300/CBP group	
			Transcription factor group: TAFII250 and TFIIIC	
			Nuclear receptors cofactors group: ACTR and SRC1	
Transcription factors (p53, E2F1-3, EKLF, TFIIEβ, TFIIF, c-Jun, TCF, GATA1, MyoD, HMGI(Y), pRb, NF-E2(MafG) and ACTR)	ε-Lys	Post?	PCAF/GCN5, p300/CBP, TAFII250, SRC1?, MOZ, Tip60? and BRCA2?	Reviewed in [8,24]
HMG proteins (HMG1 and HMG2)	ε-Lys2 and ε-Lys11	Unknown	p300/CBP and PCAF	[27,28]
Nuclear receptor HNF-4	ε-Lys	Unknown	p300/CBP	[32]
Nuclear import factors (importin-α7 and Rch1)	ε-Lys22	Post	p300/CBP	[9]
α-tubulin	ε-Lys40	Post	62-67 kDa protein	[10,41]

Abbreviations not mentioned in the text: BRCA2, breast cancer protein; Elp3, elongator protein, a subunit of the RNA polymerase II holoenzyme complex; Esa1, essential SAS2-related acetyltransferase; Gcn5, general control nonrepressible protein, a nucleosomal histone acetyltransferase; GHRF, growth-hormone-releasing factor; GNAT, Gcn5p-related amino-acetyltransferase superfamily; Hpa2, histone and other protein acetyltransferase; MOF, males absent on the first, an X-linked dosage-compensation protein in *Drosophila*; MORF, monocytic leukemia zinc-finger protein related factor; MOZ, monocytic leukemia zinc-finger protein; MYST group, named for the founding members MOZ, YBF2/SAS3 and Tip60; p53, a tumor-suppressor protein; pRb, retinoblastoma protein; Rch1, Rag1 cohort, human importin-α; Sas2, something about silencing protein, involved in silencing at telomeres and mating-type loci; SRC1, steroid nuclear receptor coactivator; Tip60, HIV Tat-interactive protein. A question mark indicates uncertainty.

amino-terminal acetylated proteins are substrates for one of NatA, NatB or NatC. Furthermore, we do not know of any acetylated proteins in yeast that could not reasonably be a NatA, NatB or NatC substrate. Nevertheless, it remains to be seen if there are other NATs that act on rarer substrates. The similarity in the pattern of amino-terminal acetylation of the proteins from higher eukaryotes and *S. cerevisiae* and the presence of genes orthologous to those encoding the three amino-terminal acetyltransferases in mammals and plants (our unpublished observations) suggest that the same systems may operate in all eukaryotes.

The biological significance of amino-terminal modification varies; some proteins require acetylation for function whereas others that are acetylated do not absolutely require the modification. The viability of yeast mutants lacking the catalytic subunits (*ard1*-Δ, *mak3*-Δ or *nat3*-Δ) or other subunits (*nat1*-Δ) of NATs suggests that the function of acetylation may be subtle and not absolute for most proteins [13]. It is possible that only a subset of proteins actually requires this modification for activity or stability, whereas the remainder are acetylated only because their termini fortuitously correspond to consensus sequences. Amino-terminal acetylation does not necessarily protect proteins from degradation, as has often been supposed, nor does it play any obvious role in protection of proteins from degradation by the 'N-end rule' pathway that determines whether to degrade proteins according to their amino-terminal residue.

Amino-terminal acetylated proteins in prokaryotes

There are few examples of amino-terminal acetylated proteins in prokaryotes; they include the *Escherichia coli* ribosomal

Table 2

The three types of yeast amino-terminal acetyltransferases

	NatA	NatB	NatC
Catalytic subunit	Ard1p	Nat3p	Mak3p
Other subunits	Nat1p	Mdm20p	Mak10p
	Others	Others	Mak31p
Substrates*	α-Ser-	α-Met-Glu-	α-Met-Ile-
	α-Ala-	α-Met-Asp-	α-Met-Leu-
	α-Gly-	α-Met-Asn-	α-Met-Trp-
	α-Thr-	α-Met-Met-	α-Met-Phe-

*Acetylation occurs on all proteins with α-Met-Glu- and α-Met-Asp- termini but only on subclasses of proteins with the other termini.

proteins S5, S18 and L12 [4,5] and mycobacterial ribosomal protein L12. These modifications probably occur post-translationally (Table 1). The corresponding *E. coli* NAT genes, *rimI*, *rimJ*, and *rimL*, have been identified but it is still unclear how related they are - structurally, functionally and evolutionarily - to eukaryotic cotranslational NATs. These *E. coli* NATs are analogous to eukaryotic NatAs, which also acetylate α-Ser and α-Ala residues of ribosomal proteins.

Amino-terminal acetylation of processed regulatory peptides and hormones

Most eukaryotic regulatory peptides, hormones, and neurotransmitters are synthesized in the cell as larger precursor proteins, which are biologically inactive and must undergo a variety of post-translational processing steps to yield the active peptides [14]. After translation, the precursor is generally cleaved by an endopeptidase; this is followed by peptide modification, for example by carboxy-terminal amidation, sulfation, phosphorylation, glycosylation or amino-terminal acetylation [6]. These modifications frequently have a profound influence on the biological activity of the peptide; for example, both sulfation on tyrosine and carboxy-terminal amidation are obligatory for full biological activity of the octapeptide cholecystokinin, a gastrointestinal hormone. Importantly, more than one biologically active peptide can be produced from the same precursor and there may be variations in the pathways of processing at any of several different levels, so that different cells expressing a particular gene may give rise to different peptide products [15,16].

The finding that amino-terminal acetylation, occurring post-translationally, increases the pigment-producing (melanotropic) effects of α-melanocyte-stimulating hormone (α-MSH) and enhances its activity in behavioral tests represents the clearest example of the biological importance of this modification [16,17]. Amino-terminal acetylation also causes a greater than 50-fold increase in the potency of growth hormone releasing factor. In contrast, amino-terminal acetylation of β-endorphin, which takes place in the 'storage' form of this peptide hormone markedly reduces its opioid activity compared with the form with a free amino terminus [18]. Acetylation can also affect protein stability: there is evidence that the half-life of nonacetylated α-MSH in rabbit plasma is one-third of that of the acetylated form [19], and the stability of acetylated synthetic peptide MART-1 (a peptide derived from human melanoma-associated Melan-A antigen) is higher than that of MART-1 with a free amino terminus. Importantly, the expression of α-MSH and β-endorphin peptides is physiologically regulated and can be induced, but little is known about the factors that govern the cell-type-specific patterns of processing and modification of regulatory peptides; elucidation of these factors is currently a major challenge.

Similarly, little is known about the regulatory peptide acetyltransferases. Although a peptide acetyltransferase activity has been partially characterized [20], the corresponding gene has not been identified. It is unlikely, but cannot be excluded in certain cases, that NATs acting cotranslationally can modify regulatory peptides or their precursors. The amino-terminal residue of β-endorphin is tyrosine, however, which is not a substrate for cotranslational NATs; this tyrosine is nevertheless normally acetylated in the storage form of β-endorphin. Also, peptide acetyltransferases probably act in cooperation with peptide secretion, in contrast to cotranslational NATs that are likely to be associated with cytoplasmic polysomes. It is possible that neuropeptides and hormones may also be modified after secretion in ways that change their biological activities. Finally, it would be interesting to identify the genes encoding acetylpeptide hydrolases, which presumably deacetylate and thus activate β-endorphin from its acetylated storage form in mammals. Studies of the regulation of function of both peptide acetyltransferases and acetylpeptide hydrolases may be of great importance for the pharmacology and molecular genetics of human diseases.

Acetylation on internal lysines

Acetylation of histones

The most studied proteins that are acetylated on ε-lysine residues include histones H2A, H2B, H3, and H4, in which the modification occurs at multiple sites in the amino-terminal tail domains, and the HMG proteins, which are found in a variety of eukaryotes from yeast to humans [7]. The important feature of acetylation of ε-lysine residues is that it is reversible. Histones are frequently subjected to post-translational modifications that include acetylation, methylation, and phosphorylation of specific arginine, lysine, histidine, serine and threonine residues [21]. These modifications, many of which are also reversible, all decrease the positive charges of histone tail structures, thereby significantly altering histone-DNA binding, and interactions between nucleosomes and between histones and regulatory proteins. The discoveries of Gcn5p, the first nuclear histone acetyltransferase (HAT), and of the first histone deacetylase (HDAC),

established that acetylation of histones is an important controlling step in transcription [22]. Some of the nuclear HATs are also well known and extensively characterized as transcription factors. Not surprisingly, histone acetylation appears to influence other processes, including cell-cycle progression, chromosome dynamics, DNA replication, recombination and repair, silencing, and apoptosis [23]. Despite significant accumulation of information on HATs, understanding of the precise molecular role of histone acetylation in the assembly of chromatin, the accessibility of transcription factors and nucleosome remodeling is still elusive.

There are over 20 HATs that fall into several families, listed in Table 1. All HATs act in a site-specific and histone-specific manner, and specificity may differ *in vivo* and *in vitro*; such diversity that may help to explain why there are so many HATs. Remarkably, some HATs are associated with other HATs and coactivators, suggesting a layer of complexity that is not yet understood. It is important to note, however, that the steady-state balance of histone acetylation appears to exert different effects on different genes in different settings. Alignment of the amino-acid sequences surrounding modified lysines in acetylated proteins and mutagenesis of the human importin-α protein Rch1 suggest that the HAT recognition motif may be G**K**XXP (in the single-letter amino-acid code, with the acetylated ε-lysine residue in bold) [24].

Histone deacetylases

A large number of HDACs have now been identified, many of which act as corepressors of transcription [23]. The yeast deacetylases Rpd3p and Hda1p are recruited by repressor proteins to promoters, causing a localized deacetylation of chromatin [25]. Specialized regions of chromatin, including telomeres, centromeres, and silent yeast mating-type loci, are transcriptionally inactive and form hypoacetylated heterochromatin-like (tightly packaged) domains. Heterochromatin formation in yeast is mediated by the silencing proteins Sir2p, Sir3p, and Sir4p; Sir2p has been found to have HDAC activity. Interestingly, deacetylases are detected in some chromatin-remodeling complexes, which regulate changes in chromatin structure, together with HATs. Little is known about the specificity of HDACs, although it has been found that HDAC1 can deacetylate not only histones but also the transcription factor E2F1 [26].

Acetylation of HMG proteins

HMG proteins are a heterogeneous family of non-histone chromosomal proteins whose function is still not completely understood, despite their abundance and ubiquity. A subset of these proteins contains the HMG domain, a DNA-binding motif that recognizes bent DNA or induces bending in linear duplex DNA. Two post-translational modifications, namely phosphorylation and acetylation, influence the DNA-binding properties of HMG1. This protein is reversibly acetylated at conserved lysines at positions 2 and 11 [27], and it has been shown that monoacetylation at lysine 2 of HMG1 increases the binding affinity of the protein for some types of distorted DNA [28]. This indicates the possible involvement of HMG1 in DNA repair, separate from its 'architectural' role in nucleoprotein complexes. Also, HMG1 and HMG2 have been implicated in protein-protein interactions and have been shown to facilitate the specific binding of regulatory proteins - such as steroid hormone receptors, Hox and POU-domain proteins (developmental transcription factors), p53 (a tumor suppressor), and the TATA box-binding basal transcription factors - to their target DNA sequences [29].

Acetylation of transcription factors

In the nucleus, DNA is tightly packaged into several orders of structure with no easy accessibility for the transcription machinery. Acetylation of lysine residues within histones, histone-like proteins, and non-histone proteins (such as transcription factors) has recently emerged as a major mechanism used by the cell to overcome repressed chromatin states [8,9]. Several transcription factors have been identified as substrates for HATs, particularly for the HATs CREB-binding protein (CBP) and its close homolog p300, which are cofactors of nuclear-receptor-activated gene transcription, and p300/CBP-associated factor (PCAF). These substrate proteins include the transcriptional activators E2F1-3 (involved in progression through G1/S cell-cycle transition), p53, c-Jun (a transcription factor involved in the response to mitogens), the erythroid Krüppel-like transcription factor (EKLF), the transcriptional coactivator GATA1 that is required for megakaryocyte and erythrocyte differentiation, the muscle-specific differentiation regulator MyoD, the product of the proto-oncogene c-*myb*, the HMG protein HMGI(Y), the T-cell factor regulated transcription activator TCF (which is downstream of Wnt signaling proteins), hepatocyte nuclear factor HNF-4, the general transcription factors TFIIEβ and TFIIF, erythrocyte transcription factor NF-E2(MafG), and the steroid hormone nuclear receptor coactivator ACTR ([9,21,24-32] and references therein). The list of the new HAT substrates is growing rapidly. Acetylation of transcription factors can alter their ability to bind DNA (in the cases of E2F1, p53, EKLF, GATA1, and HNF-4), to interact with other proteins (c-Jun, TCF, ACTR, and HNF-4), or to remain in the nucleus (HNF-4). In addition, PCAF, p300 and CBP can autoacetylate, facilitating intramolecular rearrangements between the bromodomain (which binds acetyl-lysine) and the acetylated lysine(s); this interaction may be important for HAT activity and for recruitment of remodeling complexes to acetylated chromatin [33].

The effect of acetylation on DNA-binding-protein function depends on the location of the modified site within the protein. In case of the transcription factors p53, E2F1, EKLF, and GATA-1, the acetylation site is located directly adjacent to the DNA-binding domain, and acetylation stimulates DNA binding [26,30,34]. In contrast, the lysines acetylated within HMGI(Y) are within the DNA-binding domain and

result in disruption of DNA binding. Thus, acetylation does not always stimulate transcription.

Acetylation also affects protein-protein interactions. For example, the association of nuclear steroid hormone receptors with their coactivator ACTR is inhibited by acetylation [31]. Apparently, histone acetylation generates a recognition site for the bromodomain, a motif conserved in many proteins, including HATs [33]. Histone acetylation may precede the recruitment of ATP-dependent chromatin-remodeling activities during transcriptional activation. In particular, the HAT Gcn5p is involved in stabilizing binding of the SWI/SNF chromatin-remodeling complex to a promoter, and this interaction seems to be mediated through the Gcn5p bromodomain [21]. There is some evidence, exemplified by the transcription factor E2F1, that acetylation increases the half-life of the protein [26].

Acetylation of nuclear import factors

HATs can also target other nuclear proteins. A screen of a large set of proteins involved in different cellular processes resulted in the identification of two nuclear import proteins, Rch1 and importin-α7, as substrates for the acetyltransferase CBP [9]. The reaction seemed to be specific because another nuclear import factor, importin-α3, was not a substrate for CBP. Both p300 and CBP can mediate acetylation of Rch1 and importin-α7 *in vivo*, most likely in the nucleus [9]. The acetylated residue, ε-Lys22, lies within the binding site in Rch1 for the other nuclear import factor, importin-β, and acetylation of the site promotes interaction with importin-β *in vitro* [9]. Thus, it is possible that nuclear import may be regulated by acetylation, mediated by the p300/CBP HATs.

The targeting of HAT enzymes to their substrates is likely to be important and may play a role in regulation by other signaling pathways, as indicated by the finding that phosphorylation of p53 stimulates its acetylation, probably by increasing the association of p53 with p300 [35]. Some evidence indicates that the activity of HATs is regulated by proliferation and differentiation signals [23], via phosphorylation or hormonal signaling. For example, the HAT activity of CBP is stimulated at the G1-S phase boundary of the cell cycle, and hormone-induced acetylation of ACTR represses nuclear receptor function. Together, these results have led to the hypothesis that acetylation is a regulatory modification that may rival phosphorylation in cell signaling [36].

Acetylation of tubulin

Microtubules are cylindrical cytoskeletal structures that are found in almost all eukaryotic cell types and are involved in a great variety of cellular processes, including mitosis, ciliary and flagellar motility, intracellular transport of vesicles and organelles, and possibly in determining morphology of certain cells [37]. The structural subunit of microtubules is the 100 kDa protein tubulin, which consists of α and β isoforms that form heterodimeric complexes and associate head-to-tail to form profilaments and then laterally to make up the walls of cylindrical microtubules. Several types of post-translational modification affect tubulin function, including acetylation, phosphorylation, polyglutamation, polyglycylation, and detyrosination [10]. Most of these modifications are reversible and all, except acetylation, occur at the highly variable carboxyl termini of tubulin α and β subunits.

The first evidence for acetylation of tubulins was obtained with a flagellar tubulin from the unicellular alga *Polytomella* [38]. Tubulin acetylation has since been observed in vertebrates, insects, nematodes and plants, in all of which the acetyl group is attached to the ε-amino group of lysine 40. The α-tubulin acetyltransferase was purified from the flagellated unicellular alga *Chlamydomonas* and from mammalian brain and was shown to have molecular mass of 62-67 kDa [39]. During purification of the enzyme from *Chlamydomonas*, evidence was obtained for a tubulin deacetylase and for an inhibitor of α-tubulin acetyltransferase. In *Chlamydomonas*, the tubulin acetyltransferase exhibits a two-fold preference for polymerized over soluble tubulin, but in HeLa cells the acetylation occurs mainly after polymerization [40]. Generally, acetylation can happen quickly - almost immediately - and acetylated tubulin therefore does not necessarily demarcate old microtubules. Some correlation has been found between α-tubulin acetylation and microtubule stability [40]. Acetylated microtubules commonly resist drug-induced disassembly but not cold-induced disassembly, although in some cells a subset of acetylated microtubules is cold-resistant [41]. It is still unclear, however, how the intracellular spatial organization of acetylated microtubules is determined. There may be some factors limiting acetyltransferase enzyme activity to certain cellular microtubules and to restricted regions: candidates for such factors include the microtubule-associated proteins MAP1B, MAP2 and τ, which either enhance or inhibit the interaction of the acetyltransferase with microtubules [40]. Another possibility is that the interplay of microtubules with other cytoskeletal elements or organelles regulates acetyltransferase enzyme activity.

The role of acetylated microtubules in cells remains an important unanswered question. Acetylated tubulin is not required for survival, and a mutant of the ciliate *Tetrahymena* with lysine 40 replaced with arginine is indistinguishable from the wild type [41]. Cloning and analysis of the 62-67 kDa α-tubulin acetyltransferase mentioned above will be critical for understanding the role of α-tubulin acetylation.

Diversity of acetylated proteins

Acetylated proteins are varied, and acetylation can have a range of effects on protein function. Rapidly accumulating new results of functional analysis on HATs allowed

Kouzarides [36] to suggest that, like phosphorylation, acetylation can regulate such different essential processes in the cell as transcription, nuclear import, microtubule function, and hormonal response. At the moment, there is no clear evidence that acetyltransferases act in cascade, although there are other striking similarities between phosphorylation and acetylation. Both autoacetylation [33] and autophosphorylation are known to occur, and both phosphorylation and acetylation [21,23] can be reversible. Thus, acetylation may rival phosphorylation in cell signaling. Although both phosphorylation and acetylation alter the charge of the modified protein or protein domain, the effect of acetylation is 'milder'. The spectrum of substrates for acetylation is much broader than that of phosphorylation, however, and includes proteins and polypeptides from almost all cellular compartments and involves both amino-terminal and internal ε-lysine modifications. The biological role of protein acetylation is diverse, reflecting the different acetyltransferases that have evolved to meet the requirements of individual proteins or protein families.

Acknowledgements

We thank Jeffrey J. Hayes (University of Rochester) for useful comments. This work has been supported by National Institute of Health Grant R01 GM12702. B.P. dedicates this work to the memory of his father, Wasyl Polevoda, who had an extraordinary mind and a full life.

