SUMMARY

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

STUDY TITLE

Characterization of Aryloxyalkanoate Dioxygenase-12 (AAD-12) Protein Derived from Transgenic Soybean Event DAS-68416-4

AMENDED REPORT

DATA REQUIREMENTS

Not Applicable

AUTHOR(S)

B. W. Schafer, S. K. Embrey

STUDY COMPLETED ON

18-Sep-2009

Amended Report Date: September 30, 2010

PERFORMING LABORATORY

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

LABORATORY STUDY ID

081113

Characterization of Aryloxyalkanoate Dioxygenase-12 (AAD-12) Protein Derived from Transgenic Soybean Event DAS-68416-4

SUMMARY

Soybean has been modified by the insertion of the *aad-12* gene from *Delftia acidovorans* which encodes the aryloxyalkanoate dioxygenase-12 (AAD-12) protein. The trait confers tolerance to 2,4-dichlorophenoxyacetic acid and pyridyloxyacetate herbicides and may be used as a selectable marker during plant transformation and in breeding nurseries. The expressed AAD-12 protein is approximately 32 kDa in size.

To perform various toxicology, eco-toxicology, biochemical characterization, and enzymatic activity studies, large quantities of the AAD-12 protein are required. Because it is technically infeasible to extract and purify sufficient amounts of AAD-12 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-12 protein derived from the transgenic soybean event DAS-68416-4 (event 416) and compare them with the previously characterized microbe-derived AAD-12 protein. Sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS-PAGE, stained with Coomassie blue and glycoprotein detection methods), western blot, lateral flow test strip assay (LFS), matrix assisted laser desorption/ionization time-of-flight mass spectrometry (MALDI-TOF MS) and Electrospray Ionization-Liquid Chromatography MS (ESI-LC/MS) analyses were used to characterize the biochemical properties of the protein. Using these methods, the AAD-12 proteins from *P. fluorescens* and transgenic soybean (event 416) were shown to be biochemically equivalent. These data support the use of the microbial protein for use in studies supporting the registration of transgenic soybeans expressing the AAD-12 protein.

STUDY TITLE

Characterization of Aryloxyalkanoate Dioxygenase-12 (AAD-12) Protein Derived from Transgenic Soybean Event DAS-68416-4

AMENDED REPORT

DATA REQUIREMENTS

Not Applicable

AUTHOR(S)

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

STUDY COMPLETED ON

18-Sep-2009

Amended Report Date: September 30, 2010

PERFORMING LABORATORY

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

LABORATORY STUDY ID

081113

Total number of pages is 172 including 4.1R1 of 27

STATEMENT OF NO DATA CONFIDENTIALITY CLAIMS

Compound: AAD-12

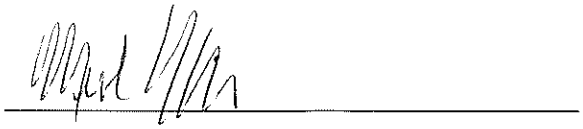
Title: Characterization of Aryloxyalkanoate Dioxygenase-12 (AAD-12) Protein Derived from Transgenic Soybean Event DAS-68416-4

STATEMENT OF NO DATA CONFIDENTIALITY CLAIMS: No claim of confidentiality, on any basis whatsoever, is made for any information contained in this document. I acknowledge that information not designated as within the scope of FIFRA sec. 10(d)(1)(A), (B), or (C) and which pertains to a registered or previously registered pesticide is not entitled to confidential treatment and may be released to the public, subject to the provisions regarding disclosure to multinational entities under FIFRA sec. 10(g).*

Company: Dow AgroSciences LLC

Company Agent: M. S. Krieger

Title: Regulatory Manager

Signature: 

Date: 30 Sep 2011

*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.

This report has been amended. This page was signed prior to or on the study completion date, but has been re-signed due to the modification.

STATEMENT OF COMPLIANCE WITH GOOD LABORATORY PRACTICE STANDARDS

Title: Characterization of Aryloxyalkanoate Dioxygenase-12 (AAD-12) Protein Derived from Transgenic Soybean Event DAS-68416-4

Study Initiation Date: 05-Dec-2008 Study Completion Date 18-Sep-2010
***Amended Report Date: September 30, 2010**

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 MALDI-TOF MS and ESI-LC/MS 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.

M. S. Krieger
Sponsor
Dow AgroSciences LLC

30 Sep 2010
Date

M. S. Krieger
Submitter
Dow AgroSciences LLC

30 Sep 2010
Date

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

30. Sep 2010
Study Completion Date

***This report has been amended. The original report was completed and signed on 18-Sep-2009. The study director's signature and date reflect the date of the report amendment.**

Dow AgroSciences Quality Assurance Unit
Good Laboratory Practice Statement Page

Compound: **Study ID:** 081113

Title: Characterization of the Aryloxyalkanoate Dioxygenase-12 (AAD-12) Protein Derived from Transgenic Soybean Event DAS-68416-4

Study Initiation Date: 05-Dec-2008 **Study Completion Date:** 18-Sep-09
***Amended Report Date:** 30-Sep-10

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
5-Dec-08	5-Dec-08	Protocol Review
14-Jan-09	14-Jan-09	Western Blot
24, 25, 26, 27-Aug-09	27-Aug-09	Raw Data and Final Report. 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.

Debra Stefanek 30-Sep-10

Debra Stefanek Date
Dow AgroSciences, Quality Assurance

***The original report was completed and signed on 18-Sep-09. The new signature and date reflect the date of the report amendment.**

Summary of Amendment

Characterization of Aryloxyalkanoate Dioxygenase-12 (AAD-12) Protein Derived from Transgenic Soybean Event DAS-684164

Summary of amendment changes:

Summary Page: Page 1**R1** of 2, the phrases AMENDED REPORT and Amended Report date were added.

Title page: Page 1**R1** of 171, the phrases AMENDED REPORT, Amended Report date.

Page 3**R1**: The report amended date and a statement describing the original signature dates were added.

Page 4**R1**: A new QA page was added

Page 4.1**R1**: This page is added to describe the changes.

Page 5**R1**: New signatures were collected.

Page 12**R1**: The seed generation was changed from T4 to T5.

Reason for Amendment:
It was discovered that the generation number of the seeds was not correct due to an error in Variety database used to determine the generation number of AAD-12 soybean (event DAS-68416-4) seeds. The correct generation number will be documented in the amended report and study files.

Impact on Study:
This amendment has a positive impact on the study as the correct generation number will be documented in the amended report and study files. The correction has no impact on the study because the nature of the seeds has not changed.

Signatures:

B. W. Schafer, Study Director — 30 Sep 2010 — Date

K. A. Clayton, Manager — 30 Sep 2010 — Date

D. Stefanek, Quality Assurance — 30 Sep 2010 — Date

SIGNATURE PAGE

B. W. Schafer
Author
Dow AgroSciences LLC

30-Sep-2010
Date

S. K. Embrey
Co-Author
Dow AgroSciences LLC

30-Sept-2010
Date

R. A. Herman
Peer-Reviewer
Dow AgroSciences LLC

30-Sept-2010
Date

G. Shan
Science Leader
Dow AgroSciences LLC

30-Sept 2010
Date

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

30 Sep 2010
Date

STUDY PERSONNEL

Title: Characterization of Aryloxyalkanoate Dioxygenase-12 (AAD-12) Protein Derived from Transgenic Soybean Event DAS-68416-4

Study Director: Barry W. Schafer

Analysts: Shawna K. Embrey

TABLE OF CONTENTS

TABLE OF CONTENTS (CONT.)

Characterization of Aryloxyalkanoate Dioxygenase-12 (AAD-12) Protein Derived from Transgenic Soybean Event DAS-68416-4

ABSTRACT

Soybean has been modified by the insertion of the *aad-12* gene from *Delftia acidovorans* which encodes the aryloxyalkanoate dioxygenase-12 (AAD-12) protein. The trait confers tolerance to 2,4-dichlorophenoxyacetic acid and pyridyloxyacetate herbicides and may be used as a selectable marker during plant transformation and in breeding nurseries. The expressed AAD-12 protein is approximately 32 kDa in size.

To perform various toxicology, eco-toxicology, biochemical characterization, and enzymatic activity studies, large quantities of the AAD-12 protein are required. Because it is technically infeasible to extract and purify sufficient amounts of AAD-12 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-12 protein derived from the transgenic soybean event DAS-68416-4 (event 416) and compare them with the previously characterized microbe-derived AAD-12 protein. Sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS-PAGE, stained with Coomassie blue and glycoprotein detection methods), western blot, lateral flow test strip assay (LFS), matrix assisted laser desorption/ionization time-of-flight mass spectrometry (MALDI-TOF MS) and electrospray ionization-liquid chromatography mass spectrometry (ESI-LC/MS) analyses were used to characterize the biochemical properties of the protein. Using these methods, the AAD-12 proteins from *P. fluorescens* and transgenic soybean (event 416) were shown to be biochemically equivalent. These data support the use of the microbial protein for use in studies supporting the registration of transgenic soybeans expressing the AAD-12 protein.

ABBREVIATIONS

AAD-12	aryloxyalkanoate dioxygenase-12 protein
C-terminus	carboxyl-terminus
ESI-LC/MS	electrospray ionization-liquid chromatography mass spectrometry
kDa	kiloDalton
L	liter
LFS	lateral flow strip assay
LSB	Laemmli sample buffer
MALDI-TOF MS	matrix assisted laser desorption/ionization time-of-flight mass spectrometry
µg	microgram
mA	milliAmps
µL	microliter
min	minutes
ng	nanogram
N-terminus	amino-terminus
PBS	phosphate buffered saline, pH 7.4
PBST	phosphate buffered saline + 0.05% Tween 20, pH 7.4
PMF	peptide mass fingerprinting
PMSF	phenylmethylsulphonyl fluoride
s	seconds
SDS-PAGE	sodium dodecyl sulfate polyacrylamide gel electrophoresis
TSN	test substance number

INTRODUCTION

Soybean has been modified by the insertion of the *aad-12* gene from *Delftia acidovorans* which encodes the aryloxyalkanoate dioxygenase-12 (AAD-12) protein. The trait confers tolerance to 2,4-dichlorophenoxyacetic acid and pyridyloxyacetate herbicides and may be used as a selectable marker during plant transformation and in breeding nurseries. The expressed AAD-12 protein is approximately 32 kDa in size.

To perform various toxicology, eco-toxicology, biochemical characterization, and enzymatic activity studies, large quantities of the AAD-12 protein are required. Because it is technically infeasible to extract and purify sufficient amounts of AAD-12 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-12 protein derived from the transgenic soybean event DAS-68416-4 (Event 416) and compare them with the previously characterized microbe-derived AAD-12 protein (Embrey and Schafer, 2009 and Kuppannan and Karnoup 2008). Sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS-PAGE, stained with Coomassie blue and glycoprotein detection methods), western blot, lateral flow test strips, matrix assisted laser desorption/ionization time-of-flight mass spectrometry (MALDI-TOF MS) and ESI-LC/MS peptide analysis 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 LFS) are widely used to identify the authenticity and relative 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. Peptide mass fingerprinting or peptide analysis (by MALDI-TOF-MS and ESI-LC/MS following

an orbital shaker for 1 hour. After the incubation, the PS beads were collected by vacuum on P8 paper and washed extensively with PBST. A final wash with 20 mM Tris, 1 M NaCl, pH 8.0 was performed to increase the conductivity and pH. After the final wash, the beads were transferred to a sintered glass funnel and the bound proteins were eluted with ten 45-mL aliquots of Milli-Q water. Each fraction was tested for the presence of the AAD-12 protein by LFS assay and the fractions that were positive were pooled, centrifuged at 6000×g and filtered through a 0.45 µ filter. The conductivity of the final pool was raised to ~34 mS/cm by adding ~9 g of NaCl and the PS bound proteins were loaded at 5 mL/min (at 4 ºC) onto an immunoaffinity column that consisted of CNBr-activated Sepharose 4B beads(GE Healthcare Cat #: 17-0430-01) conjugated with an anti-AAD-12 polyclonal antibody (DAS F1197-167.2, 18.4 mg total). The column was prepared on 15-Jul-2008, DAS Notebook F1257-120. The affinity column was washed extensively with pre-chilled PBS (Sigma Cat #:P3813) and the bound proteins were eluted with 3.5 M NaSCN (Sigma Cat #: S7757), 50 mM Tris (Sigma Cat #: T3038), pH 8.0 buffer. Eight 5-mL fractions were collected and fractions 1 – 7 were analyzed by SDS-PAGE and western blot. This was accomplished by concentrating 500 µL of each fraction to ~50 µL and washing the sample with 500 µL of PBST, and reconcentrating the sample to ~ 30 µL using pre-rinsed Millipore spin concentrators (5 kDa MWCO, Cat #: UFV5BCC00). The remaining samples were stored at 4 °C until used for subsequent analyses.

Control Substance

The control substance used in this study was a non-transgenic soybean plant extract (*Glycine max* cv Maverick). Seeds of the Maverick soybean line were planted, grown, harvested, and processed under the same conditions as the transgenic plants described above. Soybean leaf punches were harvested fresh at the Dow AgroSciences greenhouse in Indianapolis, IN and each individual plant was tested by a LFS assay (American Bionostica Inc., Cat #: 702K100) to confirm the absence of the AAD-12 protein.

an orbital shaker for 1 hour. After the incubation, the PS beads were collected by vacuum on P8 paper and washed extensively with PBST. A final wash with 20 mM Tris, 1 M NaCl, pH 8.0 was performed to increase the conductivity and pH. After the final wash, the beads were transferred to a sintered glass funnel and the bound proteins were eluted with ten 45-mL aliquots of Milli-Q water. Each fraction was tested for the presence of the AAD-12 protein by LFS assay and the fractions that were positive were pooled, centrifuged at 6000×g and filtered through a 0.45 µ filter. The conductivity of the final pool was raised to ~34 mS/cm by adding ~9 g of NaCl and the PS bound proteins were loaded at 5 mL/min (at 4 ºC) onto an immunoaffinity column that consisted of CNBr-activated Sepharose 4B beads(GE Healthcare Cat #: 17-0430-01) conjugated with an anti-AAD-12 polyclonal antibody (DAS F1197-167.2, 18.4 mg total). The column was prepared on 15-Jul-2008, DAS Notebook F1257-120. The affinity column was washed extensively with pre-chilled PBS (Sigma Cat #:P3813) and the bound proteins were eluted with 3.5 M NaSCN (Sigma Cat #: S7757), 50 mM Tris (Sigma Cat #: T3038), pH 8.0 buffer. Eight 5-mL fractions were collected and fractions 1 – 7 were analyzed by SDS-PAGE and western blot. This was accomplished by concentrating 500 µL of each fraction to ~50 µL and washing the sample with 500 µL of PBST, and reconcentrating the sample to ~ 30 µL using pre-rinsed Millipore spin concentrators (5 kDa MWCO, Cat #: UFV5BCC00). The remaining samples were stored at 4 °C until used for subsequent analyses.

Control Substance

The control substance used in this study was a non-transgenic soybean plant extract (*Glycine max* cv Maverick). Seeds of the Maverick soybean line were planted, grown, harvested, and processed under the same conditions as the transgenic plants described above. Soybean leaf punches were harvested fresh at the Dow AgroSciences greenhouse in Indianapolis, IN and each individual plant was tested by a LFS assay (American Bionostica Inc., Cat #: 702K100) to confirm the absence of the AAD-12 protein.

Reference Substances

1. Recombinant AAD-12 microbial protein, (Lot #: 466-028A), molecular weight: 31.7 kDa, 35.3% active ingredient (a.i.) AAD-12 to powder mass (Embrey and Schafer, 2009). The microbial preparation was produced in *Pseudomonas fluorescens* strain DC579 at the Dow AgroSciences Bioprocessing R&D facility in San Diego, CA and purified at the Dow AgroSciences Core Biotech R&D facility in Indianapolis, IN (Lin, *et al*., 2007). The lyophilized powder was sent to the Test Substance Coordinator at Dow AgroSciences located in Indianapolis. The material was designated TSN030732.

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

Reference Substance	Product Name	Lot Number	Assay	Reference
Soybean Trypsin Inhibitor (STI)	A component of the GelCode glycoprotein staining kit	IA110577	Glycosylation assay	Pierce Cat #: 1856274
Horseradish Peroxidase (HRP)	A component of the GelCode glycoprotein staining kit	JG124509	Glycosylation assay	Pierce Cat #: 1856273
Bovine Serum Albumin Fraction V (BSA)	Pre-diluted BSA protein assay standard set	FH71884C	SDS-PAGE & Glycosylation assay	Pierce Cat #: 23208
Molecular Weight Markers	Mark12 unstained protein standards	470020 & 399893	SDS-PAGE & Glycosylation assay	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
Prestained Molecular Weight Markers	Novex Sharp prestained protein standards	469212	Western blot & Glycosylation assay	Invitrogen Cat #: LC5800, Molecular Weight Markers of 260, 160, 110, 80, 60, 50, 40, 30, 20, 15, 10 and 3.5 kDa

SDS-PAGE, Western Blot and Lateral Flow Test Strip Assay

The soybean leaf tissues of the transgenic and nontransgenic events were harvested fresh on November 19, 2008 and were frozen, lyophilized, and stored at –80 °C until use. To confirm the presence/absence of the AAD-12 protein in the pooled tissues, approximately 15-mg samples of the lyophilized tissues were weighed out in 1.5-mL microfuge tubes and tested by the lateral flow test strip assay as described by American Bionostica, Inc (Item #: 702K100). The soluble proteins were extracted by adding 0.5 mL of extraction buffer, with one drop of additive, and grinding with a disposable pestle. The tubes were capped and further mixed by shaking the samples for ~10 seconds. The resulting supernatants were clarified by centrifuging the samples for 5 minutes at 20,000xg. The test strips were then incubated in the samples for 10 minutes to develop. After the assay was complete, the strips were removed and allowed to air dry and the results were recorded.

SDS-PAGE and western blot analysis of the transgenic AAD-12 (from T4 seed) and nontransgenic Maverick soybean extracts were performed with Bio-Rad Criterion gels (Bio-Rad Cat #: 345-0123) fitted in a Criterion Cell gel module (Cat #: 165-6001) with MES running buffer (Bio-Rad Cat #: 161-0789). Extracts were prepared by Geno-Grinding (Spex, Model #: 2000) ~75mg of tissue with steel ball bearings in a PBST based buffer (Table 2) for 3 minutes in a chilled Teflon microfuge tube holder. The supernatants were clarified by centrifuging the samples for 5 minutes at 20,000×g and 160 µL of each extract was mixed with 40 µL of 5x Laemmli sample buffer (LSB, 10% SDS, 200 mM Tris pH 6.8, 0.05% bromophenol blue, 20% (w/w) glycerol containing 10% freshly added 2-mercaptoethanol (Bio-Rad, Cat #: 161-0710)) and heated for 5 minutes at ~95 °C. After a brief centrifugation (2 min @ 20,000 ×g), 40 µL of the supernatant was loaded directly on the gel. The reference standard, microbe-derived AAD-12 (TSN030732), and control standard, BSA (Cat #: 23208), were diluted with Bio-Rad 2x Laemmli sample buffer (Bio-Rad Cat #: 161-0737 containing 5% 2-mercaptoethanol) and

processed as described earlier. The electrophoresis was conducted at a constant voltage of 150 V for ~60 minutes using MES running buffer. After separation, the gel was cut in half and one half was stained with Pierce GelCode Blue protein stain (Cat #: 24592) and scanned with a densitometer (Molecular Dynamics, Personal Densitometer Si) to obtain a permanent record of the image. The remaining half of the gel was electro-blotted to a nitrocellulose membrane (Bio-Rad, Cat #:162-0233) with a Criterion trans-blot electrophoretic transfer cell (Bio-Rad Cat#: 170-4070) for 60 minutes under a constant voltage of 100 volts. The transfer buffer contained 20% methanol and Tris/glycine buffer from Bio-Rad (Cat #: 161-0734). After transfer, the membrane was cut in half and one half was probed with an AAD-12 specific polyclonal rabbit antibody (Lot #: DAS F1197-167-2, 4.3 mg/mL) and the remaining half was probed with an AAD-12 specific monoclonal antibody (mAb 539B181, Lot #: 609.12-2-4, 1.7 mg/mL). A conjugate of goat anti-rabbit IgG (H+L) and horseradish peroxidase (Pierce Chemical, Cat #: 31460) and goat anti-mouse IgG (H+L) and horseradish peroxidase (Bio-Rad, Cat#: 170-6516) were used as the secondary antibodies respectively. GE Healthcare chemiluminescent substrate (Cat #: RPN2132) was used for development and visualization of the immunoreactive protein bands. The membranes were exposed to CL-XPosure detection film (Pierce Cat #: 34091) for various time points and subsequently developed with an All-Pro 100 Plus film developer.

Detection of Post-Translational Glycosylation

The immunoaffinity chromatography-purified, soybean-derived AAD-12 protein (Fraction #3) was concentrated ~15x with a 30-kDa MWCO filter (Millipore Cat #: UFV5BTK00) and mixed 4:1 with 5x LSB. The microbe-derived AAD-12, soybean trypsin inhibitor, bovine serum albumin, and horseradish peroxidase were diluted with 2x Bio-Rad LSB to the approximate concentration of the concentrated plant-derived AAD-12 protein. After mixing the proteins with Laemmli sample buffer, the proteins were 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 and electrophoresed as described above. 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 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 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 desktop scanner to obtain a permanent visual record of the gel.

MALDI-TOF MS and ESI-LC/MS Peptide Mass Fingerprinting and Sequence Analysis of Soybean- and Microbe-Derived AAD-12

Matrix Assisted Laser Desorption Ionization – Time-Of-Flight Mass Spectrometry (MALDI-TOF MS) and Electrospray Ionization-Liquid Chromatography Mass Spectrometry (ESI-LC/MS) analysis of the microbe- and soybean-derived AAD-12 (TSN030732 and combined immunopurified Fractions #3 - 5, respectively) were conducted at the Analytical Sciences Laboratory of The Dow Chemical Company (Midland, MI). The analytical summaries, which contain the methods and results, can be found in Appendix starting on page 36.

Statistical Treatment of Data

No statistical treatments were used during this study.

RESULTS AND DISCUSSION

Lateral Flow Test Strip Assay

The presence of the AAD-12 protein in the pooled leaf tissue (T2 – T4) of event DAS-68416-4 was confirmed using commercially prepared lateral flow test strips from American Bionostica Inc. The strips, capable of detecting between 1-10 ppb of AAD-12, easily discriminated between transgenic and nontransgenic tissue. All of the transgenic extracts were positive for AAD-12 protein and none of the non-transgenic extracts of Maverick contained detectable amounts of immunoreactive protein. This result was also confirmed by the western blot analysis using both polyclonal and monoclonal antibodies specific to the AAD-12 protein.

Purification of the AAD-12 Protein from Transgenic Event DAS-68416-4 Leaf Extracts

Hydrophobic interaction chromatography followed by immunoaffinity chromatography was conducted on an aqueous extract of ~37 grams of lyophilized DAS-68416-4 transgenic leaf (grown from T4 seed). The protein that bound to the respective columns was examined by SDS-PAGE and western blot which demonstrated that the final concentrated fractions contained the AAD-12 protein at an approximate molecular weight of 32 kDa (Figures 1 and 2). Once isolated, the soybean-derived AAD-12 was then compared with the microbe-derived protein.

SDS-PAGE Analysis

In the toxicology-lot preparation of *P. fluorescens*-produced AAD-12 (TSN030732), the major protein band, as visualized on Coomassie stained SDS-PAGE gels, was approximately 32 kDa. As expected, the corresponding soybean-derived AAD-12 protein was identical in size to the microbe-expressed proteins (Figures 1 and 2). Predictably, the plant purified fractions contained a minor amount of non-immunoreactive impurities in addition to the AAD-12 protein. The co-

eluted proteins were likely retained on the column by weak interactions with the column matrix or antibody leaching 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 immunoadsorbents (Williams, *et. al.*, 2006; Kennedy and Barnes, 1983 and Holroyde *et al*., 1976) as well as antibody leaching from the column (Vald´es, *et. al.*, 2007).

Western Blot Analysis

The microbe-derived AAD-12 protein showed a positive signal of the expected size by both polyclonal and monoclonal antibody western blot analysis (Figure 3). This was also observed in the DAS-68416-4 transgenic soybean leaf extract. In the AAD-12 western blot analysis, no immunoreactive proteins were observed in the control Maverick extract and no alternate size proteins (aggregates or degradation products) were seen in the transgenic samples (Figure 3). The monoclonal antibody did detect a small amount of the AAD-12 dimer in the microbe-derived protein. These results add to the evidence that the protein expressed in soybean is not glycosylated or post-translationally modified which would add to the overall protein molecular weight.

Detection of Glycosylation

Detection of carbohydrates, possibly covalently linked to soybean-derived AAD-12 protein, was assessed by the GelCode Glycoprotein Staining Kit from Pierce. The immunoaffinity-purified AAD-12 protein was electrophoresed simultaneously with a set of control and reference protein standards. A glycoprotein, horseradish peroxidase, was loaded as a positive indicator for glycosylation, and non-glycoproteins, microbe-derived AAD-12, soybean trypsin inhibitor and bovine serum albumin, were employed as negative controls. The results showed that both the soybean- and microbe-derived AAD-12 proteins had no detectable covalently linked carbohydrates (Figure 4).

MALDI-TOF MS and ESI LC/MS Tryptic Peptide Mass Fingerprinting

The AAD-12 protein derived from transgenic soybean tissue (event 416) was separated by SDS-PAGE (Figure 5) and the respective bands were excised and subjected to in-gel digestion by trypsin and Asp-N. The resulting peptide mixture was analyzed by MALDI-TOF MS and ESI-LC/MS to determine the peptide sequences (Kuppannan and Karnoup, 2009; see Appendix). The masses of the detected peptides were compared to those deduced based on potential trypsin or Asp-N cleavage sites in the sequence of the soybean-derived AAD-12 protein. Figure 6 illustrates the theoretical cleavage which was generated *in silico* using Protein Analysis Worksheet (PAWS) freeware from Proteometrics LLC. The theoretical and observed amino acid digest (and molecular weights) of the soybean-derived AAD-12 protein is also described in Tables I and II of Kuppannan and Karnoup, 2009 (Appendix). The AAD-12 protein, once denatured, is readily digested by endoproteases and will generate numerous peptide peaks.

In the endoproteinase digest of the transgenic-soybean-derived AAD-12 protein, the peptide sequence coverage was excellent (73.4%). The detected peptide fragments covered nearly the entire protein sequence lacking only three peptide fragments (Figure 7), one near the N-terminus (D^{29} to N^{54}), one in the middle of the protein (D^{96} to R^{142}) and one near the C-terminus (V^{275} to H^{278}). This analysis confirmed the soybean-derived protein amino acid sequence matched that of the microbe-derived AAD-12 (Karnoup and Kuppannan, 2008) protein at both the N- and C-terminus as well as a major portion of the internal sequence. In the MS chromatograms, there were several unidentified peptides detected in the enzyme 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 Asp-N, as well as random breakage of peptides during ionization. Unidentified peptides do not indicate the protein is different from the predicted amino acid sequence. Results of these analyses indicate that the amino acid sequence of the soybean-derived AAD-12 protein was equivalent to the *P. fluorescens*-expressed protein characterized earlier (Karnoup and Kuppannan, 2008 and Embrey and Schafer, 2009).

N- and C-terminal Sequence of the AAD-12 Proteins

The amino acid residues at the N-and C-termini of the soybean-derived AAD-12 protein (immunoaffinity purified from soybean event DAS-68416-4) were measured and compared with the sequence of the microbe-derived protein (Karnoup and Kuppannan, 2008). The protein sequences were measured by MALDI-TOF or ESI-LC/MS. The Asp-N digestion was performed on the soybean-derived AAD-12 protein followed by ESI-LC/MS analysis and the N-terminus was determined to be *N-acetyl*-A[2] A Q T T L Q I T P T G A T L G A T V T G V H L A T L D[27]. The N-terminal AAD-12 protein sequence is expected to be M[1] A Q T T L Q I T P T G A T L G A T V T G V H L A T L D[27]. This result showed that the N-terminal methionine of the protein had been removed (Table 3) and the second amino acid alanine was acetylated. This result is encountered frequently with eukaryotic (plant) expressed proteins since approximately 80-90% of the N-terminal residues are modified in such a way (Wellner, *et al.,* 1990 and Polevoda and Sherman, 2003). This result determined that during or after translation in soybean 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). 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).

The C-terminal sequence of the soybean- and microbe-derived AAD-12 proteins were determined essentially as described above and compared with the expected amino acid sequences (Table 4). The results indicated the measured sequences were identical to the expected sequences, and both the soybean- and microbe-derived AAD-12 proteins were indistinguishable and unaltered at the C-terminus.

CONCLUSIONS

It was demonstrated that the biochemical identity of *P. fluorescens*-produced AAD-12 protein was equivalent to the protein purified from leaf tissue of soybean event DAS-68416-4 (except for the addition of a 42-Dalton acetyl group to the N-terminus of the soybean-derived protein). Both the soybean- and microbe-derived AAD-12 proteins had an apparent molecular weight of ~32 kDa and were immunoreactive to AAD-12 protein-specific antibodies in lateral flow strip and western blot assays. The amino acid sequences were confirmed by enzymatic peptide mass fingerprinting by MALDI-TOF MS and ESI-LC/MS. In addition, the lack of glycosylation of the soybean-derived AAD-12 protein provided additional evidence that the AAD-12 protein produced by *P. fluorescens* and transgenic soybean were essentially 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 soybean- and microbe-derived AAD-12 proteins is located in The Dow Chemical Company Analytical Sciences archives in the 1897 building, Midland, MI 48667.

REFERENCES

Embrey, S. K. and Schafer, B. W. Certificate of Analysis for Test/Reference/Control Substance AAD-12 (TSN030732). Study Number: BIOT09-203009, **2009**, unpublished report of Dow AgroSciences LLC.

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 AAD-12: Batch TSN030732-0002. **2008**, Laboratory Report Code: ML-AL-MD-2008-003833, 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. and Karnoup, A. S. Analytical Characterization of Aryloxyalkanoate Dioxygenase-12 (AAD-12) expressed in Transgenic Soybean from Event DAS-68416-4. **2009**, Laboratory Report Code: ML AL MD-2009-000537, unpublished report of The Dow Chemical Company.

Lin, G., Allen, J., Chew, L., Shields, J.; Chiu, Y.; Greenwalt S.; Xu, P. and Walsh, T. Production, Purification, and Characterization of Recombinant AAD-12 Expressed in *Pseudomonas fluorescens* for Submission on Supporting Regulatory Toxicology and Eco-toxicology Study**.** Study Number: DAI0856, **2007**, unpublished report of Dow AgroSciences LLC.

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

Vald´es, R., Medina, Y., Ferro, W., Reyes, B., Geada, D., Montero, J., Alvarez, T., Leyva A., G´omeza, L., Padilla, S., Pac´ın, L., Figuero, A., Tamayoa, A., Mil´a, L., Aldamad, Y., Moya, G., Reonde, J., del Carmen Abrahantes, M. Comparison of different ligand densities in immunoaffinity chromatography of the plantibody HB-01 coupled to Sepharose CL-4B to purify the rHBsAg. **2007**, *J. Chromatogr.* B852, 1–7

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

Wellner, D., Panneerselvam, C. and Horecker, B. L. Sequencing of peptides and proteins with blocked N-terminal amino acids: N-acetylserine or N-acetylthreonine. **1990**, *Proc. Nati. Acad. Sci.* Vol. 87, pp. 1947-1949

Williams, A. J. K., Norcross, A. J., Chandler, K. A. and Bingley, P. J. Non-specific binding to protein A Sepharose and protein G Sepharose in insulin autoantibody assays may be reduced by pre-treatment with glycine or ethanolamine. **2006**, *J. Immunol. Methods* 314, 170–173

Table 1. Composition of Extraction Buffer for Large-Scale Purification of Soybean-Derived AAD-12

Amount	Ingredient	Catalog Numbers
Two foil packets	Phosphate Buffered Saline with 0.05% Tween[a] 20, pH 7.4	Sigma Cat #: P3563
25.0 mL	0.5 M EDTA	AccuGENE Cat #: 51234
1.0 mL	Tween 20	Fisher Cat #: BP337
20.0 g	PVP[b]	Sigma Cat #: PVP40
To 2000 mL	Milli-Q H_2O	NA
Added to 500 mL of extraction buffer[c] immediately before extraction		
1.5 mL	0.5 M PMSF	Sigma Cat #: P7626
0.35 mL	β-mercaptoethanol	Bio-Rad Cat#: 161-0710

Notes:

a. Tween 20 = Polyoxyethylene (20) sorbitan monolaurate

b. PVP = Polyvinylpyrrolidone

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

Table 2. SDS-PAGE and Western blot Extraction Buffer

Amount	Ingredient	Catalog Number
3 mL	Phosphate Buffered Saline with 0.05% Tween[a] 20, pH 7.4	Sigma Cat #: P3563
30 µL	0.5 M EDTA	AccuGENE Cat #: 51234
5.0 µL	β-mercaptoethanol	Bio-Rad Cat#: 161-0710
50 µL	Plant protease cocktail inhibitor	Sigma Cat #: P9599

Table 3. Summary of N-terminal Sequence Data of AAD-12 Soybean- and Microbe-Derived Proteins

Source	Expected N-terminal Sequence[1]
P. fluorescens	M^{1} A Q T T L Q I T P T G A T L G A T V T G V H L A T L D^{27}
Soybean Event DAS-68416-4	M^{1} A Q T T L Q I T P T G A T L G A T V T G V H L A T L D^{27}

Source	Detected N-terminal Sequence[2]
P. fluorescens	A^{2} Q T T L Q I T P T G A T L G A T V T G V H L A T L D^{27}
Soybean Event DAS-68416-4[3]	$^{N\text{-}Ac}A^{2}$ Q T T L Q I T P T G A T L G A T V T G V H L A T L D^{27}

[1]Expected N-terminal sequence of the first 27 amino acid residues of *P. fluorescens*- and soybean-derived AAD-12.

[2]Detected N-terminal sequences of *P. fluorescens*- and soybean-derived AAD-12.

[3]The MALDI-TOF MS data for the N-terminal peptide revealed that the soybean-derived AAD-12 protein was acetylated (*N-Acetyl*-A Q T T L Q I T P T G A T L G A T V T G V H L A T L D).