References

1. Driessen HPC, de Jong WW, Tesser GI, Bloemendal H: **The mechanism of amino-terminal acetylation of proteins.** *CRC Crit Rev Biochem* 1985, **18:**281-325.
2. Persson B, Flinta C, von Heijne G, Jörnvall H: **Structures of amino-terminally acetylated proteins.** *Eur J Biochem* 1985, **152:**523-527.
3. Polevoda B, Sherman F: **Nα-terminal acetylation of eukaryotic proteins.** *J Biol Chem* 2000, **275:** 36479-39482.
4. Yoshikawa A, Isono S, Sheback A, Isono, K: **Cloning and nucleotide sequencing of the genes *riml* and *rimJ* which encode enzymes acetylating ribosomal proteins S18 and S5 of *Escherichia coli* K12.** *Mol Gen Genet* 1987, **209:**481-488.
5. Tanaka S, Matsushita Y, Yoshikawa A, Isono K: **Cloning and molecular characterization of the gene *rimL* which encodes an enzyme acetylating ribosomal protein L12 of *Escherichia coli* K12.** *Mol Gen Genet* 1989, **217:**289-293.
6. Dockray GJ: **The biosynthesis of regulatory peptides.** *Am Rev Respir Dis* 1987, **136:**S6-S15.
7. Roth SY, Denu JM, Allis CD: **Histone acetyltransferases.** *Annu Rev Biochem* 2001, **70:**81-120.
8. Imhof A, Yang XJ, Ogryzko VV, Nakatani Y, Wolffe AP, Ge H: **Acetylation of general transcription factors by histone acetyltransferases.** *Curr Biol* 1997, **7:**689-692.
9. Bannister AJ, Mishka EA, Gorlich D, Kouzarides T: **Acetylation of importin-α nuclear import factors by CBP/p300.** *Curr Biol* 2000, **10:**467-470.
10. MacRae TH: **Tubulin post-translational modifications. Enzymes and their mechanisms of action.** *Eur J Biochem* 1997, **244:**265-278.
11. Bradshaw RA, Brickey WW, Walker KW: **Amino-terminal processing: the methionine aminopeptidase and Nα-acetyl transferase families.** *Trends Biochem Sci* 1998, **23:**263-267.
12. Augen J, Wold F: **How much sequence information is needed for the regulation of amino-terminal acetylation of eukaryotic proteins.** *Trends Bioch Sci* 1986, **11:**494-497.
13. Polevoda B, Norbeck J, Takakura H, Blomberg A, Sherman F: **Identification and specificity of amino-terminal acetyltransferases from *Saccharomyces cerevisiae*.** *EMBO J* 1999, **18:**6155-6168.
14. Harris RB: **Processing of pro-hormone precursor proteins.** *Arch Biochem Biophys* 1989, **275:**315-333.
15. Hirsch MD, O'Donohue TL: **Structural modifications of pro-opiomelanocortin-derived peptides alter their behavioral effect markedly.** *J Pharmacol Exper Ther* 1986, **237:**378-385.
16. O'Donohue TL, Handlemann GE, Miller RL, Jakobowitz DM: ***N*-acetylation regulates the behavioural activity of melanotropin in a multineurotransmitter neuron.** *Science* 1982, **215:**1125-1127.
17. Symth DG, Massey DE, Zakarian S, Finnie D: **Endorphins are stored in biologically active and inactive forms: isolation of α-*N*-acetyl peptides.** *Nature* 1979, **279:**252-254.
18. Symth DG, Zakarian S: **Selective processing of β-endorphin in regions of porcine pituitary.** *Nature* 1980, **288:**613-615.
19. Rudman D, Hollins BM, Kutner MB, Moffitt SD, Lynn MJ: **Three types of α-melanocyte-stimulating hormone: bioactivity and half-lives.** *Am J Physiol* 1983, **245:**E47-E54.
20. Glembotski CC: **Characterization of peptide acetyltransferase activity in bovine and rat intermediate pituitaries responsible for the acetylation of β-endorphin and α-melanotropin.** *J Biol Chem* 1982, **257:**10501-10509.
21. Grant PA: **A tale of histone modifications.** *Genome Biology* 2001, **2:**reviews0003.1-0003.6.
22. Struhl K: **Histone acetylation and transcriptional regulatory mechanisms.** *Genes Dev* 1998, **12:**599-606.
23. Kouzarides T: **Histone acetylases and deacetylases in cell proliferation.** *Curr Opin Genet Dev* 1999, **9:**40-48.
24. Bannister AJ, Miska EA: **Regulation of gene expression by transcription factor acetylation**. *Cell Mol Life Sci* 2000, **57:**1184-1192.
25. Wu J, Grunstein M: **25 years of the nucleosome model: chromatin modifications**. *Trends Biochem Sci* 2000, **25:**619-623.
26. Martinez-Balbas MA, Bauer U-M, Nielsen SJ, Brehm A, Kouzarides T: **Regulation of E2F1 activity by acetylation.** *EMBO J* 2000, **19:**662-671.
27. Sterner R, Vidali G, Alfrey VG: **Studies of acetylation and deacetylation in high mobility group proteins. Identification of the sites of acetylation in HMG-1.** *J Biol Chem* 1979, **254:**11577-11583.
28. Ugrinova I, Pasheva EA, Armengaud J, Pashev IG: ***In vitro* acetylation of HMG1 protein enhances its binding affinity to distorted DNA structures.** *Biochemistry* 2001, **40:**14655-14660.
29. Boonyaratankornkit V, Melvin V, Prendergast P, Altmann M, Ronfani L, Bianchi ME, Taraceviciene L, Nordeen SK, Allegretto EA, Edwards DP: **High-mobility group chromatin proteins 1 and 2 functionally interact with steroid hormone receptors to enhance their DNA binding *in vitro* and transcriptional activity in mammalian cells.** *Mol Cell Biol* 1998, **18:**4471-4487.
30. Gu W, Roeder RG: **Activation of p53 sequence-specific DNA binding by acetylation of the p53 C-terminal domain.** *Cell* 1997, **90:**595-606.
31. Chen H, Lin RJ, Xie W, Wilpitz D, Evans RM: **Regulation of hormone-induced histone hyperacetylation and gene activation via acetylation of an acetylase.** *Cell* 1999, **98:**675-686.
32. Soutoglou E, Katrakili N, Talianidis I: **Acetylation regulates transcription factor activity at multiple levels.** *Mol Cell* 2000, **5:**745-751.
33. Dhalluin C, Carston JE, Zeng L, He C, Aggarrwal A, Zhou M: **Structure and ligand of a histone acetyltransferase bromodomain**. *Nature* 1999, **399:**491-496.
34. Boyes J, Byfield P, Nakatani Y, Ogryzko V: **Regulation of activity of the transcription factor GATA-1 by acetylation.** *Nature* 1998, **396:**594-598.
35. Sakaguchi K, Herrera JE, Saito S, Miki T, Bustin M, Vassilev A, Anderson CW, Appela E: **DNA damage activates p53 through a phosphorylation-acetylation cascade.** *Genes Dev* 1998, **12:**2831-2841.
36. Kouzarides T: **Acetylation: a regulatory modification to rival phosphorylation?** *EMBO J* 2000, **19:**1176-1179.
37. Ludueña RF: **Multiple forms of tubulin: different gene products and covalent modifications**. *Int Rev Cytol* 1998, **178:**207-261.
38. L'Hernault SW, Rosenbaum JL: ***Chlamydomonas* α-tubulin is posttranslationally modified by acetylation on the ε-amino group of lysine.** *Biochemistry* 1985, **24:**473-478.
39. Greer K, Maruta H, L'Hernault SW, Rosenbaum JL: **α-tubulin acetylase activity in isolated *Chlamydomonas* flagella.** *J Cell Biol* 1985, **101:** 2081-2084.
40. Takemura R, Okabe S, Umeyama T, Kanai Y, Cowan NJ, Hirokawa N: **Increased microtubule stability and α-tubulun acetylation in cells transfected with microtubule-associated proteins MAP1B, MAP2 or τ.** *J Cell Sci* 1992, **103:**953-964.
41. Gaertig J, Cruz MA, Bowen J, Gu L, Pennok DG, Gorovsky MA: **Acetylation of lysine 40 in α-tubulin is not essential in *Tetrahymena thermophila*.** *J Cell Biol* 1995, **129:**1301-1310.

doi:10.1016/S0022-2836(02)01269-X *J. Mol. Biol.* (2003) **325**, 595–622

JMB

Available online at www.sciencedirect.com
SCIENCE DIRECT®

REVIEW

N-terminal Acetyltransferases and Sequence Requirements for N-terminal Acetylation of Eukaryotic Proteins

Bogdan Polevoda and Fred Sherman*

Department of Biochemistry and Biophysics, University of Rochester School of Medicine and Dentistry, Rochester NY 14642, USA

**Corresponding author*

N^{α}-terminal acetylation occurs in the yeast *Saccharomyces cerevisiae* by any of three N-terminal acetyltransferases (NAT), NatA, NatB, and NatC, which contain Ard1p, Nat3p and Mak3p catalytic subunits, respectively. The N-terminal sequences required for N-terminal acetylation, i.e. the NatA, NatB, and NatC substrates, were evaluated by considering over 450 yeast proteins previously examined in numerous studies, and were compared to the N-terminal sequences of more than 300 acetylated mammalian proteins. In addition, acetylated sequences of eukaryotic proteins were compared to the N termini of 810 eubacterial and 175 archaeal proteins, which are rarely acetylated. Protein orthologs of Ard1p, Nat3p and Mak3p were identified with the eukaryotic genomes of the sequences of model organisms, including *Caenorhabditis elegans*, *Drosophila melanogaster*, *Arabidopsis thaliana*, *Mus musculus* and *Homo sapiens*. Those and other putative acetyltransferases were assigned by phylogenetic analysis to the following six protein families: Ard1p; Nat3p; Mak3p; CAM; BAA; and Nat5p. The first three families correspond to the catalytic subunits of three major yeast NATs; these orthologous proteins were identified in eukaryotes, but not in prokaryotes; the CAM family include mammalian orthologs of the recently described Camello1 and Camello2 proteins whose substrates are unknown; the BAA family comprise bacterial and archaeal putative acetyltransferases whose biochemical activity have not been characterized; and the new Nat5p family assignment was on the basis of putative yeast NAT, Nat5p (YOR253W). Overall patterns of N-terminal acetylated proteins and the orthologous genes possibly encoding NATs suggest that yeast and higher eukaryotes have the same systems for N-terminal acetylation.

Keywords: acetylation; amino terminus; eukaryotic protein; N-terminal acetyltransferase; sequence requirement

Introduction

During protein synthesis and maturation, the N-terminal protein sequences of both intracellular and extracellular proteins undergo a number of modifications. Proteins from prokaryotes, mitochondria and chloroplasts initiate with formylmethionine, whereas proteins from the cytosol of eukaryotes initiate with methionine. The initial methionine may be deformylated; it may be removed; and the N-terminal residue may be acetylated or modified with another chemical group. In case of extracellular proteins and certain mitochondrial, endoplasmic reticulum, Golgi, vacuolar or vesicular proteins, i.e. proteins targeted to the specific cell compartments, a portion of the N-terminal protein sequence may be cleaved off, usually 15–30 amino acid residues long, exposing new N-terminal residues that may be further modified. Methionine cleavage and N-terminal acetylation are two major types of protein modifications.[1,2] Additional modifications of

Abbreviations used: Ac-CoA, acetyl-coenzyme A; AARE, acylamino acid-releasing enzyme; iso-1, iso-1-cytochrome *c*; MAP, methionine aminopeptidase; N-terminal, NH_2-terminal or α-amino terminal; NAT, N-terminal acetyltransferase.

E-mail address of the corresponding author: fred_sherman@urmc.rochester.edu

Table 1. Examples of altered iso-1-cytochromes *c* processed differently at the N terminus

	-1 1 2 3 4 5 6	Met cleavage	Acetylation
	(Met) Thr-Glu-Phe-Lys-Ala-Gly-		
CYC1^{+}	ATA ATG ACT GAA TTC AAG GCC GGT		
cyc1-31	ATA ATG ACT GAA TA- AAG GCC GGT		
	ATA ATA ATG TTG TTC TTG GCC GGT		
CYC1-850	Met-Leu-Phe-Leu-Ala-Gly-	0	0
	ATA ATA ATG GGT TTC TTG GCC GGT		
CYC1-841	(Met) Gly-Phe-Leu-Ala-Gly-	+	0
	ATA ATA ATG GAA TTC TTG GCC GGT		
CYC1-853	Ac-Met-Glu-Phe-Leu-Ala-Gly-	0	+
	ATA ATG TCT GAA TTC TTG GCC GGT		
CYC1-987	(Met) Ser-Glu-Phe-Leu-Ala-Gly-	+	+
	Ac		

The altered iso-1-cytochromes *c* were created by transforming the *cyc1-31* strain with synthetic oligonucleotides and selecting for functional transformants. Amino acid sequences of the N-terminal region of the iso-1-cytochromes *c* are presented along with the corresponding DNA sequences of *CYC1* alleles. Nucleotides of the transformants that differ from the *cyc1-31* sequence are designated in green. The penultimate residues are denoted in red. Cleaved N-terminal methionine residues are shown in parenthesis. The *cyc1-31* mutant completely lacks iso-1-cytochrome *c* because of the frameshift and TAA nonsense mutations, shown in blue. Altered iso-1-cytochromes *c* with four types of amino termini are illustrated, without (0) and with (+) cleavage of the N-terminal methionine and without (0) and with (+) N-terminal acetylation. (Adapted from[22,23]).

protein N termini include the following: methylation, mostly of alanine, methionine and proline residues; myristoylation of glycine residues; and the addition of more rare blocking groups, including α-amino acyl, pyroglutamate, pyruvoyl, α-ketobutyl, glucuronyl, glucose and murein.[3,4] There also some examples of double N-terminal modifications, particularly acetylation and phosphorylation, involving serine and threonine residues.[5] Many of these reactions take place cotranslationally, when the N terminus of the nascent polypeptide emerges from the ribosome and is only 20–50 residues long or still attached to the ribosome,[6,7] indicating that the susceptibility for these modifications is determined primarily by the N-terminal region of the protein.

N-terminal methionine cleavage

The methionine at N termini is cleaved from nascent chains of most prokaryotic and eukaryotic proteins. Cleavage of N-terminal methionine residues is by far the most common modification and is catalyzed cotranslationally by methionine aminopeptidases (MAP); one enzyme is described in bacteria and archaea, MAP type I and MAP type II, respectively. The archaeal MAP is not highly homologous to the bacterial enzyme. The bacteria and archaea MAPs have similar specificity, and resemble, respectively, MAP I and MAP II type enzymes found in eukaryotes.[2,8] It is also possible that eukaryotic organelles with their own translation machinery might contain different MAP isoforms, as was shown recently for chloroplasts and mitochondria of *Arabidopsis thaliana*.[9] Removal of N-terminal methionine is an essential function in yeast, as in prokaryotes, but the process can be carried out by either of two enzymes.[8] Experiments with altered iso-1-cytochromes *c* (iso-1) from yeast were the basis for the hypothesis that methionine is cleaved from penultimate residues having radii of gyration of 1.29 Å or less (glycine, alanine, serine, cysteine, threonine, proline, and valine residues),[10] a hypothesis that was confirmed in other studies with prokaryotic systems *in vivo* and *in vitro*[11,12] and other eukaryotic systems *in vivo* and *in vitro*.[13,14] The lack of methionine aminopeptidase action on proteins with large penultimate residues, as discussed above, can be now explained by steric hindrance, as deduced from the crystal structure of MAP.[15]

N-terminal acetylation

N-terminal acetylation is an enzyme-catalyzed reaction in which the protein α-amino group accepts the acetyl group from acetyl-CoA. The enzyme N-terminal acetyltransferase (NAT) has been found in all kingdoms, prokaryotes, archaea and eukaryotes, but N-terminal acetylation is likely to be cotranslational only in eukaryotes (see also below). There are some examples of viral protein acetylation but it normally employs the host cell NAT system. *In vitro* studies indicated that N-terminal acetylation of eukaryotic proteins occurs when there are between 25 and 50 residues extruding from the ribosome.[6,16] Similar to methionine cleavage, N-terminal acetylation is one of the most common protein modifications in eukaryotes, occurring on approximately 80–90%

Table 2. N-terminal sequences of acetylated and non-acetylated yeast proteins

Sequence	Acetylation[a]	NAT substrate[b]	Gene or protein	Mutants tested[c]	Method[d]	References	Footnotes
AAEKI	0	–	*RPL7A*	All	2D, MS	31,73	
Ac-AAGEQ	+	A	*GUA1*	*nat1*	2D	29,74	
AAQSK	0	–	*RPL39*	All	MS	31	
Ac-AARPQ	+	A	*CCT5*	*nat1*	2D	29,74	
Ac-ADITD	±	A	*PAB1*	*nat1*	2D	29,30,74	
ADQEN	0	–	*COX12*	None	α	33	
Ac-AEASI	+	A	*GLN1*	All	2D	24,29,30,74	
Ac-AEFLA	+	A′	*CYC1-962*	All	HPLC	24	
Ac-AEGVF	+	A	*SSB1*	All	2D	24,29,30,74	
Ac-AEGVF	+	A	*SSB2*	All	2D	24,29,30,74	
Ac-AEKEE	+	A	*VPH1*	*nat1*	2D	30	
Ac-AESHR	+	A	*RPL33A*	All	MS	31	
AFLAG	0	–	*CYC1-842*	None	HPLC	22	
AFQKD	0	–	*RPL1*	All	MS	29,31,74	
Ac-AGAIE	±	A	*VMA1*	All	2D	24,29,74	
Ac-AGETF	±	A	*HSC82*	All	2D	29,74	
AGETF	0	–	*HSC82*	All	2D	24,30	
AGGVL	0	–	*ADE3*	None	α	33	
AGKKI	0	–	*HOM2*	*nat1*	2D	29,30,74	
AGLKD	0	–	*RPL31A*	All	MS	31	
AGQVL	0	–	*ADE3*	*nat1*	2D	29,74	
AGSFL	0	–	*CYC1-13-S*	None	HPLC	21	
Ac-AGTFL	±	A	*SAM1*	All	2D	29,74	
AGTFL	0	–	*SAM1*	All	α	24,33	
Ac-AGTFL	±	A	*SAM2*	*nat1*	2D	29,74	
AGVSV	0	–	*RPS19B*	All	MS	31	
AGWDI	0	–	*SAR1*	*nat1*	2D	29,30,74	
AHENV	0	–	*RPS29*	All	MS	31	
AHIPE	0	–	*LCB1*	*nat1*	2D	30	
AIDYS	0	–	*CDC37*	*nat1*	2D	29,74	
AKEST	0	–	*CYC7*	None	α	33	
AKFLK	0	–	*RPL27A*	All	MS	31	
AKSKN	0	–	*RPL29A*	All	MS	31	
AKSKN	0	–	*RPL29B*	All	MS	31	
AKVHG	0	–	*RPS30A*	All	MS	31	
ANLRT	0	–	*RPL19A*	All	2D, MS	31,73	
ANLRT	0	–	*RPL19B*	*nat1*	2D, α	73/c >	
APGKK	0	–	*RPL8A*	All	2D, MS	31,73	
APGKK	0	–	*RPL8B*	*nat1*	2D, α	73	
APNTS	0	–	*RPL22A*	All	MS	31	
APPKK	0	–	*TIF3*	*nat1*	2D	29,74	
APSAK	0	–	*RPL25*	All	2D, MS	31,73	
APVKS	0	–	*RPL30*	All	MS	31	
APVTI	0	–	YDR380W	*nat1*	2D	29,74	
Ac-AQEEI	+	A	*SPE3*	*nat1*	2D	29,30,74	
ARDLQ	0	–	*GSY1*	None	α	33	
AREIT	0	–	*RPL38*	All	MS	31	
ARFVT	0	–	JC37	*mak3*	α	28	
ARGPK	0	–	*RPS4A*	All	MS	31	
ARRPA	0	–	*RPL10*	All	MS	31	
ARTFF	0	–	*TPI1*	All	2D	24,29,30,74	
ARYGA	0	–	*RPL17A*	All	MS	31	
Ac-ASETE	±	A	*HSP82*	*nat1*	2D	29,74	
Ac-ASIGS	+	A	*GCN20*	*nat1*	2D	29,74	
ASLPH	0	–	*RPL32*	All	MS	31	
Ac-ASNEV	+	A	*BEL1*	All	2D	24,29,30,74	
Ac-ASTAN	+	A	YDR341C	*nat1*	2D	29,30,74	
Ac-ASTDF	+	A	*EFB1*	*nat1*	2D	29,30,74	
AVGKN	0	–	*RPS1A*	All	MS, 2D	31,73	
AVGKN	0	–	*RPS1B*	*nat1*	2D, α	73	
AVKTG	0	–	*RPL36B*	All	MS	31	
AVSKV	0	–	*ENO1*	All	2D	24,29,30,74	
AVSKV	0	–	*ENO2*	All	2D	24,29,30,74	
CDSEV	0	–	*CYC1-1071*	None	HPLC	24	
CEFLA	0	–	*CYC1-1070*	None	HPLC	24	
CFLAG	0	–	*CYC1-844*	None	HPLC	22	
CGIFA	0	–	*ASN1*	*nat1*	2D	29,74	
CGIFA	0	–	*ASN2*	*nat1*	2D	29,74	
Ac-CGILG	±	(?)	*ADE4*	*nat1*	2D	29,74	
CPLAG	0	–	*CYC1-879*	None	HPLC	22	
GAYKY	0	–	*RP15A*	All	2D, MS	31,73	
GEEHK	0	–	*CDC48*	*nat1*	2D	29,30,74	

(*continued*)

Table 2 Continued

Sequence	Acetylation[a]	NAT substrate[b]	Gene or protein	Mutants tested[c]	Method[d]	References	Footnotes
Ac-GEFLA	+	A′	*CYC1-963*	All	HPLC	24	
GFLAG	0	−	*CYC1-841*	None	HPLC	22	
GGIRE	0	−	*RPLA0*	All	2D	24,30	
GISRD	0	−	*RPS8A*	All	MS	31	
GITVI	0	−	*RBK1*	*nat1*	2D	29,74	
GKNVL	0	−	*LYS9*	All	2D	24,29,30,74	
GKSHG	0	−	*RPL21A*	All	MS	31	
GLTTK	0	−	*GPP2*	*nat1*	2D	29,74	
GRMHS	0	−	*RPS13*	All	MS	31	
GRVIR	0	−	*RPL2A*	All	MS	31	
GSAFL	0	−	*CYC1-345-J*	None	HPLC	21	
Ac-GSRRY	+	A	*PRE9*	All	2D	29,32,74	
Ac-GSRSE	+	(A)	*CKB2*	None	α	75	
GVVEQ	0	−	*FBA1*	All	2D	24,29,30,74	
Ac-MDFLA	+	B	*CYC1-848*	All	HPLC	22,24	
Ac-MDIIL	+	B	*PRE1*	All	2D	32	
Ac-MDNEV	+	(B)	*SUP45*	*nat1*	MS	29,74	
Ac-MDPLA	±	(B)	*CYC1-878*	None	HPLC	22	
Ac-MDSEV	+	B	*ACT1*	All	2D	24	
Ac-MDSKT	+	B	*RPS28A*	All	MS	31	
Ac-MEAHN	+	B	*RNR4*	All	MS, 2D	24,29,74	
Ac-MEFLA	+	B	*CYC1-853*	All	HPLC	22,24	
Ac-MEEKL	+	(B)	*ADK1*	None	α	75	
Ac-MEHRY	+	(B)	*POM152*	None	α	75	
Ac-MENDK	+	B	*RPS21A*	All	MS	31	
Ac-MENDK	+	B	*RPS21B*	All	MS	31	
Ac-MERFV	+	(B)	JC38	*mak3*	α	28	
METQP	(0)	−	*CAR1*	*nat1*	2D	29,74	
MFFLA	0	−	*CYC1-856*	None	HPLC	22	
Ac-MFLTR	+	C	*PUP2*	All	2D	32	
MFNTT	(0)	−	*HAD1*	*nat1*	2D	29,74	
Ac-MFRAG	+	(C)	*IMP2*	None	α	75	
Ac-MFRNN	+	C	*PRE5*	All	2D	32	
MFRSV	(0)	−	*LYS12*	*nat1*	2D	29,74	
MFTGI	0	−	*RIB5*	None	α	33	
MHLAG	0	−	*CYC1-854*	None	HPLC	22	
Ac-MHRTY	+	(?)	YGR086C	*nat1*	2D	30	**e**
MIEFK	0	−	*CYC1-9-AU*	None	HPLC	21	
MIFLA	0	−	*CYC1-851*	None	HPLC	22	
MIFLK	0	−	*MRPL38*	None	α	33	
Ac-MIGSA	+	(C)	*AAD3*	*nat1*	2D	29,74	
MIIYK	0	−	YKL056C	All	2D	24,30	
Ac-MIKFK	+	(C)	*CYC1-9-BU*	None	HPLC	21	
Ac-MIRIL	+	(C)	*CYC1-31-Y*	None	HPLC	21	
Ac-MIRLK	+	C	*CYC1-1162*	All	HPLC	21,24	
MITEF	0	−	*CYC1-667-A*	None	HPLC	22	
MITGF	0	−	*CYC1-13-A*	None	HPLC	21	
MITKY	0	−	*MRP44*	None	α	33	
MKAIK	(0)	−	*TIF34*	*nat1*	2D	29,30,74	
MKAVV	(0)	−	YCR102C	*nat1*	2D	29,30,74	
MKDVL	(0)	−	*GYP6*	*nat1*	2D	30	
MKFLA	0	−	*CYC1-857*	None	HPLC	22	
MKFSA	(0)	−	*PDI1*	*nat1*	2D	29,30,74	
MKGAL	0	−	*CYC1-345-F*	None	HPLC	21	
MKGLI	(0)	−	*PSA1*	All	2D	24,29,30,74	
MKITE	(0)	−	*MET11*	*nat1*	2D	29,74	
MKLEN	(0)	−	*PRB1*	*nat1*	2D	29,74	
MKLQF	(0)	−	*CPR5*	*nat1*	2D	29,30,74	
MKMLT	0	−	*COP1*	None	α	33	
MKRFN	0	−	*GPP1*	All	2D	24,30	
MKSEN	0	−	*CYC2*	None	α	33	
MKTEF	0	−	*CYC1-493-A*	None	HPLC	21	
MKTLI	(0)	−	YGK037C	*nat1*	2D	29,74	
MKTYH	(0)	−	*MET16*	*nat1*	2D	29,30,74	
MKYII	(0)	−	YOR021C	*nat1*	2D	29,30,74	
MKYLA	0	−	*RPP2A*	All	2D	24,29,30,74	
MKYLA	0	−	*RPP2B*	All	2D	24,29,30,74	
MKYMV	(0)	−	YHR064C	*nat1*	2D	29,30,74	
MKYVV	(0)	−	*URA7*	*nat1*	2D	29,30,74	
MLAAK	(0)	−	*SSC1*	*nat1*	2D	29,30,74	
MLAAK	(0)	−	*SSC2*	*nat1*	2D	29,30,74	
MLAEK	(0)	−	*YHB1*	*nat1*	2D	29,74	
Ac-MLAFA	+	C	*CYC1-1286*	All	HPLC	24	

(*continued*)

Table 2 Continued

Sequence	Acetylation[a]	NAT substrate[b]	Gene or protein	Mutants tested[c]	Method[d]	References	Footnotes
Ac-MLAFV	+	C	JC32A	*mak3*	α	28	
MLDIN	(0)	−	*SES1*	*nat1*	2D	29,74	
MLEFV	0	−	JC32B	*mak3*	α	28	
MLFLA	0	−	*CYC1-850*	None	HPLC	22	
MLMPK	0	−	*RPS10A*	All	MS	31	
MLMPK	0	−	*RPS10B*	All	MS	31	
MLNIL	(0)	−	*ADE5,7*	*nat1*	2D	29,30,74	
MLNYC	0	−	*RPL26B*	All	MS	31	
MLREV	0	−	JC33B	*mak3*	α	28	
Ac-MLRAV	+	C	JC33A	*mak3*	α	28	
Ac-MLRFE	+	C	JC34A	*mak3*	α	28	
Ac-MLRFR	+	C	*CYC1-1201*	All	HPLC	24	
Ac-MLRFN	+	C	JC31	*mak3*	α	28	
Ac-MLRFV	+	C	L-A *gag*	*mak3*	α	28	
MLTEF	0	−	*CYC1-51-F*	None	HPLC	22	
Ac-MLVLS	±	(C)	*INO1*	*nat1*	2D	29,74	
MMEEF	0	−	*GRF10*	None	α	33	
MMFLA	0	−	*CYC1-855*	None	HPLC	22	
MMIMA	0	−	*CYC1-242-O*	None	HPLC	21	
Ac-MMKGS	+	(B)	*RPA14*	None	α	75	
Ac-MMNML	+	(B)	*CYC1-242-N*	None	HPLC	21	
Ac-MMNSR	+	(B)	*CYC1-183-T*	None	HPLC	21	
Ac-MMNSW	+	(B)	*CYC1-242-V*	None	HPLC	21	
Ac-MNDQT	+	(B)	*HSP104*	*nat1*	MS, 2D	29,74	
MNEKL	0	−	*CYC1-31N*	None	HPLC	21	
MNFGS	(0)	−	*CCT7*	*nat1*	2D	29,74	
Ac-MNFLA	±	(B)	*CYC1-849*	None	HPLC	22	
MNKFK	0	−	*CYC1-9-BT*	None	HPLC	21	
Ac-MNNLA	+	B	*CYC1-872*	All	HPLC	24	
Ac-MNNNL	+	(B)	*CYC1-345-H*	None	HPLC	21	
MNQFL	0	−	*CYC1-9-CB*	None	HPLC	21	
Ac-MNTDQ	+	(B)	*PAI3*	None	α	75	
MQAGL	0	−	*CYC1-345-C*	None	HPLC	21	
MQFLA	0	−	*CYC1-852*	None	HPLC	22	
MQSQD	(0)	−	*PFK1*	*nat1*	2D	29,30,74	
MREVI	(0)	−	*TUB1*	*nat1*	2D	30	
MRFLA	0	−	*CYC1-860*	None	HPLC	22	
MRFST	(0)	−	*BGL2*	*nat1*	2D	29,30,74	
MRTEF	0	−	*CYC1-133-A*	None	HPLC	21	
MTPLA	0	−	*CYC1-861*	None	HPLC	21	
MVLVQ	0	−	*RPS27A*	All	MS	31	
MVNVP	0	−	*RPL42B*	All	MS	31	
MVTEL	0	−	*CYC1-131-C*	None	HPLC	21	
MWFLA	0	−	*CYC1-859*	None	HPLC	22	
Ac-MWRFV	+	C	JC36	*mak3*	α	28	
MYFLA	0	−	*CYC1-858*	None	HPLC	22	
PALLK	0	−	YLR301W	*nat1*	2D	29,74	
PAPHG	0	−	*MET3*	*nat1*	2D	29,74	
PAPQD	0	−	*HEM13*	None	α	33	
PDYDN	0	−	*ADE13*	*nat1*	2D	29,74	
PEAKL	0	−	*PDR5*	None	α	33	
PEFLA	0	−	*CYC1-1093*	None	HPLC	24	
PFGID	0	−	*NCPR1*	*nat1*	2D	30	
PFLAG	0	−	*CYC1-846*	None	HPLC	22	
PFVKD	0	−	*OYE2*	*nat1*	2D	29,30,74	
PIDQE	0	−	*EGD1*	*nat1*	2D	29,74	
PKKVW	0	−	*WTM1*	*nat1*	2D	29,74	
PKLVL	0	−	*GPM1*	All	2D	24	
PLTTK	0	−	*GPP1*	*nat1*	2D	29,30,74	
PNASQ	0	−	*THR4*	All	2D	24,29,30,74	
PPKED	0	−	*YTA3*	*nat1*	2D	29,30,74	
PPLAG	0	−	*CYC1-881*	None	HPLC	22	
PPVSA	0	−	YER036C	*nat1*	2D	29,74	
PRVAI	0	−	YDR032C	*nat1*	2D	29,74	
PSHFD	0	−	*MET25*	All	2D	24,29,74	
PSLAE	0	−	*RPN12*	*nat1*	2D	29,74	
PSRFT	0	−	*RPL28*	All	MS	31	
PTVSV	0	−	*FRS1*	*nat1*	2D	29,74	
PYTLS	0	−	*SHM2*	All	2D	24	
Ac-SAAAD	+	A	*GPD1*	*nat1*	2D	29,74	
Ac-SAATV	+	A	*SFA1*	*nat1*	2D	29,74	
Ac-SADTG	±	A	*GND1*	*nat1*	2D	29,30,74	
Ac-SAEIE,	+	A	*SBP1*	*nat1*	2D	29,74	

(*continued*)

Table 2 Continued

Sequence	Acetylation[a]	NAT substrate[b]	Gene or protein	Mutants tested[c]	Method[d]	References	Footnotes
Ac-SAIPE	+	A	*EGD2*	*nat1*	2D	30	
Ac-SAKAQ	+	A	*RPL11A*	All	MS, 2D	31,73	
Ac-SAKSF	±	A	*DAK1*	*nat1*	2D	29,74	
Ac-SAPAA	±	A	*GSP1*	*nat1*	2D	29,30,74	
Ac-SAPAQ	+	A	*SAH1*	*nat1*	2D	29,30,74	
Ac-SAPEA	+	A	*RPS2*	All	2D, MS	31,73	
Ac-SAPQA	+	A	*RPS7A*	*nat1*	2D	73	
Ac-SAPTP	+	A	*TOM40*	*nat1*	2D	29,74	
Ac-SAQKA	+	A	*RPL6A*	All	MS	31	
Ac-SASEA	+	A	*THS1*	*nat1*	2D	29,74	
Ac-SASIP	+	A	YNL134C	*nat1*	2D	29,74	
Ac-SASKE	±	A	*SGT2*	*nat1*	2D	29,30,74	
Ac-SATLF	+	(A)	*AAT2*	None	α	75	
SATSA	(0)	−	*MOL1*	*nat1*	2D	30	
SAVNA	(0)	−	*ERG1*	*nat1*	2D	30	
Ac-SAVPS	+	A	*RPS16A*	All	MS, 2D	31,73	
Ac-SDAVT	+	A	*RPS24A*	All	MS, 2D	31,73	
Ac-SDEEH	+	A	*TIF51A*	*nat1*	α, 2D	29,30,74,75	
Ac-SDFQK	+	A′	*RPS20*	All	MS	31	
Ac-SDFQL	+	A	*FRS2*	*nat1*	2D	29,30,74	
Ac-SDGTK	+	A	*ARG4*	*nat1*	2D	29,74	
Ac-SDINE	+	A	*DYS1*	*nat1*	2D	29,74	
Ac-SDKVI	+	A	YKL117W	*nat1*	2D	30	
Ac-SDLVN	+	A	YLR109W	All	2D	24,29,30,74	
Ac-SDPSS	+	A	*PUP3*	All	2D	32	
Ac-SDPVE	+	A	*SEC17*	*nat1*	2D	29,74	
Ac-SDSQQ	+	A	*YEF3*	*nat1*	2D	29,30,74	
Ac-SDTEA	+	A	*RPS5*	All	2D, MS	29–31,73,74	
Ac-SEAQE	+	A	*NPL3*	*nat1*	2D	29,74	
Ac-SEATL	+	A	*TRP4*	*nat1*	2D	29,74	
Ac-SEEQD	+	A*	*SUG2*	*nat1*	2D	30,31	
Ac-SEFLA	+	D	*CYC1-987*	All	HPLC	24	
Ac-SEGIT	+	A	*TIF51A*	*nat1*	2D	29,74	
Ac-SEGPV	+	A	*ZWF1*	*nat1*	2D	30,75	
Ac-SEITL	±	A	*PDC1*	All	2D	24,29,30,74	
Ac-SENNE	+	A	*PUB1*	*nat1*	2D	29,30,74	
Ac-SEPAK	+	A	*TAL1*	All	2D	24,29,74	
Ac-SEPEF	+	A	*GDH1*	All	2D	24,29,30,74	
Ac-SEQLR	±	A	*TRP5*	*nat1*	2D	29,30,74	
Ac-SESPM	±	A	*ARO4*	*nat1*	2D	29,74	
Ac-SETEL	+	A	*ERG6*	*nat1*	2D	29,30,74	
Ac-SFDDL	±	A	*FSR2*	*nat1*	2D	29,30,74	
Ac-SFHQQ	+	A	YDL124W	*nat1*	2D	30	
Ac-SFKGF	+	A	*RVS167*	*nat1*	2D	29,74	
SFLAG	0	−	*CYC1-843*	None	HPLC	22	
Ac-SFLIS	+	A	*MES1*	*nat1*	2D	29,74	**f**
Ac-SFNAF	+	A	YIL041W	All	2D	24,29,30,74	
Ac-SGAAA	+	A	*SCL1*	All	2D	29,32,74	
Ac-SGYDR	+	A	*PRE6*	All	2D	32	
SHEGE	(0)	−	*SUB2*	*nat1*	2D	29,74	
SHRKY	0	−	*RPL3*	All	MS	31	
Ac-SHSLT	+	A	YFR044C	*nat1*	2D	29,74	
SIASY	0	−	*APT1*	*nat1*	2D	29,74	
Ac-SIAEF	±	A	*SEC53*	*nat1*	2D	29,30,74	**g**
Ac-SINIC	+	A	*TIF5*	*nat1*	2D	29,74	
Ac-SIPET	+	A	*ADH1*	All	2D	24,29,30,74	
Ac-SIQTS	+	A	YPL235W	*nat1*	2D	29,74	
Ac-SITKT	+	A	*ADE1*	*nat1*	2D	29,74	
Ac-SKAAV	+	A	*SSA1*	All	2D	24,29,30,74	
Ac-SKATV	+	A	*URA3*	*nat1*	2D	29,30,74	
Ac-SKAVG	+	A	*SSA2*	All	2D	24,29,30,74	
Ac-SKGKV	±	A	*ARG1*	All	2D	24,29,74	
Ac-SKITS	+	A	*RPL1A*	All	2D, MS	30,31	
Ac-SKITS	+	A	*RPL1B*	All	MS	31	
Ac-SKSKT	±	A	*SAM2*	All	2D	24,29,30,74	
Ac-SLAKS	+	(A)	*ATP7*	None	α	75	
Ac-SLERE	+	A	*SER1*	*nat1*	2D	29,30,74	
Ac-SLNIH	+	A	*DSK2*	*nat1*	2D	29,74	
Ac-SLPAT	+	A	*YST1*	All	2D, MS	29,30,74	
Ac-SLPAT	+	A	*YST2*	All	2D, MS	29–31,74	
Ac-SLQLL	+	A	*CCT6*	*nat1*	2D	29,30,74	
Ac-SLRLP	±	A	*CCT8*	*nat1*	2D	29,74	
Ac-SLSSK	+	A	*PGK1*	All	2D	24,29,74	

(*continued*)

Table 2 Continued

Sequence	Acetylation[a]	NAT substrate[b]	Gene or protein	Mutants tested[c]	Method[d]	References	Footnotes
Ac-SLTAD	±	A	*STI1*	All	2D	24,29,30,74	
Ac-SLVVQ	+	A	*RPS18A*	All	MS	31	
Ac-SNDET	+	A	*WRS1*	*nat1*	2D	29,74	
Ac-SNKLF	+	A	*PMI40*	*nat1*	2D	29,74,75	
Ac-SNNSF	±	A	*PGI1*	All	2D	24,29,30,74	
Ac-SNVVQ	+	A	*RPS14A*	All	MS	31	
SPLAG	0	−	*CYC1-880*	None	HPLC	22	
SPPVY	0	−	*POR1*	None	α	33	
SPRLK	0	−	*PPX1*	None	α	33	
Ac-SQAAK	+	(A)	*MRPL19*	None	α	33	
Ac-SQAVN	+	A	*RPS15*	All	MS	31	
Ac-SQDEN	±	A	*DPS1*	*nat1*	2D	29,74	
Ac-SQKIG	+	(A)	*MRP23*	None	α	33	
Ac-SQLFN	±	A	*CCT1*	*nat1*	2D	29,74	
Ac-SQPVV	±	A	*RPL16B*	All	MS	31	
Ac-SQQDN	±	A	*KRS1*	*nat1*	2D	29,30,74	
Ac-SQRKF	±	A	*PYC1*	*nat1*	2D	29,30,74	
Ac-SQTRE	+	A	*BMH2*	All	2D	24,29,30,74	
Ac-SQVYF	+	(A)	*CPR1*	None	α	75	
Ac-SQYAS	+	A	*KES1*	*nat1*	2D	29,30,74	
Ac-SREGF	+	(A)	*RPC10*	None	α	75	
Ac-SRFVT	+	(A)	JC35	*mak3*	α	28	
Ac-SRPQV	+	A	*RPL4A*	All	MS	31	
SRLER	0	−	*PYK1*	All	2D	24	
Ac-SRPIV	+	(A)	*GUK1*	None	α	75	
Ac-SRSGK	+	(A)	*COF1*	None	α	75	
Ac-SSAIT	±	A	*VMA4*	*nat1*	2D	29,30,74	
Ac-SSAIV	±	A	YNL179C	*nat1*	2D	29,30,74	
Ac-SSATA	+	A	*HTS1*	*nat1*	2D	29,74	
Ac-SSEDK	+	A	*YRB1*	*nat1*	2D	30	
Ac-SSGLV	+	A	*CDC60*	*nat1*	2D	29,30,74	
Ac-SSNLT	+	(A)	*CMD1*	None	α	75	
Ac-SSNNS	±	A	*UBA1*	*nat1*	2D	29,30,74	
Ac-SSSES	+	A	*ADK1*	*nat1*	2D, α	29,30,74,75	
Ac-SSSGV	+	A	YBR267W	*nat1*	2D	29,30,74	
Ac-SSSKL	±	A	*PYC2*	*nat1*	2D	29,30,74	
Ac-SSSQI	+	A	*PRO2*	*nat1*	2D	29,30,74	
Ac-SSSVA	+	A	*ARA*	*nat1*	2D	29,30,74	
Ac-SSVTG	+	A	*ADH4*	*nat1*	2D	29,74	
Ac-SSVQS	+	A	*RPS7B*	All	MS	31	
Ac-STDSI	+	A	*RPL14A*	All	MS	31	
STEKI	0	−	*RPL7B*	*nat1*	2D, α	73	
Ac-STELT	+	A	*RPS11A*	All	MS	31	
Ac-STESA	+	A	*RPP1A*	All	MS	31	
Ac-STKAQ	+	A	*RPL11B*	All	MS, 2D	31,73	
Ac-STLLK	+	(A)	*HSP10*	None	α	75	
Ac-STNFE	+	A	YJR070C	*nat1*	2D	29,30,74	
Ac-STPFG	+	A	*SSE1*	All	2D	24,29,30,74	**h**
STPFG	0	−	*SSE2*	*nat1*	2D	30	
Ac-STSRE	+	A	*BMH1*	All	2D	24,29,30,74	
Ac-STTAS	+	A	*ARG3*	*nat1*	2D	29,74	
Ac-STVNV	+	(A)	*ATP21*	None	α	75	
Ac-SVEDI	±	A	*GRS1*	*nat1*	2D	29,30,74	
Ac-SVEEV	+	A	*TIF45*	*nat1*	2D	29,30,74	
Ac-SVEPV	+	A	*RPL16A*	*nat1*	2D, α	73	
Ac-SVHAA	+	A	*TRP3*	*nat1*	2D	29,74	
Ac-SVKPI	+	(A)	*SRP21*	None	α	75	
Ac-SVQIF	+	A	*CCT2*	*nat1*	2D	29,74	
Ac-SWDDE	+	A	YLR192W	*nat1*	2D	29,74	
SYKQY	0	−	*MRPL25*	None	α	33	
TAAHP	0	−	YEL071W	*nat1*	2D	29,74	**i**
Ac-TAAKP	+	A	*LYS20*	*nat1*	2D	29,74	**i**
TADNN	0	−	*IDI1*	*nat1*	2D	29,74	
TAGSA	0	−	*ARO9*	*nat1*	2D	29,74	
TAKGL	0	−	*CYC1-31-K*	None	HPLC	21	
TAPLV	0	−	YJR105W	*nat1*	2D	30	**j**
Ac-TAPLV	±	A	YJR105W	*nat1*	2D	29,74	
TASDL	0	−	*ECM17*	*nat1*	2D	29,30,74	
TASIK	0	−	*TRP2*	*nat1*	2D	29,74	
TASLT	0	−	*URA1*	*nat1*	2D	30	
Ac-TASLT	±	A	*URA1*	*nat1*	2D	29,74	
TDILA	0	−	*CYC1-239-AB*	None	HPLC	21	
TDIQA	0	−	*CYC1-239-A*	None	HPLC	21	