Notes:

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

Amino acid residue abbreviations:

A:	alanine	D:	Aspartate	G:	glycine
H:	histidine	I:	isoleucine	L:	leucine
M:	methionine	P:	proline	Q:	glutamine
T:	threonine	V:	valine		

Table 4. Summary of C-terminal Sequence Data of AAD-12 Soybean- and Microbe-Derived Proteins

Source	**Expected C-terminal Sequence[1]**
P. fluorescens	^{281}L A G R P E T E G A A L V^{293}
Soybean Event DAS-68416-4	^{281}L A G R P E T E G A A L V^{293}

Source	**Detected C-terminal Sequence[2]**
P. fluorescens	^{281}L A G R P E T E G A A L V^{293}
Soybean Event DAS-68416-4	^{281}L A G R P E T E G A A L V^{293}

[1]Expected C-terminal sequence of the last 13 amino acid residues of *P. fluorescens*- and soybean-derived AAD-12.

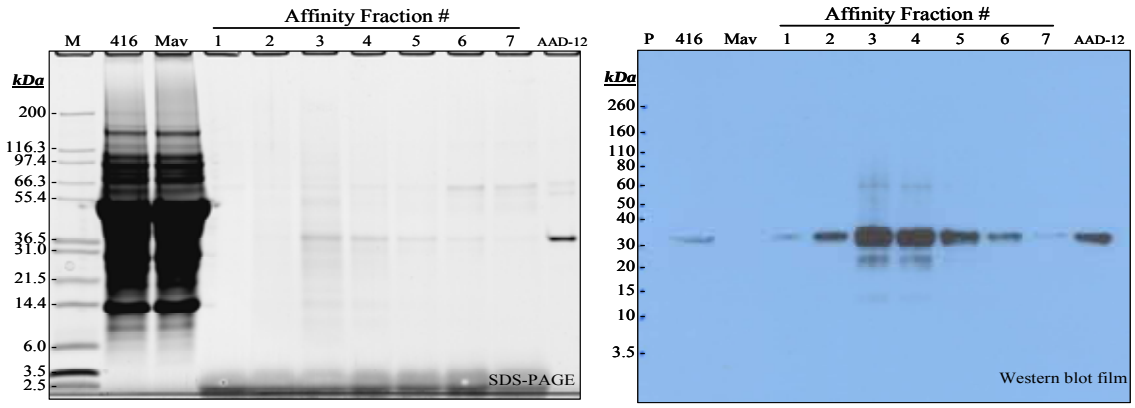
[2]Detected C-terminal sequences of *P. fluorescens*- and soybean-derived AAD-12.

Notes:

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

Amino acid residue abbreviations:

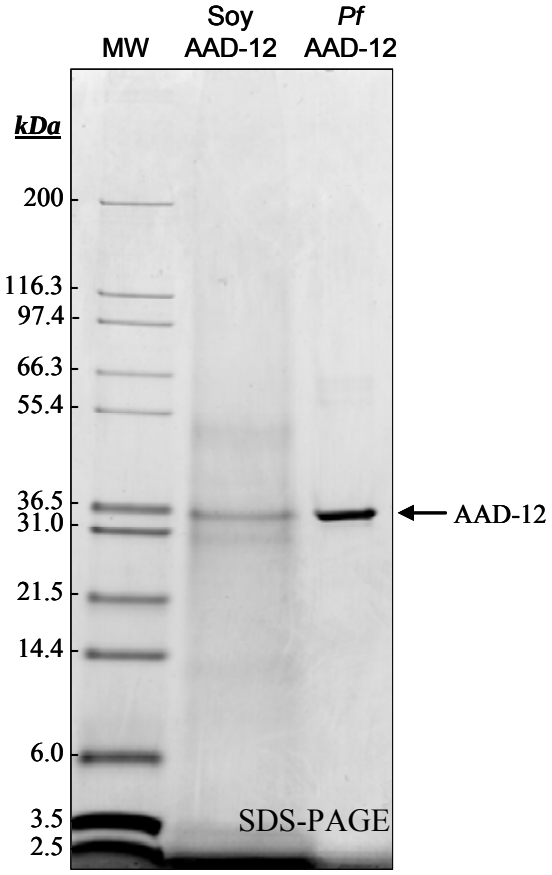
A:	alanine	E:	glutamate	G:	glycine
L:	leucine	P:	proline	R:	arginine
T:	threonine	V:	valine		



Lane	*Sample*	*Amount*
M	Invitrogen Mark12 MW Markers	10 µL
P	Novex Prestained MW Markers	10 µL
416	Transgenic Soybean Event 416	32 µL
Mav	Nontransgenic Soybean Extract	32 µL
1	Polyclonal Affinity Column Frac. #1	32 µL
2	Polyclonal Affinity Column Frac. #2	32 µL
3	Polyclonal Affinity Column Frac. #3	32 µL
4	Polyclonal Affinity Column Frac. #4	32 µL
5	Polyclonal Affinity Column Frac. #5	32 µL
6	Polyclonal Affinity Column Frac. #6	32 µL
7	Polyclonal Affinity Column Frac. #7	32 µL
AAD-12	Microbe-Derived AAD-12	~785 ng (gel)
AAD-12	Microbe-Derived AAD-12	~30 ng (blot)

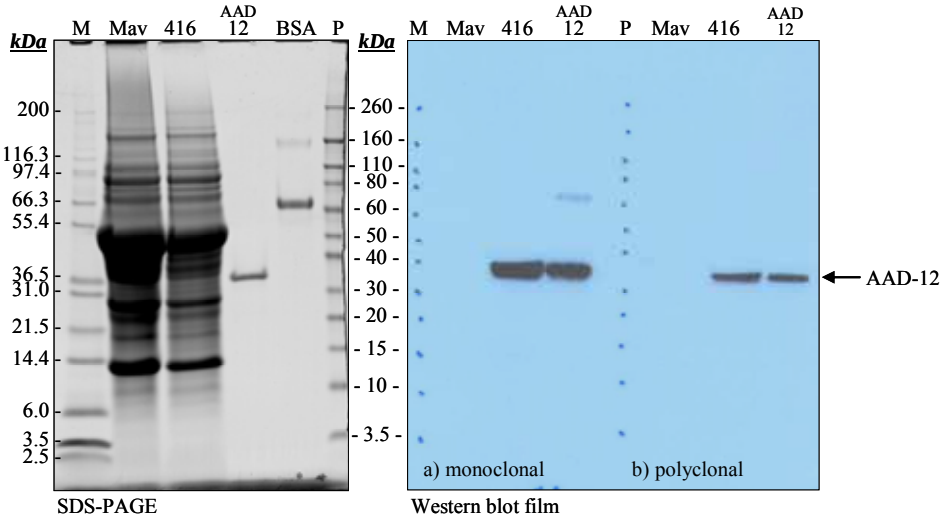
The AAD-12 protein was extracted from lyophilized leaf and stems tissue with 500 mL of PBST based buffer by grinding ~37 grams of lyophilized tissue in a chilled 1 L glass blender (Waring Blender Model #: 31BL41). The tissue was blended on high for 3 minutes and the soluble proteins were harvested by centrifuging the sample for 15 minutes at 3800×g. The pellet was re-extracted as described and the supernatants were combined and filtered through P8 grade filter paper (Fisher Cat#: 09-790-14E). The filtered supernatants were placed in a 1 L glass bottle (at 4 °C) and 25 mL of Phenyl Sepharose (PS) beads (GE Healthcare Cat #:17-0973-10) were added and allowed to incubate on a orbital shaker for 1 hour. After the incubation the PS beads were collected by vacuum on P8 paper and washed extensively with PBST. A final wash with 20 mM Tris, 1 M NaCl, pH 8.0 was performed to increase the conductivity and pH. After the final wash the beads were transferred to a sintered glass funnel and the bound proteins were eluted with ten 45-mL aliquots of Milli-Q water. Each fraction was tested for the presence of the AAD-12 protein by LFS assay and the fractions that were positive were pooled, centrifuged at 6000×g and filtered through a 0.45 µ filter. The conductivity of the final pool was raised to ~34mS/cm by adding NaCl and the PS bound proteins were loaded at 5 mL/min (at 4 ºC) onto an anti-AAD-12 immunoaffinity column that was conjugated with a polyclonal antibody (DAS F1197-167.2, 18.4 mg total) conjugated to CNBr-activated Sepharose 4B beads(GE Healthcare Cat #: 17-0430-01). The column was prepared on 15-Jul-2008, DAS Notebook F1257-120. The affinity column was washed extensively with pre-chilled PBS (Sigma Cat #:P3813) and the bound proteins were eluted with 3.5 M NaSCN (Sigma Cat #: S7757), 50 mM Tris (Sigma Cat #: T3038), pH 8.0 buffer. Eight 5-mL fractions were collected and fractions 1 – 7 were analyzed by SDS-PAGE and western blot. This was accomplished by concentrating 500 µL of each fraction to ~50 µL and washing the sample with 500 µL of PBST and reconcentrating the sample to ~ 30 µL using pre-rinsed Millipore spin concentrators (5 kDa MWCO, Cat #: UFV5BCC00).

Figure 1. SDS-PAGE and Western Blot of Immunopurified AAD-12 Protein from Soybean Event DAS-68416-4



Lane	***Sample***	***Amount***
M	Invitrogen Mark12 MW markers	10 µL
Soy	Soybean-Derived AAD-12 (Frac 3-5)	500 µL
Pf	Microbe-Derived AAD-12	1000 ng

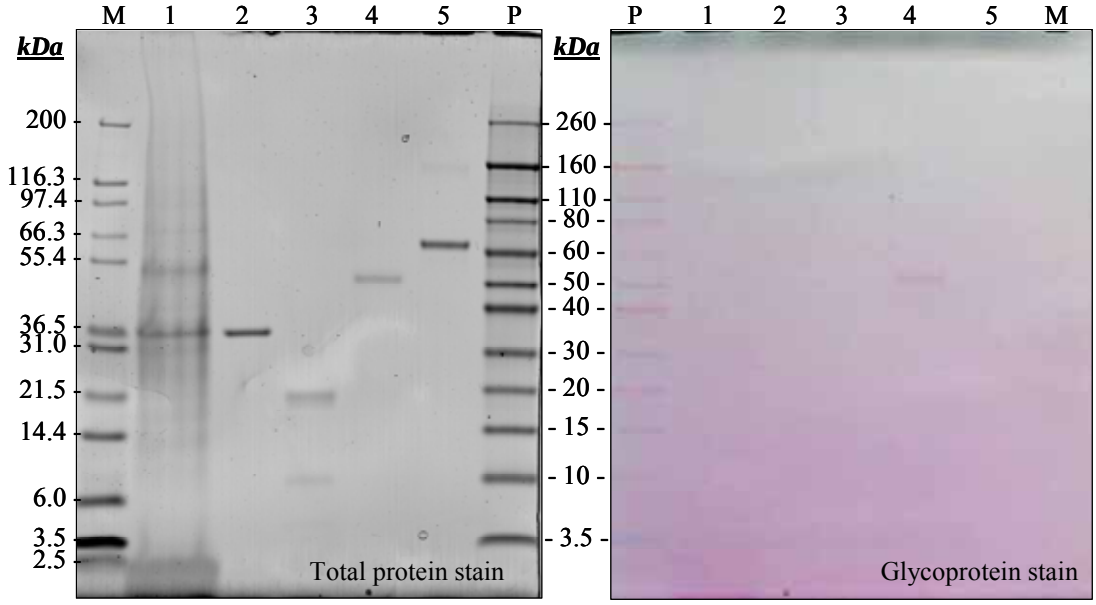
Figure 2. SDS-PAGE of the final pool of immunopurified AAD-12 from soybean event DAS-68416-4



Lane	*Sample*	*Amount*
M	Invitrogen Mark12 MW markers	10 µL
Mav	Nontransgenic Soybean Extract	32 µL
416	Transgenic Soybean Event 416	32 µL
AAD-12	Microbe-Derived AAD-12	~785 ng
BSA	Bovine Serum Albumin (BSA)	~785 ng
P	Novex Prestained MW Markers	10 µL

SDS-PAGE and western blot analysis of the transgenic AAD-12 (from T4 seed) and nontransgenic Maverick soybean extracts was performed with Bio-Rad Criterion gels (Cat #: 345-0123) fitted in a Criterion Cell gel module (Cat #: 165-6001) with MES running buffer (Cat #: 161-0789). Extracts were prepared by Geno-Grinding (Spex, Model #: 2000) ~75mg of tissue with steel ball bearings in a PBST based buffer for 3 minutes in a chilled Teflon microfuge tube holder. The supernatants were clarified by centrifuging the samples for 5 minutes at 20,000×g and 160 µL of each extract was mixed with 40 µL of 5x Laemmli sample buffer (10% SDS, 200 mM Tris pH 6.8, 0.05% bromophenol blue, 20% (w/w) glycerol containing 10% freshly added 2-mercaptoethanol (Bio-Rad, Cat #: 161-0710)) and heated for 5 minutes at ~95 °C. After a brief centrifugation (2 min @ 20,000 ×g), 40 µL of the supernatant was loaded directly on the gel. The reference standard, microbe-derived AAD-12 (TSN030732), and control standard, BSA (Pierce, Cat #: 23208), were diluted with Bio-Rad 2x Laemmli sample buffer (Cat #: 161-0737 containing 5% 2-mercaptoethanol) and processed as described earlier. The electrophoresis was conducted at a constant voltage of 150 V for ~60 minutes. After separation, the gel was cut in half and one half was stained with Pierce GelCode Blue protein stain (Cat #: 24592) and scanned with a densitometer (Molecular Dynamics, Personal Densitometer Si) to obtain a permanent record of the image. The remaining half off the gel was electro-blotted to a nitrocellulose membrane (Bio-Rad, Cat #:162-0233) with a Criterion trans-blot electrophoretic transfer cell (Cat#: 170-4070) for 60 minutes under a constant voltage of 100 volts. The transfer buffer contained 20% methanol and Tris/glycine buffer (Cat #: 161-0734). After transfer, the membrane was cut in half and one half was probed with an AAD-12 specific polyclonal rabbit antibody (Lot #: DAS F1197-167-2, 4.3 mg/mL) and the remaining half was probed with an AAD-12 specific monoclonal antibody (mAb 539B181, Lot #: 609.12-2-4, 1.7 mg/mL). A conjugate of goat anti-rabbit IgG (H+L) and horseradish peroxidase (Pierce Chemical, Cat #: 31460) and goat anti-mouse IgG (H+L) and horseradish peroxidase (Bio-Rad, Cat#: 170-6516) were used as the secondary antibodies respectively. GE Healthcare chemiluminescent substrate (Cat #: RPN2132) was used for development and visualization of the immunoreactive protein bands. The membranes were exposed to CL-XPosure detection film (Pierce Cat #: 34091) for various time points and subsequently developed with an All-Pro 100 Plus film developer.

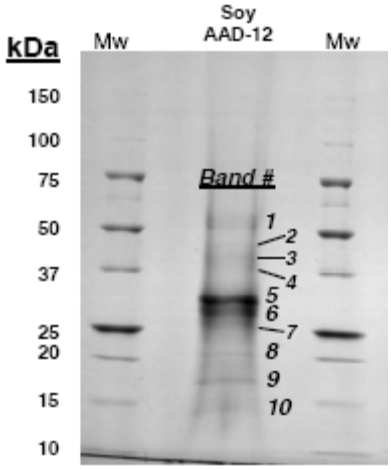
Figure 3. SDS-PAGE and Western Blot of Crude Leaf Extracts from Soybean Event DAS-68416-4



Lane	*Sample*	*Amount*
M	Invitrogen Mark12 MW markers	10 µL
1	Soybean-Derived AAD-12 (Frac 3)	500 µL
2	Microbe-Derived AAD-12	500 ng
3	Soybean Trypsin Inhibitor (STI)	500 ng
4	Horseradish Peroxidase (HRP)	500 ng
5	Bovine Serum Albumin (BSA)	500 ng
P	Novex Prestained MW markers	10 µL

The immunoaffinity chromatography-purified, soybean-derived AAD-12 protein (Fraction #3) was concentrated ~15x with a 30 kDa MWCO filter (Millipore Cat #: UFV5BTK00) and mixed 4:1 with 5x Laemmli sample buffer (LSB). The microbe-derived AAD-12, soybean trypsin inhibitor, bovine serum albumin and horseradish peroxidase were diluted with 2x Bio-Rad LSB to the approximate concentration of the concentrated plant-derived AAD-12 protein. After mixing the proteins with LSB, the proteins were 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 and electrophoresed as described above. 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 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 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 desktop scanner to obtain a permanent visual record of the gel.

Figure 4. SDS-PAGE gel stained with GelCode Blue total protein stain and GelCode Glycoprotein Stain

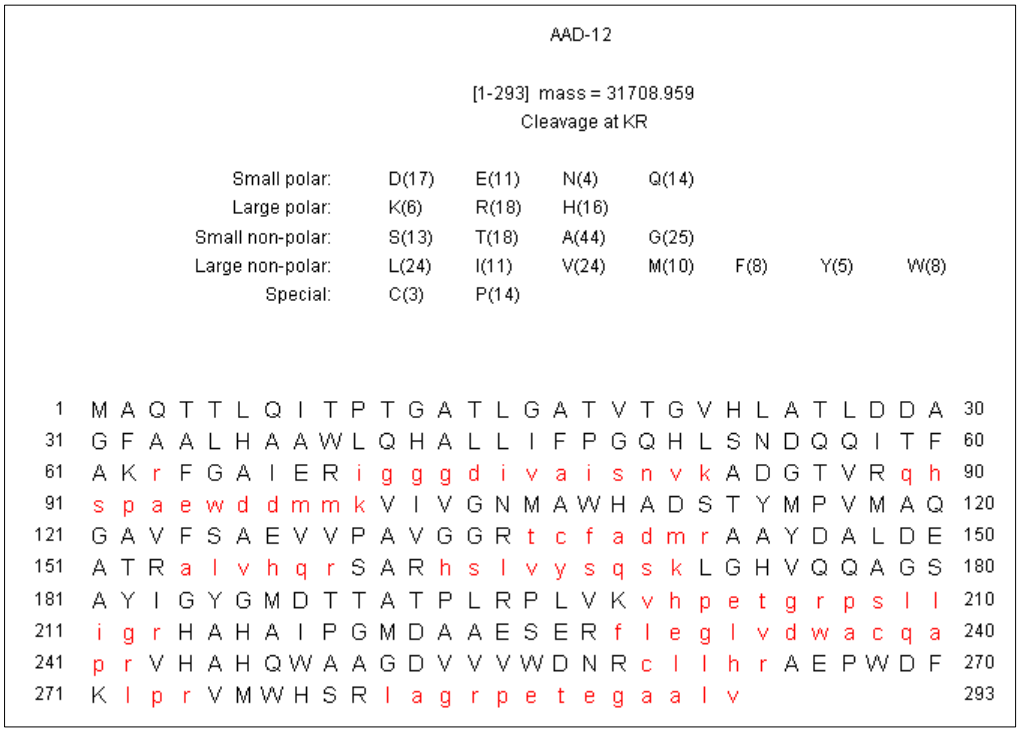


* Data from Kuppannan and Karnoup, 2009

Lane	*Sample*	*Amount*
Mw	Bio-Rad MW markers	10 µL
Soy	Soybean-Derived AAD-12 (pool frac)	~ 5000 µL

A dilute solution of AAD-12 (~ 5 mL) was concentrated using a 30kDa MWCO centrifugal membrane device. The solution was added repeatedly (0.5 mL per spin cycle) to the centrifugal membrane device, and in each cycle the device was centrifuged for 25 min at 10,000 ×g. The filtrate was removed at the end of each spin cycle and stored at approximately -20 °C. For the last spin cycle, 0.5 mL of 25 mM ammonium bicarbonate/0.1M Gu:HCl, pH 8.0, was added to the centrifugal device. The temperature in the centrifuge chamber was maintained at approximately +4 °C. The concentrated sample was transferred into a fresh 0.6-mL siliconized microcentrifuge tube. The centrifugal membrane device was rinsed with 0.1 mL of 50 mM ammonium bicarbonate, pH 7.5, buffer, and the rinse was combined with the concentrated sample. The retained sample was dried to completeness in a centrifugal evaporator. Three hundred (300) microliters of the filtrate was also dried to completeness in a separate tube. Laemmli sample buffer was prepared by adding 50 µL of β-mercaptoethanol to ~950 µL of Bio-Rad Laemmli buffer. The sample buffer was thoroughly mixed with a vortex mixer. The dried AAD-12 sample was dissolved in Laemmli sample buffer (20 µL). After briefly mixing AAD-12 sample in the Laemmli buffer, the microfuge tube was sealed with Parafilm, and placed in a pre-heated heat block set at ~ 95 °C for ~ 10 minutes. The microfuge tube was removed from the heating block and briefly centrifuged. The entire contents of the microfuge tube containing soybean AAD-12 sample was loaded on an SDS-PAGE gel in one lane.

Figure 5. SDS-PAGE gel of the immunopurified, soybean derived AAD-12 (DAS-68416-4) protein used in MALDI-TOF and ESI-LC/MS peptide sequence analysis



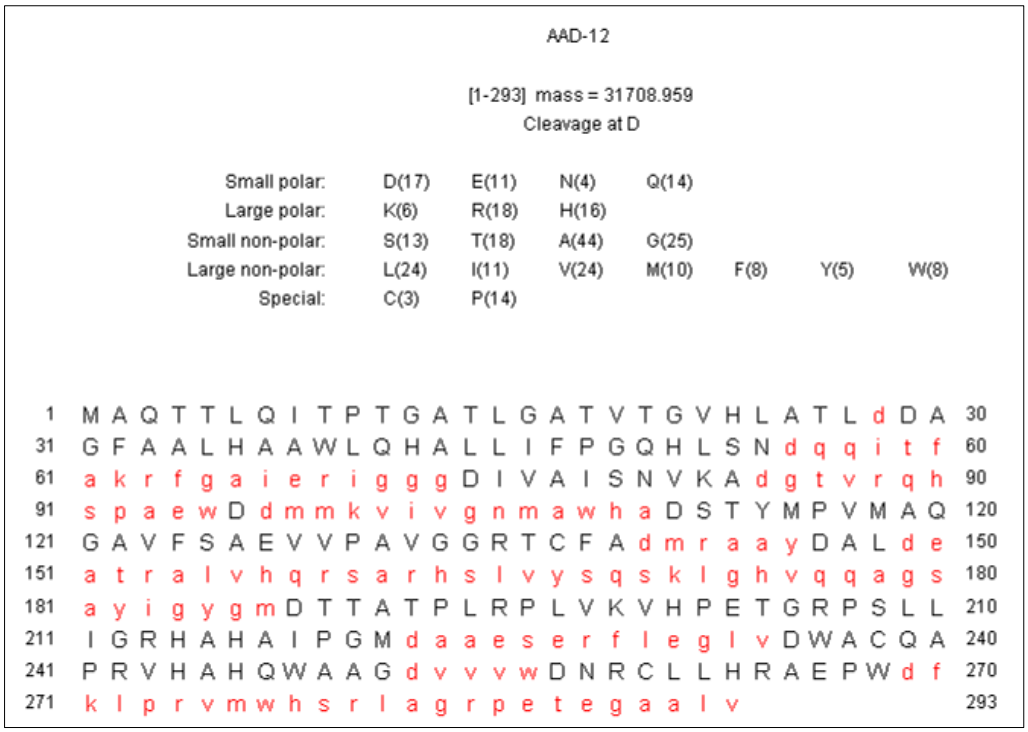
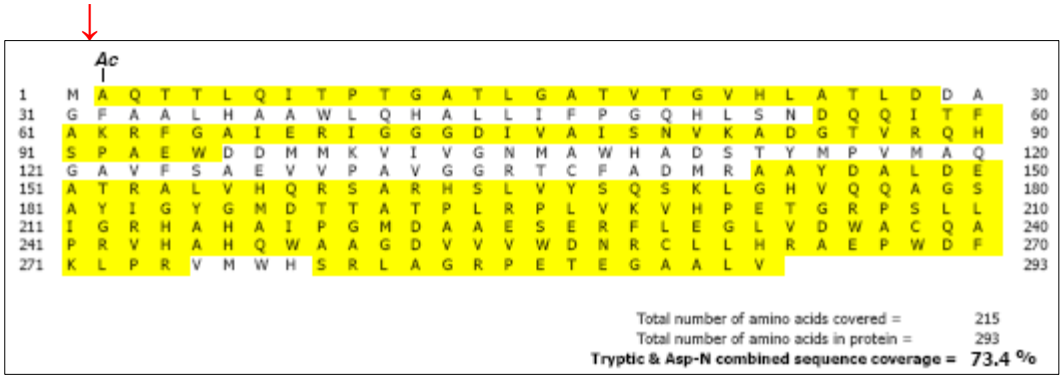


Figure 6. Theoretical cleavage of the AAD-12 protein (with trypsin [top panel] and Asp-N [bottom panel]) generated *in silico* using Protein Analysis Worksheet (PAWS) freeware from Proteometrics LLC



Note: 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 MALDI-TOF MS and ESI-LC/MS. The overall sequence coverage was 73.4%. The red arrow indicates the N-terminal methionine was removed by an aminopeptidase.

A batch of purified aryloxyalkanoate dioxygenase (AAD-12) expressed in transgenic soybean (event DAS-68416-4) was submitted by Barry Schafer of Dow AgroSciences to The Dow Chemical Company for characterization. In conjunction with Dow AgroSciences characterization, the Analytical Sciences Laboratory was requested to provide analytical data on the intact molecular weight, peptide mass fingerprinting, N-terminal and C-terminal amino acid sequences. The sample was subjected to SDS-PAGE and two primary bands (bands 5 and 6) and eight minor bands (bands 1, 2, 3, 4, 7, 8, 9 and 10) from the gel were subjected to further analyses (see Figure 5). Peptide mass fingerprinting of the main component (band 5) was accomplished by in-gel trypsin and endoproteinase Asp-N digests followed by MALDI-TOF-MS and ESI-LC/MS analyses. For the principal component (band 5), peptide mass fingerprinting resulted in 73.4% overall sequence coverage. Peptides from both N- and C-termini were detected by MALDI-TOF-MS. The N-terminal endoproteinase Asp-N fragment was consistent with the N-acetylated sequence: $^{N\text{-}Ac}$A^{2} Q T T L Q I T P T G A T L G A T V T G V H L A T L^{27}. The C-terminal tryptic fragment was consistent with the sequence: L^{281} A G R P E T E G A A L V^{293}. The data for the remaining SDS-PAGE bands (bands 1 – 4 and 6 - 10) can be found in the report by Kuppannan and Karnoup (2009) in the Appendix.

Figure 7. Amino acid sequence of the AAD-12 protein depicting the peptide masses measured by MALDI-TOF and ESI-LC/MS analysis

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

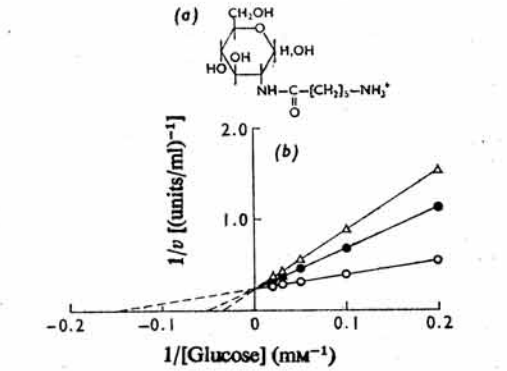


Fig. 1. *Inhibition of glucokinase activity by N-(6-aminohexanoyl)-2-amino-2-deoxy-D-glucopyranose*

(*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.

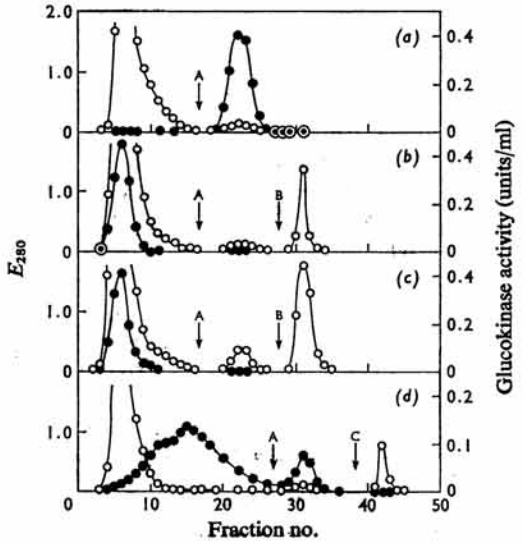


Fig. 2. *Chromatography of rat liver extracts on various Sepharose derivatives*

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 8 μmol/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

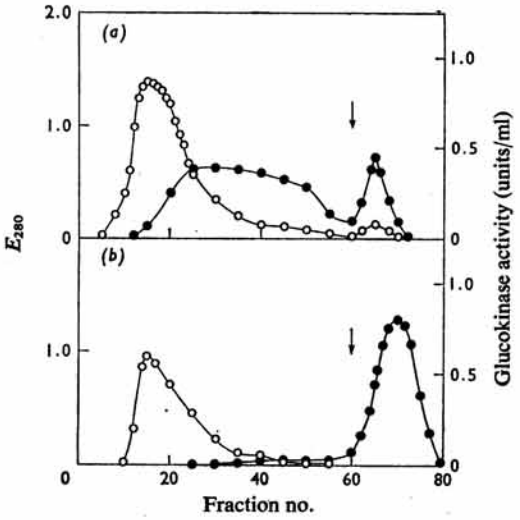


Fig. 3. *Effect of Pronase treatment of the glucosamine derivative affinity columns*

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

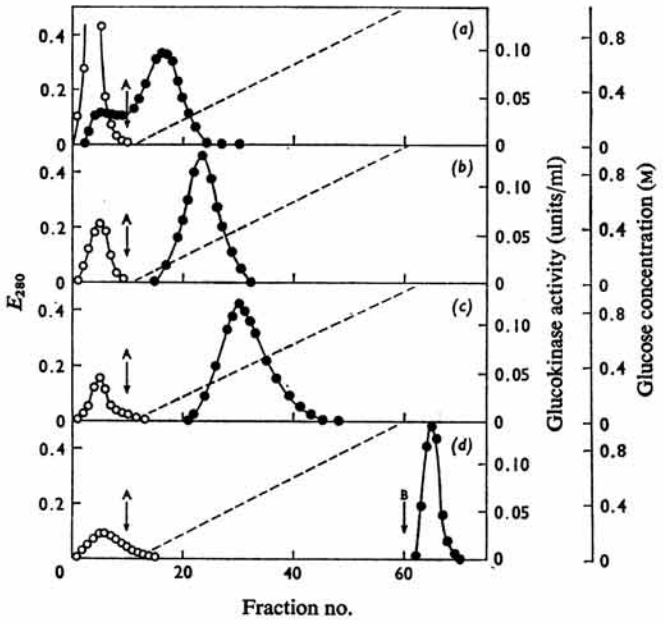


Fig. 4. *Effect of ligand concentrations on the elution of glucokinase by glucose from Sepharose–N-(6-aminohexanoyl)-2-amino-2-deoxy-D-glucopyranose*

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

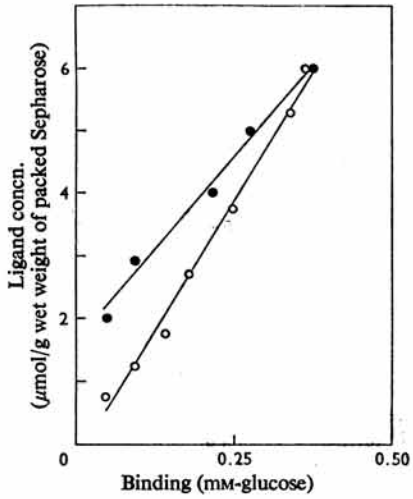


Fig. 5. *Comparison of the elution of glucokinase from affinity columns containing different concentrations of N-(6-aminohexanoyl)-2-amino-2-deoxy-D-glucopyranose prepared by either dilution or direct-linking*

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.

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

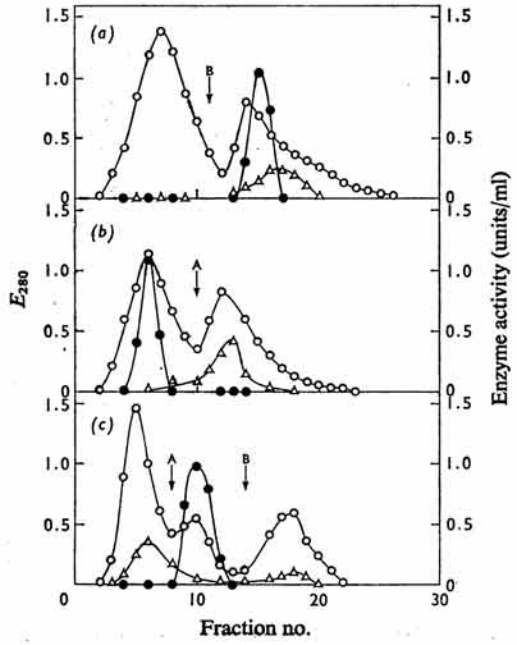


Fig. 6. *Non-specific binding of glucose 6-phosphate dehydrogenase to various Sepharose derivatives*

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.

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-2008-003833
Department	Geographic Location	Date Issued
Analytical Sciences	Midland	10/21/2008
Page Count	Protocol Study Number	Report Status
33		Final

Title

Characterization of AAD-12: Batch TSN030732-002

Author(s): Last Name and Initials (Master Numbers)	Author(s) Signature / Date
Karnoup, Anton (AS) (u369292) Kuppannan, Krishna (K) (u386368)	Karnoup 10/21/2008 10.21.2008
Reviewer Name(s)	Reviewer and/or Supervisor Signature(s)/Date
Young, Scott (SA) (u289561)	10/22/2008

Patent Status

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

A batch of purified recombinant aryloxyalkanoate dioxygenase (AAD-12) (batch TSN030732-002) 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 the intact molecular weight, peptide mass fingerprinting, and N-terminal and C-terminal sequencing.
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 31,599.2 (*des*-Met[1]). An earlier eluting peak, which accounts for ~14 % by peak area, primarily contained the protein with Met residue(s) oxidized: the base peak corresponds to *des*-Met[1] AAD-12 with two oxidized Met (or double oxidation of one Met) with a mass 31,631.2. The experimentally observed mass of *des*-Met[1] AAD-12 (31,599.2 Da) is within 0.004% of the calculated average mass of AAD-12 lacking a methionine, based on the expected amino acid sequence. Peptide mass fingerprinting was accomplished by in-solution trypsin, chymotrypsin, Arg-C, Asp-N, and Glu-C digests followed by ESI-LC/MS analysis. The peptide mass fingerprinting resulted in 100% overall mass coverage for the AAD-12 recombinant protein sample (batch TSN030732-002) (taking into account post-translational removal of Met[1]). The N-terminal and C-terminal sequences for AAD-12 (batch TSN030732-002) were determined by a combination of in-solution digestion with endoproteinases trypsin, Arg-C, and chymotrypsin, followed by tandem MS. The tandem MS data for both the N-terminal and C-terminal peptides confirmed the following sequences, ^{2}AQTTLQITPTGATLGATVTGVHLATLDDAGFAALHAAWLQHALLIFPGQHLSNDQQITFAK^{62} and ^{281}LAGRPETEGAALV^{293}, respectively.