(*continued*)

Table 2 Continued

Sequence	Acetylation[a]	NAT substrate[b]	Gene or protein	Mutants tested[c]	Method[d]	References	Footnotes
Ac-TDRYS	+	A	*PRE8*	All	2D	29,32,74	
TDSLA	0	−	*CYC1-183-U*	None	HPLC	21	
Ac-TDYIL	+	A	*ADE6*	*nat1*	2D	29,74	
TEFLA	0	−	*CYC1*	None	HPLC, α	24	
Ac-TEFEL	±	A	*GLY1*	*nat1*	2D	29,74	**k**
TEFWS	0	−	*CYC1-1371*	None	HPLC, α	24	**k**
TGEDF	0	−	*SNZ1*	*nat1*	2D	29,74	
TGILA	0	−	*CYC1-239-Y*	None	HPLC	21	
TIAPI	0	−	*COX9*	None	α	33	
TIGDK	0	−	*ALA1*	*nat1*	2D	29,74	
TIKEN	0	−	*ACS2*	*nat1*	2D	29,74	
TISLS	0	−	YNL123W	*nat1*	2D	29,74	
TISNL	0	−	*ACH1*	*nat1*	2D	29,74	
TKFLA	0	−	*CYC1-9-P*	None	HPLC	21	
Ac-TKIKV	+	(A)	*IDP2*	None	α	75	
TKLFH	0	−	*ALD6*	All	2D	24,29,30,74	
TKNFI	0	−	*PBI2*	None	α	75	
TKSEQ	0	−	*CYS4*	None	α	30,33	
Ac-TKSEQ	±	A	*CYS4*	*nat1*	2D	29,74	
TKSHS	0	−	*GAL1*	None	α	33	
TLFAG	0	−	*CYC1-9AA*	None	HPLC	21	
Ac-TLPES	±	A	*ARO8*	*nat1*	2D	29,74	**l**
TLPES	0	−	*ARO8*	*nat1*	2D	30	
Ac-TLQES	±	A	*CYS3*	*nat1*	2D	29,74	**m**
TNEKV	0	−	*LYS2*	*nat1*	2D	29,74	
TPFKA	0	−	*CYC1-9-CA*	None	HPLC	22	
TPLAG	0	−	*CYC1-861*	None	HPLC	22	
TPSTP	0	−	*SIC1*	None	α	75	
TQFTD	0	−	*TKL1*	*nat1*	2D	29,30,74	
TRSSV	0	−	*RPS22A*	All	MS	31	
TSFLA	0	−	*CYC1-9-AA*	None	HPLC	21	
Ac-TSIGT	+	A	*PRE10*	All	2D	29,32,74	
TSIYT	0	−	*CPA2*	*nat1*	2D	29,74	
Ac-TSLSS	+	(A)	*SEC26*	None	α	75	
TTAVR	0	−	*CCP1*	*nat1*	2D	29,74	
Ac-TTDNA	±	A	*TPS1*	*nat1*	2D	29,74	**n**
TTDNA	0	−	*TPS1*	*nat1*	2D	30	
Ac-TTNDT	+	A	*LYS7*	*nat1*	2D	29,74	
TVGIA	0	−	*MRPL8*	None	α	33	
Ac-TVKTG	+	A	*RPL36A*	All	MS	31	
TVPYL	0	−	*CIT2*	*nat1*	2D	29,30,74	
TVTTP	0	−	*PFK2*	*nat1*	2D	29,30,74	
TVYTA	0	−	*ERG19*	*nat1*	2D	29,74	
TYFLA	0	−	*CYC1-9-S*	None	HPLC	21	
TYTLA	0	−	*PRO3*	*nat1*	2D	29,74	
TYTTR	0	−	*IPP1*	*nat1*	2D	24,29,74	
VAFTV	0	−	*EFT1*	*nat1*	2D	29,30,74	
VAISE	0	−	YDR190C	*nat1*	2D	29,74	
VALIS	0	−	*RPS3*	All	2D, MS	31,73	
Ac-VAQVQ	±	(?)	*TSA1*	*nat1*	2D	29,74	**o**
VAQVQ	0	−	*TSA1*	*nat1*	2D	30	
VATVK	0	−	*ANC1*	*nat1*	2D	29,74	
VCDTL	0	−	*PEX11*	None	α	33	
VEFLA	0	−	*CYC1-1092*	None	HPLC	24	
VFLAG	0	−	*CYC1-847*	None	HPLC	22	
VFYKV	0	−	*MRPL33*	None	α	33	
VHLGP	0	−	*HXK1*	All	2D	24,29,30,74	
VHLGP	0	−	*HXK2*	*nat1*	2D	29,30,74	
VHNKV	0	−	*TRR1*	*nat1*	2D	29,74	
VKAVI	0	−	YNL010W	*nat1*	2D	29,30,74	
VKESI	0	−	*LEU4*	*nat1*	2D	29,74	
VKETK	0	−	*YDJ1*	*nat1*	2D	29,74	
VKVKS	0	−	*MRPL39*	None	α	33	
VLEAT	0	−	*RPN10*	*nat1*	2D	30	
VLEPS	0	−	*PPS1*	*nat1*	2D	30	
VLPIN	0	−	*HIS4*	*nat1*	2D	30	
VLSDK	0	−	*VMA2*	*nat1*	2D	29,30,74	
VPLAG	0	−	*CYC1-862*	None	HPLC	22	
VQAVA	0	−	*SOD1*	All	2D	24,29,30,74	
VQLAK	0	−	*ARO1*	*nat1*	2D	29,74	
VQSAV	0	−	*MET6*	All	2D	24,29,30,74	
VQAVA	0	−	*SOD1*	*nat1*	2D	29,74	
VRAFK	0	−	*THR1*	None	α	33	

(*continued*)

Table 2 Continued

Sequence	Acetylation[a]	NAT substrate[b]	Gene or protein	Mutants tested[c]	Method[d]	References	Footnotes
VRVAI	0	−	*TDH2*	All	2D	24,29,30,74	
VRVAI	0	−	*TDH3*	*nat1*	2D	29,30,33,74	
VSQET	0	−	*GRX1*	None	α	33	
VTEFK	0	−	*CYC1-497-B*	None	HPLC	21	
VTQQE	0	−	*SEC14*	*nat1*	2D	29,30,74	
VYTPS	0	−	*LEU1*	*nat1*	2D	29,30,74	

[a] +, Protein is acetylated; 0, protein is not acetylated; ±, protein is partially acetylated; ( ), suggested protein acetylation status.
[b] A, B, C, A′, types of NAT substrates; ( ), suggested type of NAT substrate; −, protein is not acetylated and no type of NAT substrate could be assigned.
[c] All, protein acetylation status was determined in normal, *nat1*, *nat3* and *mak3* deletion strains; *nat1*, acetylation was determined in *nat1* mutant strain and compared to normal strain; *mak3*, acetylation was determined in *mak3* mutant and normal strains only; None, only normal strain was used for determination of protein acetylation.
[d] 2D, protein acetylation status was determined by detection a shift of a corresponding protein spot on 2D-gel in mutant strain and compared to the normal strain; MS, protein acetylation was determined by mass spectrometry; HPLC, protein acetylation was determined by HPLC separation of tryptic digest of iso-1-cytochrome *c* mutant protein and by comparison its elution to the corresponding N-terminal peptide from normal strain in conjunction with peptide sequencing; α, N-terminal protein sequence was determined by Edman degradation technique. Mitochondrial proteins and other proteins with known processed N termini are excluded from this Table. There is a possibility that some data in this Table could be incorrect due to the fact that certain proteins, especially with unknown function or cellular localization, could have different sequences in their mature form. Some proteins are found as partially acetylated[74] but actual percentage of acetylation is unknown. Also, technical errors can arise because of the complexity in determining the N-terminal sequences.
[e] YGR086C was reportedly to be both acetylated and not acetylated.
[f] Mes1p was reportedly to be acetylated, as determined by mass spectrometry (MS), whereas Cyc1843p was reportedly to be not acetylated.
[g] Sec53p was reportedly to be acetylated,[29,30] Apt1p is not acetylated.[29]
[h] Sse1p was reportedly to be acetylated, whereas Sse2p was reportedly to be not acetylated.
[i] Lys20p was reportedly to be acetylated, whereas YEL071W was reportedly to be not acetylated.[29]
[j] YJR105W was reportedly to be partially acetylated[30] and not acetylated.[31]
[k] Gly1p was reportedly to be acetylated, whereas Cyc1-1371p was reportedly to be not acetylated.
[l] Aro8p was reportedly to be both acetylated,[30] and not acetylated.[31]
[m] Cys3p was reportedly to be partially acetylated,[30] and not acetylated.[31]
[n] Tps1p was reportedly to be partially acetylated,[30] and not acetylated.[31]
[o] Tsa1p was reportedly to be both acetylated,[30] and not acetylated.[31]

of the different varieties of cytosolic mammalian proteins[16,17], about 50% in yeast,[18] but rarely on prokaryotic[19] or archaeal proteins. The percentage of acetylated proteins in plants is not known, but the number of known N-terminal sequences of mature proteins in SWISS-PROT is relatively small and includes mostly cytochromes *c*, histones and some metabolic enzymes. Similarly, in invertebrates the number of characterized proteins is limited and their N-terminal acetylation appears to be rare. Although both protein sets from yeast and mammals considered herein obviously do not include the entire proteomes they nevertheless represent a large variety of all possible N termini and may be considered as a basis to generalize to all proteins.

Eukaryotic proteins susceptible to N-terminal acetylation have a variety of different N-terminal sequences, with no simple consensus motifs, and with no dependence on a single type of residue.[1,6] Proteins with serine and alanine termini are the most frequently acetylated, and these residues, along with methionine, glycine, and threonine account for over 95% of the N-terminal acetylated residues.[6,16,17,20] Our studies with N-terminally altered iso-1 from yeast and identification of three different NATs and their substrate helped to establish the basic patterns for acetylation.[1]

The iso-1-cytochrome c system

In normal yeast strains, the N-terminal methionine of iso-1-cytchrome *c* is cleaved and the newly exposed threonine residue is not acetylated. However, during the course of numerous studies spanning three decades, many mutant forms of iso-1 were found to have N termini processed in different ways,[21,22] as illustrated in Table 1. Because of the dispensability of the N-terminal region, and the ease of generating altered sequences by transformation with synthetic oligonucleotides, the iso-1 system has been used to systematically investigate N-terminal processing.[21–25] The study of mutationally altered forms of iso-1-cytochrome *c* was critical in deciphering the amino acid requirements for the two N-terminal processes, methionine cleavage and acetylation, as well as for identifying the substrate specificities for each of Ard1p, Nat3p and Mak3p, the catalytic subunits of the three N-terminal acetyltransferases.[24,26,27]

We present here a comprehensive analysis of N-terminal mature sequences for more than 450 yeast proteins, over 300 mammalian proteins, mostly human, bovine and mouse origin, and we also compare these eukaryotic proteins to the mature sequences of a large subset of over 810 eubacterial and 175 archaeal proteins. This protein database mainly constitutes cytosolic soluble

Table 3. N-terminal sequences of acetylated proteins from humans, bovine and mice

Sequence	Gene or protein; origin	Sequence	Gene or protein; origin
Ac-AAAAAI	NI8M-BOVINE	Ac-AQAFVN	GLRX-BOVINE
Ac-AAAAAS	NI8M-HUMAN	Ac-AQEFVN	GLRX-HUMAN
Ac-AAAAKP	DHSO-HUMAN	Ac-AQTPAF	ADA-BOVINE
Ac-AAADGD	2AAA-HUMAN	Ac-AQVLRG	ANX5-HUMAN
Ac-AAATGP	PEPD-HUMAN	Ac-AQVLRG	ANX5-BOVINE
Ac-AALTRN	G6PI-MOUSE	Ac-ASATRF	N4AM-BOVINE
Ac-AANATT	LSM5-HUMAN	Ac-ASATRL	N4AM-HUMAN
Ac-AAQKRP	MBP-BOVINE	Ac-ASDHQT	CRB2-HUMAN
Ac-AARRAL	DHQU-MOUSE	Ac-ASDHQT	CRB2-BOVINE
Ac-AASCVL	ALDX-HUMAN	Ac-ASGTTA	KAP0-BOVINE
Ac-AASGLR	NB4M-BOVINE	Ac-ASLGHP	DDH1-BOVINE
Ac-AAYKLV	PMGB-HUMAN	Ac-ASNNTA	GBG2-MOUSE
Ac-ACGLVA	LEG1-HUMAN	Ac-ASPDWG	CAH1-HUMAN
Ac-ACGLVA	LEG1-BOVINE	Ac-ASPPAC	KAP1-HUMAN
Ac-ACRQEP	GSHR-HUMAN	Ac-MDPNCS	MT1A-HUMAN
Ac-ADAFVG	FABH-MOUSE	Ac-MDPNCS	MT1A-BOVINE
Ac-ADDVDQ	LSM3-HUMAN	Ac-ASQKRP	MBP-HUMAN
Ac-ADEIAK	HINT-HUMAN	Ac-ASQKRP	MBP-MOUSE
Ac-ADEIAK	HINT-BOVINE	Ac-ASRLLL	ALDR-HUMAN
Ac-ADEIAK	HINT-MOUSE	Ac-ASSDIQ	STHM-HUMAN
Ac-ADGSSD	TRIC-HUMAN	Ac-ASSTGD	ALAT-HUMAN
Ac-ADKPDL	TYB9-BOVINE	Ac-ASVGEC	ATPK-HUMAN
Ac-ADNFSL	LEG3-HUMAN	Ac-ASVVPL	ATPK-BOVINE
Ac-ADNRDP	CATA-BOVINE	Ac-ATKAVC	SODC-HUMAN
Ac-ADPRQE	TAU-MOUSE	Ac-ATKAVC	SODC-BOVINE
Ac-ADPRVR	TBCA-HUMAN	Ac-ATRSPG	HPRT-HUMAN
Ac-ADPRVR	TBCA-BOVINE	Ac-AVPPTY	POR1-HUMAN
Ac-ADQLTE	CALM-HUMAN	Ac-AWKSGG	PIMT-HUMAN
Ac-ADRSGG	TRIC-BOVINE	Ac-AWKSGG	PIMT-BOVINE
Ac-AEDIQA	COXG-BOVINE	Ac-AYPLEK	S104-BOVINE
Ac-AEESSK	CYB5-BOVINE	Ac-CDFTED	MLEN-HUMAN
Ac-AEFVRN	ATPN-BOVINE	Ac-CDKEFM	V1P-MOUSE
Ac-AEGEIT	FGF1-HUMAN	Ac-GDREQL	143F-HUMAN
Ac-AEGETT	FGF1-BOVINE	Ac-GDSHVD	NCPR-HUMAN
Ac-AEGNTL	ACYO-HUMAN	Ac-GDVEKG	CYC-HUMAN
Ac-AEPRQE	TAU-HUMAN	Ac-GDVEKG	CYC-BOVINE
Ac-AEPRQE	TAU-BOVINE	Ac-GDVEKG	CYC-MOUSE
Ac-AEQATK	PA1F-HUMAN	Ac-GEVTAE	CNRA-BOVINE
Ac-AEQHGA	CRB3-HUMAN	Ac-GHFTEE[b]	HBG-HUMAN
Ac-AEQSDE	CYB5-HUMAN	Ac-GSELET	S10A-BOVINE
Ac-AEQVAL	G6PD-HUMAN	Ac-(M)DDDIA[c]	ACTB-HUMAN
Ac-AEQVAL	G6P1-MOUSE	Ac-(MC)DDEE[c]	ACTC-HUMAN
Ac-AEQVTK	PPAC-BOVINE	Ac-(MC)DEDE[c]	ACTS-HUMAN
Ac-AEQVTL	G6P2-MOUSE	Ac-(MC)EEED[c]	ACTA-HUMAN
Ac-AERVAA	NI9M-BOVINE	Ac-(MC)EEET[c]	ACTH-HUMAN
Ac-AESHLQ	CABV-BOVINE	Ac-MDAIKK	TPM1-HUMAN
Ac-AEVEQK	RS15-HUMAN	Ac-MDAIKK	TPMA-RABIT
Ac-AEVEQK	RS15-RAT	Ac-MDDIYK	TPCC-HUMAN
Ac-AFDSTW	FABI-HUMAN	Ac-MDDRED	143E-HUMAN
Ac-AFLSSG	NI2M-BOVINE	Ac-MDIAIH	CRAB-HUMAN
Ac-AFTGKF	ILBP-HUMAN	Ac-MDIAIH	CRAB-BOVINE
Ac-AGAAGL	KPBB-HUMAN	Ac-MDIAIH	CRAB-MOUSE
Ac-AGKAHR	PHS-HUMAN	Ac-MDIAIQ	CRAA-BOVINE
Ac-AGKPVL	GTC-MOUSE	Ac-MDIEAY	ARY1-HUMAN
Ac-AGLGHP	DDH1-HUMAN	Ac-MDLVKN	DIP-HUMAN
Ac-AGLLKK	NUFM-BOVINE	Ac-MDPETC	T3N-HUMAN
Ac-AGQAFR	CH10-HUMAN	Ac-MDPETC	MT3-BOVINE
Ac-AGQAFR	CH10-MOUSE	Ac-SHIQIP	KAP2-HUMAN
Ac-AGRKLA	ATPQ-BOVINE	Ac-SHIQIP	KAP2-BOVINE
Ac-AGRPAV	UCR6-BOVINE	Ac-MDPNCS	MT2-BOVINE
Ac-AGWNAY	PRO1-HUMAN	Ac-MDPNCS	MT1-MOUSE
Ac-AGWNAY	PRO1-BOVINE	Ac-MDTSRV	RS28-HUMAN
Ac-AHEHGH	NB2M-HUMAN	Ac-MDTSRV	RS28-RAT
Ac-AHGHGH[a]	NB2M-BOVINE	Ac-MDVFMK	SYUB-BOVINE
Ac-AHLTPE[b]	HBB-HUMAN	Ac-MDVTIQ	CRAA-HUMAN
Ac-AHNIVL	ALDR-BOVINE	Ac-MDVTIQ	CRAA-MOUSE
Ac-AKPAQG	ANX6-HUMAN	Ac-MEANGL	DCUP-HUMAN
Ac-AKRVAI	FMO1-HUMAN	Ac-(M)EEEIA[c]	ACTG-HUMAN
Ac-ALRACG	AP4A-HUMAN	Ac-MEEKLK	KAD1-HUMAN
Ac-AMVSEF	ANX1-HUMAN	Ac-MEEKLK	KAD1-BOVINE
Ac-ANEVIK	ADHX-HUMAN	Ac-MEELQD	B3AT-HUMAN
Ac-ANKAPS	S109-MOUSE	Ac-MEKVQY	PPLA-RABIT

(*continued*)