DISTRIBUTION LIST

CRI, 566, Midland, MI	
Young, Scott (SA)	GL Analytical Mol Spec, 1897/ E52, Midland
Schafer, Barry	RSGA, 306 – B2/782, Indianapolis
Clayton, Kathryn (KA)	Regulatory Labs, 306/A2/775, INDIANAPOLIS
O'Connor, Paul (PJ)	Analytical Sciences, 1897 Bldg., Office E21, MIDLAND

TABLE OF CONTENTS

TABLE OF TABLES

TABLE OF FIGURES

INTRODUCTION

A sample of purified recombinant aryloxyalkanoate dioxygenase (AAD-12) (Batch TSN030732-002) 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-2008-003833 [1].

EXPERIMENTAL

Sample Preparation:

A sample of microbial recombinant purified AAD-12 (Batch TSN030732-002) (dark grey lyophilized material; several mg), was submitted for analysis by Barry Schafer (Dow AgroSciences, Indianapolis, IN). Prior to analyses by mass-spectrometry, the sample was prepared as follows:

AAD-12 (Batch TSN030732-002) material (1.16 mg) was resuspended in 1.16 mL of 25 mM ammonium bicarbonate/ 0.1M Gu:HCl, pH8, buffer to a final concentration of 1 mg/mL. An aliquot of the protein solution was stored at +4 ^{o}C prior to ESI/LC/MS analysis, and the rest of the solution was frozen at -20 ^{o}C. Preparation of enzymatic digests for peptide mass fingerprinting and sequencing is described separately below.

ESI/LC-MS for Intact Protein:

Reagents and Standards:

1. Acetonitrile (HPLC grade, 99.9%, Fisher Scientific), Lot no. 082100
2. Trifluoroacetic Acid (Aldrich, 99+%), Lot no. 00339JD
3. Deionized water, 18.2 MΩcm, MilliQ gradient A10, Millipore, freshly drawn
4. Poly-DL-Alanine, Sigma, Catalog no. P9003, Lot no. 97H5912
5. Ribonuclease A (RNAse A), Sigma, Catalog no. R5000, Lot no. 122K1319
6. Bovine serum albumin (BSA), Sigma, Catalog no. A1900, Lot No. 036K7575
7. Lysozyme (from chicken egg white), Sigma, Catalog no. L7651, Lot no. 072K7062
8. Myoglobin (from horse heart), Sigma, Catalog no. M1882

Analytical Procedure:

ESI/LC/MS: The 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 mass spectrometer was calibrated prior to use in the mass range 500-1950 amu using 0.1 mg/mL Poly-DL-Alanine solution (in acetonitrile). A mixture of proteins with known molecular masses (RNAse A, BSA, Lysozyme, Myoglobin; solutions in deionized water at 10 mg/mL were used) was run as a test standard. The following mass spectrometer settings were used.

LC : Acquity UPLC system
Mobile Phase A : 0.01 % trifluoroacetic acid (TFA) in water
Mobile Phase B : 0.01 % trifluoroacetic acid (TFA) in acetonitrile (ACN)
Column : 2.1x150 mm Symmetry 300 C18 3.5 µm 300 Å; S/N: 01283608610502 Part No: 186000188
Flow rate : 100 µL/min
Column temperature : 50 °C
Injection volume : 10 µL
Injection loop : 20 µL
UV detection : 214 nm, 40 pts/sec

Gradient table:

Time, min	Flow rate, mL/min	%A	%B
Initial	0.1	90	10
3	0.1	76	24
19	0.1	44	56
21	0.3	10	90
25	0.3	10	90
26	0.3	90	10
32	0.3	90	10
33	0.1	90	10
35	0.1	90	10

Q-Tof Micro with Micromass lock-spray interface: MS Parameters:

Capillary : 2800 V
Desolvation Gas : 550 L/hr
Desolvation Temperature: 345 °C
Source Temperature : 90 °C
Sample Cone : 15 V
Extraction Cone : 0.9 V
Collision Energy : 10.0 V
MCP : 2350 V
Mode : ESI-TOF-MS +
Scan Range : 500 – 1950 amu
Scan Rate : 0.98 sec/scan

The Micromass-supplied electrospray maximum entropy algorithm (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 with a resolution of 1 Da/channel. The resulting resolution-enhanced spectral peaks were centered and integrated to display the accurate mass for intact molecular mass analysis.

Table: Molecular weight of intact standard proteins determined by ESI/LC/MS:

Protein	Theoretical Mass, Da	Observed Mass, Da	Delta Mass, Da
RNase A	13681.3	13682.4	+ 1.1
Lysozyme	14303.9	14304.5	+ 0.6
BSA	66433.2	66432.3	- 1.2
Myoglobin	16951.5	16951.7	+ 0.2

In-solution Protein Processing and enzymatic digests:

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: 2.5µL serial no. 296447, 2-20µL 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, 1.5mL, 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. 11-418-025-001 (Lot no. 13556621)
6. Roche, Chymotrypsin, cat no. 11-418-467-001 (Lot no. 13998020)
7. Roche, Asp-N, cat no. 11-054-589-001 (Lot no. 13883820)
8. Roche, Arg-C, cat no. 11-370-529-001 (Lot no. 11377132(13))
9. Roche, Glu-C, cat no. 11-047-817-001 (Lot no. 13390420)
10. Fluka, 98% Formic Acid (FA), Lot no. 1255194
11. Fisher, Trifluoroacetic acid (TFA), cat no. 04902-100
12. Milli-Q deionized water

Reagent Solution Preparation:

a. 25 mM Ammonium Bicarbonate buffer: dissolved 98.83 mg NH_4HCO_3 in 50 mL of Milli-Q water; filtered through 0.22 μm sterile syringe filter.
b. 100 mM Tris buffer: dissolved 121.1 mg Tris in 10 mL of Milli-Q water; adjusted pH to 8.11 with HCl; filtered through 0.22 μm sterile syringe filter.
c. Protein dissolution buffer (6M guanidine hydrochloride (Gu:HCl)/ 400 mM ammonium bicarbonate, pH 7.8): to 316 mg of ammonium bicarbonate, 7.5 mL of 8M Gu:HCl solution and 2.5 mL of water were added. pH was adjusted to 7.8 with NaOH. Buffer was filtered through 0.22 µm sterile syringe filter.
d. DTT solution (100 mM; prepared fresh): dissolved 15.4 mg DTT in 1 mL of water.
e. Alkylating reagent (IAA) (200 mM; prepared fresh): dissolved 37 mg IAA in 1 mL of water.
f. Trypsin solution. Step 1: Dissolved 25 µg of dried trypsin in 320 µL of 100 mM Tris buffer immediately prior to digestion procedure. Step 2: Dissolved 50 µg of dried trypsin in 320 µL of 100 mM Tris buffer immediately prior to digestion procedure.
g. Chymotrypsin solution. Step 1: Dissolved 25 µg of dried chymotrypsin in 66 µL of 1 mM HCl immediately prior to digestion procedure. Step 2: Dissolved 50 µg of dried chymotrypsin in 160 µL of 1 mM HCl immediately prior to digestion procedure.
h. Asp-N solution: Dissolved 2 µg of dried Asp-N in 50 µL of Milli-Q water immediately prior to digestion procedure.
i. Glu-C solution. Step 1: Dissolved 50 µg of dried Glu-C in 65 µL of Milli-Q water immediately prior to digestion procedure. Step 2: Dissolved 50 µg of dried Glu-C in 160 µL of Milli-Q water immediately prior to digestion procedure.
j. Arg-C solution. Step 1: Immediately prior to digestion procedure, dissolved 5 µg of dried Arg-C in 30 µL of Milli-Q water, and combine with 50 µL of activation solution (reconstituted in 100 µL of Milli-Q water, as per manufacturer's procedure). Step 2: Immediately prior to digestion procedure,

dissolved 10 µg of dried Arg-C in 30 µL of Milli-Q water, and combine with 50 µL of activation solution (reconstituted in 100 µL of Milli-Q water, as per manufacturer's procedure).

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

a. Five 180-µL aliquots of 1 mg/mL AAD-12 (Batch TSN030732-002) protein solution were dried in the centrifugal evaporator to completeness.
b. Reduction and carboxyamidomethylation (alkylation) of protein: approximately 180-µL of protein dissolution buffer, 6M Gu:HCl/ 0.4M ammonium bicarbonate, pH 7.8, was added to the dry AAD-12 [Batch TSN030732-002] samples, and samples were mixed by pipette action. Twenty 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 40 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. Eighty 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 was approximately 320 µL in each tube.
c. Desalting of the reduced/alkylated protein samples was performed using NAP-5 gravity cartridges (Sephadex G-25) as per the manufacturer's procedure. NAP-5 cartridges were pre-equilibrated with the corresponding digestion buffer (100 mM Tris buffer, pH 8.11, for tryptic and Arg-C digests; 25 mM ammonium bicarbonate, pH 7.8, for chymotryptic, Asp-N, and Glu-C digests), and protein elution was performed with the same buffer (final volume 1-mL for each sample).
d. Tryptic digestion of reduced/alkylated protein: Step 1: 100-µL of trypsin solution (25 µg in 320 µL of 100 mM Tris buffer, pH8.11) was added to the 1-mL of reduced/alkylated protein AAD-12 [Batch TSN030732-002] sample in 100 mM Tris buffer, pH8.11. The digest was incubated for 2 hours at 37 ºC in a thermomixer at 900 rpm. Step 2: 100-µL of trypsin solution (50 µg in 320 µL of 100 mM Tris buffer) was added to the digestion reaction. The digest was incubated for 16 hours at 37 ºC in a thermomixer at 900 rpm. Sample was frozen at -20 ^{o}C until ready for analysis by mass-spectrometry.
e. Arg-C digestion of reduced/alkylated protein: Step 1: 25-µL of Arg-C solution (5 µg in 30 µL of Milli-Q deionized water, combined with 50 µL of activation solution) was added to the 1-mL of reduced/alkylated protein AAD-12 [Batch TSN030732-002] sample in 100 mM Tris buffer, pH8.11. The digest was incubated for 2 hours at 37 ºC in a thermomixer at 900 rpm. Step 2: 25-µL of Arg-C solution (10 µg in 30 µL of Milli-Q deionized water, combined with 50 µL of activation solution) was added to the digestion reaction. The

digest was incubated for 16 hours at 37 ºC in a thermomixer at 900 rpm. Sample was frozen at -20 °C until ready for analysis by mass-spectrometry.

f. Chymotryptic digestion of reduced/alkylated protein: Step 1: 20-µL of chymotrypsin solution (25 µg in 66 µL of 1 mM HCl) was added to the 1-mL of reduced/alkylated protein AAD-12 [Batch TSN030732-002] sample in 25 mM ammonium bicarbonate buffer, pH7.8. The digest was incubated for 2 hours at 22 ºC in a shaker. Step 2: 50-µL of chymotrypsin solution (50 µg in 160 µL of 1 mM HCl) was added to the digestion reaction. The digest was incubated for 16 hours at 22 ºC in a shaker. Sample was frozen at -20 °C until ready for analysis by mass-spectrometry.

g. Asp-N digestion of reduced/alkylated protein: 50-µL of Asp-N solution (2 µg in 50 µL of Milli-Q deionized water) was added to the 1-mL of reduced/alkylated protein AAD-12 [Batch TSN030732-002] sample in 25 mM ammonium bicarbonate buffer, pH7.8. The digest was incubated for 16 hours at 37 ºC in a thermomixer at 900 rpm. Sample was frozen at -20 °C until ready for analysis by mass-spectrometry.

h. Glu-C digestion of reduced/alkylated protein: Step 1: 20-µL of Arg-C solution (50 µg in 65 µL of Milli-Q deionized water) was added to the 1-mL of reduced/alkylated protein AAD-12 [Batch TSN030732-002] sample in 25 mM ammonium bicarbonate buffer, pH7.8. The digest was incubated for 2 hours at 22 ºC in a shaker. Step 2: 50-µL of Glu-C solution (50 µg in 160 µL of Milli-Q deionized water) was added to the digestion reaction. The digest was incubated for 16 hours at 22 ºC in a shaker. Sample was frozen at -20 °C until ready for analysis by mass-spectrometry.

ESI-LC/MS and MS/MS of proteolytic digests

Reagents and Materials:

1. Acetonitrile (Baker analyzed HPLC solvent, JT Baker), Lot no. C10827
2. Milli-Q water
3. 98% Formic Acid (Fluka), Lot no. 1255194
4. Poly-DL-Alanine, Sigma, cat. no. P9003, Lot no. 97H5912
5. Leucine Enkephalin acetate salt, Sigma, cat. no. L-9133, Lot no. 095K5109
6. Waters total recovery HPLC vials, P/N 186000384c, lot no. 0384661180

Analytical Procedure:

ESI-LC/MS: The samples (digests) were dried to completeness in a centrifugal evaporator, resuspended in deionized water (180 µL; to a final concentration of approximately 1 mg/mL) and analyzed by LC/MS. All mass spectra were acquired on a Waters Q-Tof Micro MS system (S/N YA137). The mass spectrometer was calibrated prior to use in the mass range 350 – 1900 amu (MS for peptide mass finger printing) or 80 – 1900 amu (tandem MS) using 0.1 mg/mL Poly-DL-Alanine solution in acetonitrile. 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 :	2.1x150 mm Acquity BEH C18 1.7 µm 135 Å; S/N: 01245523640B05 Part No: 186002353
Flow rate :	100 µL/min
Column temperature :	50 °C
Injection volume :	10 µL
Injection loop :	20 µL
UV detection :	214 nm, 40 pts/sec

Gradient table:

Time, min	Flow rate, mL/min	%A	%B
Initial	0.1	95	5
5	0.1	95	5
63	0.1	60	40
63.5	0.3	60	40
69	0.3	10	90
70	0.3	10	90
71	0.3	95	5
79	0.3	95	5
80	0.1	95	5
85	0.1	95	5

MS :	QTOF-micro mass spectrometer (S/N YA137)
ESI :	Micromass lock-spray electrospray interface
Mode :	+TOFMS

MS Parameters (peptide mass fingerprinting):

Capillary	:	2850 V
Desolvation Gas	:	650 L/hr
Desolvation Temperature	:	300 °C
Source Temperature	:	110 °C
Sample Cone	:	15 V
Extraction Cone	:	0.9 V
Collision Energy	:	10.0 V
MCP	:	2350 V
Mode	:	ESI-TOF-MS +

Scan Range	:	350 – 1900 amu (PMF) or 80 – 1900 amu (tandem MS)
Scan Cycle Time	:	0.98 sec/scan
MS/MS Parameters:		
Capillary	:	2850 V
Desolvation Gas	:	650 L/hr
Desolvation Temperature	:	300 °C
Source Temperature	:	110 °C
Sample Cone	:	15 V
Extraction Cone	:	0.9 V
MCP	:	2350 V
Mode	:	ESI-TOF-MS +
Scan Range	:	80 – 1900 amu
Survey Scan		
Collision Energy	:	10.0 V
Scan Cycle Time	:	0.98 sec/scan
Precursor Selection	:	Included Masses only
Include Window	:	+/- 300 mDa
Include Retention Time	:	240 sec
Peak Detection Window	:	1 Da
MS/MS Scan		
MS to MSMS Switch Criteria	:	Intensity
MS to MSMS Switch Threshold	:	10 counts/sec
MSMS to MS Switch Criteria	:	Intensity falling below threshold
Switchback Threshold	:	3 counts/sec
MSMS Switch After Time	:	12 sec
Scan Cycle Time	:	1.98 sec/scan

Methods:

The samples were injected using a partial 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 using a 0.1 mg/mL (100 ppm) solution (in acetonitrile) of Poly-Alanine at 20 µL/min flow rate. The same instrument parameter file (with the calibration parameters) was used for MS data acquisitions. Data acquisition was performed with cycle times of 1 sec/scan (scan acquisition time: 0.88 sec; interscan delay: 0.1sec) and 2 sec/scan (scan acquisition time: 1.88 sec; interscan delay: 0.1sec) in the MS mode and MSMS mode, respectively. The lock mass data was acquired using 2.5 µM Leucine-Enkephalin peptide solution (0.1 % formic acid in 50 % acetonitrile was used as the solvent) flowing at 3 – 5 µL/min. The lock mass channel was sampled every 7 sec during MS analysis and 10 sec during MS/MS analysis. The reference ion used was the singly charged Leucine-Enkephalin ion at m/z 556.2771. The tandem MS experimental parameters used

in the analyses of N- and C- terminal peptides from both tryptic and chymotryptic digests of AAD-12 (TSN030732-002) are listed above.

Peptide mass fingerprinting of the UPLC-MS data was performed manually. The spectrum of each chromatographic peak was summed, smoothed (SG, 2x3 channels), centroided (4 channels, top 80 %, by height) and m/z error corrected (lock mass channel: 10 scans, m/z 556.2271 ± 0.5 Da). In-source fragmentation observed was used to further confirm the identity of the peptides. For some peptides eluting later in the gradient MaxEnt1 was used. The Micromass-supplied electrospray maximum entropy algorithm (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 with a resolution of 1 Da/channel. The resulting resolution-enhanced spectral peaks were centered and integrated to display the accurate mass.

The spectra from tandem MS experiments were also summed, smoothed (SG, 2x3 channels), centroided (4 channels, top 80 %, by height) and m/z error corrected (lock mass channel: 10 scans, m/z 556.2271 ± 0.5 Da). The fragments were assigned using a theoretical fragmentation ion table generated using either a local copy of Protein Prospector (v 3.2.1) or using Micromass BioLynx software.

RESULTS AND DISCUSSION

In this study, the numbering of the amino acid residues is in accordance with the theoretical sequence of the recombinant AAD-12 protein starting with Met[1] and containing a total of 293 residues (**Figure 1**).

ESI Intact Mass Spectral Characterization:

The purified AAD-12 (Batch TSN030732-002) was processed by directly solubilizing the dry protein material in PBS buffer supplemented with 0.1M Gu:HCl (to ease solubilization and to prevent immediate protein precipitation).

The solubilized proteins were then analyzed by ESI-LC/MS using a Symmetry C18 column for separation. The chromatography for AAD-12 (Batch TSN030732-002) revealed the presence of one major peak at retention time of 12.35 min, and two small satellite peaks at 11.86 min (~14% by LC-UV peak area) and 13.61 min (~4% by LC-UV peak area) (**Figure 2**). Unique features of the corresponding mass spectra were the broad charge distribution and partial resolution of the peaks. Ions related to the monomer form of AAD-12 were observed. The transformed and integrated maximum entropy spectra revealed the presence of a principal mass component at m/z 31,599.2 (LC peak II, 12.35 min). This mass is consistent with the calculated molecular weight for *des*-Met[1] AAD-12 (theor. average mass: 31,598 Da, see **Table I**).

Ions related to the monomer form of AAD-12 (*des*-Met[1]) were observed under the non-reducing conditions of the described ESI-LC/MS experiment. Thus the observed mass of the major sample component is within 0.004% of the theoretical molecular weight of AAD-12 (*des*-Met[1]) (**Table I**). Peak I, which accounts for approximately 14% by peak area, primarily contains oxidized Met residue(s) (peak I, 11.86 min; **Figure 2**). The majority of the oxidized AAD-12 protein contained two oxidation sites (either on two separate Met, or a single doubly-oxidized Met; **Table I**).

Peptide Mass Fingerprinting:

ESI-LC/MS analysis was used to generate peptide coverage maps, N-terminal and C-terminal sequences, and to determine post-translational processing sites. For that purpose, in-solution trypsin, chymotrypsin, Arg-C, Asp-N, and Glu-C digests of reduced and alkylated AAD-12 were generated and analyzed by ESI-LC/MS. **Figures 3** through **7** show the corresponding LC chromatograms of the digested AAD-12 (Batch TSN030732-002). The corresponding mass spectral data with assignments from ESI-LC/MS analyses are presented in **Tables II** through **VI** for the tryptic, Arg-C, chymotryptic, Asp-N, and Glu-C digests, respectively (in that order). The combined sequence coverage is full (100%), given the expected post-translational removal of N-terminal Met[1] (**Figure 1**). Most peptides observed in the proteolytic digests of AAD-12 (Batch TSN030732-002) exhibited in-source fragmentation patterns consistent with their expected theoretical amino acid sequences (**Tables II** through **VI**). The N- and C-terminal peptides were further analyzed by LC tandem MS to confirm their amino acid sequences.

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 are presented in **Tables VII** and **VIII**. Sequence tags were generated from the tryptic fragments with m/z 642.36 (C-terminal peptide, $[M+2H]^{2+}$) and m/z 1063.22 (N-terminal peptide, $[M+6H]^{6+}$), and chymotryptic fragments with m/z 585.82 (C-terminal peptide, $[M+2H]^{2+}$) and m/z 533.29 (N-terminal peptide, $[M+H]^{1+}$). LC tandem MS ion spectra were acquired for each individual peptide at specific retention time obtained in the preceding peptide mass fingerprinting study. Tandem MS experiments were performed with multiple collision energies for each peptide. The tandem MS fragments observed for N-terminal peptides from both tryptic and chymotryptic digests were consistent with the N-terminal peptide sequence, AQTTLQITPTGATLGATVTGVHLATLDDAGFAALHAAWLQHALLIFPGQHLSNDQQITFAK (**Table VIII**). The tandem MS fragments observed for C-terminal peptides from both tryptic and chymotryptic digests were consistent with the C-terminal peptide sequence, LAGRPETEGAALV (**Table VII**).

REFERENCES

1. Raw data packet ML-AL MD-2008-003833

Table I: Molecular weight of intact AAD-12 (Batch TSN030732-002) determined by ESI-LC/MS

Sample Lot # / Peak #	Residues	Theoretical Average Mass	Observed	Comment
TSN030732-002/ I	2-293 (des-Met[1])	31598.0	31594.6	
	2-293 (des-Met[1])		31612.3	One oxidation site
	2-293 (des-Met[1])		31631.2*	Two oxidation sites
	2-293 (des-Met[1])		31645.7	Three oxidation sites
	2-293 (des-Met[1])		31661.7	Four oxidation sites
TSN030732-002/ II	2-293 (des-Met[1])	31598.0	**31599.2***	
	2-293 (des-Met[1])		31615.1	One oxidation site
	4-293 (des-Met[1])	31398.8	31399.1	
			31579.8	unidentified
TSN030732-002/ III	2-293 (des-Met[1])	31598.0	31596.1	
	2-293 (des-Met[1])		31612.1	One oxidation site
			31714.3	unidentified
			31720.9	unidentified
			31733.4	unidentified

* Primary mass component within the LC peak

Table II: Tryptic digest peptide mass fingerprinting of AAD-12 (Batch TSN030732-002).

Peptide	# of missed cleavages	a.a. ##	Sequence	m/z theor.							m/z obs. (ESI-LC/MS)	UPLC elution time, min	In-source fragments observed (matching amino-acid sequence)	Comment
				$(M+H)^{1+}$	$(M+2H)^{2+}$	$(M+3H)^{3+}$	$(M+4H)^{4+}$	$(M+5H)^{5+}$	$(M+6H)^{6+}$	$(M+7H)^{7+}$				
T1	0	2-62	(-) AQTTLQITPTGATLGATVTGVHLATLDDAGFAALHAAWLQHALLIFPGQHLSNDQQITFAK (R)	6366.22	3183.62	2122.75	1592.31	1274.05	1061.88	910.32	1592.09 (4+), 1273.90 (5+), 1061.84 (6+), 910.28 (7+)	63.42		processed N-terminus; no oxidation observed; average masses
				6365.22 [0 charge, ave.]										
T3	0	64-69	(R) FGAIER (I)	692.37	346.69	231.46	173.85	139.28			692.37 (1+)	21.69	545.31 (1+, y5), 488.28 (1+, y4), 417.25 (1+, y3)	
T4	0	70-82	(R) IGGGDIVAISNVK (A)	1242.71	621.86	414.91	311.43	249.35			1242.70 (1+), 621.86 (2+)	34.84	1129.64 (1+, y12), 1072.59 (1+, y11), 843.55 (1+, y8), 796.45 (1+, b9), 631.38 (1+, y6), 612.33 (1+, b7), 560.33 (1+, y5), 513.27 (1+, b6), 447.26 (1+, b4), 400.19 (1+, b5), 360.23 (1+, y3)	
T5	0	83-88	(K) ADGTVR (Q)	618.32	309.66	206.78	155.34	124.47			618.32 (1+)	4.70		
T6	0	89-100	(R) QHSPAEWDDMMK (V)	1474.61	737.81	492.21	369.41	295.73			737.81 (2+), 492.21 (3+)	31.31	954.37 (1+, y7), 825.33 (1+, y6), 650.29 (1+, b6), 539.25 (1+, y5), 524.23 (1+, y4), 521.24 (1+, b5), 504.24 (2+, y8-H2O)	
T6-Met-ox1	0	89-100	(R) QHSPAEWDDMMK (V)	1490.61	745.81	497.54	373.40	298.92			ND	ND		
T6-Met-ox2	0	89-100	(R) QHSPAEWDDMMK (V)	1506.61	753.81	502.87	377.40	302.12			ND	ND		
T7	0	101-135	(K) VIVGNMAWHADSTYMPVMAQGAVFSAEVVPAVGGR (T)	3617.77	1809.39	1206.60	905.20	724.36			1206.60 (3+), 905.22 (4+)	53.51	838.87 (2+, b15), 655.39 (1+, y7), 556.32 (1+, y6)	
T7-Met-ox1	0	101-135	(K) VIVGNMAWHADSTYMPVMAQGAVFSAEVVPAVGGR (T)	3633.77	1817.39	1211.92	909.19	727.55			ND	ND		
T7-Met-ox2	0	101-135	(K) VIVGNMAWHADSTYMPVMAQGAVFSAEVVPAVGGR (T)	3649.77	1825.39	1217.26	913.19	730.75			ND	ND		
T8	0	136-142	(R) TCFADMR (A)	900.37	450.69	300.80	225.85	180.88			900.37 (1+), 450.69 (2+)	24.42	639.29 (1+, y5), 595.22 (1+, b5), 492.22 (1+, y4), 421.19 (1+, y3)	
T8-Met-ox1	0	136-142	(R) TCFADMR (A)	916.37	458.69	306.12	229.84	184.07			ND	ND		
T9	0	143-153	(R) AAYDALDEATR (A)	1195.56	598.28	399.19	299.65	239.92			1195.67 (1+), 598.29 (2+)	26.67	1053.49 (1+, y9), 1021.45 (1+, b10), 920.44 (1+, b9), 890.44 (1+, y8), 849.33 (1+, b8), 775.40 (1+, y7), 720.33 (1+, b7), 704.34 (1+, y6), 591.27 (1+, y5), 527.25 (2+, y9), 492.22 (1+, b5), 476.24 (1+, y4), 421.18 (1+, b4)	
T10	0	154-159	(R) ALVHQR (S)	723.43	362.22	241.81	181.61	145.49			723.43 (1+)	7.82	440.24 (1+, y3), 539.30 (1+, y4)	
T12	0	163-171	(R) HSLVYSQSK (L)	1048.54	524.78	350.19	262.89	210.51			1048.54 (1+), 524.78 (2+)	15.99		
T13	0	172-199	(K) LGHVQQAGSAYIGYGMDTTATPLRPLVK (V)	2944.54	1472.77	982.18	736.89	589.71			1472.64 (2+), 982.19 (3+), 736.90 (4+)	40.74	919.99 (2+, y17), 860.46 (2+, y16), 831.96 (2+, y15), 750.41 (2+, y14), 723.39 (2+, b14), 641.82 (2+, b13), 613.33 (2+, b12), 556.77 (2+, b11), 475.25 (2+, b10)	
T13-Met-ox1	0	172-199	(K) LGHVQQAGSAYIGYGMDTTATPLRPLVK (V)	2960.54	1480.77	987.51	740.89	592.91			ND	ND		
T14	0	200-213	(K) VHPETGRPSLLIGR (H)	1531.87	766.44	511.30	383.72	307.18			766.45 (2+), 511.30 (3+)	29.31	777.41 (1+, b7), 648.37 (2+, y12), 594.34 (2+, b11), 537.80 (2+, b10), 481.26 (2+, b9), 458.32 (1+, y4)	
T15	0	214-228	(R) HAHAIPGMDAAESER (F)	1591.73	796.37	531.25	398.69	319.15			796.38 (2+), 531.26 (3+)	21.95	1062.50 (1+, y10), 965.41 (1+, y9), 908.38 (1+, y8), 777.35 (1+, y7), 684.37 (1+, b7), 662.32 (1+, y6), 591.28 (1+, y5), 530.28 (1+, b5), 520.24 (1+, y4), 465.72 (2+, b9), 417.20 (1+, b4), 391.20 (1+, y3)	
T15-Met-ox1	0	214-228	(R) HAHAIPGMDAAESER (F)	1607.73	804.37	536.58	402.68	322.35			ND	ND		
T16	0	229-242	(R) FLEGLVDWACQAPR (V)	1661.81	831.41	554.61	416.21	333.17			831.41 (2+), 554.61 (3+)	49.66	774.41 (1+, b7), 702.34 (1+, y6), 695.83 (2+, b12), 659.39 (1+, b6), 631.30 (1+, y5), 560.30 (1+, b5), 502.23 (2+, y8), 471.27 (1+, y4), 390.21 (1+, b3)	
T17	0	243-259	(R) VHAHQWAAGDVVVWDNR (C)	1959.96	980.48	653.99	490.75	392.80			980.49 (2+), 654.00 (3+)	34.84	1172.54 (1+, b11), 1073.49 (1+, b10), 788.41 (1+, y6), 788.89 (2+, b14), 689.34 (1+, y5), 685.85 (2+, b13), 636.33 (2+, b12), 590.27 (1+, y4), 586.78 (2+, b11), 572.80 (2+, a11), 537.26 (2+, b10), 404.19 (1+, y3), 394.72 (2+, y6)	
T18	0	260-264	(R) CLLHR (A)	698.38	349.69	233.46	175.35	140.48			698.37 (1+)	17.57	425.27 (1+, y3)	
T19	0	265-271	(R) AEPWDFK (L)	892.42	446.71	298.15	223.86	179.29			892.42 (1+), 446.72 (2+)	34.41	692.34 (1+, y5), 595.30 (1+, y4), 571.26 (1+, a5), 409.21 (1+, y3)	
T20	0	272-274	(K) LPR (V)	385.26	193.13	129.09	97.07	77.86			385.26 (1+)	8.57		
T21	0	275-280	(R) VMWHSR (L)	815.40	408.20	272.47	204.61	163.89			815.39 (1+), 408.21 (2+)	18.07	716.32 (1+, y5), 585.29 (1+, y4), 399.22 (1+, y3)	
T21-Met-ox1	0	275-280	(R) VMWHSR (L)	831.40	416.20	277.80	208.60	167.08			ND	ND		
T22	0	281-293	(R) LAGRPETEGAALV (-)	1283.70	642.35	428.57	321.68	257.55			1283.71 (1+), 642.36 (2+)	30.86	1053.54 (1+, b11), 982.50 (1+, b10), 886.46 (1+, y9), 854.43 (1+, b8), 583.82 (2+, b12), 527.27 (2+, b11), 491.77 (2+, b10), 398.25 (1+, b4)	**C-terminus**
T1-2	1	2-63	(-) AQTTLQITPTGATLGATVTGVHLATLDDAGFAALHAAWLQHALLIFPGQHLSNDQQITFAKR (F)	6522.41	3261.71	2174.81	1631.36	1305.29	1087.91	932.64	1304.92 (5+), 1087.71 (6+), 932.46 (7+)	62.27		Processed N-term.; missed cleavage
T2-3	1	63-69	(K) RFGAIER (I)	848.47	424.74	283.50	212.87	170.50			424.74 (2+)	19.93		missed cleavage
non-spec1		106-135	(N) MAWHADSTYMPVMAQGAVFSAEVVPAVGGR (T)	3135.49	1568.25	1045.83	784.63	627.90			1568.19 (2+), 1045.85 (3+), 784.68 (4+)	51.04	1290.64 (2+, b24), 860.72 (3+, b24), 761.34 (2+, b13), 754.44 (1+, y8), 655.38 (1+, y7), 597.74 (2+, b10), 556.32 (1+, y6),	non-specific cleavage
non-spec1-Met-ox1		106-135	(N) MAWHADSTYMPVMAQGAVFSAEVVPAVGGR (T)	3151.49	1576.25	1051.16	788.62	631.10			ND	ND		
non-spec1-Met-ox2		106-135	(N) MAWHADSTYMPVMAQGAVFSAEVVPAVGGR (T)	3167.49	1584.25	1056.50	792.62	634.30			ND	ND		
non-spec1-Met-ox3		106-135	(N) MAWHADSTYMPVMAQGAVFSAEVVPAVGGR (T)	3183.49	1592.25	1061.83	796.62	637.50			ND	ND		

"Met-ox#" =number of possible oxidations of Met residue(s) in a peptide

Cys residues are reduced and carboxyamidomethylated

Table III: Arg-C digest peptide mass fingerprinting of AAD-12 (Batch TSN030732-002).