Table 3 Continued

Sequence	Gene or protein; origin	Sequence	Gene or protein; origin
Ac-ANKGPS	TAGL-HUMAN	Ac-MELLQV	N7BM-BOVINE
Ac-ANKGPS	TAGL-MOUSE	Ac-MELSPA[b]	HBA-HUMAN
Ac-MEMEKE	PTN1-HUMAN	Ac-SLSNKL	PGK1-HUMAN
Ac-METQAE	CRBA-HUMAN	Ac-SLTKTE	HBAZ-HUMAN
Ac-METQTV	CRBA-BOVINE	Ac-SMAEGD	ACYO-BOVINE
Ac-MMCGAP	CYTB-HUMAN	Ac-SMTDLL	PRVA-HUMAN
Ac-MMCGGT	CYTB-BOVINE	Ac-SMTDVL	PRVA-MOUSE
Ac-MMTGRQ	N4BM-BOVINE	Ac-SNKFLG	MYP2-HUMAN
Ac-MNFSGK	FABL-BOVINE	Ac-SNKFLG	MYP2-BOVINE
Ac-MNGTEG	OPSD-HUMAN	Ac-SNTQAE	S107-HUMAN
Ac-MNGTEG	OPSD-BOVINE	Ac-SQAAKA	CRB1-HUMAN
Ac-MNLEPP	CNRG-BOVINE	Ac-SQAEFD	ACBP-BOVINE
Ac-MQPPIR	PDI2-MOUSE	Ac-SQAEFE	CBP-HUMAN
Ac-MWGPNL	CYTX-BOVINE	Ac-SQPAAK	CRB1-BOVINE
Ac-MYRALR	FUMH-HUMAN	Ac-SSGIHV	DHCA-HUMAN
Ac-SAEVPE	ARPP-HUMAN	Ac-SSKTAS	GBGC-BOVINE
Ac-SAEVPE	ARPP-BOVINE	Ac-SSQIRQ	FRIL-HUMAN
Ac-MDPGAG	PPCT-BOVINE	Ac-SSSAMP	DHAC-BOVINE
Ac-MDPKDR	IPPD-BOVINE	Ac-SSSGTP	DHAC-HUMAN
Ac-SESSGT	F13A-BOVINE	Ac-STAGKV	ADHA-HUMAN
Ac-SESSSK	HMGI-HUMAN	Ac-STAQSL	ACYM-HUMAN
Ac-SETAPA	H1A-HUMAN	Ac-STGRPL	ACYM-BOVINE
Ac-SETAPA	H11-BOVINE	Ac-STVHEI	ANX2-HUMAN
Ac-SETAPL	H1C-HUMAN	Ac-STVHEI	ANX2-BOVINE
Ac-SETSRT	F13A-HUMAN	Ac-STVHEI	ANX2-MOUSE
Ac-SETVPA	H1T-HUMAN	Ac-TANGTA	FASC-MOUSE
Ac-SFPKYE	NB5M-BOVIEN	Ac-TDQQAE	TPCS-HUMAN
Ac-SFSGKY	FABL-HUMAN	Ac-TENSTS	H10-HUMAN
Ac-SFTTRS	K1CR-HUMAN	Ac-TKGLVL	AMPL-BOVINE
Ac-SGGGVI	SFR1-HUMAN	Ac-TLQCTK	CRBD-HUMAN
Ac-SGPRPV	KGUA-HUMAN	Ac-TMDKSE	143B-HUMAN
Ac-SGRGKG	H4-HUMAN	Ac-TMDKSE	143B-BOVINE
Ac-SGRGKQ	H2AA-HUMAN	Ac-TSALEN	LSM8-HUMAN
Ac-SGRGKT	H2AX-HUMAN	Ac-VDAFLG	FABH-HUMAN
Ac-SGWESY	KU70-HUMAN	Ac-VDAFVG	FABH-BOVINE
Ac-SGYTPE	NB7M-BOVINE	Ac-VDPEQL	143G-HUMAN
Ac-SHHWGY	CAH2-HUMAN	Ac-VDREQL	143G-BOVINE
Ac-SHHWGY	CAH2-BOVINE	Ac-VDSVYR	GLMT-HUMAN
Ac-SAKKSP	S10D-BOVINE		
Ac-SATAAT	VAM2-HUMAN		
Ac-SATAAT	VAM2-MOUSE		
Ac-SDAAVD	THYA-BOVINE		
Ac-SDAAVD	THYA-MOUSE		
Ac-SDKPDL	TYBN-HUMAN		
Ac-SDKPDM	TYB4-HUMAN		
Ac-SDKPDM	TYB4-MOUSE		
Ac-SEGAGT	CRAL-BOVINE		
Ac-SEGVGT	CRAL-HUMAN		
Ac-SEKSVE	THYP-BOVINE		
Ac-SELEED	KGPA-HUMAN		
Ac-SELEED	KGPA-BOVINE		
Ac-SELEKA	S10B-BOVINE		
Ac-SESGSK	HMGY-MOUSE		
Ac-SESLVV	GLMB-HUMAN		
Ac-SIEIPA	KAP3-HUMAN		
Ac-SIEIPA	KAP3-BOVINE		
Ac-SIEIPA	KAP3-MOUSE		
Ac-SKTGTK	CRBS-HUMAN		
Ac-SLLPVP	LPPL-HUMAN		

List of mammalian acetylated proteins, gene names and their origins are taken from SWISS-PROT protein bank. Ac-, acetyl group; (), amino acid cleavage.
[a] Variants of hemoglobins (alpha chain) that are acetylated.
[b] Actin Ac- proteins with unique N-terminal processing.
[c] Partial acetylation.

proteins or other cellular proteins whose N termini are not processed in a manner of cotranslocational cleavage of signal sequences or secretion. Also by using BLAST programs and amino acid sequence alignment, we have identified the orthologs of N-terminal acetyltransferases from the genomes of the completely or partially sequenced model organisms representing all live kingdoms; and we have constructed a NAT phylogenetic tree. We have uncovered six NAT protein families: three of

Table 4. N-terminal sequences of acetylated bacterial and archaeal proteins

A. Acetylated protein N termini from *E. coli* (ECOLI, left column) and from *M. luteus* (MICLU), both bacteria; and from archaea: *S. solfataricus* (SULSO) and *H. marismortui* (HALMA) (right column)			
Ac-AHIEKQ	RS5_ECOLI	Ac-MNKEQI	RL7_MICLU
Ac-ARYFRR	RS18_ECOLI	Ac-MEEVLS	DHE3_SULSO
Ac-SITKDQ	RL7_ECOLI	Ac-SAEDTP	RS7_HALMA
Ac-SKEKFE	EFTU_ECOLI	Ac-SASDFE	RL31_HALMA
B. Sequences of *E. coli* proteins (left column) and archaeal proteins (right column) with N termini determined as blocked (not specified)			
AKKVQA	RL11_ECOLI	MIAGQP	THSB_THEAC
MNRLIE	BLR_ECOLI	MLQSGM	BACR_HALSR
MTAINRI	ATOC_ECOLI	MMTGQV	THSA_THEAC
MWDVID	APPC_ECOLI	MQGRLE	ENDA_HALVO
MYEALL	SECG_ECOLI	MQSLSD	TOPG_SULAC
TTNTHT	FABZ_ECOLI	SNVPKE	PPSA_STAMA
		STTATV	THSB_SULS7

them, Ard1p, Nat3p and Mak3p, are related to each of the yeast catalytic subunits, Ard1p, Nat3p and Mak3p, respectively; a fourth, CAM, is composed of Camello1 and Camello2 putative acetyltransferase proteins, which most likely is evolutionarily related to the Mak3p family; a fifth, BAA family, is composed of diverged bacterial and archaeal NATs, some being related to *Escherichia coli* Rim acetyltransferases, which act on certain ribosomal subunits; and a new, hypothetical Nat5p family, with unknown substrate specificity.

Sequence requirements for N-terminal acetylation; characterized proteins with known N-terminal mature sequences

N-terminal amino acid sequences of yeast proteins, presented in Table 2, were taken from the results of acetylation analysis of mutationally altered iso-1-cytochromes *c*;[21–25] mutationally altered β-galactosidases;[28] abundant proteins;[29,30] ribosomal proteins;[31] and 19 S and 20 S proteasome subunits.[32] Also, some sequences, mostly for non-acetylated N termini, were taken from Proteome, Inc. database (now Incyte Genomics)[33] for individual proteins with N termini determined experimentally and from the original published reports. N-terminal protein sequences of known acetylated mammalian, bacterial and archaeal proteins are shown in Tables 3 and 4, respectively, and were taken from SWISS-PROT database[34] and from the original literature references. To search the acetylated protein subset from each organism in SWISS-PROT, FtDescription (Feature) option from Sequence Retrieval System has been used. Since N-terminal processing is determined predominantly by the sequence at the beginning of the protein, only the first six amino-terminal residues were taken into account. Proteins that were assumed to be acetylated on the basis of similarity or homology to orthologous acetylated proteins were not included. In isolated cases, we included the proteins that have been described as N-terminally blocked, although the nature of the blocking group was not identified. In these cases,

Table 5. Frequencies of acetylated and non-acetylated yeast proteins with various N-terminal sequences

Acetylation	+	0	%		+	0	%
NatA substrates				NatB substrates			
Ser-	124	12	91	Met-Glu- and Met-Asp-	13[a]	0	100
Ala-	19	44	30	Met-Lys- and Met-Arg-[b]	0	13	0
Gly-	3	13	23	NatC substrates			
Thr-	15	44	25	Met-Leu-	8	7	53
Cys-	1	6	14	Met-Phe-	3	2	60
Val-	1	32	3	Met-Ile-	4	7	36
Pro-[b]	0	24	0	Met-Trp-	1	1	50
Ser-Glu- and Ser-Asp-	25	0	100	Met-Lys-	0	9	0
Ser-Pro-[b]	0	3	0	Met-Gln-	0	2	0
Ser-Lys- and Ser-Arg	8	5	62	Met-Arg-	0	2	0
Ala-Glu- and Ala-Asp-	6	1[c]	86				
Ala-Pro-[b]	0	7	0				
Ala-Lys- and Ala-Arg-	0	12	0				
Thr-Glu- and Thr-Asp-	3	5	38				
Thr-Pro-[b]	0	3	0				
Thr-Lys- and Thr-Arg-	2	6	25				

+, Complete or partial acetylation; 0, no acetylation; %, percent acetylated proteins. Proteins whose acetylation status is questionable are not included in this Table.

[a] One of the proteins having N-terminal sequence MDPLA is partially acetylated.

[b] Not a NAT substrate; placed in NAT substrate section only for comparison purpose.

[c] Cox12p, a mitochondrial protein apparently lacking any processing except cleavage of the N-terminal methionine residue.[76]

Table 6. Frequencies of acetylated and non-acetylated mammalian proteins with various N-terminal sequences

Acetylation	+	0	%		+	0	%
NatA substrates				NatB substrates			
Ser-	67	0	100	Met-Glu- and Met-Asp-	33	0	100
Ala-	103	1	99	Met-Lys-[a]	0	2	0
Gly-	8	4	67	NatC substrates			
Thr-	8	2	80	Met-Leu-	0	4	0
Cys-	2	0	100	Met-Phe-	0	0	N/A
Val-	5	6	45	Met-Ile-	0	1	0
Pro-[a]	0	7	0	Met-Trp- and Met-Tyr-	2	1	67
Asp- and Glu-[b]	5	0	100	Gly-Asp- and Gly-Glu-	6	0	100
Ser-Glu- and Ser-Asp-	20	0	100	Gly-Lys-	0	3	0
Ser-Pro-	0	0	N/A				
Ser-Lys-	1	0	100				
Ala-Glu- and Ala-Asp-	33	0	100				
Ala-Pro-	0	1	0				
Ala-Lys-	2	0	100				
Thr-Glu- and Thr-Asp-	2	0	100				
Thr-Lys-	1	0	100				

+, Complete or partial acetylation; 0, no acetylation; %, percent acetylated proteins, N/A, Not available.
[a] Not a NAT substrate; placed in NAT substrate section only for comparison purpose.
[b] Unique processing of actins.

the proteins are designated in the Tables by a footnote. The protein N-terminal sequences in Tables 2–4 are organized in alphabetical order. We combined acetylated and non-acetylated protein sequences in one table for *Saccharomyces cerevisiae*, representing lower eukaryotes, or grouped the proteins from related species of the same kingdom, such as human, bovine and mouse, into one mammalian proteins subset, acetylated or non-acetylated. The summary tables for both, yeast and mammalian proteins, are also presented (Tables 5 and 6, respectively).

In order to avoid duplications for different isomers, we considered all proteins from one organism with the same protein name and having the same N-terminal sequence as one unique entry (for example, numerous human MTA1, MTA2 or HTA2 proteins, were counted as one unique sequence for each case). It should be mentioned also that multiple isoforms for many eukaryotic proteins often are observed due to the differential gene expression, splicing, post-translational modifications, like phosphorylation and glycosylation, or partial modification and this fact complicates acetylation analysis, especially in higher eukaryotes. Finally, we did not take into account the acetylation of regulatory peptides or hormones, like β-endorphin and melanotropic hormone, α-MSH, or other small popypeptides because they normally undergo extra proteolytic cleavage steps and their acetylation is posttranslational. Some of these regulatory macromolecules are synthesized enzymatically without ribosomes.

Sequence requirements for N-terminal acetylation in eukaryotes

The analysis of the mature N termini of yeast proteins presented in Table 2 indicates that 43% of all proteins are acetylated, which is comparable to about the 50% estimate made previously by 2D-gel technique for cytosolic soluble proteins.[18] The small difference could be due to the fact that, in our protein set, the abundant proteins are over-represented and might not reflect the random protein population. In addition to amino acid sequences, Table 2 contains the identified or suggested NAT substrate types for all acetylated proteins, the NAT deficient mutants used in the analysis, the method used to detect acetylation, and the original reference. The data presented in Table 2 and summarized in Table 5 showed that in N-terminal sequences of yeast, acetylated proteins have termini predominantly of serine (124), methionine (29), alanine (19) and threonine (15) residues. Serine and alanine residues together contribute more than 74% of all acetylated proteins. Besides the four mentioned amino acid residues, only a few examples are found for other acetylated N termini; three for glycine and one each for cysteine and valine, with the two latter residues most likely being only partially modified. Notably, methionine is clearly the second, after serine, most common acetylated residue in yeast, in contrast to mammalian proteins (see below) where serine and alanine are the most preferentially modified. Also, the effect of penultimate residue on acetylation is profound; acidic aspartic or glutamic residues stimulate acetylation, whereas proline inhibits acetylation and positively charged lysine and arginine usually but not always inhibit acetylation (Table 5). All methionine residues of Met-Asp- or Met-Glu- N termini (NatB substrates) were acetylated, as well as all serine and alanine residues in the same context. Hydrophobic aromatic or branched residues like leucine, isoleucine, phenyalanine, and tryptophan at penultimate position cause the methionine acetylation in about 50% of the cases (Nat C substrates); other

structural features may interfere with the NAT action. However, we observed that in such sequences the presence of an acidic residue at the third position often inhibits acetylation.[1,25]

The majority of mammalian proteins presented in Table 3 are acetylated, totaling about 89%, which is in good agreement with an earlier estimate.[6,16] The data provided in Table 6 summary shows that in N-terminal sequences of mammalian acetylated proteins, alanine (103), serine (67) and methionine (33) are predominant terminal residues following much smaller numbers for glycine, threonine, valine and cysteine residues. In the entire set of mammalian proteins, a substantially larger number of mature sequences begin with serine, alanine, or methionine residues, which are most often acetylated, 78% compared to 58% in yeast. Actin proteins with N-terminal glutamic acid and aspartic acid are acetylated by a unique protein processing system and will be considered in a separate section.

N-terminal serine residues are almost equally well acetylated in lower and higher eukaryotes. However, a significantly higher number of alanine residues are acetylated in the mammalian protein set compared to the yeast set, 99% *versus* 30%, respectively. The same is true for glycine and threonine residues. Also, while in yeast cysteine and valine residues are rarely modified by an acetyl group, in mammals it occurs more often. Both yeast and mammalian Met-Glu- and Met-Asp- proteins are always acetylated, but in mammals a variety of other types of N-terminal sequences with retained initial methionine (shown in Tables 5 and 6) is much less, with only ten such proteins as compared to 33 in yeast. This is consistent with an earlier view that retention of methionine and the lack of its acetylation are more characteristic of evolutionarily simpler genomes, especially bacterial and archaeal.[6] Particularly, prokaryotic proteins with retained methionine often have Met-Lys- sequences that are not observed frequently in mammals. Overall, eukaryotic proteins appear less prone to retain their N-terminal methionine residues. The stimulating effect of the acidic residues, like aspartic and glutamic acids, on N-terminal acetylation and inhibitory effect of basic residues, like lysine, arginine and proline residues, at penultimate position of mammalian proteins can be clearly seen from the Tables for both lower and higher eukaryotes.

It is also interesting to note that the larger proportion of acetylated proteins in higher eukaryotes could be explained, at least in part, by higher representation of acidic residues at the penultimate position. Specifically, in yeast the N-terminal sequences X-Glu-, or X-Asp-, where X designates Ser-, Ala-, Thr- or Met- termini, are almost always acetylated except for a few cases of Thr-Glu-, Thr-Asp- proteins; these X-Glu-, and X-Asp- termini comprise only 17% of all mature Ser-, Ala-, Thr- and Met- N termini in yeast, but that number is more than 39% in mammals (Tables 5 and 6). More frequent acetylation of N-terminal cysteine and valine residues in mammals may occur by the same reason. On the other hand, in yeast the number of N-terminal X-Lys-, X-Arg- or X-Pro- sequences, where X designates Ser-, Ala-, Thr- or Met-, are seldom acetylated, and these comprise about 21% of all Ser-, Ala-, Thr- and Met- N termini, while in mammals they comprise only 3%. Nevertheless, the NAT substrate specificities for yeast and mammals still appear to be the same.

In general, the acetylation patterns in yeast and mammals are very similar and may be evolutionarily conserved. However, a greater number of N-terminal protein sequences from higher eukaryotes are acetylated, probably reflecting some form of selection during evolution. Three lines of evidence, discussed above, support this conclusion, in which mammalian proteins contain the following: (1) a higher representation of most likely acetylated residues, serine, alanine, or methioine, at the first position; (2) a much higher representation of stimulating acidic residues at the penultimate position; and (3) a significantly lower representation of inhibitory basic residues at the penultimate position. The biological significance of such evolved difference remains to be determined.

Protein N-terminal acetylation in prokaryotes

The N-terminal amino acid composition of the soluble proteins from a cell-free extract of *E. coli* determined by dinitrophenyl- and phenyl-thiohydantoin methods showed that methionine, alanine, serine, threonine and aspartic and glutamic acid residues, with the latter in minor amounts, account for close to 95% of the end groups recovered.[19] N-terminal acetylation does not appear to be widespread in prokaryotes. However, systematic N-terminal characterization of bacterial and archaeal proteins has not been undertaken and the counterparts to eukaryotic NATs have not been identified. In *E. coli*, three NATs, RimI, RimJ and RimL, specifically modify single ribosomal proteins S18, S5 and L12, with Ala-Arg-, Ala-His- and Ser-Ile- N termini, respectively,[35,36] but there is no evidence that they act cotranslationally. For example, the family of large subunit ribosomal proteins L7/L12 is present in each 50 S subunit in four copies organized as two dimers and together with L10 is assembled in *E. coli* ribosomes on the conserved region of 23 S rRNA, termed the GTPase-associated domain.[37] The L7/L12 dimer probably interacts with elongation factor EFTu. Because the L7 and L12 proteins have identical amino acid sequences in the N-terminal region, and because only L7 is N-terminally acetylated, this modification apparently occurs post-translationally after partial or complete ribosome assembly, and RimL most likely recognizes some certain protein structure and not just the very N terminus. It is also not known if Rim enzymes act on other substrates *in vivo*.

We have searched the SWISS-PROT database for *E. coli* and other bacterial N-terminally acetylated proteins by the same procedure that was applied for mammalian proteins. We found only five such examples, which include the three ribosomal proteins mentioned above, *E. coli* EFTu elongation factor and one acetylated ribosomal protein, L7, from *Micrococcus luteus* that normally is present in the cell in two almost identical forms, one of which is acetylated and second which is not[38] (Table 4). This *M. luteus* protein probably corresponds to L7/L12 protein of *E. coli*. It is interesting to note that its acetylated N-terminal sequence, Met-Asn-Lys-Glu-Gln, is different from the *E. coli* L7/L12, Ser-Ile-Thr-Lys-Asp, suggesting that L7 protein acetylation is a conserved function in bacteria, but it does not depend on the first few N-terminal amino acid residues. We also found only the following three archaeal N-terminally acetylated proteins in SWISS-PROT: ribosomal proteins S7 and L31e from *Haloarcula marismortui*; and glutamate dehydrogenase, DHE2, from *Sulfolobus solfataricus* (Table 5). Eight other archaeal and six *E. coli* proteins were annotated as N-terminally blocked but the nature of the block groups was not known. Some of the blocking groups definitely could be other than acetyl modification, for example, L11 protein of *E. coli* is actually trimethylated.[39]

However, we were able to find a relatively large number of archaeal proteins with experimentally determined and non-acetylated N-terminal sequences, and many of them are ribosomal proteins. From a total of 97 mature N-terminal sequences, 28 were started with alanine, 26 with methionine, 16 with serine, ten with proline, seven with threonine, six with glycine and four with valine. More importantly, 807 out of 810 *E. coli* proteins with verified N-terminal sequences and listed in EcoGene Web Site† were not acetylated. Thus, most bacterial and archaeal proteins with characterized N-terminal sequences obviously are not acetylated, even though their counterparts are acetylated in eukaryotes. The few acetylated bacterial and archaeal proteins probably reflect an important functional requirement of resulting charge at the amino terminus.

In addition, the SWISS-PROT database was specifically searched for N-terminal acetylation of Met-Glu- and Met-Asp- proteins from bacteria, archaea and eukaryotes. As stressed above, all Met-Glu- and Met-Asp- proteins from eukaryotes are acetylated. The search, presented in Table 7, revealed that out of 47 mature N-terminal protein sequences from bacteria and archaea only one protein was found acetylated, DHE2 from archaea *S. solfataricus* with the sequence Met-Glu-Glu-Val-Leu-. In contrast, all 13 yeast and 51 mammalian proteins with Met-Glu- and Met-Asp- termini were acetylated. These results add to the conviction that N-terminal acetylation of eukaryotic proteins fundamentally differs from the N-terminal acetylation of bacteria and archaea proteins.