Peptide	# of missed cleavages	a.a. ##	Sequence	m/z theor.							m/z obs. (ESI-LC/MS)	UPLC-MS elution time (min)	In-source fragments observed (matching amino-acid sequence)	Comment
				$(M+H)^{1+}$	$(M+2H)^{2+}$	$(M+3H)^{3+}$	$(M+4H)^{4+}$	$(M+5H)^{5+}$	$(M+6H)^{6+}$	$(M+7H)^{7+}$				
				6522.41	3261.71	2174.71	1631.36	1305.29	1087.91	932.64				
1	0	2-63	(-) AQTTLQITPTGATLGATVTGVHLATLDDAGFAALHAAWLQHALLIFPGQHLSNDQQITFAKR (F)	6521.40 [0 charge]							6523.47 (± 0.03) (z=0)	61.96		average mass, transformed
2	0	64- 69	(R) FGAIER (I)	692.37	346.69	231.46	173.85	139.28			692.37 (1+)	21.60	545.30 (1+, y5), 488.28 (1+, y4), 417.25 (1+, y3)	
3	0	70- 88	(R) IGGGDIVAISNVKADGTVR (T)	1842.01	921.51	614.67	461.26	369.21			921.51 (2+), 614.68 (3+)	33.83	836.45 (2+, y17), 807.93 (2+, y16), 779.45 (2+, y15), 721.91 (2+, y14), 683.37 (1+, b8), 665.38 (2+, y13), 580.32 (2+, y11), 523.78 (2+, y10), 513.26 (1+, b6), 432.26 (1+, y4), 400.18 (1+, b5)	
5	0	136-142	(R) TCFADMR (A)	900.37	450.69	300.79	225.85	180.88			900.37 (1+), 450.70 (2+)	24.41	639.29 (1+, y5), 595.22 (1+, b5), 492.23 (1+, y4), 421.18 (1+, y3), 409.16 (1+, b3), 400.17 (2+, y6)	
6	0	143-153	(R) AAYDALDEATR (A)	1195.56	598.28	399.19	299.65	239.92			1195.56 (1+), 598.28 (2+)	26.68	1053.47 (1+, y9), 1021.48 (1+, b10), 920.35 (1+, b9), 890.42 (1+, y8), 849.35 (1+, b8), 775.39 (1+, y7), 720.33 (1+, b7), 704.36 (1+, y6), 591.27 (1+, y5), 527.25 (2+, y9), 492.21 (1+, b5), 476.25 (1+, y4), 421.17 (1+, b4)	
7	0	154-159	(R) ALVHQR (H)	723.43	362.22	241.81	181.61	145.49			723.42 (1+), 362.22 (2+)	6.71	539.30 (1+, y4), 440.24 (1+, y3),	
10	0	214-228	(R) HAHAIPGMDAAESER (F)	1591.73	796.37	531.25	398.69	319.15			796.37 (2+), 531.25 (3+)	21.93	1062.44 (1+, y10), 965.40 (1+, y9), 777.34 (1+, y7), 684.35 (1+, b7), 662.32 (1+, y6), 601.27 (2+, b12), 591.26 (1+, y5), 530.28 (1+, b5), 520.24 (1+, y4), 501.24 (2+, b10), 465.72 (2+, b9), 417.20 (1+, b4), 408.20 (2+, b8), 391.20 (1+, y3)	
11	0	229-242	(R) FLEGLVDWACQAPR (V)	1661.81	831.41	554.61	416.21	333.17			1661.80 (1+), 831.41 (2+), 554.61 (3+)	49.58	1390.66 (1+, b12), 1319.57 (1+, b11), 1191.54 (1+, b10), 1031.52 (1+, b9), 1003.42 (1+, y8), 960.51 (1+, b8), 774.40 (1+, b7), 702.34 (1+, y6), 701.34 (2+, y12), 895.82 (2+, b12), 659.37 (1+, b6), 631.30 (1+, y5), 596.27 (2+, b10), 560.30 (1+, b5), 551.76 (2+, y9), 516.27 (2+, b9), 502.22 (2+, y8), 471.26 (1+, y4), 444.71 (2+, y7), 390.21 (1+, b3)	
12	0	243-259	(R) VHAHQWAAGDVVVWDNR (C)	1959.96	980.48	653.99	490.74	392.80			980.49 (2+), 653.99 (3+)	34.78	1172.57 (1+, b11), 1073.44 (1+, b10), 901.45 (1+, b8), 887.50 (1+, y7), 830.41 (1+, b7), 788.40 (1+, y6), 778.89 (2+, b14), 689.34 (1+, y5), 685.86 (2+, b13), 636.32 (2+, b12), 590.27 (1+, y4), 586.79 (2+, b11), 572.78 (2+, a11), 537.25 (2+, b10), 451.23 (2+, b8), 424.55 (3+, b12), 417.71 (2+, b7), 404.19 (1+, y3), 394.71 (2+, y6)	
13	0	260-264	(R) CLLHR (A)	698.38	349.69	233.46	175.35	140.48			698.37 (1+), 349.69 (2+)	17.33	425.26 (1+, y3)	
14	0	265-274	(R) AEPWDFKLPR (V)	1258.66	629.83	420.22	315.42	252.54			629.83 (2+), 420.22 (3+)	40.60	961.54 (1+, y7), 775.43 (1+, y6), 660.41 (1+, y5), 529.79 (2+, y8), 513.34 (1+, y4), 481.26 (2+, y7), 385.26 (1+, y3), 399.16 (1+, PWD int fr)	
15	0	275-280	(R) VMWHSR (L)	815.40	408.20	272.47	204.60	163.89			815.40 (1+), 408.21 (2+)	17.92	716.34 (1+, y5), 585.29 (1+, y4), 399.21 (1+, b3-H2O), 358.58 (2+, y5)	
16	0	281-293	(R) LAGRPETEGAALV (-)	1283.70	642.35	428.57	321.68	257.54			1283.69 (1+), 642.35 (2+)	30.75	1166.62 (1+, b12), 1053.52 (1+, b11), 982.47 (1+, b10), 911.44 (1+, b9), 886.44 (1+, y9), 854.42 (1+, b8), 624.35 (1+, b6), 583.81 (2+, b12), 569.80 (2+, a12), 574.80 (2+, b12-H2O), 527.27 (2+, b11), 513.27 (2+, a11), 518.26 (2+, b11-H2O), 491.75 (2+, b10), 398.25 (1+, b4)	
8-16	8	160-293	(R) SARHSLVYSQSKLGHVQQAGSAYIGYGMDTTATPLRPLVKVHPETGRPSLLIGRHAHAIPGMDAAESERFLEGLVDWACQAPRVHAHQWAAGDVVVWDNRCLLHRAEPWDFKLPRVMWHSRLAGRPETEGAALV (-)	14946.03 [0 charge]							**14946.03** (± 0.02) (z=0)	52.73		Incomplete C-term. cleavage; average masses, transformed; ave. 17+ charge; traces of Na- & K-salts are observed
non-spec1		196-206	(R) PLVKVHPETGR (P)	1232.71	616.86	411.58	308.93	247.35			616.86 (2+), 411.58 (3+)	19.11	795.44 (1+, y7), 696.32 (1+, y6), 674.43 (1+, b6), 568.34 (2+, y10), 559.28 (1+, y5), 511.79 (2+, y9), 462.25 (2+, y8)	unusual R-P cleavages
non-spec2		163-171	(R) HSLVYSQSK (L)	1048.54	524.78	350.19	262.89	210.51			1048.53 (1+), 524.77 (2+)	15.75	612.29 (1+, y5), 449.23 (1+, y4)	tryptic-like cleavage
non-spec3		172-195	(K) LGHVQQAGSAYIGYGMDTTATPLR (P)	2507.24	1254.12	836.42	627.56	502.25			1254.11 (2+), 836.42 (3+)	36.51	1282.63 (1+, y12), 1225.64 (1+, y11), 1167.05 (2+, b23), 1112.50 (1+, b11), 1061.99 (2+, b21), 1011.48 (2+, b20), 976.00 (2+, b19), 926.42 (2+, b18), 817.43 (2+, b16), 759.43 (1+, y7), 751.90 (2+, b15), 723.37 (2+, b14), 658.37 (1+, y6), 641.81 (2+, b13), 613.31 (2+, b12), 556.79 (2+, b11), 486.30 (1+, y4), 385.26 (1+, y3)	tryptic-like & R-P cleavage; trace of K-salt also observed

Cys residues are reduced and carboxyamidomethylated

Table IV: Chymotrypsin digest peptide mass fingerprinting of AAD-12 (Batch TSN030732-002).

Peptide	# of missed cleavages	a.a. ##	Sequence	m/z theor.					m/z obs. (ESI-LC/MS)	UPLC-MS elution time (min)	In-source fragments observed (matching amino-acid sequence)	Comment
				$(M+H)^{1+}$	$(M+2H)^{2+}$	$(M+3H)^{3+}$	$(M+4H)^{4+}$	$(M+5H)^{5+}$				
Y1	0	2-6	(-) AQTTL (Q)	533.29	267.15	178.44	134.08	107.46	533.29 (1+)	18.93	515.29 (1+, pep - H2O), 402.20 (1+, b4), 384.19 (1+, b4-H2O)	
Y2	0	7-15	(L) QITPTGATL (G)	901.50	451.25	301.17	226.13	181.11	901.50 (1+)	31.96	770.39 (1+, b8), 752.38 (1+, b8-H2O), 669.36 (1+, b7), 651.35 (1+, b7-H2O), 598.30 (1+, b6), 559.31 (1+, y6), 385.70 (2+, b8), 376.70 (2+, b8-H2O)	
Y3	0	16- 24	(L) GATVTGVHL (A)	854.47	427.74	285.50	214.37	171.70	854.48 (1+), 427.74 (2+)	31.73	726.41 (1+, y7), 625.37 (1+, y6), 586.32 (1+, b7), 526.30 (1+, y5), 390.20 (2+, y8-H2O), 363.72 (2+, y7), 354.71 (2+, y7-H2O)	
Y5	0	28- 32	(L) DDAGF (A)	524.20	262.60	175.40	131.81	105.65	524.20 (1+)	19.39	506.19 (1+, pep - H2O), 359.13 (1+, b4)	
Y7	0	36- 39	(L) HAAW (L)	484.23	242.62	162.08	121.81	97.65	484.23 (1+)	18.67	466.22 (1+, pep - H2O)	
Y11	0	46- 52	(L) IFPGQHL (S)	811.45	406.23	271.15	203.62	163.10	811.44 (1+), 406.23 (2+)	31.29	698.37 (1+, y6), 551.30 (1+, y5), 533.28 (1+, y5-H2O), 454.24 (1+, y4), 397.22 (1+, y3), 380.19 (1+, y3-17)	
Y12	0	53- 60	(L) SNDQQITF (A)	952.44	476.72	318.15	238.86	191.29	952.45 (1+)	31.96	787.37 (1+, b7), 769.34 (1+, b7-H2O), 686.31 (1+, b6), 573.22 (1+, b5), 433.25 (2+, y7), 394.18 (2+, b7), 380.22 (2+, a7)	
Y13	0	61- 64	(F) AKRF (G)	521.32	261.16	174.44	131.09	105.07	521.32 (1+)	6.00		
Y24	0	168-172	(Y) SQSKL (G)	562.32	281.66	188.11	141.34	113.27	562.32 (1+)	9.15		
Y27	0	186-194	(Y) GMDTTATPL (R)	906.42	453.72	302.81	227.36	182.09	906.43 (1+)	30.58	718.37 (1+, y7), 700.34 (1+, y7-H2O), 678.28 (1+, b7), 660.27 (1+, b7-H2O), 577.23 (1+, b6), 559.21 (1+, b6-H2O), 506.20 (1+, b5), 416.15 (2+, y8-H2O), 405.15 (1+, b4)	
Y27-Met-ox1	0	186-194	(Y) GMDTTATPL (R)	922.42	461.71	308.14	231.36	185.28	ND	ND		
Y28	0	195-197	(L) RPL (V)	385.26	193.13	129.09	97.07	77.86	385.26 (1+)	11.42		
Y29	0	198-209	(L) VKVHPETGRPSL (L)	1319.74	660.38	440.59	330.69	264.75	660.34 (2+), 440.57 (3+)	20.35		
Y36	0	249-256	(W) AAGDVVVW (D)	816.43	408.72	272.81	204.86	164.09	816.43 (1+)	41.12	698.37 (1+, x6), 657.36 (1+, y6-17), 612.34 (1+, b7), 594.33 (1+, b7-H2O), 584.33 (1+, a7), 513.27 (1+, b6), 495.26 (1+, b6-H2O), 485.27 (1+, a6), 414.21 (1+, b5), 403.25 (1+, y3), 396.20 (1+, b5-H2O), 386.20 (1+, a5)	
Y39	0	263-268	(L) HRAEPW (D)	795.39	398.20	265.80	199.60	159.88	795.40 (1+), 398.20 (2+)	20.76	658.32 (1+, y5), 591.30 (1+, b5), 494.24 (1+, b4), 485.09 (1+, y4-17), 466.26 (1+, a4)	
Y41	0	271-277	(F) KLPRVMW (H)	929.54	465.27	310.52	233.14	186.71	929.53 (1+), 465.27 (2+)	37.65	688.36 (1+, y5)	
Y41-Met-ox1	0	271-277	(F) KLPRVMW (H)	945.54	473.27	315.85	237.14	189.91	ND	ND		
Y42	0	278-281	(W) HSRL (A)	512.29	256.65	171.44	128.83	103.26	512.29 (1+)	5.50		
Y43	0	282-292	(L) AGRPETEGAAL (V)	1071.54	536.28	357.85	268.64	215.11	1071.53 (1+), 536.28 (2+)	21.89	940.44 (1+, b10), 869.42 (1+, b9), 811.38 (1+, x8), 470.73 (2+, b10), 435.22 (2+, b9)	
Y23-24	1	166-172	(L) VYSQSKL (G)	824.45	412.73	275.49	206.87	165.70	824.46 (1+), 412.73 (2+)	26.92		
Y29-30	1	198-210	(L) VKVHPETGRPSLL (I)	1432.83	716.92	478.28	358.96	287.37	1432.80 (1+), 716.93 (2+), 478.28 (3+)	28.13	969.53 (1+, y9), 872.45 (1+, y8), 810.46 (1+, c7), 667.39 (2+, y12), 651.38 (2+, b12), 642.38 (2+, b12-H2O), 603.34 (2+, y11), 553.82 (2+, y10), 485.28 (2+, y9), 464.30 (1+, b4), 434.59 (3+, b12), 425.25 (3+, a12)	
Y37-38	1	257-262	(W) DNRCLL (H)	790.39	395.70	264.13	198.35	158.88	790.39 (1+), 395.71 (2+)	28.80	659.30 (1+, b5), 561.32 (1+, y4), 546.20 (1+, b4)	
Y38-39	1	262-268	(L) LHRAEPW (D)	908.47	454.74	303.50	227.87	182.50	908.48 (1+), 454.75 (2+)	23.51	658.34 (1+, y5), 607.32 (1+, b5), 352.70 (2+, b6)	
Y43-44	1	282-293	(L) AGRPETEGAALV (-)	1170.61	585.81	390.88	293.41	234.93	1170.60 (1+), 585.82 (2+)	27.64	1053.54 (1+, b11), 940.45 (1+, b10), 886.46 (1+, y9), 869.41 (1+, b9), 741.37 (1+, b7), 527.27 (2+, b11), 518.27 (2+, b11-H2O), 513.28 (2+, a11), 470.73 (2+, b10), 435.21 (2+, b9)	
Y12-14	2	53- 95	(L) SNDQQITFAKRFGAIERIGGGDIVAISNVKADGTVRQHSPAEW (D)	4639.38	2320.19	1547.13	1160.60	928.68	1547.31 (3+), 1160.54 (4+), 928.64 (5+)	40.87		
Y32-34	2	230-236	(F) LEGLVDW (A)	831.42	416.22	277.81	208.61	167.09	831.43 (1+)	47.16	672.33 (1+, EGLVDW-28 int.fr.), 627.33 (1+, b6), 512.35 (1+, b5), 484.31 (1+, a5), 419.19 (1+, y3), 413.24 (1+, b4), 385.25 (1+, a4)	
non-spec1		125-135	(F) SAEVVPAVGGR (T)	1041.57	521.29	347.86	261.15	209.12	1041.56 (1+), 521.29 (2+)	25.28	833.51 (1+, y9), 754.45 (1+, y8), 655.39 (1+, y7), 556.32 (1+, y6), 486.26 (1+, b5), 458.26 (1+, a5), 442.25 (2+, y9), 387.19 (1+, b4), 359.20 (1+, a4)	partially-tryptic cleavage
non-spec2		125-136	(F) SAEVVPAVGGRT (C)	1142.62	571.81	381.54	286.41	229.33	1142.57 (1+), 571.82 (2+)	25.98	855.49 (1+, y9), 756.44 (1+, y8), 657.37 (1+, y7), 486.25 (1+, b5), 390.21 (1+, y4), 387.20 (1+, b4)	
non-spec3		65-80	(F) GAIERIGGGDIVAISN (V)	1541.83	771.42	514.61	386.21	309.17	1541.78 (1+), 771.43 (2+)	36.01	705.40 (2+, b15), 661.88 (2+, b14), 605.34 (2+, b13), 404.22 (1+, y4)	
non-spec4		65-78	(F) GAIERIGGGDIVAI (S)	1340.75	670.88	447.59	335.94	268.96	1340.80 (1+), 670.89 (2+)	42.20	1138.64 (1+, b12), 1039.58 (1+, b11), 926.53 (1+, b10), 640.38 (1+, b6), 605.34 (2+, b13), 569.82 (2+, b12), 555.82 (2+, a12), 520.28 (2+, b11)	
non-spec5		269-275	(W) DFKLPRV (M)	874.52	437.76	292.18	219.38	175.71	874.51 (1+)	51.35	583.38 (1+, b5-H2O), 484.31 (1+, y4), 466.30 (1+, y4-H2O), 391.20 (1+, b3), 371.23 (1+, y3), 353.22 (1+, y3-H2O)	

Cys residues are reduced and carboxyamidomethylated

Table V: Asp-N digest peptide mass fingerprinting of AAD-12 (Batch TSN030732-002).

Peptide	# of missed cleavages	a.a. ##	Sequence	m/z theor.					m/z obs. (ESI-LC/MS)	UPLC elution time, min	In-source fragments observed (matching amino-acid sequence)
				$(M+H)^{1+}$	$(M+2H)^{2+}$	$(M+3H)^{3+}$	$(M+4H)^{4+}$	$(M+5H)^{5+}$			
D1	0	2-27	(-) AQTTLQITPTGATLGATVTGVHLATL (D)	2536.40	1268.70	846.14	634.85	508.09	1268.72 (2+), 846.15 (3+)	51.58	1152.63 (2+, b24), 802.47 (3+, b25),
D9	0	111-139	(A) DSTYMPVMAQGAVFSAEVVPAVGGRTCFA (D)	3018.42	1509.71	1006.81	755.36	604.49	1006.84 (3+)	52.53	598.24 (1+, b5), 518.28 (2+, y10)
D10	0	140-145	(A) DMRAAY (D)	726.32	363.67	242.78	182.34	146.07	726.34 (1+)	19.17	545.25 (1+, b5), 480.26 (1+, y4)
D17	0	257-268	(W) DNRCLLHRAEPW (D)	1566.76	783.88	522.92	392.45	314.16	783.88 (2+), 522.93 (3+)	30.73	633.32 (2+, b10)
D18	0	269-293	(W) DFKLPRVMWHSRLAGRPETEGAALV (-)	2836.50	1418.76	946.17	709.88	568.11	946.17 (3+), 709.89 (4+)	40.57	907.13 (3+, b24), 869.47 (3+, b23), 845.77 (3+, b22), 680.87 (4+, b24), 652.35 (4+, b23), 634.61 (4+, b22), 413.25 (1+, y5-17)
D4-5	1	55- 83	(N) DQQITFAKRFGAIERIGGGDIVAISNVKA (D)	3074.67	1537.84	1025.56	769.42	615.74	1025.57 (3+), 769.45 (4+)	44.31	
D14-15	1	222-251	(M) DAAESERFLEGLVDWACQAPRVHAHQWAAG (D)	3377.59	1689.30	1126.53	845.15	676.32	845.19 (4+), 676.56 (5+)	54.79	
non-spec1		23-100	(V) HLATLDDAGFAALHAAWLQHALLIFPGQHLSNDQQITFAKRFGAIERIGGGDIVAISNVK ADGTVRQHSPAEWDDMMK (V)	8492.61 [0 charge]					8492.65 [0 charge; transformed, average mass]	60.95	

Cys residues are reduced and carboxyamidomethylated

Table VI: Glu-C digest peptide mass fingerprinting of AAD-12 (Batch TSN030732-002).

Peptide	# of missed cleavages	a.a. ##	Sequence	m/z theor.					m/z obs. (ESI-LC/MS)	UPLC elution time, min	In-source fragments observed (matching amino-acid sequence)
				$(M+H)^{1+}$	$(M+2H)^{2+}$	$(M+3H)^{3+}$	$(M+4H)^{4+}$	$(M+5H)^{5+}$			
A11	0	112-127	(D) STYMPVMAQGAVFSAE (V)	1688.77	844.89	563.59	422.95	338.56	844.89 (2+)	44.29	1206.57 (1+, y12), 810.35 (1+, b7), 679.34 (1+, b6), 483.20 (1+, b4)
A12	0	128-140	(E) VVPAVGGRTCFAD (M)	1348.67	674.84	450.23	337.92	270.54	674.84 (2+)	29.45	575.78 (2+, y11)
A13-14	1	141-150	(D) MRAAYDALDE (A)	1154.52	577.76	385.51	289.38	231.71	577.76 (2+)	28.09	892.43 (1+, b8)
A17-18	1	204-225	(E) TGRPSLLIGRHAHAIPGMDAAE (S)	2270.18	1135.60	757.40	568.30	454.84	1135.63 (2+), 747.40 (3+), 568.30 (4+)	31.43	790.97 (2+, b15), 734.42 (2+, b14), 708.39 (3+, b21), 698.93 (2+, b13), 684.71 (3+, b20), 675.37 (3+, a20), 661.03 (3+, b19), 630.34 (2+, b12), 622.69 (3+, b18), 579.01 (3+, b17), 531.55 (4+, b21), 513.78 (4+, b20), 496.02 (4+, b19), 489.95 (3+, b14), 466.27 (3+, b13)
A1-4	3	2-68	(-) AQTTLQITPTGATLGATVTGVHLATLDDAGFAALHAAWLQHALLIFPGQHLSNDQQITFA KRFGAIE (R)	7039.99	3520.50	2347.33	1760.75	1408.80	1408.67 (5+), 1174.11 (6+), 1006.52 (7+), 880.81 (8+)	62.58	
				7038.99 [0 charge]					7040.57 [0 charge; transformed]		
A12-14	3	128-150	(E) VVPAVGGRTCFADMRAAYDALDE (A)	2484.17	1242.59	828.73	621.80	497.64	1245.55 (2+), 828.74 (3+), 621.82 (4+)	40.82	1169.07 (2+, b22), 1143.52 (2+, y21), 1111.55 (2+, b21), 1055.00 (2+, b20), 1019.50 (2+, b19), 1009.98 (2+, y18), 961.98 (2+, b18), 880.42 (2+, b17), 779.71 (3+, b22), 762.69 (3+, y21), 741.37 (3+, b21), 725.30 (1+, y6), 703.68 (3+, b20), 562.24 (1+, y5), 466.29 (1+, b5), 447.21 (1+, y4), 376.17 (1+, y3)

Cys residues are reduced and carboxyamidomethylated

Table VII:

(A) Amino acid sequence obtained by in-source fragmentation for the C-terminal peptide ^{281}LAGRPETEGAALV^{293} (m/z 642.35) from Arg-C digest of sample AAD-12 (Batch TSN030732-002) eluting at 30.7 min.

Sequence: LAGRPETEGAALV
Fragment ion masses: monoisotopic
Peptide mass $[M+2H]^{2+}$: 642.35

Ion		*L*	*A*	*G*	*R*	*P*	*E*	*T*	*E*	*G*	*A*	*A*	*L*	*V*
y	theoretical	-	1170.61	1099.57	1042.55	886.45	789.40	660.36	559.31	430.27	373.25	-	-	-
	experimental					**886.41**				**430.26**				
y^{2+}	theoretical		585.81^{+2}	550.29^{+2}	521.78^{+2}	443.73^{+2}	395.20^{+2}	-	-	-	-	-	-	-
	experimental													
b	theoretical	-	-	-	398.25	495.30	624.35	725.39	854.44	911.46	982.50	1053.53	1166.62	-
	experimental				**398.25**		**624.35**		**854.40**	**911.44**	**982.47**	**1053.52**	**1166.61**	
b^{2+}	theoretical	-	-	-	-	-	-	363.20^{+2}	427.72^{+2}	456.23^{+2}	491.75^{+2}	527.27^{+2}	583.81^{+2}	-
	experimental										**491.75**	**527.27**	**583.81**	
a	theoretical	-	-	-	370.26	467.31	596.35	697.40	826.44	883.46	954.50	1025.54	1138.62	-
	experimental					**467.22**								
a^{2+}	theoretical	-	-	-	-	-	-	-	413.73^{+2}	442.24^{+2}	477.75^{+2}	513.27^{+2}	569.81^{+2}	
	experimental											**513.27**	**569.80**	

** only ions with m/z between 350 and 1900 were recorded in this experiment (LC/MS).*

(B) Amino acid sequence obtained by tandem MS for the C-terminal peptide ^{281}LAGRPETEGAALV^{293} (m/z 642.36) from tryptic digest of sample AAD-12 (Batch TSN030732-002) eluting at 29.65 min.

Sequence: LAGRPETEGAALV
Fragment ion masses: monoisotopic
Theoretical monoisotopic peptide mass $[M+2H]^{2+}$: 642.35

Ion		*L*	*A*	*G*	*R*	*P*	*E*	*T*	*E*	*G*	*A*	*A*	*L*	*V*
y	theoretical	-	1170.61	1099.57	1042.55	886.45	789.40	660.36	559.31	430.27	373.25	302.21	231.17	118.09
	experimental												**231.18**	**118.09**
b	theoretical	114.09	185.13	242.15	398.25	495.30	624.35	725.39	854.44	911.46	982.50	1053.53	1166.62	-
	experimental		**185.14**	**242.17**	**398.23**		**624.37**	**725.41**	**854.43**	**911.43**	**982.51**	**1053.55**	**1166.66**	
b^{2+}	theoretical	-	93.07^{+2}	121.58^{+2}	199.63^{+2}	248.16^{+2}	312.68^{+2}	363.20^{+2}	427.72^{+2}	456.23^{+2}	491.75^{+2}	527.27^{+2}	583.81^{+2}	-
	experimental									**456.25**	**491.77**	**527.29**		
a	theoretical	86.10	157.13	214.16	370.26	467.31	596.35	697.40	826.44	883.46	954.50	1025.54	1138.62	-
	experimental	**86.10**	**157.14**	**214.15**	**370.27**		**596.38**		**826.46**		**954.54**	**1025.55**		
a^{2+}	theoretical	-	-	107.58^{+2}	185.63^{+2}	234.16^{+2}	298.68^{+2}	349.20^{+2}	413.73^{+2}	442.24^{+2}	477.75^{+2}	513.27^{+2}	569.81^{+2}	
	experimental										**477.76**			

** only ions with m/z between 80 and 1900 were recorded in this experiment (MS/MS).*

Table VII: (continued)

(C) Amino acid sequence obtained by tandem MS for the C-terminal peptide ^{282}AGRPETEGAALV^{293} (m/z 585.82) from chymotryptic digest of sample AAD-12 (Batch TSN030732-002) eluting at 26.6 min.

Sequence: AGRPETEGAALV
Fragment ion masses: monoisotopic
Theoretical monoisotopic peptide mass $[M+2H]^{2+}$: 585.81

Ion		*A*	*G*	*R*	*P*	*E*	*T*	*E*	*G*	*A*	*A*	*L*	*V*
y	theoretical	-	1099.57	1042.55	886.45	789.40	660.36	559.31	430.27	373.25	302.21	231.17	118.09
	experimental											**231.18**	**118.09**
b	theoretical	-	129.07	285.17	382.22	511.26	612.31	741.35	798.37	869.41	940.45	1053.53	-
	experimental			**285.18**	**382.18**	**511.26**	**612.30**	**741.37**	**798.38**	**869.42**	**940.46**	**1053.56**	
b^{2+}	theoretical	-	-	143.09^{+2}	191.61^{+2}	256.14^{+2}	306.66^{+2}	371.18^{+2}	399.69^{+2}	435.21^{+2}	470.73^{+2}	527.27^{+2}	-
	experimental								**399.68**	**435.20**	**470.72**		
a	theoretical	-	101.07	257.17	354.23	483.27	584.32	713.36	770.38	841.42	912.45	1025.54	-
	experimental		**101.07**	**257.18**		**483.25**		**713.40**	**770.39**	**841.42**	**912.46**	**1025.54**	
a^{2+}	theoretical	-	-	129.09^{+2}	177.62^{+2}	242.14^{+2}	292.66^{+2}	357.18^{+2}	385.69^{+2}	421.21^{+2}	456.73^{+2}	513.27^{+2}	
	experimental									**421.20**			

** only ions with m/z between 80 and 1900 were recorded in this experiment (MS/MS).*

Table VIII:

(A) Amino acid sequence obtained by tandem MS for the N-terminal peptide ^{2}AQTTLQITPTGATLGATVTGVHLATLDDAGFAALHAAWLQHALLIFPGQHLSNDQQITFAK^{62} (m/z 1063.22) from tryptic digest of sample AAD-12 (Batch TSN030732-002) eluting at 59.93 min.

Sequence: AQTTLQITPTGATLGATVTGVHLATLDDAGFAALHAAWLQHALLIFPGQHLSNDQQITFAK
Fragment ion masses: monoisotopic
Theoretical peptide average mass $[M+6H]^{6+}$: 1061.88

Ion		A	Q	T	T	L	Q	I	T	P	T	G	A	T	L	G	A	T	V	T	G	V
y^{5+}	theoretical	-	1259.06	1233.45	1213.24	1193.03	1170.41	1144.80	1122.19	1101.98	1082.56	1062.35	1050.95	1036.74	1016.53	993.92	982.51	968.30	948.09	928.28	908.07	896.67
	experimental									**1102.02**												
y^{4+}	theoretical	-	1573.58	1546.57	1516.30	1491.04	1462.77	1430.76	1402.49	1377.22	1352.96	1327.70	1313.44	1295.68	1270.42	1242.15	1227.90	1210.14	1184.87	1160.11	1134.84	1120.59
	experimental															**1242.18**						
b	theoretical	-	200.10	301.15	402.20	515.28	643.34	756.43	857.47	954.53	1055.57	1112.60	1183.63	1284.68	1397.76	1454.79	1525.82	1626.87	1725.94	1826.99	1884.01	1938.08
	experimental		**200.12**	**301.17**		**515.30**		**756.51**														
b^{2+}	theoretical	-	100.56	151.08	201.60	258.15	322.17	378.72	429.24	477.77	528.29	556.80	592.32	642.84	699.39	727.90	763.42	813.94	863.47	914.00	942.51	992.04
	experimental																					
a	theoretical	-	172.11	273.16	374.20	487.29	615.35	728.43	829.48	926.53	1027.58	1084.60	1155.64	1256.69	1369.77	1426.79	1497.83	1598.88	1697.94	1798.99	1856.01	1955.08
	experimental																					

Ion		H	L	A	T	L	D	D	A	G	F	A	A	L	H	A	A	W	L	Q	H
y^{2+}	theoretical	2190.64	2122.11	2065.57	2030.05	1979.52	1922.98	1865.47	1807.96	1772.44	1743.93	1670.39	1634.87	1599.35	1542.81	1474.28	1438.76	1403.25	1310.21	1253.66	1189.63
	experimental																				
b^{2+}	theoretical	1060.57	1117.11	1152.63	1203.16	1259.70	1317.21	1374.72	1410.24	1438.75	1512.29	1547.81	1583.33	1639.87	1708.40	1743.92	1779.43	1872.47	1929.02	1993.04	2061.57
	experimental																				

Ion		A	L	L	I	F	P	G	Q	H	L	S	N	D	Q	Q	I	T	F	A	K
y	theoretical	2241.20	2170.17	2057.08	1944.00	1830.91	1683.85	1586.79	1529.77	1401.71	1264.65	1151.57	1064.54	950.49	835.47	707.41	579.35	466.27	365.22	218.15	147.11
	experimental																				
y^{2+}	theoretical	1121.11	1085.59	1029.04	972.50	915.96	842.43	793.90	765.39	701.36	632.83	576.29	532.77	475.75	418.24	354.21	290.18	233.64	183.11	109.58	-
	experimental																				

** only ions with m/z between 80 and 1900 were recorded in this experiment (MS/MS).*

Table VIII: (continued)

(B) Amino acid sequence obtained by tandem MS for the N-terminal peptide ^{2}AQTTL^{6} (m/z 533.29) from chymotryptic digest of sample AAD-12 (Batch TSN030732-002) eluting at 18.52 min.

Sequence: AQTTL
Fragment ion masses: monoisotopic
Theoretical peptide average mass $[M+H]^{1+}$: 533.29

Ion		*A*	*Q*	*T*	*T*	*L*
y	theoretical	-	462.26	334.20	233.15	132.10
	experimental				**233.16**	**132.11**
y°	theoretical	-	444.25	316.19	215.14	114.09
	experimental				**215.13**	
b	theoretical	-	200.10	301.15	402.20	-
	experimental		**200.11**	**301.15**		
b°	theoretical	-	182.09	283.14	384.19	-
	experimental			**283.13**	**384.16**	
b*	theoretical	-	183.08	284.12	385.17	-
	experimental		**183.08**	**284.11**		
a	theoretical	-	172.11	273.16	374.20	-
	experimental					

** only ions with m/z between 80 and 1900 were recorded in this experiment (MS/MS).*

Figure 1: Theoretical amino acid sequence of AAD-12 with sequence coverage for AAD-12 (Batch TSN030732-002) by ESI-LC/MS. Cys residues were alkylated with Iodoacetamide. Overall sequence coverage was 100%. Complete removal of N-terminal Met[1] was observed (shown with an arrow). No Asn deamidation was observed.
Theoretical average mass of processed (*des*-Met[1]) AAD-12 is 31769.17 Da (Cys reduced and carboxyamidomethylated) [31598.02 Da for unmodified reduced Cys].