† http://bmb.med.miami.edu/Eco Gene/Eco Web/CESSPages /VerifiedProts.htm

Table 7. Proteins with Met-Asp-, Met-Glu- N-terminal sequences

	Acetylated	Non-acetylated
Prokaryotes (*E. coli*)	0	3[a]
Prokaryotes (Bacteria)	0	37[a]
Archaea	1	6
Yeast (*S. cerevisiae*)	13	0
Mammals	51	0

[a] N-terminal protein sequences of ATP synthase *c* chain from *E. coli*, cytochrome *c* oxidase polypeptide IIa from *Thermus aquaticus* (subsp. *thermophilus*) and rubredoxin from *Desulfovibrio gigas* are formylated; three others from *Synechococcus vulcanus* were blocked, but the nature of blocking group is unknown.

Three different NATs in yeast; are there more?

Studies with yeast *S. cerevisiae* so far revealed three different N-terminal acetyltransferases, NatA, NatB and NatC, that act on groups of substrates, with each group containing degenerate motifs.[1] Polevoda *et al.*[24] characterized their substrate specificity *in vivo* by investigation of acetylation of several subsets of yeast proteins from various NAT deletion mutants. As described above, Ard1p, Nat3p and Mak3p are related to each other by amino acid sequence, and are believed to be the catalytic subunits of three NATs, NatA, NatB, and NatC, respectively, with each NAT acting on different sets of proteins having different N-terminal regions (Table 8). NatA is a major NAT in yeast cells with multiple substrates *in vivo*.[18] Ard1p activity requires at least two subunits, Ard1p itself, and Nat1p.[26] The *MAK3* gene encodes a NAT that is required for the N-terminal acetylation of the killer viral major coat protein, *gag*, with a Met-Leu-Arg-Phe-terminus,[28] two subunits of the 20 S proteasome[32] and probably some mitochondrial proteins. The co-purification of Mak3p, Mak10p and Mak31p suggests that these three subunits form a complex

Table 8. The three types of NATs

Type	NatA	NatB	NatC
Catalytic subunit	Ard1p	Nat3p	Mak3p
Other subunits	Nat1p Other?	Mdm20p Other	Mak10p Mak31p
Substrates[a]	Ser- Ala- Gly- Thr- Cys- ? Val- ?	Met-Glu- Met-Asp- Met-Asn- Met-Met-	Met-Ile- Met-Leu- Met-Trp- Met-Phe-

[a] Acetylation occurs only on subclasses of proteins containing the indicated termini, except for Met-Glu- and Met-Asp-, which are apparently always acetylated.

that is required for N-terminal acetylation. Recently we have shown that all three subunits are required for NatC activity but not for acetylation of NatA or NatB substrate types.[25] Nat3p was originally identified on the basis of similarities of its amino acid sequence to those of Ard1p and Mak3p, and Nat3p complex contains three other subunits, Mdm20p and proteins with molecular masses about 47 kDa and 16 kDa (B.P., T. Cardillo, G. Bedi & F.S., unpublished results). NatB substrates *in vivo* include actin, Act1p, and Rnr4p,[25] two ribosomal proteins[31] and three subunits of 26 S proteasome.[32] All acetylated proteins in yeast can be assigned to one of the NatA, NatB or NatC substrates. Furthermore, we do not know of any acetylated proteins in yeast that could not reasonably be a NatA, NatB or NatC substrate. Nevertheless, it remains to be seen if there are other NATs, acting on rarer substrates.

Generality of N-terminal acetylation of eukaryotic proteins

The similarity in the pattern of N-terminally acetylated proteins from higher eukaryotes and *S. cerevisiae* suggest that the same systems may operate in all eukaryotes, including the presence of homologous N-terminal acetyltransferases that are the members of a larger acetyltransferase family, PF00583 (GNAT).[40] Although three different NATs in yeast are not highly similar in their amino acid sequences, the similarity in the regions of their putative Ac-CoA-binding motifs A–D is much stronger, indicative of a conserved protein function. On the other hand, the protein sequences of the yeast NATs are sufficiently diverged to allow the identification of proteins corresponding to sets of the same ortholog from other species. We have used the general BLAST server from the National Center for Biotechnology Information (NCBI) to identify such orthologs in different model organisms. In some cases, to limit the search options or to identify the candidates with the highest similarity, we ran BLAST searches against individual organism proteomes, which were completely or incompletely sequenced. Protein sequence alignments and phylogenetic analysis were undertaken after the candidate proteins with the closest homology to a particular NAT were identified. If necessary, some corrections were made at this point and less likely candidate proteins were discarded. Multalin program[41] was used for protein alignment and the MegAlin program from LaserGene99 package (DNAStar, Madison, WI) was used for phylogenetic analysis.

The presence of the orthologous genes encoding the three different N-terminal acetyltransferases in worms, flies, plants and mammals serves as an additional evidence that the same or similar N-terminal acetylation system may be operating in higher eukaryotes as in yeast. Species containing orthologs of the yeast Ard1p include *Schizosaccharomyces pombe*, *Caenorhabditis elegans*, *Drosophila melanogaster*, *A. thaliana*, *Trypanosoma brucei*, *Dictyostelium discoideum*, *Mus musculus* and *Homo sapiens*; of the yeast Nat3p include *S. pombe*, *C. elegans*, *D. melanogaster*, *A. thaliana*, *Leishmania donovani*, *M. musculus* and *H. sapiens*; and of the yeast Mak3p include *S. pombe*, *C. elegans*, *D. melanogaster* and *A. thaliana* (Figure 1). Several highly homologous proteins, the so-called Camello proteins, from rat, mouse and human form a special NAT group, that evolutionarily could be linked to Mak3p. Bacterial and archaeal proteins are generally not very similar to eukaryotic NATs and are even more diverged between themselves. The presence of multiple bacterial enzymes for antibiotic inactivation by acetylation, for example chloramphenycol acetyltransferases, sometimes complicates the NAT homology searches because the amino acid sequences of motifs A–D responsible for acetyl-CoA binding in such proteins are very close, as was noted earlier.[27]

The identified NATs from different species were also analyzed by a phylogenetic approach. The following six NAT families were detected on the basis of their protein similarity: Ard1p, the Ard1p related group; Nat3p, the Nat3p related group; Mak3p, the Mak3p related group; CAM, the Camello1 and Camello2 related group;[42] BAA, bacterial and archaeal putative acetyltransferases; and Nat5p, the newly uncovered hypothetical yeast Nat5p (YOR253W) related group (Figure 2). All of these groups are distantly related to each other, except for the CAM family, which is phylogeneticly related more closely to the Mak3p family and which most likely diverged from an

Table 9. Amino acid sequence similarities (%) of yeast and *E. coli* NATs

	A				B			C	
	Nat3p	Mak3p	Ard1p	RimI	RimJ	RimL		Mak3p	Ard1p
P46854	15.4	14.8	16.0	18.2	15.4	18.5	Nat3p	13.1	15.4
RimI	16.2	19.6	16.2	–	14.9	13.5	Mak3p	–	18.8
RimJ	10.3	12.5	13.9	–	–	17.3	Ard1p	–	–
RimL	11.7	10.2	11.2	–	–	–			

A, *E. coli* putative NATs, RimI, RimJ, RimL and P46854 to the three major yeast *S. cerevisiae* NATs, Ard1p, Mak3p and Nat3p; B, similarities among the *E. coli* putative NATs; and C, similarities among the yeast NATs. Sequence pair distances were made by using the Clustal method with PAM250 residue weight table (MegAlign, DNAStar, Inc.).

```
           1                                                                                  80
 MYlQ49857 MSIVLAIDTA TAAVTAGIVA FDGHDCFTLA ERVTVDAKAH VERLTPNVLV ALADAELAMC ELDAVVVGCG PGPFTGLRVG
   At79611                                                                        MRST PLGTTAVSPA
    ScNAT3                                              MYGCLE MFGCYFMHVL FININVEAIS NEVYIVHNIA
AtCAB96691                                                                              MG
  CeT16306                                  KSDAENT EAVSTPEKVS DSVEPEKQDQ VDQLVPLINK FDVNGKALKT
 ScYOR253w                                                                    MGR DICTLDNVYA
  MtH69037                                MHLDLNHK NPLILSKEDT MRVHPFTPTP SSRRSVSHLI YESDRQLFDW
  ApF72584                                              MNREA QGSQEGVRIR KARAQDIPIV MEINLESL--
HsBAB14853                                            MTEVVP SSALSEVSLR LLCHDDIDTV KHLCGDWF--
   At26961                                MSRFPRRF IDDSSMEDAG ISLCPSIHYR PINPNDLDRL EQIHRDIF--
AtCAC01844                                MSRFPRGF VENSSMEEPK IARRPTICFR PINPSDLERL EQIHRDIF--
CcoAAB5344                                                              MITEMKAEHL KDIDKPSE--

           81                                                                                 160
CeAAB65989                        MAIGVQKKT STPLDVGTEQ EPTNLEKTIG TLRRCLQIAG --TSNKPGSR SAKNSISEES
DmAAF47497                                                                        TE TIFNDFQELS
 MYlQ49857 MATAAAYGHA LGIPVHGVCS LDAIGVRTTG DTLVVTDARR HEVYWARYRD GVRIAGPAVG SPTDVDPGTA LTVAGSPEHA
  SsS76652                              MEKSP LSPPGYCLYK GTTRDRRLLV DFLERTYH-- --SLFPELCD FPH---LEQT
   At79611 SIKHNSYGYG RFVSSSGVSN F-SIHRRRRH SSFSISQAPS QINSGACNAS QIVDLFPAVS PEIVVREARL EDCWEVAETH
 MjaQ58604              MRVQLILH L-EIPKEVCR SLEVDNYINN SIEINLKCEK KPTLYIKTHS --IGSLKSIL DDFFRCQNAA
HsAAF22299                                 MAPYHIRK YQESDRKSVV GLLSGGMAEH APATFRRLLK LPRTLILLLG
HsAAF22303                                 MAPCHIRK YQESDRQWVV GLLSRGMAEH APATFRQLLK LPRTLILLLG
MmAAF80483                                 MVPYHIRQ YQDSDHKRVV DVFTKGMEEY IPSTFRHMLM LPRTLLLLLG
RnAAF22297                                 MAPYHIRQ YQDSDHKSVV DVFTKGMEEH IPSTFRHMLM LPRTLLLLLG
RrAAF80487                                  VSYHICE YQDSDYKSVV DVFTKGAEEY IPSTFRHLLL LPRTLLLLLG
MMAAF80486                                 MASFRIRQ FQERDYKQVV DVFSRGMEEH IPTAFRHLLT LPRTLLLLAV
RnAAF22298                                 MASFHIRQ FQERDYEQVV DMFSRGMKEH IPTAFRHLLL LPRTLLLLLG
MmAAF80484                                 MAAYHIRQ YQEKDHKRVL ELFSSGMKEL IPAAIRQMLT LPHSLLLLPG
MmAAF80485                                 MAPYHIRK YQDSDHRSVV DLFRRGMEEH IPATFRHMLL LPRTLLLLLG
RnAAF22304                                           SDHRSVV NLFCRGTEEH ISASFRYMLL LPGTLLILLG
 SspS75213                                                                            MVKST
  DrF75274                                                                              M
 StyQ08021                                                                              M
    ScARD1                      M PINIRRATIN DIICMQNANL HNLPENYMMK YYMYHILSWP EASFVATTTT LDCEDSDEQD
  SpT37723                         MDIRPARIS DLTGMQNCNL HNLPENYQLK YYLYHAISWP MLSYVA---- --------TD
  HsP41227                         MNIRNARPE DLMNMQHCNL LCLPENYQMK YYFYHGLSWP QLSYIA---- --------ED
MmAAF22155                         MNIRNARPE DLMNMQHCNL LCLPENYQMK YYFYHGLSWP QLSYIA---- --------ED
DmAAF50178                         MNIRCAKPE DLMTMQHCNL LCLPENYQMK YYFYHGLTWP QLSYVA---- --------VD
  CeT33023                         MNIRCARVD DLMSMQNANL MCLPENYQMK YYFYHALSWP QLSYIA---- --------ED
TbCAB89123                         MQVRRATME DMYQMQHCNL RCLPENYNLR YYLYHILSWP QLLYVQ---- --------ED
AtBAB10599                       MVCIRRATVD DLLAMQACNL MCLPENYQMK YYLYHILSWP QLLYVA---- --------ED
  DdP36416                       MVSIRPCQIG DLMSMQNANL TCLPENYQMK YYLYHFLTWP QTSFVA---- --------ED
    ScNAT3  YVKDSGTLRI FKERDRRIRE MTTIQPFEPV DLFKTNNVNL DILTENFPLE FYFEYMIIWP DLFF------ --KSSEMTVD
HsAAD40190                       MTTLRAFTCD DLFRFNNINL DPLTETYGIP FYLQYLAHWP EYF------- -----IVAEA
DmAAF48987                       MTTLRPFTCD DLFKFNNVNF DPLTETYGLS FYTQYLAKWP EYF------- -----QLAES
AtAAD25793                       MTTIRRFSCN DLLRFTSVNL DHLTETFNMS FYMTYLARWP DYF------- -----HVAEG
  SpT41102                       MTDTRKFKAT DLFSFNNINL DPLTETFNIS FYLSYLNKWP SLC------- -----VVQES
  LdQ05885                       MTTYRRMTLC DTLQFNFVNL DQLTETYNTS FYGEYVTHWP EYQ------- -----RMCVH
DmAAF53273                     V MTSPRLFVLE DLFKFNNIVM DPLAEVYSLP FLLPKILEHP EL-------- ------VLAA
SpCAA20373       MIELD AINPNNLKIL EVINEKCFDP EIIIFPTSFY KDTISVGPLA QYAYFNQVCV GAVR------ --CK----KE
DmAAF34715  MTRSSIELG DVTPHNIKQL KKLNTVVFP- --VSYNDKFY VDVLEAGELA KLAYYNDIVV GAVC------ --CRI---DN
MmAAF22301  MTRSSIELG DVTPHNIKQL KKLNTVVFP- --VSYNDKFY VDVLEAGELA KLAYYNDIVV GAVC------ --CRI---DN
HsBAB14397  MKGSRIELG DVTPHNIKQL KRLNQVIFP- --VSYNDKFY KDVLEVGELA KLAYFNDIAV GAVC------ --CRV---DH
AtCAB96691 AGREVSVSLD GVRDKNLMQL KILNTVLFP- --VRYNDKYY ADAIAAGEFT KLAYYNDICV GAIA------ --CRL---EK
  CeT16306 IAGHGTVHLG EITPHNILQL KKLNEDVFP- --IAYNDKFY VEARYCGELG RLAYYNDVVV GAVC------ --CRI---DD
 ScYOR253w NNLGMLTKLA HVTVPNLYQD AFFSALFAED SLVAKNKK-- PSSKKDVHFT QMAYYSEIPV GGLV------ --AKLVPKKQ
  MtH69037 FFTDPDLIIE KLIAHGLYLD PDRIVLASRG PEVIGALVYE LKGAEISR-L DLIRCLGPLD FLRFSA---- MDLIDSLSTL
  ApF72584 ----PENYWY GFYKYILDNW GEAFLVAEVG GEIVGYAMSR VE--QTSD-P VLLGMKDELE GDKSVI---- DKILDAIRNQ
DmAAF50213          L SWYEDITSST RFFALAAVYN LAIIGLIVAE IKPYRNVN-K EVIANMSDSD ELYTRLSGFP MQDKGILPDS
HsBAB14853 ----PIECOSTANZO SWYRDITSNK KFFSLAATYR GAIVGMIVAE IKNRTKIH-K E--------- ---------- --DGDILASN
   At26961 ----PIKYES EFFQSVVNGV DIVSWAAVDR SRPDDHSDEL IGFVTA---K FVLAKDSEI- ---------- --DDLIHYDS
AtCAC01844 ----PIRYES EFFQNVVNGG DIVSWAAVDR SRPDGHSEEL IGFVTA---K IVLAKESEI- ---------- --SDLIRYDS
CcoAAB5344 ----PFEVIG KIIPRYENEN WTFTELLYEA PYLKSYQDEE DEEDEEADCL EYIDNTDKII YLYYQ----- DDKCVGKVKL
 Consensus                                                                                  ..
```

Figure 1 (*legend on page 614*)

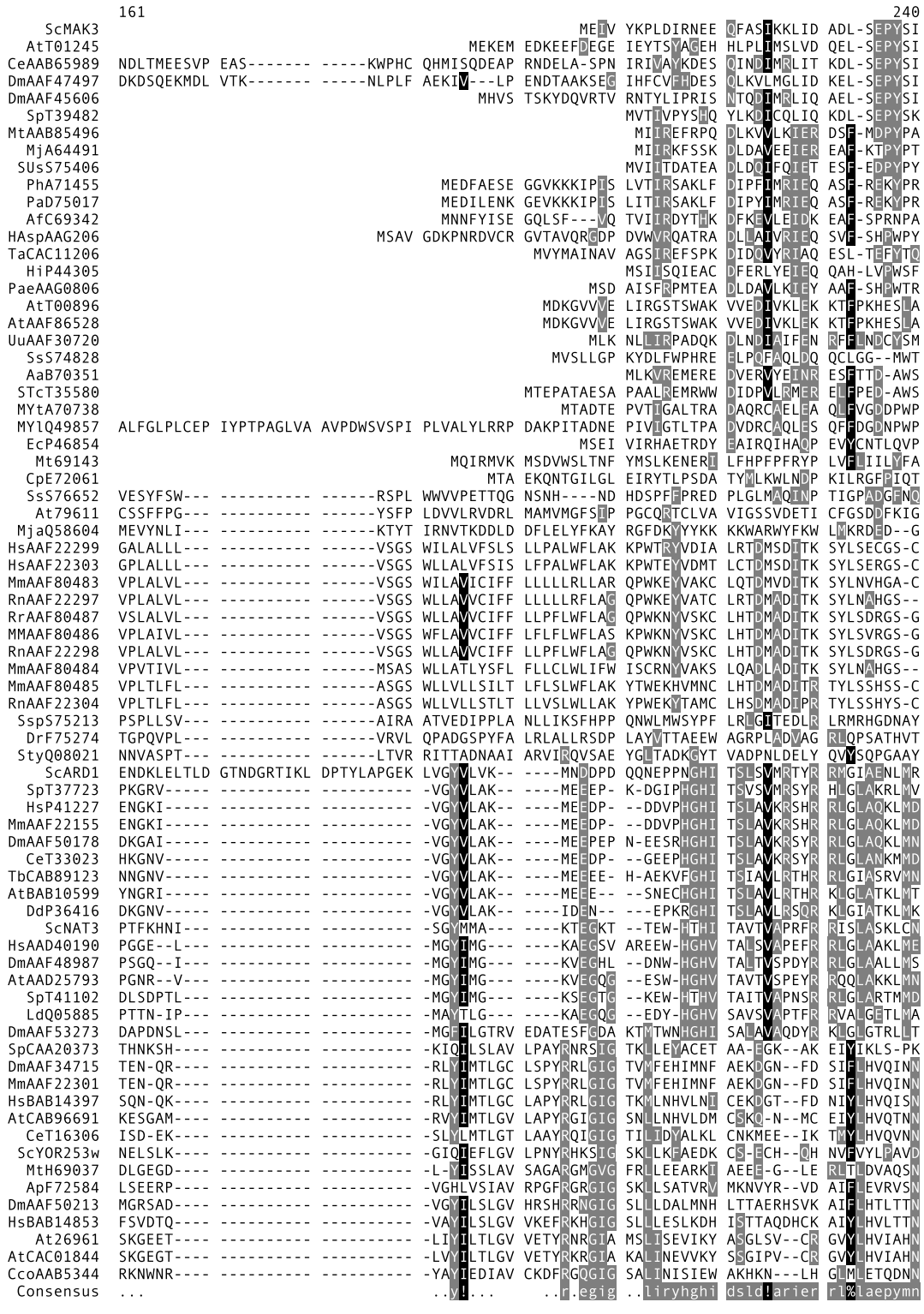


Figure 1 (*legend on page 614*)

Figure 1 (*legend on page 614*)

Figure 1. Protein amino acid sequence alignment of known and putative NATs. Multiple sequence alignment was made by using Multalin version 5.4.1.[41] Highly conserved residues are highlighted in black, where multiple designations are as follows: ! = I or V; % = F or Y; # = N, D, Q, E, B, or Z. Moderately conserved residues are highlighted in gray. The protein accession numbers refer to proteins derived from the following species: *Aeropyrum pernix* (strain K1) ApF72584, archaea; *Aquifex aeolicus* AaB70351, bacteria; *A. thaliana* AtT01245, AtAAF86528, AtC69711, AtAAD25793, AtCAC01844, AtCAB96669, plant; *Archaeoglobus fulgidus* AfC69342, archaea; *C. elegans* CeT33023, CeAAB65989, CeT16306, invertebrate; *Campylobacter coli* CcoAAB5344, bacteria; *Chlamydia pneumoniae* (strain CWL029) CpE72061, bacteria; *Deinococcus radiodurans* (strain R1) DrF75274, bacteria; *D. discoideum* DdP36416, fungus; *D. melanogaster* DmAAF34715, DmAAF47497, DmAAF45606, DmAAF50178, DmAAF48987, DmAAF53273, DmAAF50213, invertebrate; *E. coli* EcP46854, bacteria; *Haemophilus influenzae* HiP44305, bacteria; *Halobacterium* sp. NRC-1 HAspAAG20621, archaea, *H. sapiens* HsAAF22299, HsAAF22303, HsP41227, HsAAD40190, HsBAB14397, HsBAB14853, mammal; *L. donovani* LdQ05885, protozoa; *Methanobacterium thermoautotrophicum* MtAAB85496, MtH69037, archaea; *Methanococcus jannaschii* MjA64491, MjaQ58604, MtB69143, archaea; *M. musculus* MmAAF22155,

ancestral Mak3p. Although it was recently shown that Camello proteins play an essential role in embryo development in *Xenopus levis*,[43] no substrate for Camello enzymes has been so far identified.