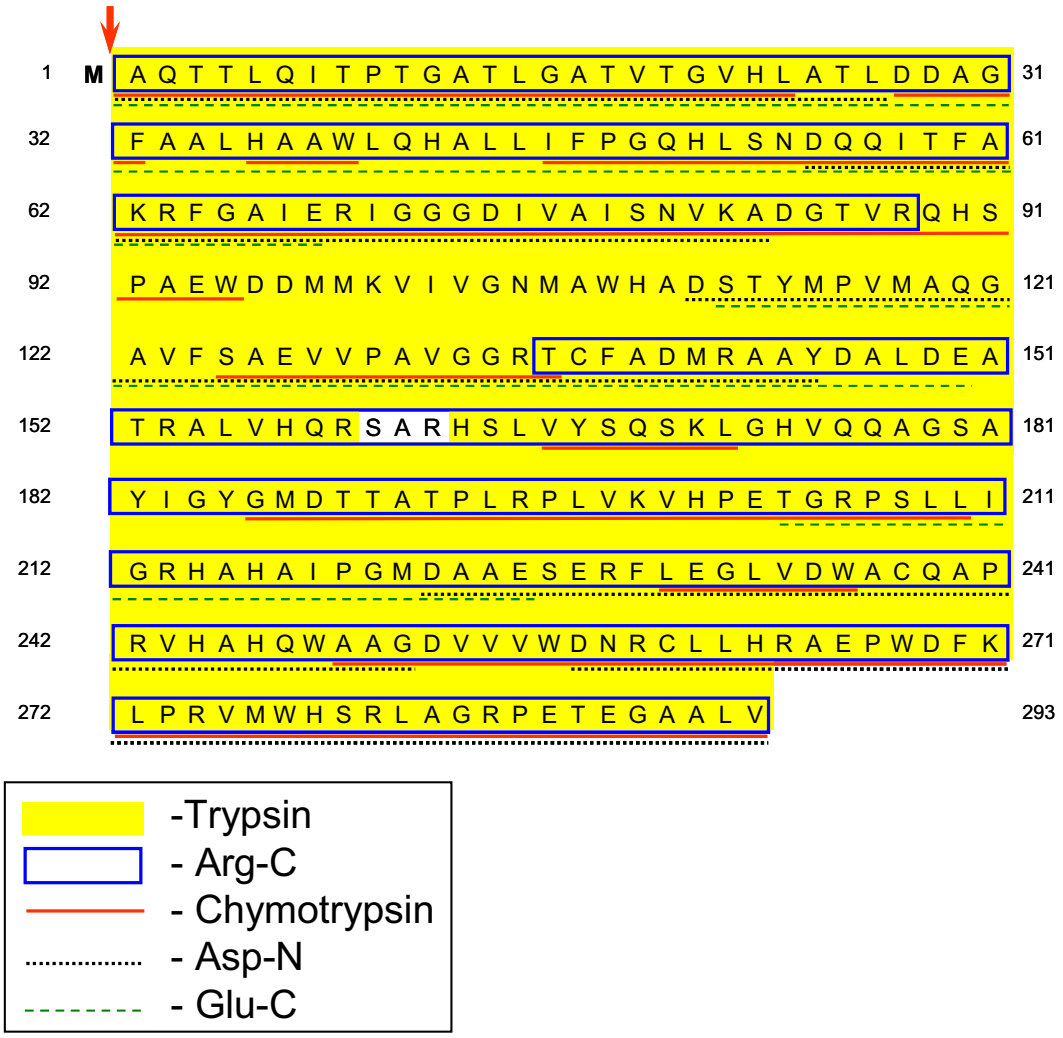


Figure 2: AAD-12 sample (Batch TSN030732-002) prepared in PBS/0.1M Gu:HCl buffer, analyzed by UPLC-MS. (**a**) Chromatograph (UV 214 nm trace); (**b**) multiple charge envelope mass spectrum (top; m/z axis) of component in Peak I, and the corresponding deconvoluted and centered mass spectrum (bottom; true mass axis); (**c**) multiple charge envelope mass spectrum (top; m/z axis) of component in Peak II, and the corresponding deconvoluted and centered mass spectrum (bottom; true mass axis); (**d**) multiple charge envelope mass spectrum (top; m/z axis) of component in Peak III, and the corresponding deconvoluted and centered mass spectrum (bottom; true mass axis). See **Table I** for peak assignments.

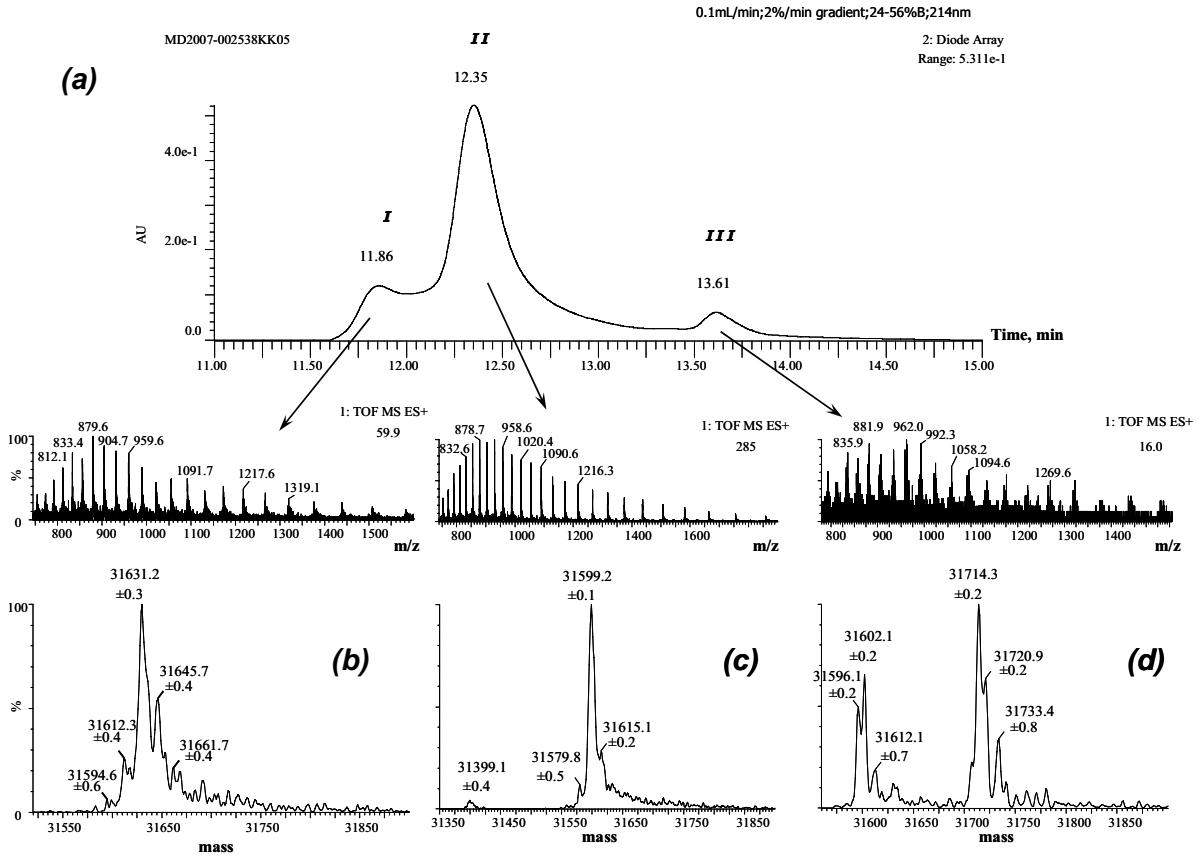


Figure 3: ESI-LC/MS chromatogram (MS TIC) for AAD-12 (Batch TSN030732-002) tryptic digest.

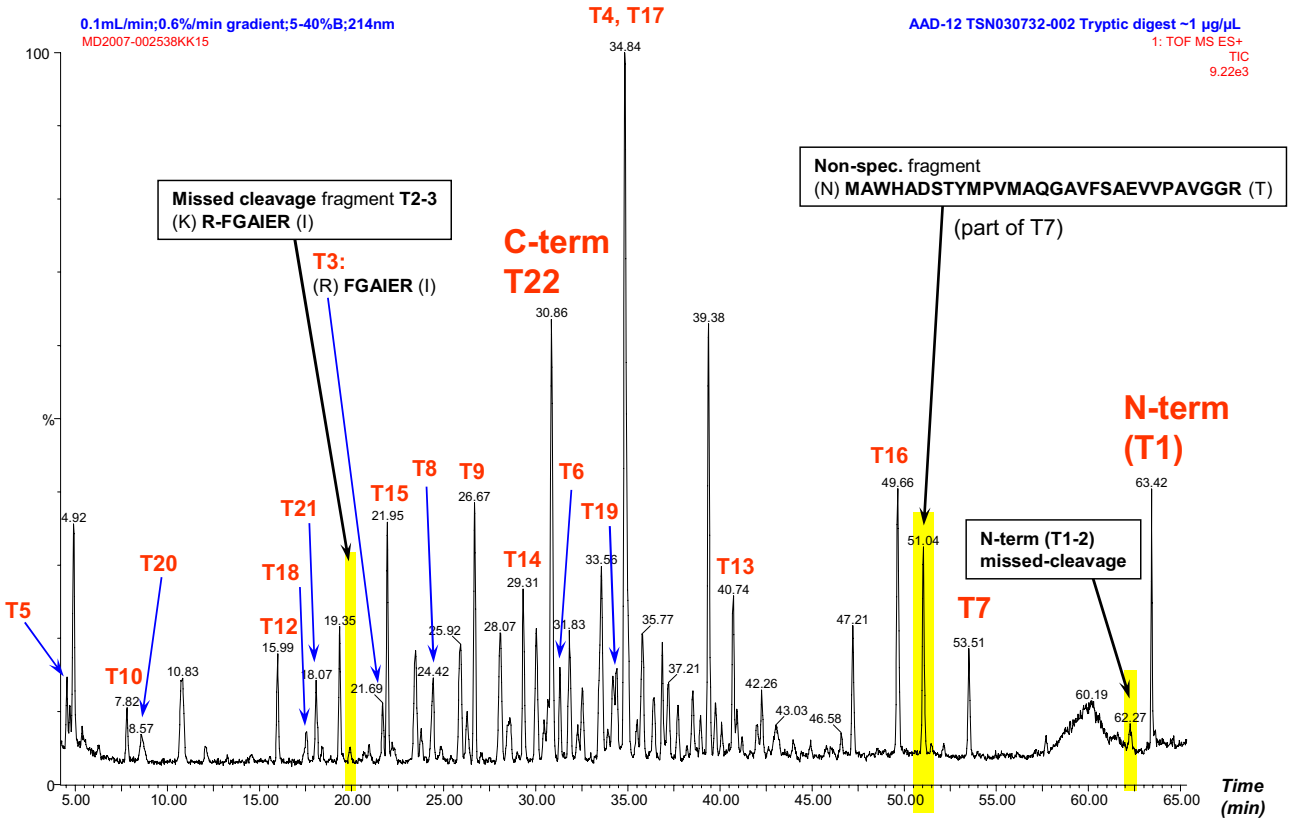


Figure 4: ESI-LC/MS chromatogram (MS TIC) for AAD-12 (Batch TSN030732-002) Arg-C digest.

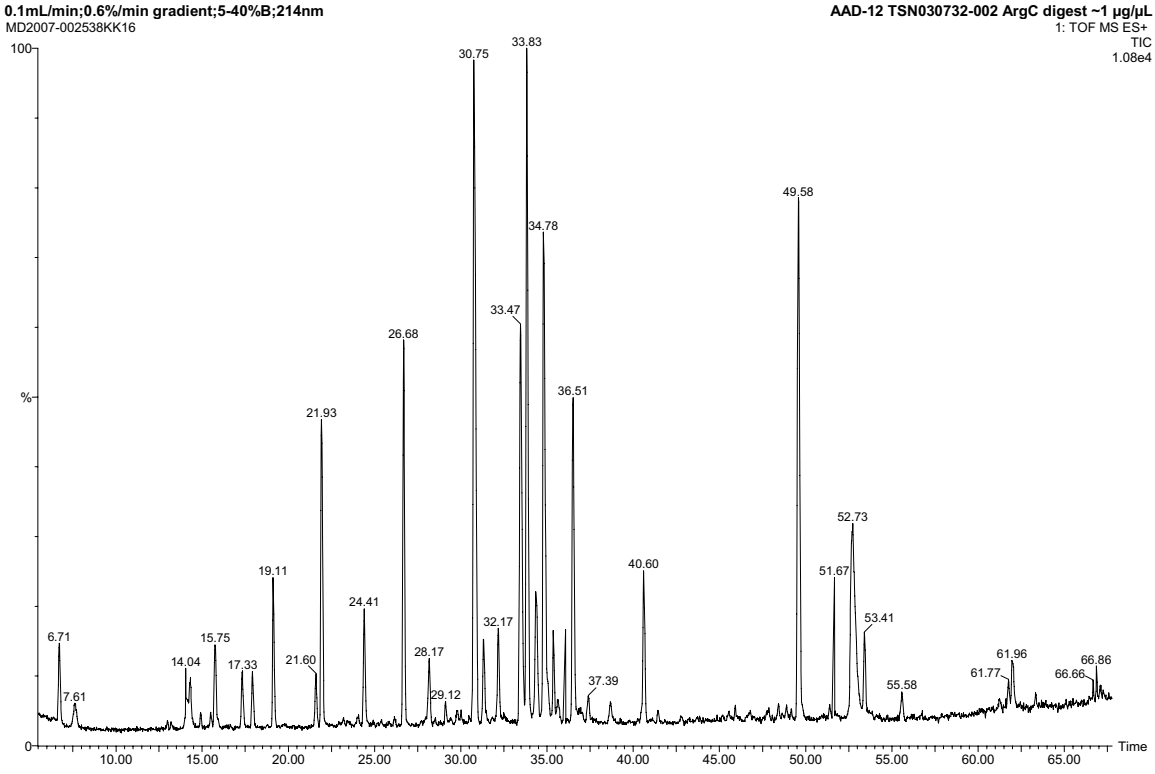


Figure 5: ESI-LC/MS chromatogram (MS TIC) for AAD-12 (Batch TSN030732-002) chymotryptic digest.

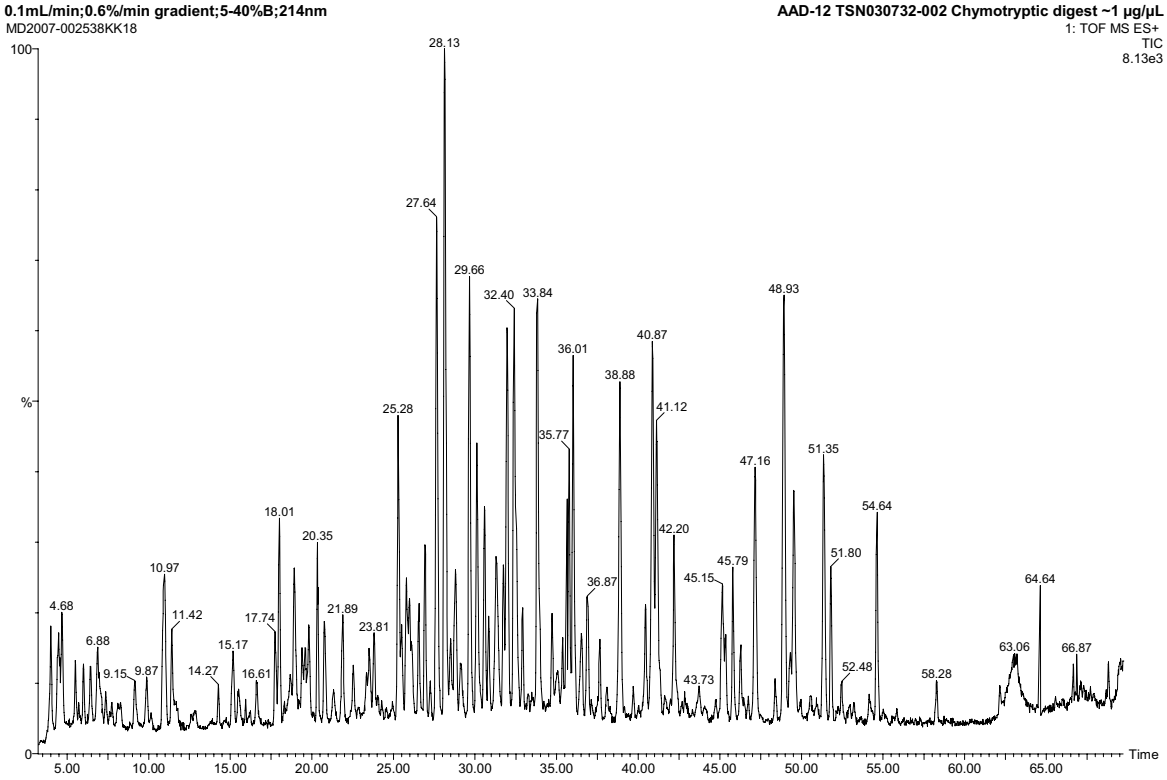


Figure 6: ESI-LC/MS chromatogram (MS TIC) for AAD-12 (Batch TSN030732-002) Asp-N digest.

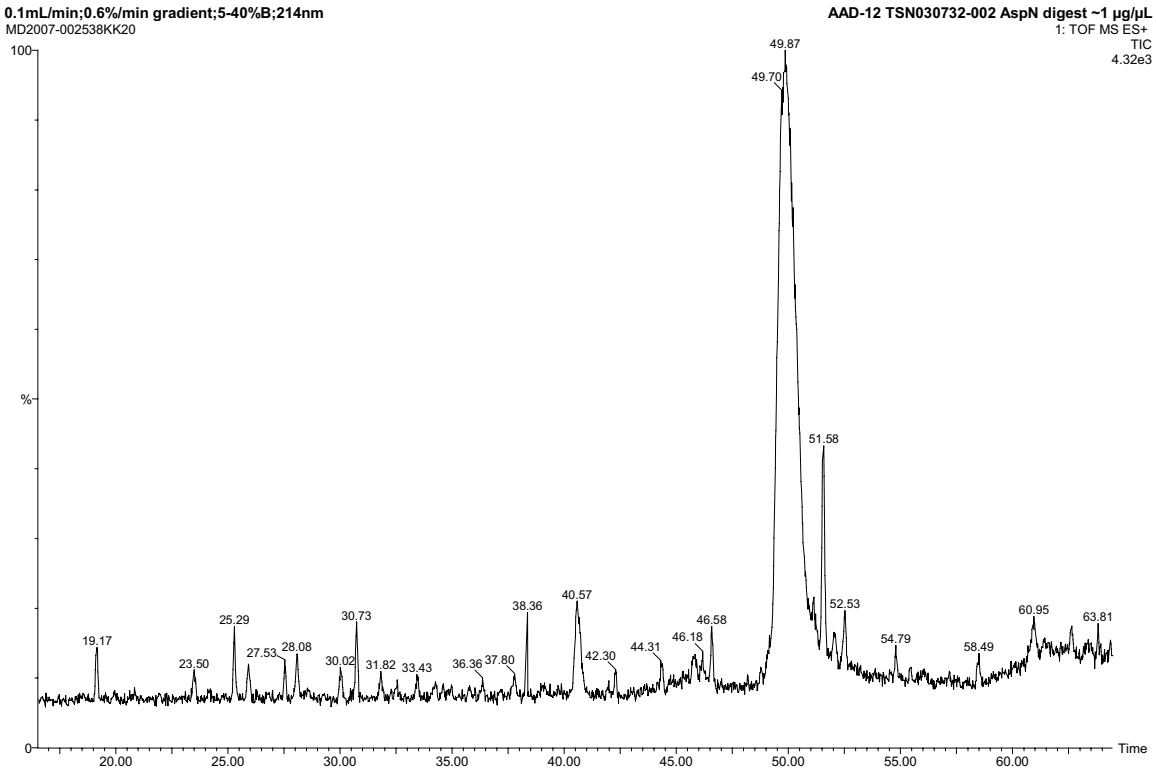


Figure 7: ESI-LC/MS chromatogram (MS TIC) for AAD-12 (Batch TSN030732-002) Glu-C digest.

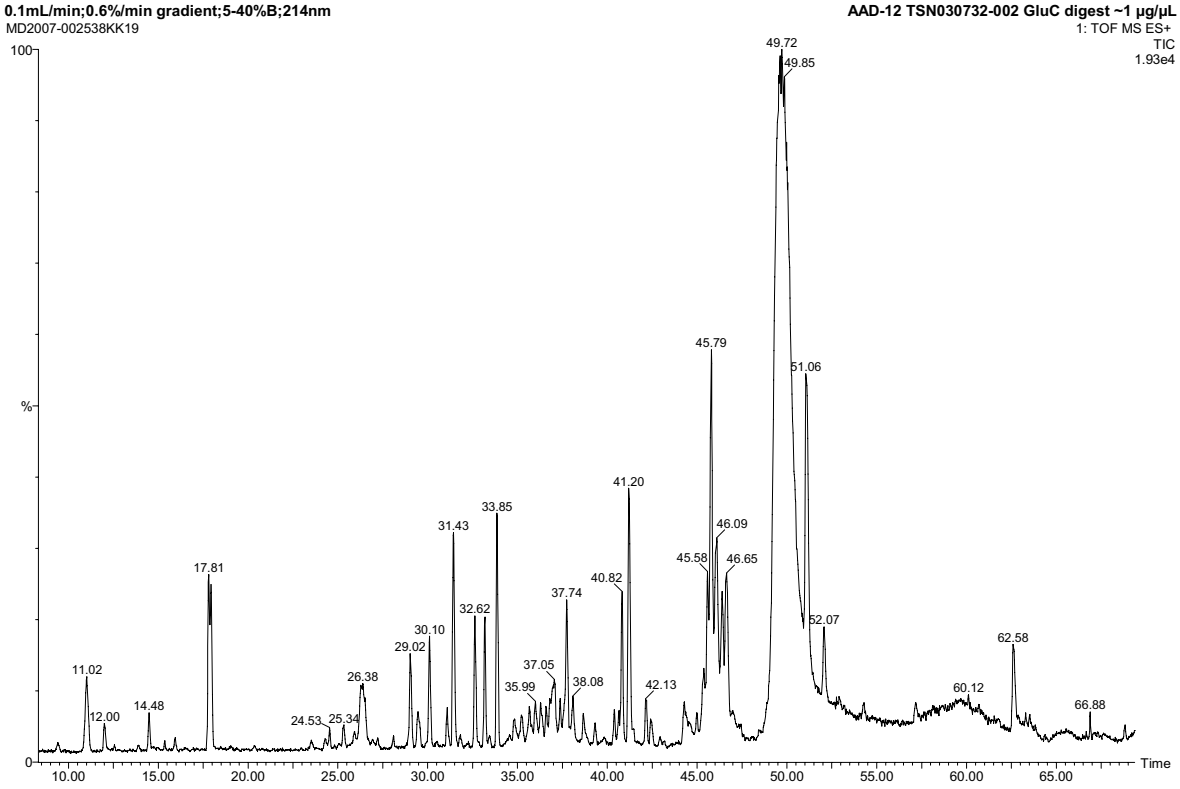
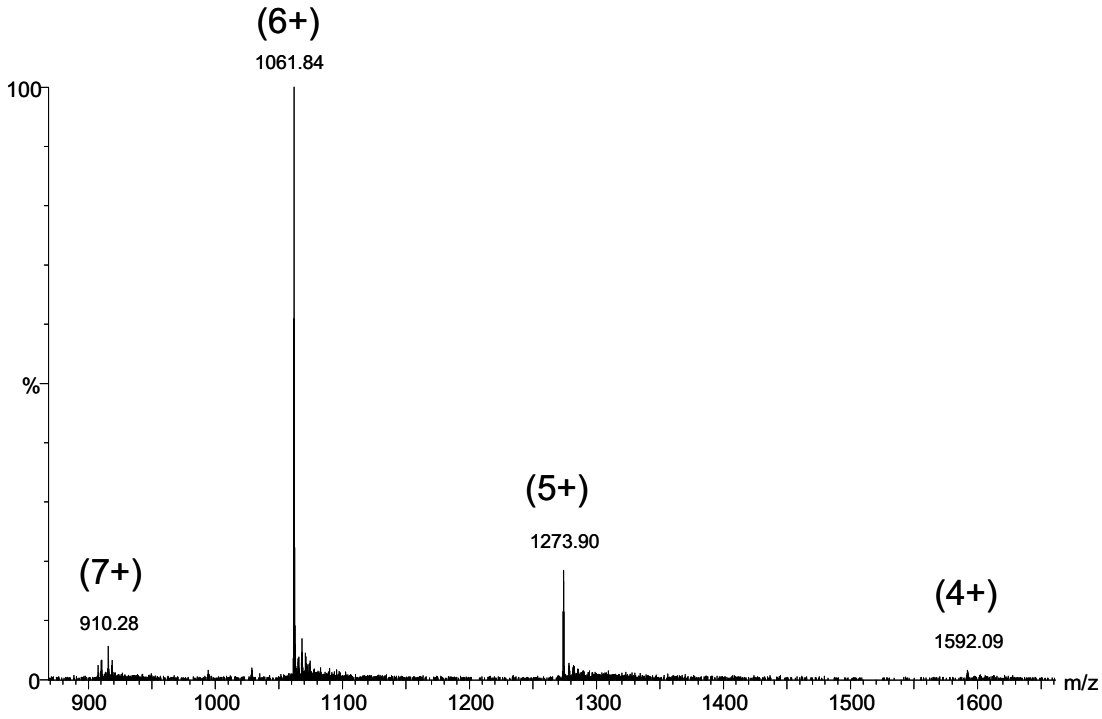


Figure 8: A representative ESI-MS spectrum of N-terminal tryptic fragment T1 (*des*-Met[1]): [2]AQTTLQITPTGATLGATVTGVHLATLDDAGFAALHAAWLQHALLIFPGQHLSNDQQITFAK[62]. Charged states and m/z values are indicated above peaks (see **Figure 3**, and **Tables II** and **VIII** for details). Sample: AAD-12 (Batch TSN030732-002) tryptic digest. Similar spectra of N-terminal fragments were also observed in Arg-C and Glu-C digests (not shown; see **Tables III** and **VI**).



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%).

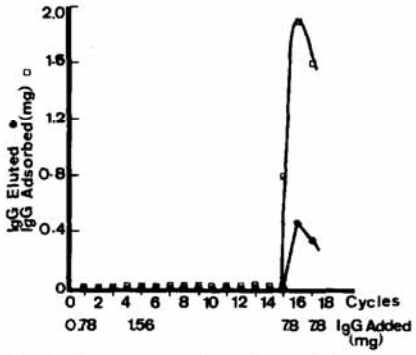


Fig. 1. Non-specific adsorption and elution profiles of human serum IgG on ethanolamine-substituted CNBr-activated Sepharose 4B.

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

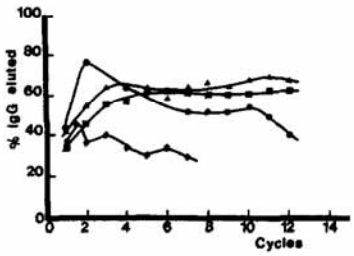
phenomenon of non-specific adsorption, sheep IgG was employed as a model non-specific ligand instead of the specific sheep anti-human IgG ligand. When pooled whole human serum (0.2 ml, 2.32 mg IgG) was loaded onto a 3-ml (10.9 mg sheep IgG/ml gel) control adsorbent column (cyanogen bromide-activated 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



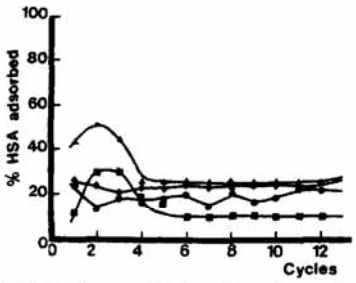


Fig. 2. Composite plots of the percentage human serum IgG eluted on serial adsorption–desorption cycles when increasing amounts of antigen in the form of whole human serum (WHS) were loaded on (CNBr-activated Sepharose 4B–sheep anti-human IgG) immunoadsorbent columns (3 ml gel) of the same potential 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

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-2009-000537
Department	Geographic Location	Date Issued
Analytical Sciences	Midland	June 23, 2009
Page Count	Protocol Study Number	Report Status
28		Final

Title
Analytical Characterization of Aryloxyalkanoate Dioxygenase-12 (AAD-12) expressed in Transgenic Soybean from Event DAS-68416-4

Author(s): Last Name and Initials (Master Numbers)	Author(s) Signature / Date
* Kuppannan, Krishna (K) (U386368) Karnoup, Anton (AS) (U369292)	
Reviewer Name(s)	**Reviewer and/or Supervisor Signature(s)/Date**
Young, Scott (SA) (U289561) Deshmukh, Balasaheb (BK) - U099523	

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

Descriptive Summary and Conclusions

A batch of purified aryloxyalkanoate dioxygenase (AAD-12) expressed in transgenic soybean (event DAS-68416-4) 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 the intact molecular weight, peptide mass fingerprinting, N-terminal and C-terminal amino acid sequences. The sample was subjected to SDS-PAGE. Two primary bands (bands 5 and 6) and eight minor bands (bands 1, 2, 3, 4, 7, 8, 9 and 10) from the gel were subjected to further analyses. The principal component (band 5) apparent molecular weight was estimated to be ~ 30 kDa (theoretical average Mw = 31.6 kDa for *des*-Met[1] protein). Peptide mass fingerprinting of the main components (bands 5 and 6) was accomplished by in-gel trypsin and endoproteinase Asp-N digests followed by MALDI-TOF-MS and ESI-LC/MS analyses. For the principal component (band 5), peptide mass fingerprinting resulted in 73.4% overall sequence coverage. Peptides from both N- and C-termini were detected by MALDI-TOF-MS. The N-terminal endoproteinase Asp-N fragment was consistent with the N-acetylated sequence Ac-A^{2}QTTLQITPTGATLGATVTGVHLATL^{27}. The C-terminal tryptic fragment was consistent with the sequence L^{281}AGRPETEGAALV^{293}. For the lower MW SDS-PAGE band (band 6), based on MALDI-TOF-MS analysis, the N –terminal endoproteinase Asp-N fragment was consistent with a truncated version P^{10}TGATLGATVTGVHLATL^{27}. The C- terminal tryptic peptide from band 6 was consistent with the sequence L^{281}AGRPETEGAALV^{293}. Peptide mass fingerprinting of band 6 resulted in 64.2% overall sequence coverage. Due to very weak MALDI-TOF-MS signal intensity from the minor components (bands 1, 2, 3, 4, 7, 8, 9 and 10), unequivocal identification of the components was not possible.

DISTRIBUTION LIST

Schafer, Barry - U097380
O'Connor, Paul (PJ) - U077484
Deshmukh, Balasaheb (BK) - U099523
Hung, Shao-Ching SC. - u382160
Young, Scott (SA) - u289561

TABLE OF CONTENTS

INTRODUCTION

A sample of purified aryloxyalkanoate dioxygenase (AAD-12) from transgenic soybean (event DAS-68416-4) 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, N-terminal and C-terminal sequences. Original experimental data are stored in the raw data packet ML-AL MD-2009-000537[(1)].

EXPERIMENTAL

Sample Preparation:

A sample of soybean-expressed recombinant affinity-purified AAD-12 from event DAS-68416-4 (approximately 5 mL of AAD-12 solution in 3.5M NaSCN/ 50 mM Tris, pH8.0, buffer), was submitted for analysis by Barry Schafer (Dow AgroSciences, Indianapolis, IN) (Peer-to-peer request 209347). Prior to analyses by mass-spectrometry, the sample was prepared as described in the following section. A brief summary is given below.

Briefly, in order to remove excess buffer components from the AAD-12 material, the original AAD-12 sample (~ 5 mL) was concentrated using a 30kDa MWCO centrifugal membrane device. The concentrated sample was run on an SDS-PAGE gel under denaturing/reducing conditions. The SDS-PAGE protein bands were excised from the gel and processed for in-gel proteolysis. The main protein bands (bands 5 and 6) were subjected to digestion by trypsin and endoproteinase Asp-N (in parallel). All the minor bands (bands 1, 2, 3, 4, 7, 8, 9 and 10) were subjected to digestion by trypsin only. The in-gel digests were analyzed by MALDI-TOF mass-spectrometry and ESI/LC-MS. The sample preparation and analysis procedures are described in detail below.

SDS-PAGE analysis:

Equipment:

a) Biomax-30 Membrane centrifugal device (MWCO 30 kDa), Millipore, cat. no. UFV5 BTK 25
b) Centrifugal evaporator (Centrivap), Labconco, cat. no. 7812013, S/N 051146935 A
c) Criterion Cell, Bio-Rad, cat. no. 165-6001
d) PowerPac 1000, Bio-Rad, cat. no. 165-5054
e) Traceable Digital Thermometer, model NEW 15-078J
f) Heating Block, Fisher brand
g) Centrifuge, Eppendorf , model 5415D, serial no. 5425 17645
h) Adjustable pipettes, Eppendorf: 2-20 µL, and 10-100 µL
i) Aros 160 Orbital Shaker
j) Vortex Genie 2, Fisher, serial no. 2-156856

k) Safe-lock microfuge tubes 1.5 mL, Eppendorf, cat. no. 22 36 332-8
l) Gel loading tips, Bio-Rad, cat. no. 223-9917
m) Parafilm®
n) Graduated cylinder, 1000 mL
o) Fluor-S MultiImager, Bio-Rad, cat. no. 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. Guanidine Hydrochloride (Gu:HCl), Pierce, cat. no. 24110
4. 4-20% Tris-HCl Criterion Precast Gel, Bio-Rad, lot no. C053101C1
5. Tris/Glycine/SDS Running Buffer, Bio-Rad, lot no. 68199A
6. Coomassie Stain Solution, Bio-Rad, lot no. 68198A
7. Destaining solution I (Methanol/Water/Acetic acid 45/45/10)
8. Destaining solution II (Methanol/Water/Acetic acid 5/88/7)
9. Certified Unstained Precision Protein Standards, Bio-Rad, cat. no. 161-0362, lot no. 90310
10. Certified Precision All-Blue Protein Standards, Bio-Rad, cat. no. 161-0373

Analytical Procedure:

The apparent molecular weight of AAD-12 and minor impurities contained in the sample was determined by high-resolution SDS-PAGE analysis. Preparation of reagents, samples, and standards are shown below:

a. Dilute solution of AAD-12 (~ 5 mL) was concentrated using a 30kDa MWCO centrifugal membrane device. The solution was added repeatedly (0.5 mL per spin cycle) to the centrifugal membrane device, and in each cycle the device was centrifuged for 25 min at 10,000 ×g. The filtrate was removed at the end of each spin cycle and stored at approximately -20 °C. For the last spin cycle, 0.5 mL of 25 mM ammonium bicarbonate/ 0.1M Gu:HCl, pH 8.0, was added to the centrifugal device. The temperature in the centrifuge chamber was maintained at approximately +4 °C.
b. The concentrated sample was transferred into a fresh 0.6-mL siliconized microcentrifuge tube. The centrifugal membrane device was rinsed with 0.1 mL of 50 mM ammonium bicarbonate, pH 7.5, buffer, and the rinse was combined with the concentrated sample. The sample was dried to completeness in a centrifugal evaporator. Three hundred (300) microliters of the filtrate was also dried to completeness in a separate tube.
c. Laemmli sample buffer was prepared by adding 50 µL of β-mercaptoethanol to ~950 µL of Bio-Rad Laemmli buffer. The sample buffer was thoroughly mixed with a vortex mixer.
d. The dried AAD-12 sample was dissolved in Laemmli sample buffer (20 µL). After briefly mixing AAD-12 sample in the Laemmli buffer, the microfuge tube was sealed with

Parafilm, and placed in a pre-heated heat block set at approximately 95 °C for approximately 10 minutes. The microfuge tube was removed from the heating block and briefly centrifuged. The entire contents of the microfuge tube containing soybean AAD-12 sample was loaded on an SDS-PAGE gel in one lane.

e. The dried filtrate material was dissolved in Laemmli sample buffer (20 µL) and prepared as described above, in parallel with the soybean AAD-12 material.

f. Known amounts of *Pseudomonas fluorescence (P.f.)*-expressed AAD-12 reference material (batch TSN030732-0002[2)], further referred to as *P.f.* AAD-12-0002) were also prepared as follows: a 1.0 mg/mL fresh stock solution in 10 mM PBST/ 5% glycerol, pH 7.4, buffer was prepared; the stock solution was diluted 4 times to a final concentration 0.25 mg/mL; aliquots (in duplicates) of the *P.f.* AAD-12-0002 were taken (0.25 µg = 1 µL, 0.5 µg = 2 µL, 1 µg = 4 µL, 1.5 µg = 6 µL, 2 µg = 8 µL, 3 µg = 12 µL, 5 µg = 20 µL), dried, each dissolved in 20 µL of the final Laemmli sample buffer, and further prepared for SDS-PAGE as described above.

Instrumental Conditions:

A 4 - 20 % Tris-HCl Criterion pre-cast 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 cassette was inserted into one of the slots in the Criterion electrophoresis tank. The upper buffer chamber of the Criterion gel was filled with 1X Tris/Glycine/SDS (100 mL of 10X Tris/Glycine/SDS mixed with 900 mL of deionized water) premixed running buffer. The remaining running buffer was added to the lower buffer chamber. Approximately 15 µL of certified unstained precision protein standards was loaded into far right and far left wells with a pipette using gel-loading tips. After applying the samples and reference materials to the gels, the lid was placed on the tank, the electrical leads were plugged into the power supply, and the power was turned on. Constant current of 30 mA was applied to the Criterion cell for 1 hour, followed by 45 mA constant current for 40 min, 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 and placed in the gel cassette tray with ~ 40-mL of Coomassie Stain Solution. The gel and staining solution were incubated at room temperature for approximately 30 minutes on an orbital shaker set at 35 rpm. The Coomassie Stain Solution was discarded and replaced with approximately 40-mL of Destaining Solution I and placed on an orbital shaker at 35 rpm for approximately 30 minutes. The Destaining Solution I was replaced with Destaining Solution II and the gel was destained for approximately 16 hours.

Methods for determining protein molecular weight: After destaining of the SDS-PAGE gel was complete, gel images were acquired using 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 tools for determining molecular weight. The molecular weight value was determined relative to the certified Bio-Rad protein

standards defined for the gel, and the band's position in the lane. To estimate of the quantity of the soybean AAD-12 material on the SDS-PAGE gel, a calibration curve was obtained from the measured band densities and the known µg amounts for the *P.f.* AAD-12-0002 reference material on the same gel. The calibration curve was applied to the lane containing soybean AAD-12 material to estimate the amount of soybean AAD-12 on the gel.

In-gel protein processing and enzymatic digests:

Equipment:

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

Reagents and Standards:

1. Acetonitrile, Fisher, cat. no. A998-1, Lot no. 080757
2. Ammonium bicarbonate, Sigma, cat. no. A-6141
3. Iodoacetamide (IAA), Sigma, cat. no. I-1149
4. Dithiothreitol (DTT), Pierce, cat. no. 20290
5. Trypsin, Roche, cat. no. 11-418-025-001 (Lot no. 13556621)
6. Endoproteinase Asp-N, Roche, cat. no. 11-054-589-001 (Lot no. 14971420)
7. Formic Acid (FA), 98%, Fluka, Lot no. 1255194
8. Trifluoroacetic acid (TFA), Fisher, cat. no. 04902-100
9. Milli-Q deionized water

Reagent Solution Preparation:

a. 25 mM Ammonium Bicarbonate buffer: Solution was prepared by dissolving 98.83 mg NH_4HCO_3 in 50 mL of Milli-Q water and filtering through 0.22-µm sterile syringe filter.
b. DTT solution (100 mM; prepared fresh): A 100 mM solution was prepared by dissolving 15.4 mg DTT in 1 mL of water.
c. Alkylating reagent (IAA) (200 mM; prepared fresh): Solution was prepared by dissolving 37 mg IAA in 1 mL of water.

d. Trypsin solution: Dissolved 50 µg of dried trypsin (contents of 2 vials) in 600 µL of 25 mM ammonium bicarbonate buffer immediately prior to digestion procedure.
e. Endoproteinase Asp-N solution: Dissolved 2 µg of dried endoproteinase Asp-N in 50 µL of deionized water (Milli-Q) immediately prior to digestion procedure. Prepared 2 vials of endoproteinase Asp-N for two separate digests.

In-gel Protein Processing:

a. The following protein bands from "soybean AAD-12" lane on the SDS-PAGE gel were excised and placed into 1.5-mL siliconized microcentrifuge tubes:

   Band 1: ~ 52-55 kDa – weak band
   Band 2: ~ 44.4 kDa – weak band
   Band 3: ~ 40.4 kDa – weak band
   Band 4: ~ 38.1 kDa – weak band
   Band 5: ~ 30.5 kDa – main soybean AAD-12 band
   Band 6: ~ 28.6 kDa – satellite lower-MW AAD-12 band
   Band 7: ~ 26.1 kDa – weak band
   Band 8: ~ 20.3 kDa – weak band
   Band 9: ~ 17.1 kDa – weak band
   Band 10: ~ 14.1 kDa – weak band

   Bands 5 and 6 were each split into two equal pieces for separate processing for tryptic and Asp-N digests. Bands 1, 2, 3, 4, 7, 8, 9, and 10 were processed for only tryptic digest. Two sections of blank gel were also cut out and processed (one for tryptic digest "blank" and one for Asp-N digest "blank") along with the protein bands as a control/ background.
b. Gel destaining: Four hundred µL of 25 mM ammonium bicarbonate buffer and 400 µL of acetonitrile were added to each tube containing gel pieces, and the gel pieces were crushed and macerated using sterile micro-pestles. The gel pieces were shaken at room temperature at 1400 rpm for 30 min, and then centrifuged for 1 min. The supernatant was discarded. The destaining procedure was repeated 2 more times.
c. Protein alkylation: Twenty microliters of freshly prepared DTT solution was added to each tube and the tubes were incubated at room temperature for 30 min. Seventy-one microliters of freshly prepared IAA solution was added to each tube, and the tubes were shaken (900 rpm) at room temperature for 1 hour. Fifty microliters of DTT solution was added to each tube and the tubes were let stand at room temperature for 10 min. The IAA/ DTT solution was removed, and the gel pieces were rinsed twice with 1 mL of a 1:1 mixture of 25 mM ammonium bicarbonate buffer and acetonitrile.
d. Gel pieces were dried to completeness in a centrifugal evaporator.

e. Proteolytic cleavage with either trypsin or endoproteinase Asp-N was performed next. Approximately 50 µL of trypsin solution was added to each tube containing the dehydrated gel pieces (bands 1 through 10) destined for tryptic digest, and the tubes were incubated at 37 °C for approximately 16 hours (900 rpm shaking). Approximately 50 µL of Asp-N solution was added to each tube with the dehydrated gel pieces destined for Asp-N digest (Bands 5 and 6), and the tubes were incubated at 37 °C for approximately 16 hours (900 rpm shaking).
f. Proteolytic peptides were extracted from gel sequentially with 500 µL of 50 % acetonitrile/ 0.1 % TFA for 30 min, 500 µL of 50 % acetonitrile/ 5 % FA for 30 min, and 500 µL of 70 % acetonitrile/ 5 % FA for 30 min. The three extracts from each digested band were combined and dried in a centrifugal evaporator.
g. The dried extracts were reconstituted in 20 µL of 0.1 % aqueous TFA prior to analysis by MALDI-TOF MS or ESI/LC-MS. The peptide extracts were stored at approximately –20 °C prior to analysis by mass-spectrometry.

ESI-LC/MS of proteolytic digests (for SDS-PAGE bands 5 and 6):

Reagents and Materials:

1. Acetonitrile with 0.1 % formic acid (LC/MS reagent), J.T. Baker, Lot no. H02E29
2. Milli-Q water, freshly drawn 04.27.09
3. Formic Acid 98%, Fluka, Lot no. 1363869
4. Poly-DL-Alanine, Sigma, cat. no. P9003, Lot no. 97H5912
5. Leucine Enkephalin acetate salt, Sigma, cat. no. L-9133, Lot no. 095K5109
6. Total recovery HPLC vials, Waters, P/N 186000384c, lot no. 0384661180

Analytical Procedure:

ESI-LC/MS: Tryptic and endoproteinase Asp-N digests from bands 5 and 6 were analyzed by LC/MS. All mass spectra were acquired on a Waters Q-Tof Micro MS system (S/N YA137). The mass spectrometer was calibrated prior to use in the mass range 350 – 1900 amu (MS for peptide mass finger printing) using 0.1 mg/mL Poly-DL-Alanine solution in acetonitrile at 20 µL/min flow rate. 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 :	1x100 mm Acquity BEH C18 1.7 µm 135 Å; S/N: 115B1511210N18 Part No: 186002346
Flow rate :	25 µL/min
Column temperature :	50 °C
Injection volume :	10 µL

Injection loop : 20 µL

UV detection : 214 nm, 40 pts/sec

Gradient table:

Time, min	Flow rate, mL/min	%A	%B
Initial	0.025	95	5
5	0.025	95	5
40	0.025	60	40
45	0.025	10	90
45.1	0.1	10	90
47.5	0.1	10	90
48	0.1	95	5
52	0.1	95	5
52.1	0.025	95	5
54	0.025	95	5

MS : QTOF-micro mass spectrometer (S/N YA137)

ESI : Micromass lock-spray electrospray interface

Mode : +TOFMS

MS Parameters (peptide mass fingerprinting):

Capillary	:	2850 V
Desolvation Gas	:	250 L/hr
Desolvation Temperature	:	150 ℃
Source Temperature	:	90 ℃
Sample Cone	:	20 V
Extraction Cone	:	1.0 V
Collision Energy	:	10.0 V
MCP	:	2350 V
Mode	:	ESI-TOF-MS +
Scan Range	:	350 – 1900 amu
Scan Cycle Time	:	0.98 sec/scan
Reference Scan Frequency	:	7 sec

Methods:

The samples were injected using the partial loop configuration under the conditions described above.

Data acquisition was performed with a cycle time of 1 sec/scan (scan acquisition time: 0.88 sec; interscan delay: 0.1sec). The lock mass data was acquired using 2.5 µM Leucine-Enkephalin peptide solution (0.1 % formic acid in 50 % acetonitrile was used as the solvent) flowing at 3 – 5 µL/min. The lock mass channel was sampled every 7 sec during MS analysis. The reference ion used was the singly charged Leucine-Enkephalin ion at m/z 556.2771.

Peptide mass fingerprinting of the UPLC-MS data was performed manually. The spectrum of each chromatographic peak was summed, smoothed (SG, 2x3 channels), centroided (4 channels, top 80 %, by height) and m/z error corrected (lock mass channel: 10 scans, m/z 556.2271 ± 0.5 Da). In-source fragmentation observed was used to further confirm the identity of the peptides.

MALDI-TOF MS of soybean AAD-12 proteolytic digests:

Reagents and Materials:

1. Acetonitrile, Fisher, cat. no. A998-1, Lot no. 080757
2. Milli-Q water (18 $M\Omega cm^{-1}$, TOC 30-20 ppb)
3. Siliconized 0.6-ml microcentrifuge tubes, Fisher, cat. no. 02-681-330
4. Pipette tips 0.2-10 µL, Eppendorf
5. Zip Tips C18, Millipore, cat. no. ZTC18S096
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

Analytical Procedure:

MALDI-TOF MS: The in-gel tryptic and Asp-N digests were purified using C18 ZipTips, according to the manufacturer's procedure. For each sample, 4-µL fractions eluting off C18 ZipTips in 10 %, 25 %, 50 % and 70 % acetonitrile/0.1 % TFA were deposited onto a MALDI plate, mixed with 1 µL of CHCA matrix, air-dried, and analyzed by MALDI-TOF MS. All mass spectra were acquired on a Voyager DE STR MALDI-TOF mass spectrometer (S/N 4260). 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: 62%
Mirror voltage ratio: 1.12
Extraction delay time: 200 nsec
Acquisition mass range: 500-7000 Da
Number of laser shots: 500/spectrum
Laser intensity: 1600 – 2000 (varied)
Low mass gate: 500-1900 Da (varied)
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 $[M+H]^{1+}$ m/z values used): Bradykinin (fragment 1-7), m/z 757.3997; Angiotensin II, m/z 1046.5423; P14R synthetic peptide, m/z 1533.8582; ACTH (fragment 18-39), m/z 2465.1989; and Insulin oxidized B chain (bovine), m/z 3494.6513.

The list of ions observed in "blank" tryptic and Asp-N digests were removed from lists of the observed ions from the protein digests prior to subjecting the data to manual peptide mass-fingerprinting. The proteolytic fragments for all SDS-PAGE bands were assigned using theoretical ion tables generated using Micromass BioLynx software and a local copy of Protein Prospector (v 3.2.1).

RESULTS AND DISCUSSION

In this study, the numbering of the amino acid residues is in accordance with the theoretical sequence of the recombinant AAD-12 protein starting with Met[1] and containing a total of 293 residues (**Figure 1**).

SDS-PAGE:

Figure 2 A shows SDS-PAGE image containing soybean-expressed AAD-12 protein in lane 2 and *P.f.*-expressed AAD-12-0002 reference material[(2)] in lanes 4-13. The reference material was used for *(a)* positive identification of the main AAD-12 band in the soybean-expressed AAD-12 sample, and *(b)* for an estimate of the AAD-12 quantity in the soybean-expressed sample.

From the band mobility of the main band (band 5) of the soybean-expressed AAD-12 on the SDS-PAGE, the apparent MW of soybean-expressed AAD-12 (main component) was estimated at ~30 kDa (theoretical average MW = 31.6 kDa for *des*-Met[1] protein). The MW estimate obtained by SDS-PAGE is an apparent and not accurate molecular weight, because factors such as pH, buffer composition, non-specific interactions between the protein and the gel medium may influence the migration of the standards and test protein. Both the soybean-expressed AAD-12 (band 5) and *P.f.*-expressed AAD-12-0002 reference material migrated with nearly identical mobility on the gel, and the *P.f.*-expressed AAD-12-0002 reference material was previously accurately characterized and its intact MW was determined to be 31.6 kDa, as expected for *des*-Met[1] protein[2)].

The amount of protein contained in SDS-PAGE band 5 of soybean-expressed AAD-12 was estimated to be approximately 4 µg, based on the known amounts of *P.f.*-expressed AAD-12-002 reference material run on the same gel (**Figure 2 B**). Ten bands (**Figure 2 A**) were cut out of the gel and processed for peptide mass fingerprinting. Due to the limited amount of protein in each band, only the most intense bands (bands 5 and 6) were digested with both trypsin and endoproteinase Asp-N; the other eight bands were processed with trypsin only.

Peptide Mass Fingerprinting:

MALDI-TOF MS analysis was used to analyze all of the SDS-PAGE bands of the soybean-expressed AAD-12 sample. ESI LC/MS was also used for peptide mass fingerprinting of SDS-PAGE bands 5 and 6. In-gel trypsin (for all bands) and endoproteinase Asp-N (for bands 5 and 6 only) digests of reduced and alkylated protein bands in the soybean-expressed sample were generated. In-gel trypsin and endoproteinase Asp-N digests of the blank sections of the gel were also generated and analyzed in order to remove artifacts (auto-tryptic and auto-Asp-N peaks, as well as other "spurious peaks" coming from the

gel background itself) from the mass spectra of the digests of the protein samples. Due to close proximity and partial overlap of some of the SDS-PAGE bands, certain level of sample carryover, and therefore presence of some m/z peaks in the corresponding mass-spectra, is to be expected for neighboring bands.

Tables I and **II** show the results of tryptic and endoproteinase Asp-N peptide mass-fingerprinting of soybean-expressed AAD-12 band 5, respectively. **Figure 1 A** graphically represents the combined sequence coverage (73.4%) for band 5. Band 5 was identified as *des*-Met[1] AAD-12 protein with N-terminal Ala N-acetylated. The N-terminal endoproteinase Asp-N fragment was consistent with the N-acetylated sequence Ac-A[2]QTTLQITPTGATLGATVTGVHLATL[27]. The C-terminal tryptic fragment was consistent with the sequence L[281]AGRPETEGAALV[293]. This is the principal protein component in the soybean-expressed AAD-12 sample.

Tables III and **IV** show the results of tryptic and endoproteinase Asp-N peptide mass-fingerprinting of soybean-expressed AAD-12 in band 6, respectively. **Figure 1 B** graphically represents the combined sequence coverage (64.2%) for band 6. For the lower MW SDS-PAGE band (band 6), based on MALDI-TOF-MS analysis, the N –terminal endoproteinase Asp-N fragment was consistent with a truncated version P[10]TGATLGATVTGVHLATL[27]. The C-terminal tryptic fragment was consistent with the sequence L[281]AGRPETEGAALV[293].

The list of ions observed from the MALDI-TOF MS analyses of the tryptic digests from bands 1, 2, 3, 4, 7, 8, 9, 10) are listed in **Table V**. Due to the low intensity of these peaks and the number of peaks observed per band being limited, peptide mass fingerprinting of these bands to identify the protein(s) was not attempted.

REFERENCES

1. Raw data packet ML-AL MD-2009-000537
2. A. Karnoup, K. Kuppannan, ML-AL MD-2008-003833

Table I: Peptide Mass Fingerprinting of soybean AAD-12 (Band 5) from In-gel Tryptic Digest. Highlighted are the observed ions.

Fragment	Residues	Sequence	Charge State	Theoretical monoisotopic m/z	Observed m/z MALDI-MS	MALDI: Eluted off C18-ZipTip in % Acetonitrile	Observed m/z LC-MS	Retention Time, min	Comment
T1	1-62	(-) MAQTTLQITPTGATLGATVTGVHLATLDDAGFAALHAAWLQHA LLIFPGQHLSNDQQITFAK(R)	1	6362.34					
			2	3181.67					
			3	2121.45					
			4	1591.34					
			5	1273.27					
T2	63-63	(K) R(F)	1	175.12					
T3	64-69	(R) FGAIER(I)	1	692.37					
T3	64-69	(R) FGAIER(I)	1	692.37	692.21	25%			
T2-3	63-69	(K) RFGAIER(I)	1	848.47	848.26	25%			
T4	70-82	(R) IGGGDIVAISNVK(A)	1	1242.71	1242.39	25%			
			2	621.86			621.82	22.8	513.31 (1+, b6), 796.44 (1+, b9); minor
			3	414.91					
T5	83-88	(K) ADGTVR(Q)	1	618.32					
T6	89-100	(R) QHSPAEWDDMMK(V)	1	1474.61					
			2	737.81					
			3	492.21					
T7	101-135	(K) VIVGNMAWHADSTYMPVMAQGAVFSAEVVPAVGGR(T)	1	3617.77					
			2	1809.39					
			3	1206.60					
			4	905.20					
			5	724.36					
T8	136-142	(R) TCFADMR(A)	1	900.37					
			2	450.69					
T9	143-153	(R) AAYDALDEATR(A)	1	1195.56	1195.31	10, 25, 50, 70%	1195.52		
			2	598.28			598.25	18.7	421.21 (1+, b4), 591.28 (1+, y5), 704.37 (1+, y6), 775.42 (1+, y7); minor
			3	399.19					
T10	154-159	(R) ALVHQR(S)	1	723.43	723.25	25%	723.41	4.5	early eluting
			2	362.22					
T9-10	143-159	(R) AAYDALDEATRALVHQR(S)	1	1899.97	1900.46	25%			
T11	160-162	(R) SAR(H)	1	333.19					
T10-11	154-162	(R) ALVHQRSAR(H)	1	1037.60			1037.46	16.8	minor peak
			2	519.30					
T12	163-171	(R) HSLVYSQSK(L)	1	1048.54	1048.30	10, 25%			
			2	524.77			524.77	9.9	
			3	350.19					
NS	163-183	(R) HSLVYSQSKLGHVQQAGSAYI(G)	1	2273.17					
			2	1137.09			1137.15	26.2	591.80(2+, b11-H2O), 600.77 (2+, b11), 1071.46 (2+, b20)
			3	758.39			758.38		

Table I continued

Fragment	Residues	Sequence	Charge State	Theoretical monoisotopic m/z	Observed m/z MALDI-MS	MALDI: Eluted off C18-ZipTip in % Acetonitrile	Observed m/z LC-MS	Retention Time, min	Comment
T13	172-199	(K) LGHVQQAGSAYIGYGMDTTATPLRPLVK(V)	1	2944.54	2943.88	25, 50, 70%			singly and doubly oxidized forms also detected in MALDI (obs. m/z 2959.84, 2975.87)
			2	1472.77					
			3	982.19					
			4	736.89					
			5	589.71					
T14	200-213	(K) VHPETGRPSLLIGR(H)	1	1531.87	1531.54	10, 25, 50, 70%			
			2	766.44					
			3	511.30					
T15	214-228	(R) HAHAIPGMDAAESER(F)	1	1591.73	1591.38	25%			
			2	796.37					
			3	531.25					
T16	229-242	(R) FLEGLVDWACQAPR(V)	1	1661.81	1661.46	25, 50, 70%			
			2	831.41					
			3	554.61					
T17	243-259	(R) VHAHQWAAGDVVVWDNR(C)	1	1959.96	1959.54	25, 50, 70%			oxidized form also detected in MALDI (obs. m/z 1975.53)
			2	980.48					
			3	653.99					
			4	490.75					
T18	260-264	(R) CLLHR(A)	1	698.38					
T19	265-271	(R) AEPWDFK(L)	1	892.42					
			2	446.71					
T20	272-274	(K) LPR(V)	1	385.26					
T19-20	265-274	(R) AEPWDFKLPR(V)	1	1258.66	1258.40	25, 50, 70%			
NS	279-284	(H) SRLAGR(P)	1	659.40			659.38	11.7	485.33 (1+, b5), 572.37 (1+, y5)
T21	275-280	(R) VMWHSR(L)	1	815.40					
			2	408.20					
T22	281-293	(R) LAGRPETEGAALV(-)	1	1283.70	1283.43	10, 25, 50, 70%	1283.69	20.3	527.27 (2+, b11), 577.29 (2+, y12-NH3), 583.82 (2+, b12), 585.33 (2+, y12), 982.55 (1+, b10_, 1053.51 (1+, b11), 1153.65 (1+, y12-NH3); minor
			2	642.35			642.35		

NS: Non-specific cleavage

Table II: Peptide Mass Fingerprinting of soybean AAD-12 (band 5) from In-gel endoproteinase Asp-N Digest. Highlighted are the detected ions.

Fragment	Residues	Sequence	Charge State	Theoretical monoisotopic m/z	Observed m/z MALDI-MS	MALDI: Eluted off C18-ZipTip in % Acetonitrile	Observed m/z LC-MS	Retention Time, min	Comment
D1-Met1	1-27	(-) MAQTTLQITPTGATLGATVTGVHLATL(D)	1	2667.44					
			2	1334.22					
			3	889.82					
			4	667.62					
D1-des-Met1 (H)	2-27	(-) AQTTLQITPTGATLGATVTGVHLATL(D)	1	2536.40	2536.06	50, 70%			*trace amount*
			2	1268.70					
			3	846.14					
			4	634.86					
D1-des-Met1-N-Ac	2-27	**Ac**-AQTTLQITPTGATLGATVTGVHLATL(D)	1	2578.41	2578.07	50, 70%			
			2	1289.71					
			3	860.14					
			4	645.36					
D1-2-Met1	1-28	(-) MAQTTLQITPTGATLGATVTGVHLATLD(D)	1	2782.47					
			2	1391.74					
			3	928.16					
			4	696.37					
D1-2-des Met1 (H)	2-28	(-) AQTTLQITPTGATLGATVTGVHLATLD(D)	1	2651.43					
			2	1326.22					
			3	884.48					
			4	663.61					
D1-2-des-Met1-N-Ac	2-28	**Ac**-AQTTLQITPTGATLGATVTGVHLATLD(D)	1	2693.44	2693.05	50, 70%			
			2	1347.22					
			3	898.48					
			4	674.11					
D3	29- 54	(D) DAGFAALHAAWLQHALLIFPGQHLSN(D)	1	2798.45					
			2	1399.73					
			3	933.49					
			4	700.37					
D4	55- 73	(N) DQQITFAKRFGAIERIGGG(D)	1	2064.10					
			2	1032.55					
			3	688.70					
			4	516.78					
D4-5	55- 83	(N) DQQITFAKRFGAIERIGGGDIVAISNVKA(D)	1	3074.67	3074.26	50, 70%			
			2	1537.84					
			3	1025.56					
			4	769.42					
D5	74- 83	(G) DIVAISNVKA(D)	1	1029.59					
			2	515.30					
			3	343.87					
			4	258.15					

Table II continued

Fragment	Residues	Sequence	Charge State	Theoretical monoisotopic m/z	Observed m/z MALDI-MS	MALDI: Eluted off C18-ZipTip in % Acetonitrile	Observed m/z LC-MS	Retention Time, min	Comment
D6	84- 95	(A) DGTVRQHSPAEW(D)	1	1382.64	1382.65	25%			
			2	691.83					
			3	461.55					
			4	346.42					
D8	97-110	(D) DMMKVIVGNMAWHA(D)	1	1602.76					
			2	801.88					
			3	534.92					
			4	401.45					
D9	111-139	(A) DSTYMPVMAQGAVFSAEVVPAVGGRTCFA(D)	1	3018.42					
			2	1509.71					
			3	1006.81					
			4	755.36					
D10	140-145	(A) DMRAAY(D)	1	726.32					
			2	363.67					
			3	242.78					
			4	182.34					
D11	146-148	(Y) DAL(D)	1	318.17					
			2	159.59					
D12	149-187	(L) DEATRALVHQRSARHSLVYSQSKLGHVQQAGSAYIGYGM(D)	1	4272.15					
			2	2136.58					
			3	1424.72					
			4	1068.79					
D13	188-221	(M) DTTATPLRPLVKVHPETGRPSLLIGRHAHAIPGM(D)	1	3639.01					
			2	1820.01					
			3	1213.67					
			4	910.51					
D14	222-234	(M) DAAESERFLEGLV(D)	1	1435.71	1435.53	50, 70%			
			2	718.36					
			3	479.24					
			4	359.68					
D15	235-251	(V) DWACQAPRVHAHQWAAG(D)	1	1960.90	1960.69	25, 50%			
			2	980.95					
			3	654.30					
			4	490.98					
D16	252-256	(G) DVVVW(D)	1	617.33					
			2	309.17					
			3	206.45					
			4	155.09					
D17	257-268	(W) DNRCLLHRAEPW(D)	1	1566.76	1566.54	25, 50%			
			2	783.88					
			3	522.92					
			4	392.45					
D18	269-293	(W) DFKLPRVMWHSRLAGRPETEGAALV(-)	1	2836.50					
			2	1418.76					
			3	946.17					
			4	709.88					

Table III: Peptide Mass Fingerprinting of soybean AAD-12 (Band 6) from In-gel Tryptic Digest. Highlighted are the detected ions.

Fragment	Residues	Sequence	Charge State	Theoretical monoisotopic m/z	Observed m/z MALDI-MS	MALDI: Eluted off C18-ZipTip in % Acetonitrile	Observed m/z LC-MS	Retention Time, min	Comment
T1	1-62	(-) MAQTTLQITPTGATLGATVTGVHLATLDDAGFAALHAAWLQHALL IFPGQHLSNDQQITFAK(R)	1	6362.34					
			2	3181.67					
			3	2121.45					
			4	1591.34					
			5	1273.27					
NS	50-53	(G) QHLS(N)	1	484.25			484.24	12.3	397.22 (1+, b3); minor
T2	63-63	(K) R(F)	1	175.12					
T3	64-69	(R) FGAIER(I)	1	692.37					
T4	70-82	(R) IGGGDIVAISNVK(A)	1	1242.71					
			2	621.86			621.87		
			3	414.91				22.6	631.38 (1+, y6); minor
T5	83-88	(K) ADGTVR(Q)	1	618.32					
NS	85-89	(D) GTVRQ(H)	1	560.32			560.31	13.3	402.21 (1+, y3); minor
T6	89-100	(R) QHSPAEWDDMMK(V)	1	1474.61					
			2	737.81					
			3	492.21					
T7	101-135	(K) VIVGNMAWHADSTYMPVMAQGAVFSAEVVPAVGGR(T)	1	3617.77					
			2	1809.39					
			3	1206.60					
			4	905.20					
			5	724.36					
T8	136-142	(R) TCFADMR(A)	1	900.37					
			2	450.69					
T9	143-153	(R) AAYDALDEATR(A)	1	1195.56	1195.48	10, 25%	1195.57		
			2	598.28			598.27		591.24 (1+, y5), 890.39 (1+, y8);
			3	399.19				18.5	minor
T10	154-159	(R) ALVHQR(S)	1	723.43					
			2	362.22					
T11	160-162	(R) SAR(H)	1	333.19					
T10-11	154-162	(R) ALVHQRSAR(H)	1	1037.60			1037.48		
			2	519.30				16.5	minor peak
T12	163-171	(R) HSLVYSQSK(L)	1	1048.54					
			2	524.77					
			3	350.19					
NS	163-183	(R) HSLVYSQSKLGHVQQAGSAYI(G)	1	2273.17					437.24 (1+, b4), 537.34 (2+, y10),
			2	1137.09			1137.15		591.77 (2+, b11-H2O), 600.78 (2+,
			3	758.39			758.41	26.1	b11), 718.87 (2+, b13), 1071.46 (2+,

Table III continued

Fragment	Residues	Sequence	Charge State	Theoretical monoisotopic m/z	Observed m/z MALDI-MS	MALDI: Eluted off C18-ZipTip in % Acetonitrile	Observed m/z LC-MS	Retention Time, min	Comment
T13	172-199	(K) LGHVQQAGSAYIGYGMDTTATPLRPLVK(V)	1	2944.54	2944.41	25, 50%			singly and doubly oxidized forms also detected in MALDI (obs. m/z 2959.81, 2975.80)
			2	1472.77					
			3	982.19					
			4	736.89					
			5	589.71					
T14	200-213	(K) VHPETGRPSLLIGR(H)	1	1531.87	1531.85	10, 25, 50, 70%			
			2	766.44					
			3	511.30					
T15	214-228	(R) HAHAIPGMDAAESER(F)	1	1591.73					
			2	796.37					
			3	531.25					
T16	229-242	(R) FLEGLVDWACQAPR(V)	1	1661.81	1661.79	25, 50, 70%			
			2	831.41					
			3	554.61					
T17	243-259	(R) VHAHQWAAGDVVVWDNR(C)	1	1959.96	1959.92	25, 50, 70%			oxidized peptide at m/z 1975.81 was also observed in MALDI
			2	980.48					
			3	653.99					
			4	490.75					
T18	260-264	(R) CLLHR(A)	1	698.38					
T19	265-271	(R) AEPWDFK(L)	1	892.42					
			2	446.71					
T20	272-274	(K) LPR(V)	1	385.26					
T19-20	265-274	(R) AEPWDFKLPR(V)	1	1258.66					
NS	279-284	(H) SRLAGR(P)	1	659.40			659.39	10.5	485.27 (1+, b5), 572.36 (1+, y5)
T21	275-280	(R) VMWHSR(L)	1	815.40					
			2	408.20					
T22	281-293	(R) LAGRPETEGAALV(-)	1	1283.70	1283.72	25, 50%			982.45 (1+, b10_, 1053.59 (1+, b11); minor
			2	642.35			642.38	20.1	

NS: Non-specific cleavage

Table IV: Peptide Mass Fingerprinting of soybean AAD-12 (band 6) from In-gel endoproteinase Asp-N Digest. Highlighted are the detected ions.

Fragment	Residues	Sequence	Charge State	Theoretical monoisotopic m/z	Observed m/z MALDI MS	MALDI: Eluted off C18-ZipTip in % Acetonitrile	Observed m/z LC-MS	Retention Time, min	Comment
D1-Met1	1-27	(-) MAQTTLQITPTGATLGATVTGVHLATL(D)	1	2667.44					
			2	1334.22					
			3	889.82					
			4	667.62					
D1-des-Met1 (H)	2-27	(-) AQTTLQITPTGATLGATVTGVHLATL(D)	1	2536.40					
			2	1268.70					
			3	846.14					
			4	634.86					
D1-des-Met1-N-Ac	2-27	**Ac**-AQTTLQITPTGATLGATVTGVHLATL(D)	1	2578.41	2578.19	50, 70%			weak signal in MALDI; likely to be a carryover from Band-5
			2	1289.71					
			3	860.14					
			4	645.36					
D1-2-Met1	1-28	(-) MAQTTLQITPTGATLGATVTGVHLATLD(D)	1	2782.47					
			2	1391.74					
			3	928.16					
			4	696.37					
D1-2-des Met1 (H)	2-28	(-) AQTTLQITPTGATLGATVTGVHLATLD(D)	1	2651.43					
			2	1326.22					
			3	884.48					
			4	663.61					
D1-2-des-Met1-N-Ac	2-28	**Ac**-AQTTLQITPTGATLGATVTGVHLATLD(D)	1	2693.44	2693.19	50%			weak signal in MALDI; likely to be a carryover from Band-5
			2	1347.22					
			3	898.48					
			4	674.11					
truncated D1	10-27	(-) PTGATLGATVTGVHLATL(D)	1	1679.93	1679.63	25, 50%			could be present; not confirmed by LC/MS
D3	29- 54	(D) DAGFAALHAAWLQHALLIFPGQHLSN(D)	1	2798.45					
			2	1399.73					
			3	933.49					
			4	700.37					
D4	55- 73	(N) DQQITFAKRFGAIERIGGG(D)	1	2064.10					
			2	1032.55					
			3	688.70					
			4	516.78					

Table IV continued

Fragment	Residues	Sequence	Charge State	Theoretical monoisotopic m/z	Observed m/z MALDI MS	MALDI: Eluted off C18-ZipTip in % Acetonitrile	Observed m/z LC-MS	Retention Time, min	Comment
D4-5	55- 83	(N) DQQITFAKRFGAIERIGGGDIVAISNVKA(D)	1	3074.67	3074.48	50, 70%			
			2	1537.84					
			3	1025.56					
			4	769.42					
D5	74- 83	(G) DIVAISNVKA(D)	1	1029.59					
			2	515.30					
			3	343.87					
			4	258.15					
D6	84- 95	(A) DGTVRQHSPAEW(D)	1	1382.64	1382.54	25%			
			2	691.83					
			3	461.55					
			4	346.42					
D6-7	84-96	(A) DGTVRQHSPAEWD(D)	1	1497.67	1497.17	25%			
D8	97-110	(D) DMMKVIVGNMAWHA(D)	1	1602.76					
			2	801.88					
			3	534.92					
			4	401.45					
D9	111-139	(A) DSTYMPVMAQGAVFSAEVVPAVGGRTCFA(D)	1	3018.42					
			2	1509.71					
			3	1006.81					
			4	755.36					
D10	140-145	(A) DMRAAY(D)	1	726.32					
			2	363.67					
			3	242.78					
			4	182.34					
D11	146-148	(Y) DAL(D)	1	318.17					
			2	159.59					
D12	149-187	(L) DEATRALVHQRSARHSLVYSQSKLGHVQQAGSAYIGYGM(D)	1	4272.15					
			2	2136.58					
			3	1424.72					
			4	1068.79					
D13	188-221	(M) DTTATPLRPLVKVHPETGRPSLLIGRHAHAIPGM(D)	1	3639.01					
			2	1820.01					
			3	1213.67					
			4	910.51					

Table IV continued

Fragment	Residues	Sequence	Charge State	Theoretical monoisotopic m/z	Observed m/z MALDI MS	MALDI: Eluted off C18-ZipTip in % Acetonitrile	Observed m/z LC-MS	Retention Time, min	Comment
D14	222-234	(M) DAAESERFLEGLV(D)	1	1435.71	1435.65	25, 50, 70%			
			2	718.36					
			3	479.24					
			4	359.68					
D15	235-251	(V) DWACQAPRVHAHQWAAG(D)	1	1960.90	1960.83	25, 50%			
			2	980.95					
			3	654.30					
			4	490.98					
D16	252-256	(G) DVVVW(D)	1	617.33					
			2	309.17					
			3	206.45					
			4	155.09					
D17	257-268	(W) DNRCLLHRAEPW(D)	1	1566.76					
			2	783.88					
			3	522.92					
			4	392.45					
D18	269-293	(W) DFKLPRVMWHSRLAGRPETEGAALV(-)	1	2836.50					
			2	1418.76					
			3	946.17					
			4	709.88					

Table V: MALDI-TOF MS peak lists for tryptic digests of SDS-PAGE bands 1, 2, 3, 4, 7, 8, 9, and 10. Observed SDS-PAGE apparent molecular weights of the bands are indicated on the right.