The BAA family form a well isolated branch in the NAT phylogenetic tree with broader diversity of eubacterial and archaeal NATs, but none of the proteins has been shown to have biochemical activity. Thus, substrate specificity of those proteins also is unknown. Although some members of the BAA family are annotated in databases as acetyltransferases related to *E. coli* Rim proteins, primarily RimI, which act on ribosomal proteins, none of the Rim proteins themselves is present in our NAT phylogenetic tree. Instead, another *E. coli* protein, accession number P46854, was identified phylogenetically as the closest to the three eukaryotic NATs (Figure 1). Although initially we included all three Rim proteins on the basis of amino acid sequence alignment, only the RimI protein showed significant homology to the eukaryotic NATs. Both RimI and P46854 are more similar to each other, but are relatively dissimilar to the three major yeast NATs (Table 9), although P46854 protein had a higher match in the conserved NAT region. It appears as if known eukaryotic NATs evolved from primordial forms of RimI and P46854. The analysis of the putative bacterial acetyltransferases was strengthened by the fact that the overall homology between three major eukaryotic NATs is low and may reflect the diversity of the substrates they act on. There is no information on which domains or residues are involved in protein substrate binding or if any other subunit of NAT complexes specifies the substrate recognition. Although we have considered the BAA family as putative acetyltransferases, obviously further analyses are required for definitive conclusions concerning their activity, function and relationship to eukaryotic NATs.

Nat5p represents a family of the putative NATs with orthlogous proteins identified in yeast, *S. pombe*, *C. elegans*, *D. melanogaster*, *A. thaliana* and *H. sapiens*. The finding of this new family is only based on sequence similarity of Nat5p (YOR253Wp) to other NATs. Our attempts to detect any Nat5p substrates in yeast by 2D-gel electrophoresis has been so far unsuccessful, but this may reflect the rarity of the substrates *in vivo* or that Nat5p is acting on the smaller polypetides with mobility parameters undetectable by our regular 2D-gel procedure (R. Svensson, B.P., F.S. & A. Blomberg, unpublished result).

Are NATs present in the cell organelles?

With availability of the increased number of completely sequenced eukaryotic genomes and powerful computer search programs, it is now possible to search for the presence of NAT isoforms for particular organisms. Recently such an approach was applied for identification of MAPs in the *A. thaliana* genome.[9] Six new MAP cDNAs were found, MAP1A–MAP1D, which are located at different genomic loci, and which are closely related to yeast Map1p (and *E. coli* MetAP) in their protein sequences; and the duplicated MAP2A and MAP2B, which are closely related to yeast Map2p and nearly identical in protein sequences, but are located on different chromosomes. Three MAP isoforms were expressed and localized in cytoplasm, MAP1A and both MAP2s; one, MAP1B, was detected exclusively in plastids; and the others, MAP1C and MAP1D, localized in both mitochondria and plastids. The three MAP1B–MAP1D enzymes that localize to organelles possess the unique N-terminal pre-sequences to direct each protein to its proper cell compartment, but otherwise they are very similar to each other in catalytic domain. Multiple isoforms of another N-terminal processing enzyme, protein deformylase, that localize to mitochondria and plastids, also were detected in the *A. thaliana* genome.[9]

These findings with *A. thaliana* encouraged us to search for NAT isoforms located in cellular compartments where *de novo* protein synthesis occurs, even though eukaryotic organelles were derived from ancestral endosymbiotic eubacteria that lacked cotranslational N-terminal acetylation. Using regular BLAST searches, we were unable to find NAT isoforms in human or mouse genomes, unlike those multiple MAPs in *A. thaliana*. However, it is still possible that distinct NATs may be found in mammalian and plant organelles that acetylate individual proteins posttranslationally, similar, for example, to *E. coli* Rim enzymes. In support of this, three proteins synthesized in spinach chloroplasts were described as both N-terminally acetylated and phosphorylated.[5]

MmAAF80483, AAF80484, MmAAF80485, MmAAF80486, MmAAF22155, MmAAF22301, mammal; *Mycobacterium leprae* MylQ49857; *Mycobacterium tuberculosis* (strain H37RV) MYtA70738, bacteria; *Pseudomonas aeruginosa* PaeAAG08065, bacteria; *Pyrococcus abyssi* (strain Orsay) PaD75017; *Pyrococcus horikoshii* PhA71455, archaea; *Rattus norvegicus* RnAAF22297, RnAAF22298, RnAAF22302, RnAAF22304, mammal; *Rattus rattus* RrAAF80487, mammal; *S. cerevisiae* Ard1p, Nat3p, Mak3p, Nat5p, (ScYOR253w), fungus; *Salmonella typhimurium* StyQ08021, bacteria; *S. pombe* SpT39482, SpT37723, SpT41102, SpCAA20373, fungus; *Streptomyces coelicolor* STcT35580, bacteria; *S. solfataricus* SUsS75406, archaea; *Synechocystis* sp. (strain PCC 6803) SspS74828, SspS75213, SspS76652, bacteria; *Thermoplasma acidophilum* TaCAC11206, archaea; *T. brucei* TbCAB89123, protozoa; and *Ureaplasma urealyticum* UuAAF30720, bacteria.

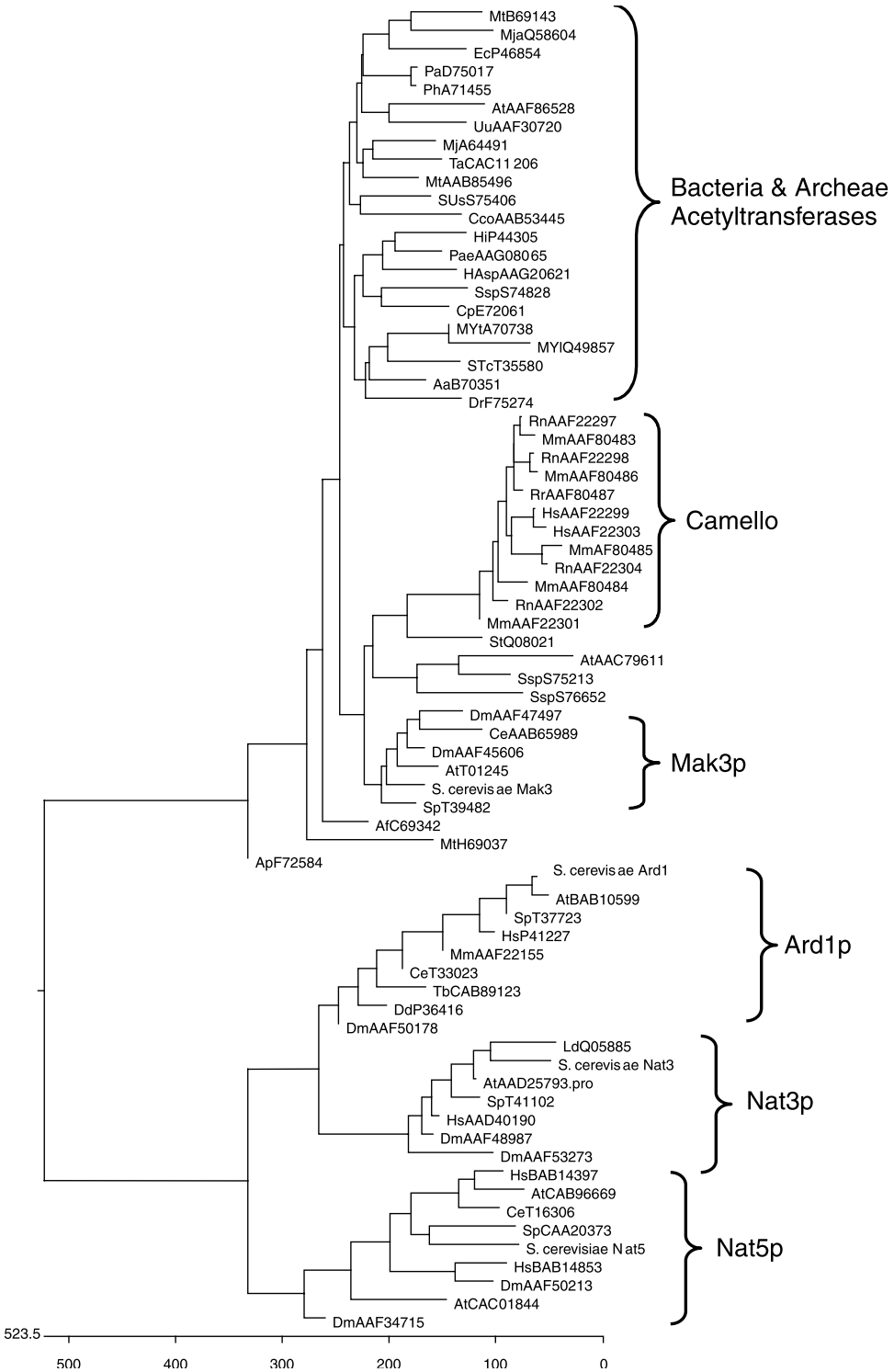


Figure 2. NAT proteins phylogenetic tree. Protein accesion numbers are listed in the legend to Figure 1.

NATs substrate specificities and prediction of N-terminal acetylation

Previous attempts to predict N-terminal acetylation on the basis of the properties of amino acid residues distributed along the N-terminal region were unsuccessful. A computer program, Pattern Learn, was used in an attempt to distinguish the patterns in 56 Ac-Ala- acetylated and 104 Ala- non-acetylated eukaryotic proteins by comparing the first 40 amino acid residues for their statistical assignment as secondary structure formers, breakers or neutrals.[44] Some distinguishing features were found in the sequences mainly between 1–10 residues, smaller features at 16–24 and 30–40 residues, but the precise nature of these features was not determined. However, new insight on this problem has been provided by using yeast mutants deleted in one or another NAT genes. The substrate specificities for each of the Ard1p, Nat3p and Mak3p enzymes were deduced from considering the lack of acetylation of the different protein subsets and the corresponding substrate types were designated NatA, NatB, NatC.[1,23] As was summarized earlier,[1] subclasses of proteins with Ser-, Ala-, Gly- or Thr termini were not acetylated in *ard1*-Δ mutants (NatA substrates); proteins with Met-Glu- or Met-Asp- termini and subclasses of proteins with Met-Asn- and Met-Met- N termini were not acetylated in *nat3*-Δ mutants (NatB substrates); and subclasses of proteins with Met-Ile-, Met-Leu-, Met-Trp- or Met-Phe- termini were not acetylated in *mak3*-Δ mutants (NatC substrates). In addition, a special subclass of NatA substrates with Ser-Glu-, Ser-Asp-, Ala-Glu-, or Gly-Glu- termini, designated NatA′ substrates, were also only partially acetylated in *nat3*-Δ and *mak3*-Δ mutants.

The NatA substrates appear to be the most degenerate, encompassing a wide range of sequences, especially those with N-terminal residues of serine or alanine. Nevertheless, it has not been excluded that new NATs may be discovered, especially for proteins with unusual and rare N-terminal sequences that are not substrates for NatA, NatB, or NatC. For example, the acetylation of Cys-Asp- actin in yeast[45] is not, as expected, a NatA substrate.

Generally, acetylation cannot be definitively predicted from the primary amino acid sequence. Only the NatB substrates have common sequences that can be easily deciphered and normally are acetylated. But even NatB substrate acetylation could be diminished by the presence of inhibitory residues. For example, altered iso-1, Ac-Met-Asp-Pro- was only 67% acetylated and one can assume that adjacent proline residue diminished the action of Nat3p. While the reason for the lack of acetylation is unclear, the N-terminal region of many of the non-acetylated proteins related to both NatA and NatB substrates contain basic residues, lysine, arginine, and histidine, as well as proline residues. At the same time, the N termini of non-acetylated proteins related to yeast NatC substrate contain acidic residues, such as glutamic acid at their N termini. As we mentioned above, normally acidic residues stimulate acetylation of substrates NatA and B. Moreover, the stimulating and inhibitory residues may occupy sites further than the fifth amino acid position from the N terminus.[1]

We suggested earlier that NATs act on substrates with specific but degenerate sequences, and that the activities can be diminished by suboptimal residues.[1,22] We further suggested that acetylation can be diminished by the inhibitory residues situated anywhere on the nascent chain at the time of this addition. Thus, the degree of acetylation is the net effect of positive optimal or suboptimal residues, and negative inhibitory residues. Furthermore, this lack of acetylation could be due to the absence of required residues or the presence of inhibitory residues. Because the identities of required and inhibitory residues are not completely understood, the ability of a protein to be acetylated cannot be definitively predicted from the primary sequence. Because the required and inhibitory residues may affect acetylation to various degrees, and because inhibitory residues may possibly occupy various sites in the nascent chain, predicting acetylated and non-acetylated sequences is still not absolutely reliable; however, considering our new studies presented herein, the acetylation of many proteins can now be predicted with a high degree of accuracy.

The biological significance of N-terminal acetylation

The biological significance of N-terminal modification varies with the particular protein, with some proteins requiring acetylation for function, whereas others do not. For example, the 30-fold increased dissociation of HbF_1 form of human fetal hemoglobin compared with normal HbF is most likely due to the presence of N-terminal acetyl group at the juncture where $\alpha\gamma$ dimers assemble to form tetramer.[46] Also, N-terminal acetylation of tropomyosin is required for its binding to actin (also see below).[47] The recombinant enzyme rat glycine *N*-methyltransferase (GNMT) expressed in *E. coli* and lacking an N-terminal acetyl group exhibited similar kinetic patterns to the GNMT purified from liver but showed hyperbolic kinetics at low pH in contrast to the sigmoidal behavior of native protein.[48] In some cases, a loss of acetylation leads to decreased thermal stability of protein, kinetic parameters or less efficiency in the complex assembly. An earlier suggestion was made that N-terminal acetylation protects protein from degradation, but in those examples, the proteins lacking acetylated termini also had other differences in amino acid sequences. Clearly, N-terminal acetylation does not necessarily protect proteins from degradation, as often supposed, nor does it play any obvious role in protection of proteins from degradation by the "N-end rule" pathway.

A significant means for assessing the general importance of N-terminal acetylation comes from the phenotypic defects in mutants lacking one or another of the NATs. The lack of N-terminal acetylation of the viral major coat protein, *gag*, in *mak3*$^-$ strains prevents assembly or maintenance of the viral particle.[27] Also *mak3*$^-$ strains do not utilize non-fermentatable carbon sources at 37 °C, probably because of the lack of acetylation of a still unidentified mitochondrial protein.[25,27]

We have previously reported that *nat3*-Δ mutants exhibit multiple defective phenotypes, including slow growth, lack of growth on YPG medium at 37 °C, reduced growth on medium containing NaCl, and reduced mating.[24] Such defects could arise from the lack of acetylation of any number of proteins essential for different processes. While the unacetylated proteins responsible for these defects are not easily identified, the temperature and NaCl sensitivity could be attributed to lack of acetylation of actin (Act1p), which contains a normal N-terminal sequence Ac-Met-Asp-Ser-Glu-. In addition, it has been shown that acetylation at the N terminus of actin strengthens weak interaction between actin and myosin.[49]

Actin cable formation requires tropomyosin for stability. The N-terminal tail of tropomyosin and its acetylation status is very important for protein function.[47] Furthermore, yeast tropomyosins Tpm1p and Tpm2p, with N termini Met-Asp- and Met-Glu-, respectively, very likely are the substrates for NatB. It was found that Mdm20p, a NatB subunit (Table 8) is necessary for actin–tropomyosin interaction but the protein role was not determined.[50] Previous work by Hermann *et al*.[51] revealed that *mdm20*-Δ strains were defective in mitochondrial inheritance and actin cables (bundles of actin filaments), and that extra copies of *TPM1*, a gene encoding the actin filament-binding protein tropomyosin, suppress mitochondrial inheritance defects and partially restore actin cables in *mdm20*-Δ cells. Synthetic lethality was also observed between *mdm20* and *tpm1* mutant strains, and certain dominant alleles of *ACT1* and *TPM1* suppressed *mdm20*-Δ. Interestingly, one of the *mdm20* deletion mutant suppressors was *TPM1-5* allele containing altered protein N terminus with extended seven amino acid residues and utilizing earlier ATG start, resulting in Met-His-, instead of the native Met-Asp- terminus. Although Mdm20p does not co-localize with actin or tropomyosin in the growing cables,[50] it is nevertheless required for association of these proteins.[52] Using the TAP-protocol, we recently found that Mdm20p is a subunit of NatB (Table 8) (B.P., T. Cardillo, G. Bedi & F.S., unpublished results) and we suggest that protein acetylation is required for proper actin–tropomyosin interaction.

In contrast, many non-acetylated recombinant proteins are fully active. For example, the N-terminal acetylation of chaperonin Hsp10 protein is not necessary for the correct folding of the protein and also is not important for chaperonin activity or mitochondrial import.[53] Similarly, other proteins that normally contain an acetylated N terminus, such as alcohol dehydrogenase, are stable and fully functional.[54] Results with annexin II tetramer (AIIt) indicate that N-terminal acetylation does not affect the *in vitro* activities or conformational stability of the protein.[55] The number of examples of proteins either requiring or not requiring N-terminal acetylation undoubtedly will continue to be augmented. Not only can the lack of acetylation result in various defects, but abnormal acetylation also can prevent normal functions. For example, the acetylation of the N-terminal catalytic threonine residue of various 20 S proteosome subunits causes the loss of specific peptidase activities.[56]

Obviously, both N-terminal acetylation and the lack of N-terminal acetylation have evolved to meet the individual requirements of specific proteins. The viability of *ard1*-Δ, *nat1*-Δ, *mak3*-Δ and *nat3*-Δ mutants lacking NATs suggests that the role of acetylation may be subtle and not absolute for most proteins. Possibly only a subset of proteins actually requires this modification for activity or stability, whereas the remainder are acetylated only because their amino termini fortuitously correspond to consensus sequences.

Unique N-terminal processing of eukaryotic actins

Actin is a major contractile protein in both muscle and non-muscle eukaryotic cells. All actins are highly homologous and contain several acidic amino acid residues at N termini, which are required for function (see above). Apparently all actin isoforms from all eukaryotes undergo the normal cotranslational processing of methionine cleavage and acetylation as described above for typical proteins. However, extensive studies by Rubenstein and colleagues revealed that at least some actins from many eukaryotes undergo additional specific posttranslational processing, including actins from the slime mold *D. discoideum*,[57] the fruit fly *D. melanogaster*,[58] birds, and mammals.[59–61] However, additional post-translational processing of actin does not occur in the protozoa *Acanthamoeba castellanii*,[62] or in the fungi *S. cerevisiae*, *Aspergillus nidulans*, *S. pombe*, and *Candida albicans*.[45] The posttranslational processing of actin requires an *N*-acetylaminopeptidase (ANAP), which specifically removes N-terminal Ac-Met or Ac-Cys from actin to leave an acidic N-terminal residue, and which has been isolated from rat liver and partially characterized.[63,64]

The specific posttranslational processing of actin can now be assigned to the following two types, Type I and Type II, which are shown above, and which consider the more recently studied general cotranslational systems: (While only single examples of the actins are depicted, Type I actins include both Met-Asp- and Met-Glu- proteins.)