Band-1, Tryptic digest .-- Blank gel digest peaks filtered out **MW ~ 55 kDa**

m/z MH+			Comment	Theor m/z	Comment
C18 Zip-tip elution conditions:					
25% ACN	50% ACN	70% ACN			
		1256.60			
1465.46			weak		
		1565.67			
		1743.88	weak		
		2675.17	weak		
3437.89			very weak		

Band-2, Tryptic digest .-- Blank gel digest peaks filtered out **MW ~ 44 kDa**

m/z MH+			Comment	Theor m/z	Comment
C18 Zip-tip elution conditions:					
25% ACN	50% ACN	70% ACN			
1010.28			weak		
1023.21			weak		
1298.26	1298.46				
1616.38			weak		
2959.64			very weak	2960.54	may come from AAD12
	3659.01				

Band-3, Tryptic digest .-- Blank gel digest peaks filtered out **MW ~ 40 kDa**

m/z MH+			Comment	Theor m/z	Comment
C18 Zip-tip elution conditions:					
25% ACN	50% ACN	70% ACN			
1015.54			weak		
	1275.65		weak		
1465.67			weak		
1475.66			weak		
1487.66			weak		
	1502.66		weak		
1524.74					
1613.62			weak		
1623.64	1623.61		weak		
	1681.81		weak		
	1882.82	1882.76			
1959.83			weak	1959.96	may come from AAD12
2022.79					
2028.84					
	2082.85		weak		
	2152.88		weak		
2208.86					
2744.04	2743.98				
	2840.08	2840.01	weak		
	3888.50		very weak		
	4042.53		very weak		
	5971.19		very weak		

Table V (continued).

Band-4, Tryptic digest .-- Blank gel digest peaks filtered out **MW ~ 38 kDa**

m/z MH+			Comment	Theor m/z	Comment
C18 Zip-tip elution conditions:					
25% ACN	50% ACN	70% ACN			
2028.75			weak		
		3180.90	very weak		
		4576.50	very weak		

Band-7, Tryptic digest .-- Blank gel digest peaks filtered out **MW ~ 26 kDa**

m/z MH+			Comment	Theor m/z	Comment
C18 Zip-tip elution conditions:					
25% ACN	50% ACN	70% ACN			
1195.43			weak	1195.56	may come from AAD12 (T9)
1258.52	1258.49		weak	1258.66	may come from AAD12 (T19-20)
1283.55				1283.70	may come from AAD12 (T22 -- C-term)
1360.52					
1531.70	1531.64			1531.87	may come from AAD12 (T14)
	1661.60	1661.53		1661.81	may come from AAD12 (T16)
1959.73	1959.70		weak	1959.96	may come from AAD12 (T17)
2944.17			trace	2944.54	may come from AAD12 (T13)
2960.18	2960.07			2960.54	may come from AAD12 (oxidized T13)
3374.21			weak		
3415.14			weak		

***** This band may represent an AAD12 degradation product (truncated from N-terminus)***

Band-8, Tryptic digest .-- Blank gel digest peaks filtered out **MW ~ 20 kDa**

m/z MH+			Comment	Theor m/z	Comment
C18 Zip-tip elution conditions:					
25% ACN	50% ACN	70% ACN			
	860.60		weak		
888.37					
905.39	904.62				
944.38			weak		
	948.64		weak		
1017.47	1017.52				
1195.48			weak	1195.56	may come from AAD12 (T9)
1258.54	1258.62		weak	1258.66	may come from AAD12 (T19-20)
1304.56			weak		
1527.65	1526.73				
	1661.75		weak	1661.81	may come from AAD12 (T16)
	1670.82		weak		
1682.63					
1939.78	1939.87	1939.76	weak		
1959.78			weak	1959.96	may come from AAD12 (T17)
2813.17			weak		
2960.23			weak	2960.54	may come from AAD12 (oxidized T13)
	3093.39	3093.23			
	4177.70	4177.60			
	5330.10		very weak		

***** There may also bea carry-over from bands 6 & 7***
Or it may be an AAD-12 degradation product

Table V (continued).

Band-9, Tryptic digest .-- Blank gel digest peaks filtered out **MW ~ 17 kDa**

m/z MH+			Comment	Theor m/z	Comment
C18 Zip-tip elution conditions:					
25% ACN	50% ACN	70% ACN			
	881.28	881.21			
906.45			weak		
944.43			weak		
1010.53			very weak		
1195.51				1195.56	may come from AAD12 (T9)
1257.59	1258.59			1258.66	may come from AAD12 (T19-20)
1283.64	1283.56			1283.70	may come from AAD12 (T22 -- C-term)
	1471.64		weak		
1475.68			weak		
	1487.66				
1531.80				1531.87	may come from AAD12 (T14)
1657.71					
	1661.68	1661.62		1661.81	may come from AAD12 (T16)
1707.69			weak		
1838.80			weak		
1959.86			weak	1959.96	may come from AAD12 (T17)
2565.08			weak		
2877.32					
2907.32			weak		
2960.35			weak	2960.54	may come from AAD12 (oxidized T13)
3340.09			weak		
5000.54			weak		
	5330.09		very weak		

***** There may also be a carry-over from bands 6 & 7***
Or it may be an AAD-12 degradation product

Band-10, Tryptic digest .-- Blank gel digest peaks filtered out **MW ~ 14 kDa**

m/z MH+			Comment	Theor m/z	Comment
C18 Zip-tip elution conditions:					
25% ACN	50% ACN	70% ACN			
944.36			weak		
1181.46			weak	1181.64	may come from AAD12 (T20-21)
1363.51			very weak		
1707.62			very weak		
	2328.92	2328.74	weak		
		2708.01	very weak		
	2721.20	2721.00	weak		
	3164.11	3163.90	weak		
	3181.13	3180.90	weak		
		4576.38	very weak		
	4965.76		very weak		
5000.31			weak		
	5330.99		very weak		

Figure 1: Theoretical amino acid sequence of the recombinant soybean-expressed AAD-12 protein starting with Met[1] and containing a total of 293 residues. ***(A)*** Sequence coverage (highlighted in yellow) obtained from tryptic and Asp-N peptide mass fingerprinting for SDS-PAGE Band 5 (see **Figure 2**). For Band 5, the N-terminal Ala was found to be N-acetylated. ***(B)*** Sequence coverage (highlighted) obtained from tryptic and Asp-N peptide mass fingerprinting for SDS-PAGE Band 6 (see **Figure 2**). For Band 6, a m/z consistent with a truncated version of the N-terminal peptide was observed (the first detected Asp-N fragment is highlighted in light blue).

(A) Band 5:

		Ac																										
1	M	A	Q	T	T	L	Q	I	T	P	T	G	A	T	L	G	A	T	V	T	G	V	H	L	A	T	L	D	D	A	30
31	G	F	A	A	L	H	A	A	W	L	Q	H	A	L	L	I	F	P	G	Q	H	L	S	N	D	Q	Q	I	T	F	60
61	A	K	R	F	G	A	I	E	R	I	G	G	G	D	I	V	A	I	S	N	V	K	A	D	G	T	V	R	Q	H	90
91	S	P	A	E	W	D	D	M	M	K	V	I	V	G	N	M	A	W	H	A	D	S	T	Y	M	P	V	M	A	Q	120
121	G	A	V	F	S	A	E	V	V	P	A	V	G	G	R	T	C	F	A	D	M	R	A	A	Y	D	A	L	D	E	150
151	A	T	R	A	L	V	H	Q	R	S	A	R	H	S	L	V	Y	S	Q	S	K	L	G	H	V	Q	Q	A	G	S	180
181	A	Y	I	G	Y	G	M	D	T	T	A	T	P	L	R	P	L	V	K	V	H	P	E	T	G	R	P	S	L	L	210
211	I	G	R	H	A	H	A	I	P	G	M	D	A	A	E	S	E	R	F	L	E	G	L	V	D	W	A	C	Q	A	240
241	P	R	V	H	A	H	Q	W	A	A	G	D	V	V	V	W	D	N	R	C	L	L	H	R	A	E	P	W	D	F	270
271	K	L	P	R	V	M	W	H	S	R	L	A	G	R	P	E	T	E	G	A	A	L	V								293

Total number of amino acids covered =	215
Total number of amino acids in protein =	293
Tryptic & Asp-N combined sequence coverage =	**73.4 %**

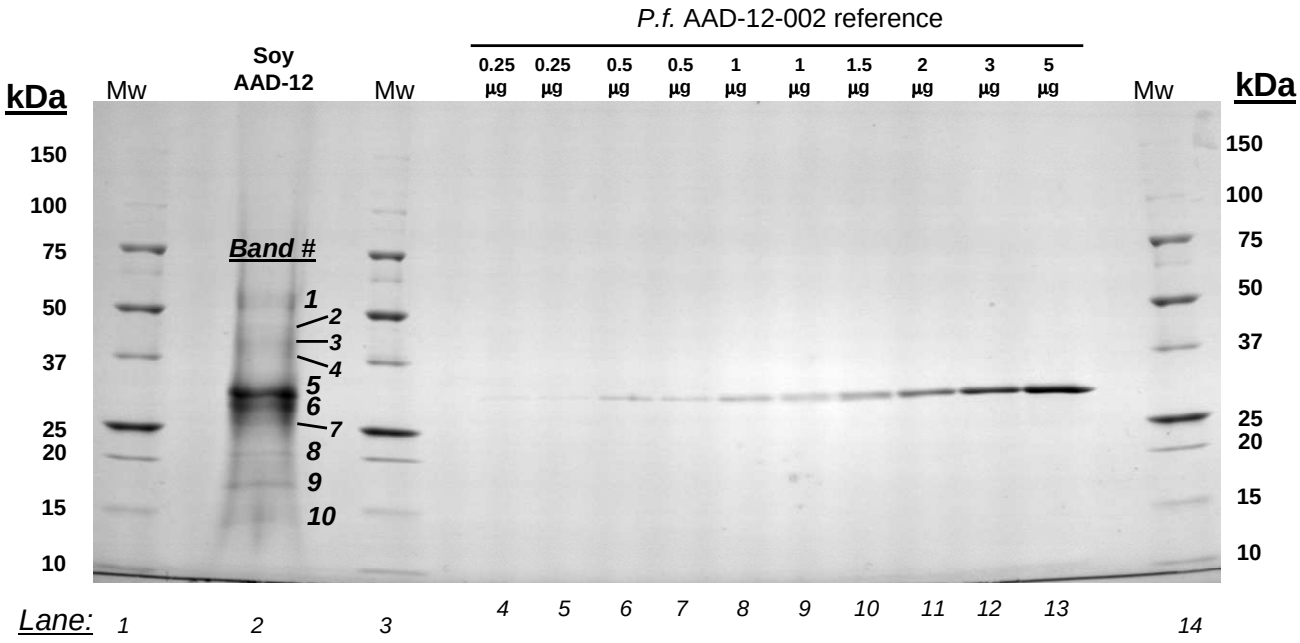
(B) Band 6:

1	M	A	Q	T	T	L	Q	I	T	P	T	G	A	T	L	G	A	T	V	T	G	V	H	L	A	T	L	D	D	A	30
31	G	F	A	A	L	H	A	A	W	L	Q	H	A	L	L	I	F	P	G	Q	H	L	S	N	D	Q	Q	I	T	F	60
61	A	K	R	F	G	A	I	E	R	I	G	G	G	D	I	V	A	I	S	N	V	K	A	D	G	T	V	R	Q	H	90
91	S	P	A	E	W	D	D	M	M	K	V	I	V	G	N	M	A	W	H	A	D	S	T	Y	M	P	V	M	A	Q	120
121	G	A	V	F	S	A	E	V	V	P	A	V	G	G	R	T	C	F	A	D	M	R	A	A	Y	D	A	L	D	E	150
151	A	T	R	A	L	V	H	Q	R	S	A	R	H	S	L	V	Y	S	Q	S	K	L	G	H	V	Q	Q	A	G	S	180
181	A	Y	I	G	Y	G	M	D	T	T	A	T	P	L	R	P	L	V	K	V	H	P	E	T	G	R	P	S	L	L	210
211	I	G	R	H	A	H	A	I	P	G	M	D	A	A	E	S	E	R	F	L	E	G	L	V	D	W	A	C	Q	A	240
241	P	R	V	H	A	H	Q	W	A	A	G	D	V	V	V	W	D	N	R	C	L	L	H	R	A	E	P	W	D	F	270
271	K	L	P	R	V	M	W	H	S	R	L	A	G	R	P	E	T	E	G	A	A	L	V								293

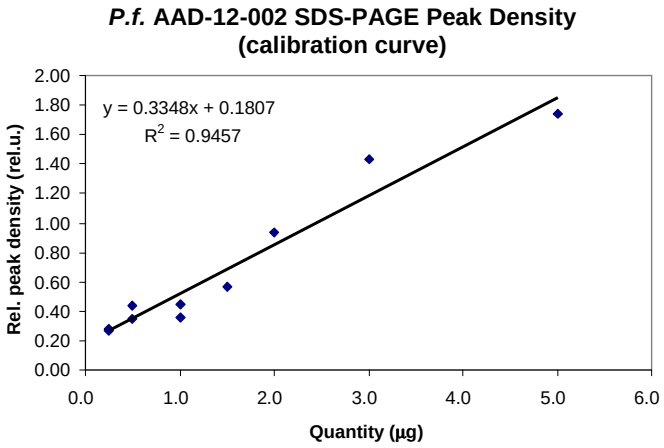
Total number of amino acids covered =	188
Total number of amino acids in protein =	293
Combined Asp-N & Tryptic sequence coverage =	**64.2 %**

Figure 2: *(A)* SDS-PAGE of soybean-expressed AAD-12 protein. Loaded lanes are numbered at the bottom. Lanes containing Bio-Rad molecular weight standards are labeled "Mw" (lanes 1, 3, and 14) and their Mw values are indicated on the far left and far right sides of the image. Lane 2: Soy-expressed AAD-12. Lanes 4 through 13: known quantities (indicated above the lanes) of *P.f.*-expressed AAD-12-002 reference material. *(B)* Estimate of the soybean-expressed AAD-12 quantity from SDS-PAGE based on known amounts of *P.f.*-expressed AAD-12-002 reference material. Band 5 was estimated to contain approximately 4 µg of protein.

(A)



(B)



Soy AAD-12 sample:

-- 5mm band-width

Band #	Mw (kDa)	Peak Density (rel.u.)	Quantity estimate (µg)
5	**30.02**	**1.56**	**4**
6	28.58	1.34	3
7	26.06	1.05	3
8	20.32	0.67	1
9	17.07	0.73	2
10	14.05	0.63	1

Minireview

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

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.

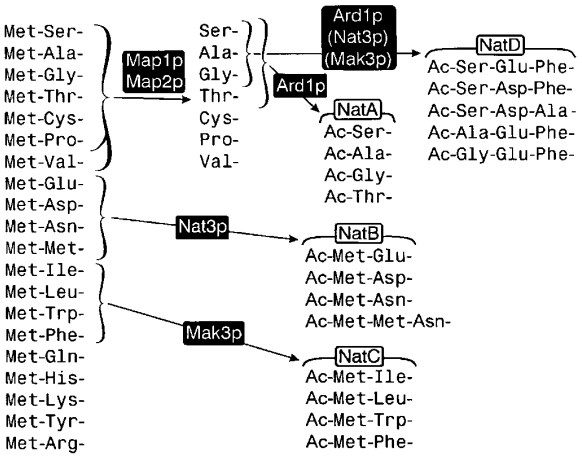


FIG. 1. **The pathways of N-terminal processing.** The two methionine aminopeptidases, Map1p and Map2, cleave N-terminal methionine residues that have small side chains (glycine, alanine, serine, cysteine, threonine, proline, and valine), although methionine is retained on some proteins having penultimate 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 actin-omyosin (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 post-translationally 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

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

*Corresponding author

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*)

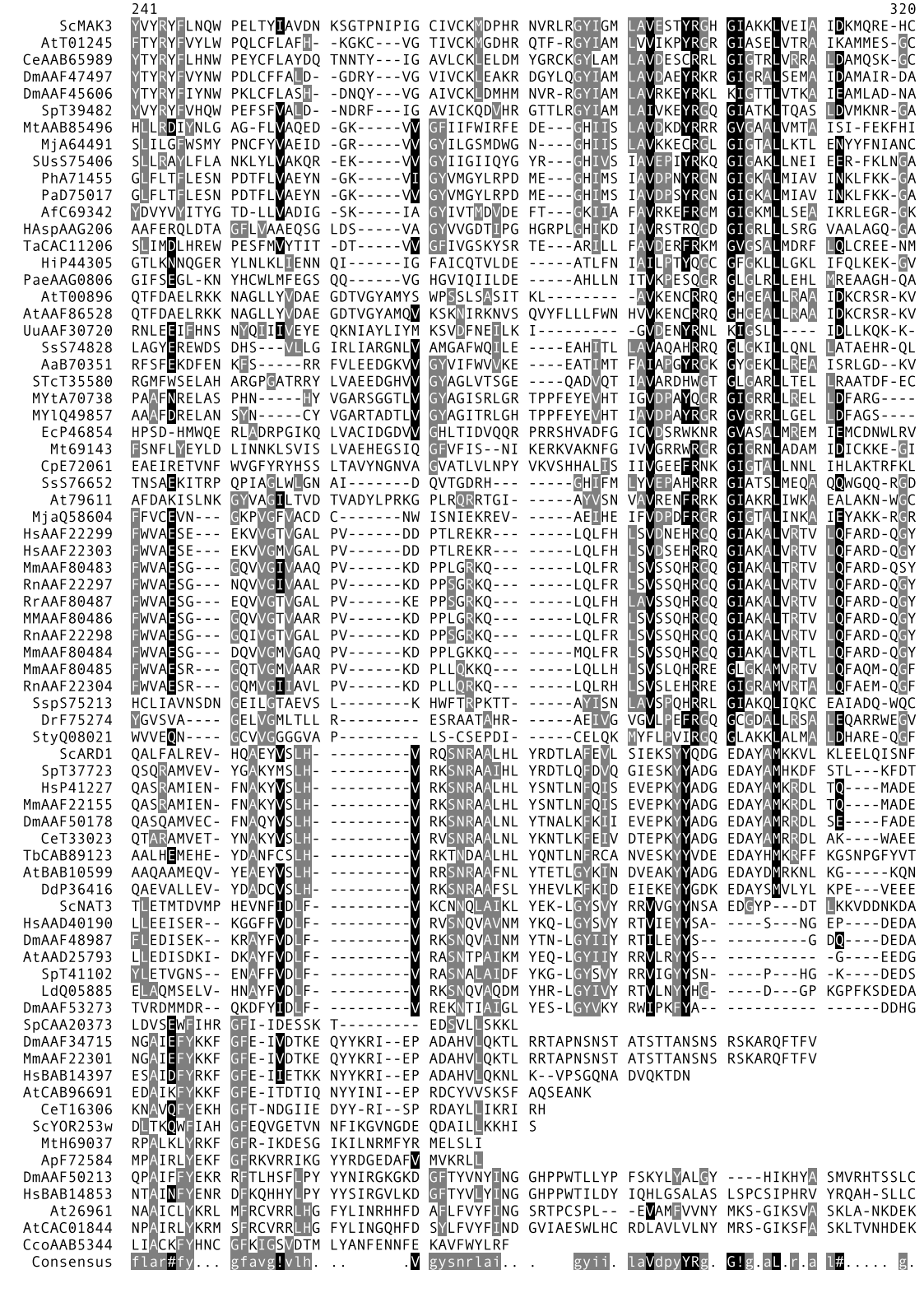


Figure 1 (*legend on page 614*)

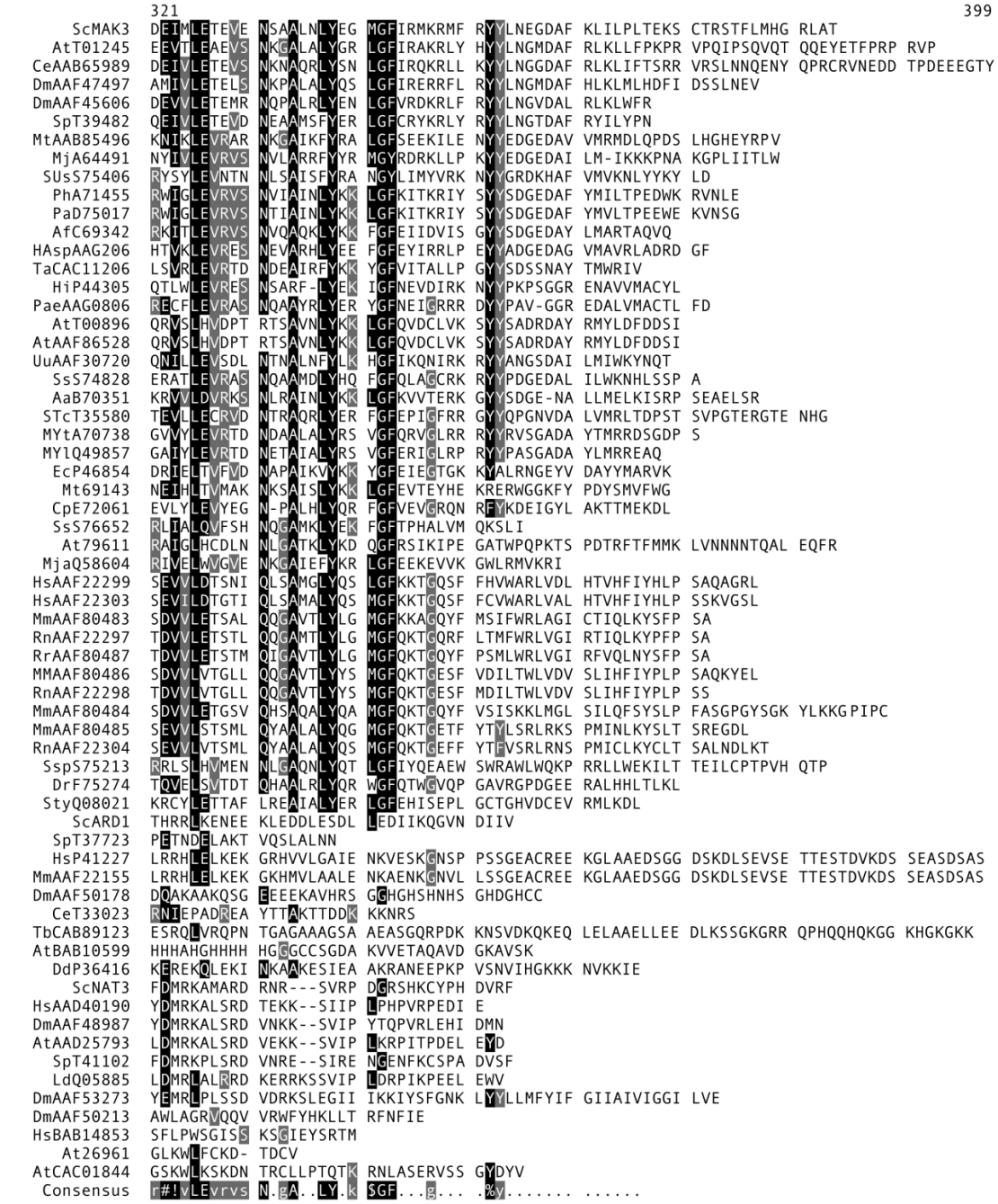


Figure 1. Protein amino acid sequence alignment of known and putative NATs. Multiple sequence alignment was made by using Multalin version 5.4.1.[41] Highly conserved residues are highlighted in black, where multiple designations are as follows: ! = I or V; % = F or Y; # = N, D, Q, E, B, or Z. Moderately 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.

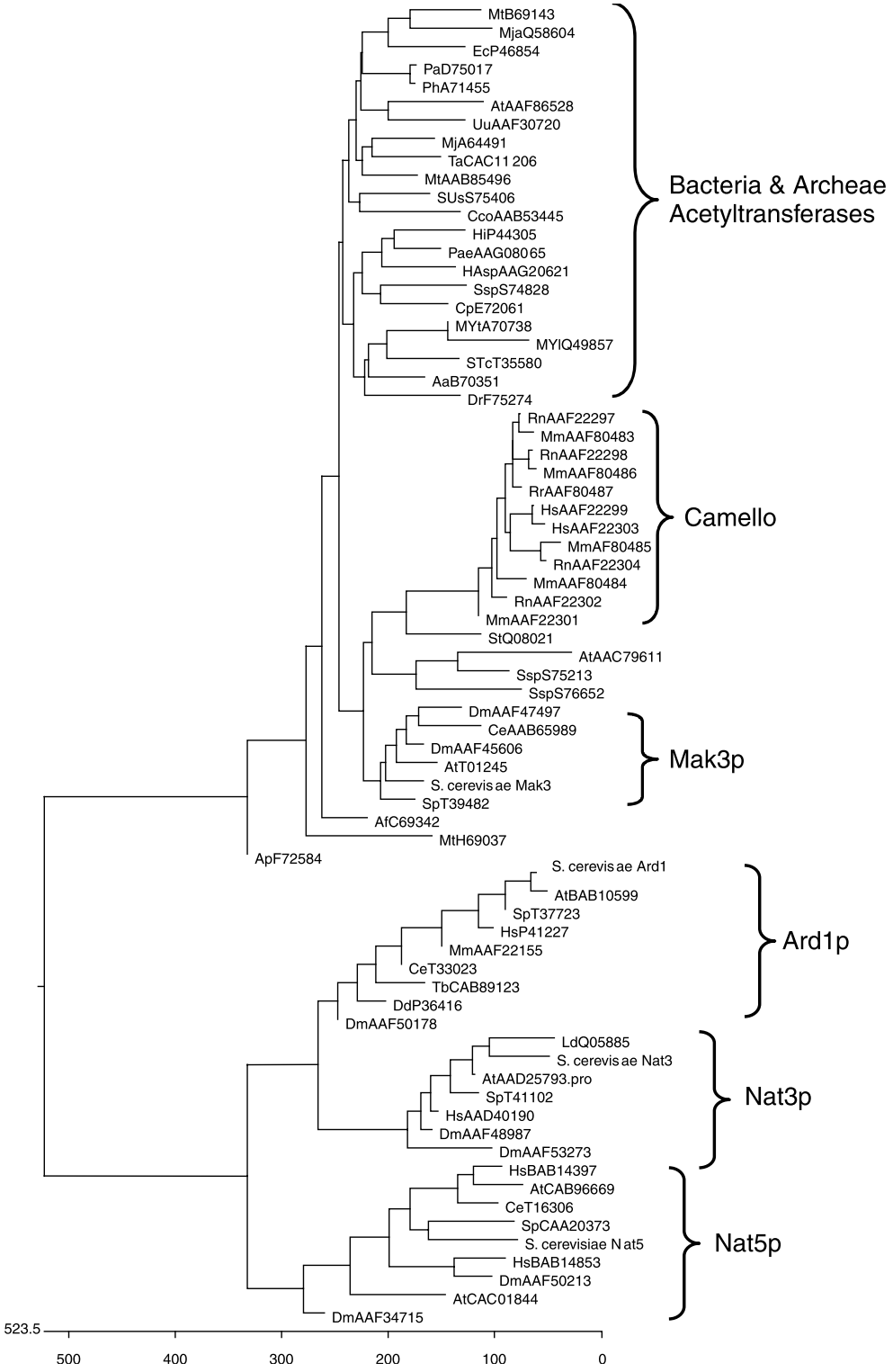


Figure 2. NAT proteins phylogenetic tree. Protein accesion numbers are listed in the legend to Figure 1.

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*-acetylamino-peptidase (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). *N*α-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)

Available online at www.sciencedirect.com
ScienceDirect
Journal of Chromatography B, 852 (2007) 1–7
JOURNAL OF CHROMATOGRAPHY B
www.elsevier.com/locate/chromb

Comparison of different ligand densities in immunoaffinity chromatography of the plantibody HB-01 coupled to Sepharose CL-4B to purify the rHBsAg

Rodolfo Valdés [a], Yenisley Medina [a], William Ferro [a], Biunayki Reyes [e,*], Déborah Geada [b], José Montero [a], Tatiana Alvarez [a], Alberto Leyva [c], Leonardo Gómez [a], Sigifredo Padilla [a], Leonardo Pacín [a], Alejandro Figueroa [a], Andrés Tamayo [a], Lorely Milá [d], Yurisleydi Aldama [d], Galina Moya [d], Jorge Reonde [d], María del Carmen Abrahantes [e]

[a] *Monoclonal Antibody Production Department, Center for Genetic Engineering and Biotechnology, P.O. Box 6162, Havana 10600, Cuba*
[b] *Tobacco Research Institute, Aguiar #360 be/Obispo and Obrapía, Havana 10100, Cuba*
[c] *Process Control Department, Center for Genetic Engineering and Biotechnology, P.O. Box 6162, Havana 10600, Cuba*
[d] *Quality Control Direction, Center for Genetic Engineering and Biotechnology, P.O. Box 6162, Havana 10600, Cuba*
[e] *Process Development Direction, Center for Genetic Engineering and Biotechnology, P.O. Box 6162, Havana 10600, Cuba*

Received 10 May 2006; accepted 3 December 2006
Available online 17 December 2006

Abstract

This paper evaluates the immunopurification behavior of a plantibody HBsAg specific plantibody coupled to Sepharose CL-4B at different ligand densities. Results show no significant differences in the adsorption and elution capacities, and rHBsAg recovery of immunosorbents at 3.43, 4.45, and 5.31 mg/mL of ligand densities compared to its mouse-derived mAb counterpart consistently used in the rHBsAg purification process. Therefore, plantibody ligand densities higher than 3.43 mg/mL do not improve the immunopurification behavior of this immunosorbent, but increase the antibody consumption and the Hepatitis B vaccine cost. Immunosorbent of 2.23 mg/mL of ligand density demonstrated a poor performance. The IgG leached detectable level never exceeded the approved limit (3 ng IgG/μg rHBsAg). Values close to this limit were only observed at the ligand density of 5.31 and 2.27 mg/mL. In the case of the ligand density of 2.23 mg/mL the IgG leached value was high (2.90 ng IgG/μg rHBsAg) due to a low level of eluted antigen. In conclusion, it supports feasibility of using this plantibody at 3.43 mg/mL of ligand density for large-scale immunopurification of rHBsAg for human use, avoiding the biosafety and ethical concerns of the massive use of animals for this purpose.

Keywords: Plant-derived antibody; Plantibody; Monoclonal antibody; Hepatitis B virus surface antigen; Immunoaffinity; Immunopurification

1. Introduction

Hepatitis B virus infection is the most common cause of liver disease worldwide [1,2]. Vaccines against this virus have been available since the 1980s and more than 1 billion doses have been employed [3–5]. The immense majority of these vaccines use the rHBsAg as the active pharmaceutical ingredient, which is purified using several procedures, such as immunopurification chromatography based on mouse mAb [6,7]. This affinity separation is a well-established technique for the purification and recovery of biological molecules [8,9], but the regulatory and ethical constrains of mAb production in animals break the massive application of this technique for large-scale production of biological products, forcing the researchers to evaluate other production alternatives.

Many pharmaceutical proteins of mammalian origin have been synthesized in plants [10]. It has gained a great importance in the last years [11] since it represents a cost-effective system for the large-scale production of pharmaceutical antibodies and

Abbreviations: rHBsAg, recombinant Hepatitis B virus surface antigen; mAb, mouse-derived monoclonal antibody; plantibody, plant-derived antibody

* Corresponding author. Fax: +537 271 3208.
E-mail address: biunayki.reyes@cigb.edu.cu (B. Reyes).

1570-0232/$ – see front matter © 2006 Elsevier B.V. All rights reserved.
doi:10.1016/j.jchromb.2006.12.018

provides an additional level of safety compared with animals, because plant viruses are not known to infect humans [12].

Within plants the main advantages of tobacco plant include the nature technology for gene transfer and expression, the high biomass yield, the potential for rapid scale-up owing to prolific seed production, and the availability of large-scale biomass infrastructure for processing [11]. In this work, different ligand densities of the plant-derived antibody HB-01 expressed in *Nicotiana tabacum* plants specific for the aminoacid sequence CKTCTT of the HBsAg "a" determinant were evaluted [13,14]. The determination of the optimal ligand density contributes valuable information for reducing the product cost and the ligand level leached from the matrix and for making a precise estimation of the required production capacity.

2. Experimental

2.1. Antigen source

A recombinant strain of *Pichia pastoris* was fermented in common inorganic salts (e.g. $(NH_4)_2SO_4$, KH_2PO_4, $MgSO_4$) supplemented with glycerol and vitamins (e.g. biotin, riboflavin). Methanol was continuously added at a flow of 2.9–10.9 g L^{-1} h^{-1} to induce the synthesis of rHBsAg. The level of methanol never reached values higer than 2% because it is a toxic level for this yeast. The rHBsAg was purified according to the procedure described by Pentón et al. [5].