Type I

Met-Asp-
↓ Cotranslational acetylation (NatB substrate)
Ac-Met-Asp-
↓ Posttranslational cleavage of Ac-Met (ANAP)
Asp-
↓ Posttranslational acetylation of aspartic acid (NAT unknown)
Ac-Asp-

Type II

Met-Cys-Asp-
↓ Cotranslational methionine cleavage (MAP1 or MAP2)
Cys-Asp-
↓ Cotranslational acetylation (NAT unknown)
Ac-Cys-Asp-
↓ Posttranslational cleavage of Ac-Cys (ANAP)
Asp-
↓ Posttranslational acetylation of aspartic acid (NAT unknown)
Ac-Asp-

These specific posttranslational processing events have obviously evolved to produce actin with Ac-Asp- or Ac-Glu- termini, reflecting the requirement for an acidic amino acid at the N terminus. It should be noted that proteins with just Asp- or Glu- at the N terminus would be unstable, as they would be degraded by the N-end rule degradation system;[65] thus, acetylation may be required in part for stabilization of the actins in some but not all organisms with acidic residues at the termini (see below). So far, no NAT specifically acting on actins with aspartic acid, glutamic acid or cysteine termini have been identified. On the other hand, actins from *S. cerevisiae* (Ac-Met-Asp-Ser-Glu-), other fungi, and *A. castellanii* (Ac-Gly-Asp-Glu-) have evolved without requiring acidic residues at the immediate N terminus and without requiring posttranslational processing, although nearby acidic residues are required for normal function. Thus, the different actins have different N-terminal sequence requirements. In this regard as discussed above, we previously suggested that the slow growth phenotype, lack of growth on non-fermentable carbon sources, temperature and salt sensitivity in *nat3*-Δ yeast mutants, lacking Nat B, could all be attributed primarily to the lack of actin acetylation.[24]

However, the ACT88F actin isoform from *D. melanogaster* is normally N-terminally processed *in vivo* by the cleavage of Ac-Cys, but the resulting N-terminal aspartic acid residue is not acetylated.[66] Nevertheless, the actin with the free α-amino aspartic acid residue is stable. Furthermore, Schmitz *et al*.[66] reported that *D. melanogaster* carrying the *mod*$^-$ mutation failed to complete posttranslational processing of the ACT88F actin. They proposed that the *mod* gene product is normally responsible for removing Ac-Cys from actin, and may correspond to an ANAP. The biological significance of this process was demonstrated by observations that retention of the Ac-Cys- at the terminus of ACT88F affected the flight muscle function of *mod*$^-$ flies. Clearly, the N terminus requirement varies with different actins.

Deacetylases and acylamino acid-releasing enzyme (AARE)

In addition to the N-terminal acetylation occurring cotranslationally, there are numerous examples of acetylation of the ε-amino group of lysine residues at various positions occurring posttranslationally.[67] The most studied example is histones H2A, H2B and H3, in which the modification occurs at multiple sites in the N-terminal domains. In contrast to N-terminal acetylation, ε-Lys acetylation of histones is reversible, due to the action of histone deacetylases.[68] There is no evidence for deacetylases that act on N-terminal acetylated proteins.

However, there are acylamino acid-releasing enzymes (AARE) (also designated acylaminoacyl-peptide hydrolase), which cleave Ac-Ala, Ac-Thr, Ac-Met, Ac-Gly, and Ac-Ser from the N-terminal end of short peptides, but are not known to act on N-terminal acetylated proteins.[69–71] AARE have been isolated from eukaryotes and an archaeon, but not from prokaryotes.[72] On the basis of their *in vitro* properties, AARE have been suggested to possibly act on short nascent chains during translation, although their physiological function is unknown. It is also unknown how AARE is related to the acetyl-Met and acetyl-Cys hydrolase, which are involved in type I and type II actin processing, although they are clearly different. We favor the

view that AARE play an important role in the recycling of amino acid residues for protein synthesis, but are not involved in cotranslational or posttranslational processing of N-terminal acetylated proteins.

Acknowledgements

Supported by National Institute of Health Grant R01 GM12702.

References

1. Polevoda, B. & Sherman, F. (2000). *N*-terminal acetylation of eukaryotic proteins. *J. Biol. Chem.* **275**, 36479–36482.
2. Bradshaw, R. A., Brickey, W. W. & Walker, K. W. (1998). *N*-terminal processing: the methionine aminopeptidase and Nα-acetyl transferase families. *Trends Biochem. Sci.* **23**, 263–267.
3. Wold, F. (1981). *In vivo* chemical modifications of proteins. *Annu. Rev. Biochem.* **50**, 783–814.
4. Han, K.-K. & Martinage, A. (1993). Post-translational chemical modifications of proteins. III. Current developments in analytical procedures of identification and quantitation of post-translational chemically modified amino acid(s) and its derivatives. *Int. J. Biochem.* **25**, 957–970.
5. Michel, H., Hunt, D. F., Shabanowitz, J. & Bennett, J. (1988). Tandem mass spectrometry reveals that three photosystem II proteins of spinach chloroplasts contain *N*-acetyl-O-phosphothreonine at their NH_2 termini. *J. Biol. Chem.* **263**, 1123–1130.
6. Driessen, H. P., de Jong, W. W., Tesser, G. I. & Bloemendal, H. (1985). The mechanism of *N*-terminal acetylation of proteins. *CRC Crit. Rev. Biochem.* **18**, 281–325.
7. Kendall, R. L., Yamada, R. & Bradshaw, R. A. (1990). Cotranslational amino-terminal process. *Methods Enzymol.* **185**, 398–407.
8. Li, X. & Chang, Y.-H. (1995). Amino-terminal protein processing in *Saccharomyces cerevisiae* is an essential function that requires two distinct methionine aminopeptidases. *Proc. Natl Acad. Sci.* **92**, 12357–12361.
9. Giglione, C., Serero, A., Pierre, M., Boisson, B. & Meinnel, T. (2000). Identification of eukaryotic peptide deformylases reveals universality of *N*-terminal protein processing mechanisms. *EMBO J.* **19**, 5916–5929.
10. Sherman, F., Stewart, J. W. & Tsunasawa, S. (1985). Methionine or not methionine at the beginning of a protein. *BioEssays*, **3**, 27–31.
11. Hirel, H-P., Schmitter, J-M., Dessen, P., Fayat, G. & Blanquet, S. (1989). Extent of *N*-terminal methionine excision from *Escherichia coli* proteins is governed by the side-chain length of the penultimate amino acid. *Proc. Natl Acad. Sci. USA*, **86**, 8247–8251.
12. DalbÏge, H., Bayen, S. & Pedersen, J. (1990). *In vivo* processing of *N*-terminal methionine in *E. coli*. *FEBS Letters*, **266**, 1–3.
13. Huang, S., Elliott, R. C., Liu, P. S., Koduri, R. K., Weickmann, J. L., Lee, J. H. *et al.* (1987). Specificity of cotranslational amino-terminal processing of proteins in yeast. *Biochemistry*, **26**, 8242–8246.
14. Boissel, J. P., Kasper, T. J. & Bunn, H. F. (1988). Cotranslational amino-terminal processing of cytosolic proteins: cell-free expression of site-directed mutants of human hemoglobin. *J. Biol. Chem.* **263**, 8443–8449.
15. Lowther, W. T. & Matthews, B. W. (2000). Structure and function of the methionine aminopeptidases. *Biochim. Biophys. Acta*, **1477**, 157–167.
16. Persson, B., Flinta, C., von Heijne, G. & Jörnvall, H. (1985). Structures of *N*-terminally acetylated proteins. *Eur. J. Biochem.* **152**, 523–527.
17. Jönvall, H. (1973). Acetylation of protein *N*-terminal amino groups: structural observations on α-amino acetylated proteins. *J. Theor. Biol.* **55**, 1–12.
18. Lee, F. J., Lin, L. W. & Smith, J. A. (1989). N^{α}-acetyltransferase deficiency alters protein synthesis in *Saccharomyces cerevisiae*. *FEBS Letters*, **256**, 139–142.
19. Walker, J.-P. (1963). The NH_2-terminal residues of the proteins from cell-free extracts of *E. coli*. *J. Mol. Biol.* **7**, 483–496.
20. Flinta, C., Persson, B., Jörnvall, H. & von Heijne, G. (1986). Sequence determinants of cytosolic *N*-terminal protein processing. *Eur. J. Biochem.* **154**, 193–196.
21. Tsunasawa, S., Stewart, J. W. & Sherman, F. (1985). Amino-terminal processing of mutant forms of yeast iso-1-cytochrome *c*: the specificities of methionine aminopeptides and acetyltransferase. *J. Biol. Chem.* **260**, 5382–5391.
22. Moerschell, R. P., Hosokawa, Y., Tsunasawa, S. & Sherman, F. (1990). The specificities of yeast methionine aminopeptidase and acetylation of amino-terminal methionine *in vivo*: processing of altered iso-1-cytochromes c created by oligonucleotide transformation. *J. Biol. Chem.* **265**, 19638–19643.
23. Sherman, F., Moerschell, R. P., Tsunasawa, S. & Sternglanz, R. (1993). *N*-terminal acetylation of mutationally altered form of iso-1-cytochromes *c* in normal and *nat1*$^{-}$ strains deficient in the major *N*-terminal acetyl transferase of the yeast *Saccharomyces cerevisiae*. In *Methods in Protein Sequence Analysis* (Imahori, K. & Sakiyama, F., eds), pp. 173–181, Plenum Publishing Corporation, New York, NY.
24. Polevoda, B., Norbeck, J., Takakura, H., Blomberg, A. & Sherman, F. (1999). Identification and specificities of *N*-terminal acetyltransferases from *Saccharomyces cerevisiae*. *EMBO J.* **18**, 6155–6168.
25. Polevoda, B. & Sherman, F. (2001). NatC N-terminal acetyltransferase of yeast contains three subunits, Mak3p, Mak10p, and Mak31p. *J. Biol. Chem.* **276**, 20154–20159.
26. Mullen, J. R., Kayne, P. S., Moerschell, R. P., Tsunasawa, S., Gribskov, M., Colavito-Shepanski, M. *et al.* (1989). Identification and characterization of genes and mutants for an *N*-terminal acetyltransferase from yeast. *EMBO J.* **8**, 2067–2075.
27. Tercero, J. C. & Wickner, R. B. (1992). *MAK3* encodes an *N*-acetyltransferase whose modification of the L-A *gag* NH_2 terminus is necessary for virus particle assembly. *J. Biol. Chem.* **267**, 20277–20281.
28. Tercero, J. C., Dinman, J. D. & Wickner, R. B. (1993). Specificity of the yeast *MAK3 N*-acetyltransferase that modifies gag of the L-A dsRNA virus. *J. Bacteriol.* **175**, 3192–3194.
29. Perrot, M., Sagliocco, F., Mini, T., Monbirot, C., Schneider, U., Shevchenko, A. *et al.* (1999).

Two-dimentional gel protein database of *Saccharomyces cerevisiae* (update 1999). *Electrophoresis*, **20**, 2280–2298.
30. Garrels, J. I., McLaughlin, C. S., Warner, J. R., Futcher, B., Latter, G. I., Kobayashi, R. *et al.* (1997). Proteome studies of *Saccharomyces cerevisiae*: identification and characterization of abundant proteins. *Electrophoresis*, **18**, 1347–1360.
31. Arnold, R., Polevoda, B., Reilly, J. P. & Sherman, F. (1999). The action of *N*-terminal acetyltransferases on yeast ribosomal proteins. *J. Biol. Chem.* **274**, 37035–37040.
32. Kimura, Y., Takaoka, M., Tanaka, S., Sassa, H., Tanaka, K., Polevoda, B. *et al.* (2000). Nα-Acetylation and proteolytic activity of the yeast 20 S proteasome. *J. Biol. Chem.* **275**, 4635–4639.
33. Costanzo, M. C., Crawford, M. E., Hirschman, J. E., Kranz, J. E., Olsen, P., Robertson, L. S. *et al.* (2001). YPD™, PombePD™, and WormPD library, an integrated resource for protein information. *Nucl. Acids Res.* **29**, 75–79.
34. Bairoch, A. & Apweiler, R. (2000). The SWISS-PROT protein sequence database and its supplement TrEMBL in 2000. *Nucl. Acids Res.* **28**, 45–48.
35. Yoshikawa, A., Isono, S., Sheback, A. & Isono, K. (1987). Cloning and nucleotide sequencing of the genes rimI and rimJ which encode enzymes acetylating ribosomal proteins S18 and S5 of *Escherichia coli* K12. *Mol. Gen. Genet.* **209**, 481–488.
36. Tanaka, S., Matsushita, Y., Yoshikawa, A. & Isono, K. (1989). Cloning and molecular characterization of the gene rimL which encodes an enzyme acetylating ribosomal protein L12 of *Escherichia coli* K12. *Mol. Gen. Genet.* **217**, 289–293.
37. Uchiumi, T., Hori, K., Nomura, T. & Hachimori, A. (1999). Replacement of L7/L12.L10 protein complex in *Escherichia coli* ribosomes with the eukaryotic counterpart changes the specificity of elongation factor binding. *J. Biol. Chem.* **274**, 27578–27582.
38. Itoh, T. (1981). Primary structure of an acidic ribosomal protein from *Micrococcus lysodeikticus*. *FEBS Letters*, **127**, 67–70.
39. Dognin, M. J. & Wittman-Liebold, B. (1980). Purification and primary structure determination of the *N*-terminal blocked protein, L11, from *Escherichia coli* ribosomes. *Eur. J. Biochem.* **112**, 131–151.
40. Neuwald, A. F. & Landsman, D. (1997). GCN5-related histone *N*-acetyltransferases belongs to a diverse superfamily that includes the yeast SPT10 protein. *Trends Biochem. Sci.* **22**, 154–155.
41. Corpet, F. (1988). Multiple sequence alignment with hierarchical clustering. *Nucl. Acids Res.* **16**, 10881–10890.
42. Ozaki, K., Fujiwara, T., Nakamura, Y. & Takahashi, E. (1998). Isolation and mapping of a novel human kidney- and liver-specific gene homologous to the bacterial acetyltransferases. *J. Hum. Genet.* **43**, 255–258.
43. Popsueva, A. E., Luchinskaya, N. N., Ludwig, A. V., Zinovjeva, O. Y., Poteryaev, D. A., Feigelman, M. M. *et al.* (2001). Overexpression of camello, a member of a novel protein family, reduces blastomere adhesion and inhibits gastrulation in *Xenopus laevis*. *Dev. Biol.* **234**, 483–496.
44. Augen, J. & Wold, F. (1986). How much sequence information is needed for the regulation of amino-terminal acetylation of eukaryotic proteins? *Trends Biochem. Sci.* **11**, 494–497.
45. Cook, R. K., Sheff, D. R. & Rubenstein, P. A. (1991). Unusual metabolism of the yeast actin amino terminus. *J. Biol. Chem.* **266**, 16825–16833.
46. Manning, L. R. & Manning, J. M. (2001). The acetylation state of human fetal hemoglobin modulates the strength of its subunit interactions: long-range effects and implications for histone interactions in the nucleosome. *Biochemistry*, **40**, 1635–1639.
47. Urbancikova, M. & Hitchcock-DeGregori, S. E. (1994). Requirement of amino-terminal modification for striated muscle α-tropomyosin function. *J. Biol. Chem.* **269**, 24310–24315.
48. Ogawa, H., Gomi, T., Takata, Y., Date, T. & Fujioka, M. (1997). Recombinant expression of rat glycine *N*-methyltransferase and evidence for contribution of *N*-terminal acetylation to co-operative binding of *S*-adenosylmethionine. *Biochem. J.* **327**, 407–412.
49. Abe, A., Saeki, K., Yasunaga, T. & Wakabayashi, T. (2000). Acetylation at the *N*-terminus of actin strengthens weak interaction between actin and myosin. *Biochem. Biophys. Res. Commun.* **268**, 14–19.
50. Singer, J. M., Hermann, G. J. & Shaw, J. M. (2000). Suppressors of *mdm20* in yeast identify new alleles of *ACT1* and *TPM1* predicted to enhance actin-tropomysin interactions. *Genetics*, **156**, 523–534.
51. Hermann, G. J., King, E. J. & Shaw, J. M. (1997). The yeast gene, *MDM20*, is necessary for mitochondrial inheritance and organization of the actin cytoskeleton. *J. Cell Biol.* **137**, 141–153.
52. Evangelista, M., Pruyne, D., Amberg, D. C., Boone, C. & Bretscher, A. (2002). Formins direct Arp2/3-independent actin filament assembly to polarize cell growth in yeast. *Nature Cell Biol.* **4**, 32–41.
53. Ryan, M. T., Naylor, D. J., Hoogenraad, N. J. & Hoj, P. B. (1995). Affinity purification, overexpression, and characterization of chaperonin 10 homologues synthesized with and without *N*-terminal acetylation. *J. Biol. Chem.* **270**, 22037–22043.
54. Hoog, J. O., Weis, M., Zeppezauer, M., Jomvall, H. & von Bahr-Lindstrom, H. (1987). Expression in *Escherichia coli* of active human alcohol dehydrogenase lacking *N*-terminal acetylation. *Biosci. Rep.* **7**, 969–974.
55. Kang, H.-M., Kassam, G., Jarvis, S. E., Fitzpatrick, S. L. & Waisman, D. M. (1997). Characterization of human recombinant annexin 11 tetramer purified from bacteria: role of *N*-terminal acetylation. *Biochemistry*, **36**, 2041–2050.
56. Arendt, C. S. & Hochstrasser, M. (1999). Eukaryotic 20 S proteasome catalytic subunit propeptides prevent active site inactivation by *N*-terminal acetylation and promote particle assembly. *EMBO J.* **18**, 3575–3585.
57. Rubenstein, P. A., Redman, K., Solomon, L. & Martin, D. (1987). Amino-terminal processing of *Dictyostelium discoideum* actin. *Methods Cell Biol.* **28**, 231–243.
58. Rubenstein, P. A. & Martin, D. J. (1983). NH_2-terminal processing of *Drosophila melanogaster* actin. Sequential removal of two amino acids. *J. Biol. Chem.* **258**, 11354–11360.
59. Solomon, L. R. & Rubenstein, P. A. (1985). Correct NH_2-terminal processing of cardiac muscle α-isoactin (class II) in a nonmuscle mouse cell. *J. Biol. Chem.* **260**, 7659–7664.
60. Strauch, A. R., Offord, J. D., Chalkley, R. & Rubenstein, P. A. (1986). Characterization of actin mRNA levels during BC3H1 cell differentiation. *J. Biol. Chem.* **261**, 849–855.

61. Martin, D. & Rubenstein, P. A. (1987). Alternate pathways for removal of the class II actin initiator methionine. *J. Biol. Chem.* **262**, 6350–6356.
62. Redman, K. L., Martin, D. J., Korn, E. D. & Rubenstein, P. A. (1985). Lack of NH_2-terminal processing of actin from *Acanthamoeba castellanii*. *J. Biol. Chem.* **260**, 14857–14861.
63. Sheff, D. R. & Rubenstein, P. A. (1992). Isolation and characterization of the rat liver actin *N*-acetylaminopeptidase. *J. Biol. Chem.* **267**, 20217–20224.
64. Sheff, D. R. & Rubenstein, P. A. (1992). Amino-terminal processing of actins mutagenized at the Cys^{-1} residue. *J. Biol. Chem.* **267**, 2671–2678.
65. Varshavsky, A. (2000–2001). Recent studies of the ubiquitin system and the N-end rule pathway. *Harvey Lect.* **96**, 93–116.
66. Schmitz, S., Clayton, J., Nongthomba, U., Prinz, H., Veigel, C., Geeves, M. & Sparrow, J. (2000). *Drosophila* ACT88F indirect flight muscle-specific actin is not *N*-terminally acetylated: a mutation in *N*-terminal processing affects actin function. *J. Mol. Biol.* **295**, 1201–1210.
67. Polevoda, B. & Sherman, F. (2002). Diversity of acetylated proteins. *Genome Biol.* **3**, reviews0006.1–0006.
68. Kouzarides, T. (1999). Histone acetylases and deacetylases in cell proliferation. *Curr. Opin. Genet. Dev.* **9**, 40–48.
69. Krishna, R. G. & Wold, F. (1992). Specificity determinants of acylaminoacyl-peptide hydrolase. *Protein Sci.* **1**, 582–958.
70. Sokolik, C. W., Liang, T. C. & Wold, F. (1994). Studies on the specificity of acetylaminoacyl-peptide hydrolase. *Protein Sci.* **3**, 126–131.
71. Raphel, V., Giardina, T., Guevel, L., Perrier, J., Dupuis, L., Guo, X. J. & Puigserver, A. (1999). Cloning, sequencing and further characterization of acylpeptide hydrolase from porcine intestinal mucosa. *Biochim. Biophys. Acta*, **1432**, 371–381.
72. Ishikawa, K., Ishida, H., Koyama, Y., Kawarabayasi, Y., Kawahara, J., Matsui, E. & Matsu, I. (1998). Acylamino acid-releasing enzyme from the thermophilic archaeon *Pyrococcus hoirkoshii*. *J. Biol. Chem.* **273**, 17726–17731.
73. Takakura, H., Tsunasawa, S., Miyagi, M. & Warner, J. R. (1992). NH_2-terminal acetylation of ribosomal proteins of *Saccharomyces cerevisiae*. *J. Biol. Chem.* **267**, 5442–5445.
74. Boucherie, H., Sagliocco, F., Joubert, R., Maillet, I., Labarre, J. & Perrot, M. (1996). Two-dimensional gel protein database of *Saccharomyces cerevisiae*. *Electrophoresis*, **17**, 1683–1699.
75. Apweiler, R., Attwood, T. K., Bairoch, A., Bateman, A., Birney, E., Biswas, M. *et al.* (2001). The InterPro database, an integrated documentation resource for protein families, domains and functional sites. *Nucl. Acids Res.* **29**, 37–40.
76. LaMarche, A. E., Abate, M. I., Chan, S. H. & Trumpower, B. L. (1992). Isolation and characterization of *COX12*, the nuclear gene for a previously unrecognized subunit of *Saccharomyces cerevisiae* cytochrome *c* oxidase. *J. Biol. Chem.* **267**, 22473–22480.

Edited by S. Reed

(Received 26 August 2002; received in revised form 4 November 2002; accepted 4 November 2002)