2.2. Monoclonal antibody source

The murine B-lymphocyte hybridoma (CB.Hep-1) was previously generated by Fontirrochi et al. [15]. It was obtained by a fusion of Sp2/0-Ag14 myeloma cells and spleen cells of a BALB/c mouse immunized with a natural HBsAg. The hybridoma CB.Hep-1 secretes an IgG-2b mAb, directed against the "a" determinant of the HBsAg [16]. This mAb is routinely produced by ascites to be employed as an immunoligand at 3.8 mg/mL of ligand density in the rHBsAg purification process. The antigen obtained by this procedure is the active pharmaceutical ingredient of the commercially available HB vaccine (Heberbiovac^{TM} HB, HeberBiotec S.A., Cuba) [5].

2.3. Plant-derived antibody HB-01 production

Transgenic tobacco plants expressing the antiHBsAg plantibody HB-01 generated as described by Ramírez et al. [13] were used for this study. Seedlings were transplanted in Zeolite and leaves were harvested and wetted with phosphate buffered saline (PBS), pH 8.0 (400 mL of PBS 150 mM/Kg of leave) and 0.56 mM ascorbic acid at 4 °C to be ground by a Fitzmill Comminutor (The Fitzpatrick Company, USA). The green fraction was removed from the supernatant by filtration using a Rina basket centrifuge at 1051.38 × *g* (Riera Nadeu S.A., Spain) and the supernatant was centrifuged again at 19,635 × *g* in a CEPA tubular centrifuge (Carl Padberg, Germany). The purification of plantibody HB-01 was carried out by means of an affinity chromatography using a recombinant protein A Streamline column (Amersham-Biosciences, Uppsala, Sweden) and finally a buffer exchange from affinity elution buffer to 20 mM Tris-HCl/150 mM NaCl pH 7.2 was carried out by gel filtration chromatography using Sephadex G-25 (Amershan-Biosciences, Uppsala, Sweden) [14].

2.4. Immunoaffinity matrix

Sepharose CL-4B (Amersham-Biosciences, Uppsala, Sweden) was activated by the CNBr method [17]. The mouse-derived antibody and the plantibody were coupled as recommended by the manufacturer (Amersham-Biosciences, Uppsala, Sweden). The amount of coupled antibody was determined by measuring the total protein before and after the coupling reaction. In all cases, the antibody concentration to start the immobilization process was similar to the expected ligand densities.

2.5. Immunoaffinity chromatography

Gels (12.1 mL) were packed into PD10 columns (Amersham-Bioscences, Uppsala, Sweden) and equilibrated with the buffer containing 20 mM Tris–HCl/3 mM EDTA/3 mM NaCl pH 7.2. Adsorption and elution flow rates were 20 and 35 mL h^{-1}, respectively. These flow rates were chosen to improve the adsorption and to speed up the elution. The columns were loaded with 7 mg of a purified rHBsAg preparation diluted in the equilibrium buffer in each purification cycle. After washing, the bound antigen was eluted with 20 mM Tris/3 M potassium thiocyanato/3 mM EDTA, pH 7.0, and monitored at 280 nm. All experiments were done at 22 ±2 °C.

2.6. Gel filtration chromatography

The eluted antigen buffer exchange to water was performed by gel filtration chromatography in a PD10 column with Sephadex G-25 (Amersham-Biosciences, Uppsala, Sweden). The column equilibrium was done with 3-column volumes of purified water and the volume of the applied samples was 2.5 mL. The buffer exchange to purified water was done because this antigen is quite stable under these conditions, allowing the evaluation of the concentration and purity by SDS-PAGE without the interference of any buffer.

2.7. Estimation of antigen-specific antibody activity by enzyme linked immunoassay

A polystyrene (PE) microplate (Costar, Cambridge, USA) was coated with 10 μg per well of HBsAg in 0.1 M $NaHCO_3$ buffer for 20 min at 50 °C. After this step samples were added to the plate in 0.05% Tween 20 in PBS and incubated for 1 h at 37 °C. After several washes with 0.05% Tween 20/PBS the plate was incubated for 1 h at 37 °C with an anti-mouse IgG-horseradish peroxidase conjugate (Sigma Chemical, St. Louis, USA). The reaction was revealed using 100 μL/well of 0.05% orthophenylenediamine and 0.015% H_2O_2 in citrate buffer, pH 5.0, and stopped with 50 μL/well of 1.25 M H_2SO_4. The

absorbance was measured in a Multiskan ELISA reader (Labsystems, Helsinki, Finland) using a 492 nm filter [18].

2.8. Estimation of leached antibody by enzyme linked immunoassay

A sandwich ELISA determined the concentration of IgG leached from the matrixes. Briefly, a plate (Costar, Cambridge, USA) was coated with sheep anti-mouse polyclonal immunoglobulin overnight at 4 °C. The plate was blocked with 1% fat-free milk in PBS for 30 min at 37 °C. Wells were washed and the eluted samples from the immunosorbents were added and incubated for 3 h at 37 °C with 1% non-fat milk dried powder in PBS. After washing the plate, it was incubated with 100 μL/well of a goat anti-mouse polyclonal immunoglobulin-horseradish peroxidase conjugate at a work dilution 1:9000 (Sigma Chemical, St. Louis, USA). The reaction was then revealed using 100 μL/well of 0.05% orthophenylenediamine and 0.015% H_2O_2 in citrate buffer (pH 5.0) and stopped with 50 μL/well of 1.25 M H_2SO_4. The absorbance was measured in a Multiskan ELISA reader (Labsystems, Helsinki, Finland) using a 492 nm filter.

2.9. Protein determination

Protein concentration was performed by the method described by Lowry et al. [19], using bovine serum albumin as standard material. Absorbance measurement at 280 nm was used for the quantification of the purified antibodies and antigens. The molar extinction coefficients used were 1.37 for mAb CB.Hep-1 and plantibody and 5.0 for rHBsAg [20].

2.10. SDS-PAGE

Samples were analyzed by electrophoresis on 12.5% sodium dodecyl sulfate (SDS)–polyacrylamide gels as described by Laemmli [21]. Separated proteins were stained with Coomassie blue R-250 and then analyzed by gel densitometry.

2.11. Western blot

Proteins were transferred by a semi-dry electrophoretic transfer with 25 mM Tris/192 mM glycine/20% methanol at 25 V for 20 min onto a nitrocellulose membrane (Scheiler & Schuell, Daseel, Germany) [22]. After incubating with 1% non-fat milk dried powder in PBS for 1 h at 37 °C the membrane was washed three times in PBS and incubated for 1 h at room temperature with 100 μL of a goat anti-mouse IgG conjugate to horseradish peroxidase (Sigma Chemical, St. Louis, USA). Bands were visualized by reacting with the substrate solution (5 mg of 3,3-diaminobenzidine/10 μL of 30% H_2O_2/10 mL PBS). The reaction was stopped with deionized water.

2.12. HPLC gel filtration

A HPLC-GF column TSK G3000 PW was used to determine the purity of the mAb and also of the plantibody. The mobile phase of the chromatography was PBS pH 7.0 and the volumetric flow rate employed was 0.2 mL min^{-1}. In both cases, the buffer of the samples was previously exchanged from 20 mM Tris-HCl/150 mM NaCl pH 7.2 to PBS pH 7.0 using a PD 10 column loaded with Sephdex G-25 (Amersham-Biosiences/Uppsala, Sweden).

2.13. Isoelectrofocussing

A high-resolution electrophoretic technique was applied to resolve proteins and peptides based on their isoelectrofocussing. The conditions used were a linear pH gradient, ranging from 5.0 to 8.0 (PHASTGEL IEF, Amershan-Biosciences, Uppsala, Sweden); and a homogeneous polyacrylamide gel (5%T, 3%C). An inner standard of mAb CB.Hep-1 was used as a reference material.

2.14. Statistical analysis

Adsorption and elution capacities, recovery and leached IgG were evaluated by ANOVA (simple factor). The significance level (a) was 0.05, and the STATISTICA for Window application was used.

3. Results and discussion

Immunoaffinity chromatography offers a high yield and purity in a single purification step simplifying a further downstream process. In this case, this technique acquires a higher meaning because this antibody is directed against the "a" determinant of the HBsAg, which is important to raise a protective human immune response against this virus [5,14,16,23]. Therefore, it is an additional guarantee of the immunological properties of the purified active pharmaceutical ingredient of the Hepatitis B vaccine.

In order to purify the rHBsAg, the plantibody HB-01 was obtained from tobacco leaves with a yield of up to 12 mg IgG/Kg of biomass, SDS-PAGE (Fig. 1) and HPLC-GF (Fig. 2) purity over 95%. Differences in the SDS-PAGE and HPLC-GF purity were not observed between mAb CB.Hep-1 and plantibody HB-01. Differences in the isoelectrofocussing pattern were observed (Fig. 3). The pattern was characterized by the presence of 7–8 bands with isoelectric pH value ranged from 7.5 to 6.8. The average theoretical isoelectric pH was 6.9, and this value was obtained from the analysis of the amino acid sequence (Fig. 4).

The nature of the matrix to which proteins are attached is important in several respects. It must be physically and chemically stable under the experimental conditions, must have satisfactory flow properties and must be free from non-specific adsorption effects. The open pore structure of Sepharose-4B allows the use of this matrix for gel filtration chromatography of proteins up to 20 million Daltons and also displays virtually all the desirable features of a matrix for protein immobilization. The immunopurification behavior of the immunosorbents was evaluated using 12.1 mL of matrix (IgG-Sepharose CL-4B) that represents the 0.1% of the real production scale (12 L) and several experiments were previously done at this scale (0.1%)

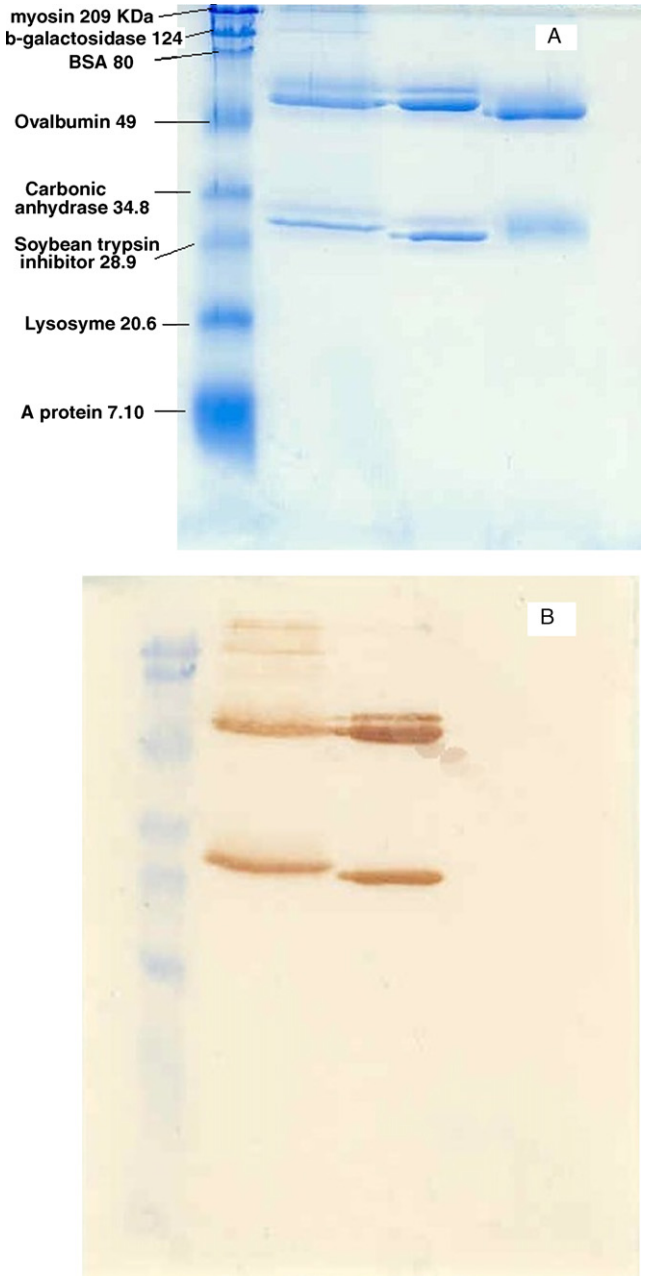


Fig. 1. Coomassie blue stained SDS-PAGE of plantibody HB-01 purified by expanded bed adsorption method and mAb CB.Hep-1 purified by packed bed adsorption. (A) Lane 1, molecular weight marker lane 2, plantibody HB-01 (5 μg); lane 3, mAb CB.Hep-1 (5 μg), lane 4, goat IgG (5 μg). (B) Immunoblot analysis, samples were transferred onto nitrocellulose and antibodies were detected with a horseradish peroxidase labeled goat anti-mouse IgG. Lane 1, pre-stained molecular weight marker; lane 2, plantibody HB-01 (5 μg); lane 3, mAb CB.Hep-1 (5 μg), lane 4, goat IgG (5 μg, negative control).

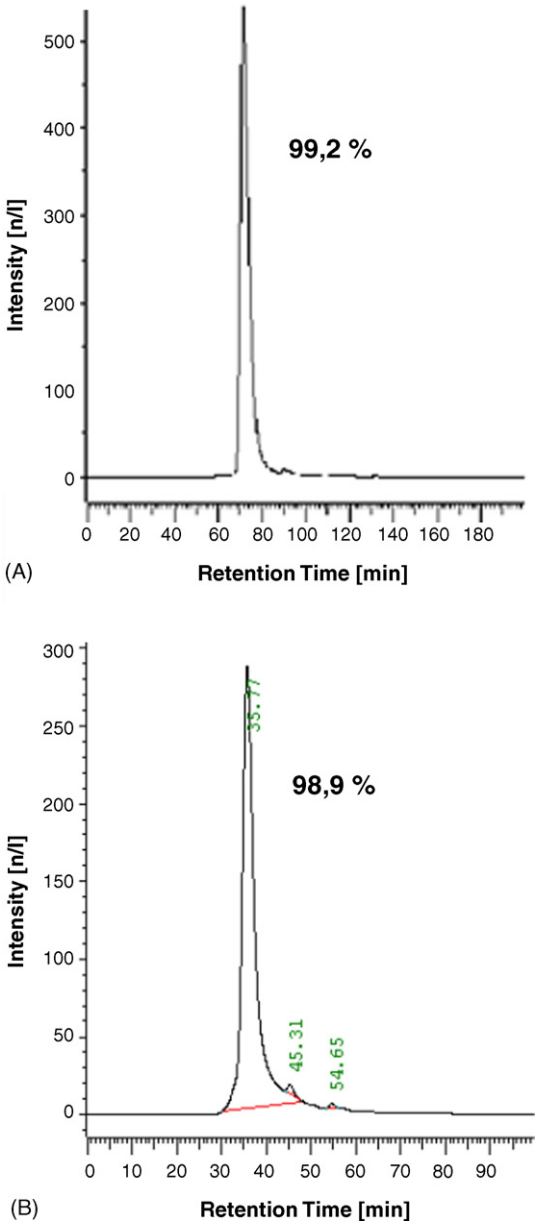


Fig. 2. HPLC gel-filtration profile of the mAb CB.Hep-1 (A) and plantibody HB-01 (B).

demonstrating a rHBsAg recovery of about 30–50% and biocomparability between antigens purified at both scales (data not shown).

On the other hand, the most popular, randomly oriented immobilized ligand activation procedure for agarose is the cyanogen bromide method. A large number of reaction schemes have been proposed for the coupling of proteins and other biopolymers to insoluble matrices, and the cyanogen bromide method results in an active product to which substances may be covalently coupled easily under mild conditions. Cyanogen bromide reacts with the hydroxyl groups of Sepharose to form imidocarbonate and carbamate groups. During the subsequent coupling of the protein to the active product, the imidocarbonate groups react with amino groups belonging to the protein with the formation of stable covalent linkages [17].

In this study, we select the cyanogen bromide method to immobilize the mAb CB.Hep-1 and the plantibody HB-01; this selection was made based on the number of (k) (lysine) present in the amino acid sequence of the mAb CB.Hep-1 and plantibody HB-01 molecules (Fig. 4). The coupling efficiency was over 92% for a final ligand density of 2.23, 3.41, 3.82, 4.45 and 5.31 mg/mL that is coincident with the coupling efficiency reported for this method. Nevertheless, this parameter

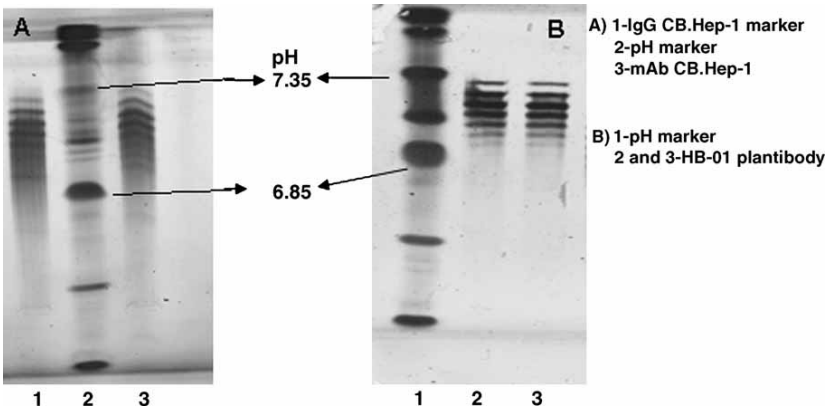


Fig. 3. Isoelectrofocussing pattern of the mAb CB.Hep-1 (A) and the plantibody HB-01(B). (A) Sample per lanes: (1) IgG CB.Hep-1 marker; (2) pH marker; and (3) CB.Hep-1 mouse antibody; (B) sample per lanes: (1) pH marker; (2 and 3) plantibody HB-01.

MDIVMSQSPSSLAVSVGE**K**VALSC**K**SSQSLLYLNNH**K**NYLAWFQQ**K**PGQSP**K**LLIYWASTRDSGVPDR
FTGSGSGTDFTLMISSV**K**AEDLAVYYCQQYYNYPYTFGGGT**K**LEI**K**RADAAPTVSIFPPSSEQLTSGGA
SVVCFLNNFYP**K**DINV**K**W**K**IDGSERQNGVLNSWTDQDS**K**DSTYSMSSTLTLT**K**DEYERHNSYTCEATH
KTSTSPIV**K**SFNRNECSR**K**DEL

(A) Plant-derived antibody HB-01 Light Chain

MEV**K**LDETGGGLVQPGRPM**K**LSBVASGFTFSDFWMNWVRQSPE**K**GLEWVAQIRD**K**PDNYAIYYSESV
KGRFTISRDDSRSSVFLQMNSLRPEDTGIYYBTAGFDYWGQGTTLTVSSA**K**TTPPSVYPLAPGBGDTT
GSSVTLGBLV**K**GYFPESVTVTWNSGSLSSSVHTFPALLQSGLYTMSSSVTVPSSTWPSQTVTBSVAHP
ASSTTVD**K**K**K**LEPSGPISTINPBPPBPPB**K**EBH**K**BPAPNLEGGPSVFIFPPNI**K**DVLMISLTP**K**VTBVVVDV
SEDDPDVQISWFVNNVEVHTAQTQTHREDYNSTIRVVSTLPIQHQDDWMSG**K**EF**K**B**K**VNN**K**DLPSPIE
RTIS**K**I**K**GLVRAPQVYILPPPAEQLSR**K**DVSLTBLVVGFNPGDISVEWTSNGHTEENY**K**DTAPVLDSDGS
YFIYS**K**LNM**K**TS**K**WE**K**TDSFSBNVRHEGL**K**NYYL**K**K**K**TISRSPG**K**SR**K**DEL

(B) Plant-derived antibody HB-01 Heavy Chain

Fig. 4. Aminoacid sequence of plantibody HB-01. In bold (K) the potential linkage points with CNBr activated matrix. (A) Aminoacid sequence of light chain with 16 potential linkage points, 6 of them inside the variable region of each chain. (B) Aminoacid sequence of heavy chain with 30 potential linkage points, 6 potential linkage points inside the variable region of each chain.

was always inferior in the case of the plantibody HB-01 in comparison with the mAb purified from ascitic fluid (Table 1); thus, further experiments should be conducted to investigate this difference.

Each column was purified between 4.03 and 30.69 mg of rHBsAg in 12 purification cycles (Table 1) for a rHBsAg recovery between 4.52 and 59.37%. The higher recovery values were observed at the ligand densities of 3.41, 4.45, 5.31 mg/mL and the control (mouse mAb column, 3.82 mg/mL). The statistical analysis of the immunopurification parameters (adsorption capacity, elution capacity, and recovery) did not show significant differences ($p = 0.89$, 0.74 and 0.51, respectively) corroborating the similitude between the mouse-derived immunoaffinity column and the plantibody immunoaffinity columns behavior at these ligand densities. Conversely, only 8.11 mg rHBsAg were adsorbed and 4.03 mg eluted from the matrix coupled with plantibody HB-01 at a ligand density of 2.23 mg/mL showing significant differences ($p = 5.96715\text{E-}11$) with the rest of ligand densities evaluated. This may be explained by the fact that a low antibody amount in regard to the number of active group could compromise also the antigen recognition site with the matrix where there are several potential linkage points with the activated groups (Fig. 4).

The reason why the rHBsAg recovery is relatively low (30–50%) is still unclear. We hypothesize that rHBsAg particles size makes this immunochromatography inefficient. As

Table 1
Immunopurification behavior of the plantibody HB-01columns in 12 purification cycles

Ligand density (mg/mL)	Coupling efficiency (%)	Total adsorbed rHBsAg (mg)	Total purified rHBsAg (mg)	Recovery of rHBsAg (%)	Average of leached IgG (ngIgG/μgrHBsAg)	IgG leached (%)
2.23	92.21	8.11 ± 0.26	4.03 ± 0.16	4.52 ± 2.26	2.90 ± 2.93	0.098
3.4	92.48	65.55 ± 1.81	30.68 ± 0.89	46.80 ± 12.79	0.13 ± 0.08	0.023
4.45	94.62	45.69 ± 0.83	24.93 ± 0.85	54.56 ± 12.32	0.79 ± 0.50	0.009
5.31	98.05	42.09 ± 1.03	24.99 ± 0.32	59.37 ± 4.94	3.29 ± 0.99	0.031
3.82 (mAb column)	98.10	63.46 ± 2.20	30.69 ± 1.26	48.33 ± 17.91	0.88 ± 0.50	0.037

can be extrapolated from the Table 1, about 45% of the applied antigen pass through the column and can be quantified in the non-bound fraction, even using a low linear flow rate during the adsorption process. We have also demonstrated that there is rHBsAg which is retained into the matrix in every purification cycle and that the matrix regeneration buffers (Buffer 1: 0.1 M Tris-HCL + 0.5 M NaCl pH 8.5 and Buffer 2: 0.1 M Tris + 0.5 M NaCl + 0.1 M AcNa pH 4.0) are strong to eliminate unspecifically bound proteins, carbohydrates and lipids but not the retained rHBsAg.

Several factors can cause a loss of column capacity during a repeated operation. One of the most important is the irreversible denaturalization of the antibody, usually caused by harsh elution conditions [24]. The plantibody HB-01 columns (ligand density: 3.41, 4.45, and 5.31 mg/mL) showed rapid elution capacity decrease similar to that of the mAb column. These rapid decreases seem to depend on the use of the chaetropic agent in the elution buffer and on the following of the matrix with non-specific adsorbed antigen, perhaps provoked by the lower ion exchange character of the adsorbent. Other factors, such as contaminants cannot be considered because the applied rHBsAg was always pure (≥95%).

Ligand leached contributes impurities to the product. Taking into consideration that the antigen eluted from these columns is intended for pharmaceutical use, this parameter raises great importance. This immunoaffinity chromatography is the first step of the chromatographic purification process of the rHBsAg, for that reason an IgG level ≤3 ngIgG/μgrHBsAg has been established as the approved maximum co-elution level, because the subsequent downstream process is able to assure a non-immunogenic level of IgG in the Hepatitis B vaccine [22]. In this work, the average IgG detectable level never exceeded the approved limit. Table 1 and Fig. 5 show clearly the average amount of the IgG leached from the matrixes. A value close to the limit was observed only at the ligand density of 5.31 mg/mL, constituting this another aspect against the use of high ligand densities and also in the immunosorbent of ligand density 2.23 mg/mL, but in such case the ratio IgG/rHBsAg was affected by the low level of antigen eluted from the column. The percentage of IgG leached for this immunosorbent was only 0.098% (Table 1). In addition, the ligand leached was not the major cause of the elution capacity decrease because the total leached IgG was lower than 0.029% of the total coupled IgG

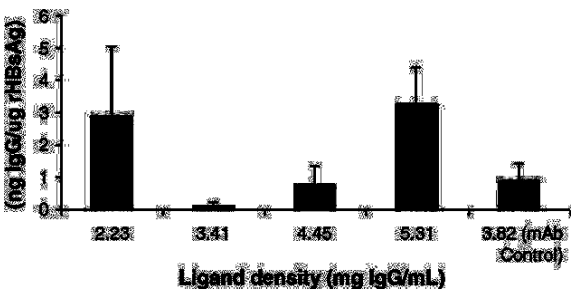


Fig. 5. Antibody leached from the immunochromatography matrixes. The approved value of IgG leached for the immunochromatography eluates in 3 ngIgG/μgrHBsAg.

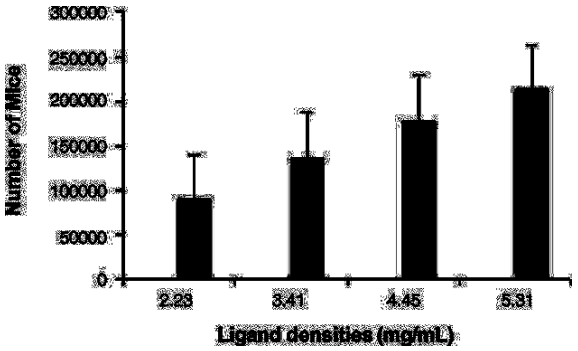


Fig. 6. Total amount of mice required for the production of 200 L of immunosorbents. This volumen correspond with our installed production capacity.

for plantibody HB-01 columns and 0.037% for mAb CB.Hep-1 column (Table 1).

These results validate the use of this plantibody for the immunopurification of the rHBsAg for human use, opening a new alternative to overcome constrains of mAb CB.Hep-1 production in mice. Fig. 6 shows the amount of mice required for the production of 200 L of this immunosorbent. Therefore, as soon as the plantibody HB-01 is introduced in the large-scale production of the active pharmaceutical ingredient of this Hepatitis B vaccine, thousands of mice will not be required for this purpose.

4. Conclusion

These results support the feasibility of using the plantibody HB-01 for the large-scale immunopurification of the rHBsAg for human use at a ligand density of 3.47 mg/mL because it demonstrated the same immunopurification behavior of the mouse-derived antibody column routinely used for this purpose.

Acknowledgment

Authors greatly acknowledge the rHBsAg Production Department for providing the purified recombinant Hepatitis B surface antigen.

References

[1] D.S. Dane, C.H. Cameron, M. Briggs, Lancet 1 (1970) 7651.
[2] H.D. Shen, K.B. Choo, S.D. Lee, Y.T. Tsai, S.H. Han, J. Med. Virol. 18 (1986) 201.
[3] G.A. Bitter, K.M. Egan, W.N. Burnette, B. Samal, J.C. Fieschko, D.L. Peterson, M.R. Downing, J. Med. Virol. 25 (1988) 123.
[4] S. Cadranel, S. Zeglhache, S. Fernández, A. Safary, F. André, Postgrad. Med. J. 63 (2) (1987) 59.
[5] E. Pentón, L. Herrera, V. Muzio, V. Ramírez, A. García, C. Duarte, C. Ruiz, M. Izquierdo, L. Pérez, G. Fontirrochi, M. González, M. Nazabal, A. Beldarían, G. Padrón, J. García, G. de la Riva, A. Santiago, F. Ayán, R. Páez, A. Agraz, R. Díaz, Y. Quiñónez, Eur. Pat. Apl. 480 (1992) 525.
[6] D. Escors, A. Izeta, C. Capiscol, L. Enjuanes, J. Virol. 77 (14) (2003) 7890.
[7] V. Ossipow, P. Fonjallaz, U. Schibler, Mol. Cell. Biol. 19 (2) (1999) 1242.
[8] H.A. Chase, J. Mol. Recognit. 11 (1998) 217.
[9] R.H. Clemmitt, H.A. Chase, Biotechnol. Bioeng. 67 (2) (2000) 206.
[10] J.K.-C. Ma, P.M.W. Drake, P. Christou, Nat. Rev. Genet. 4 (2003) 794.
[11] E.E. Hood, J.M. Jilka, Curr. Opin. Biotechnol. Aug. 10 (4) (1999) 382.

[12] E. Stoger, M. Sack, R. Fischer, P. Christou, Curr. Opin. Biotechnol. 13 (2002) 161.
[13] N. Ramírez, M. Rodríguez, M. Ayala, J. Cremata, M. Pérez, A. Martínez, M. Linares, Y. Hevia, R. Páez, R. Valdés, J.V. Gavilondo, G. Selman-Housein, Biotechnol. Appl. Biochem. 38 (Pt 3) (2003) 223.
[14] R. Valdés, L. Gómez, S. Padilla, J. Brito, B. Reyes, T. Álvarez, O. Mendoza, O. Herrera, W. Ferro, M. Pujol, V. Leal, M. Herrera, Y. Hevia, C. García, L. Milá, O. García, R. Sánchez, A. Acosta, D. Geada, R. Páez, J.L. Vega, C. Borroto, Biochem. Biophys. Res. Commun. 308 (2003) 94.
[15] G. Fontirrochi, M. Dueñas, M.E. Fernzández de Cossio, P. Fuentes, M. Pérez, D. Mainet, M. Ayala, J. Gavilondo, C. Duarte, Biotecnología Aplicada 10 (1993) 24.
[16] M.E. Fernández de Cossío, T. Díaz, A. Galván, R. Valdés, E. González, M. Ayala, J. Díaz, M. Bestagno, O. Burrone, J. Gavilondo, J. Biotechnol. 56 (1997) 69.
[17] J. Kohn, M. Wilchek, Appl. Biotechnol. 9 (1984) 285.
[18] R. Valdés, L. Leyva, E. González, D. Mainet, L. Costa, Biotecnología Aplicada 11 (1994) 219.
[19] D.H. Lowry, N.J. Rosembrough, A.L. Farr, R.J. Randal, J. Biol. Chem. 193 (1951) 265.
[20] C.N. Pace, F. Vajdos, L. Fee, G. Grimsley, T. Gray, Protein Sci. 4 (1995) 2411–2423.
[21] U.K. Laemmli, Nature 227 (1970) 680.
[22] W.N. Burnette, Anal. Biochem. 112 (1981) 195–203.
[23] E. Hardy, E. Martínez, D. Diago, R. Díaz, D. González, L. Herrera, J. Biotechnol. 77 (2000) 157.
[24] A. Agraz, C. Duarte, L. Costa, L. Pérez, R. Páez, V. Pujol, G. Fontirrochi, J. Chromatogr. A 672 (1994) 25.

Proc. Natl. Acad. Sci. USA
Vol. 87, pp. 1947–1949, March 1990
Biochemistry

Sequencing of peptides and proteins with blocked N-terminal amino acids: *N*-Acetylserine or *N*-acetylthreonine

(Edman degradation/amino acid sequence/deblocking/N-acetylated amino acids/phenyl isothiocyanate)

DANIEL WELLNER, C. PANNEERSELVAM*, AND B. L. HORECKER

Department of Biochemistry, Cornell University Medical College, New York, NY 10021

Contributed by B. L. Horecker, December 29, 1989

ABSTRACT **Many proteins cannot be directly sequenced by Edman degradation because they have a blocked N-terminal residue. A method is presented for deblocking such proteins when the N-terminal residue is *N*-acetylserine (which occurs frequently in eukaryotic proteins) or *N*-acetylthreonine. The method has been applied successfully to the determination of the N-terminal amino acid sequence of human, bovine, and rat parathymosins. Prothymosin α and other blocked proteins and peptides were also readily deblocked and sequenced by this procedure. It is proposed that the mechanism of the deblocking reaction involves an acid-catalyzed N → O shift of the acetyl group followed by a β-elimination.**

Although notable improvements have been made recently in the instrumentation available for automated sequencing of proteins and peptides, many proteins still present a challenging problem to investigators who attempt to determine their sequence. One problem frequently encountered is that the N-terminal residue is modified in such a way that it does not react with the Edman reagent phenyl isothiocyanate. For example, the blocked N-terminal residue may be an *N*-acetyl amino acid (1), a glycosylated amino acid (2), or a pyrrolidone carboxylate group (3). Of these, proteins with an N-acetylated amino acid are encountered most frequently. Evidence has been presented that about 80% of the soluble proteins in mammalian cells have acetylated N-terminal amino acids (4).

Previous attempts to remove the blocking group have used enzymatic or limited acid hydrolysis. Nakamura *et al.* (5) have used a rat liver peptidase to remove *N*-acetylserine from the N-terminal peptide released from thrombin by bovine Factor XIII. Both the rat liver peptidase (6) and a similar enzyme from human erythrocytes (7) were shown to split off *N*-acetylserine from α-melanocyte-stimulating hormone. However, the applicability of this method is limited by the restricted specificity of these enzymes. In particular, they do not appear to act efficiently on large peptides or proteins. Proteins blocked with a pyrrolidone carboxylate group often can be unblocked by treatment with pyrrolidone carboxylate peptidase (8). Some success has been reported in removing *N*-acetyl groups from peptides by limited HCl hydrolysis (9, 10).

In this report, a nonenzymatic method is presented for deblocking proteins and peptides having N-terminal acetylserine or acetylthreonine residues. This method has been applied successfully to the elucidation of the primary structure of prothymosin α and parathymosin, acidic polypeptides found in the thymus gland and other mammalian tissues.

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. §1734 solely to indicate this fact.

MATERIALS AND METHODS

Materials. Human, bovine, and rat parathymosins were isolated from liver as described (11). Bovine and rat prothymosins α were isolated from thymus glands (11). Thymosin β_4 was purified from rat thymus (12). Horse heart cytochrome *c* was from Schwarz/Mann. *Drosophila* aldolase (13) was a gift from O. Brenner-Holzach, and the synthetic N-acetylated peptides were gifts from E. P. Heimer of Hoffmann–La Roche. The reagents and supplies used for sequencing were purchased from Applied Biosystems.

Deblocking Proteins for Sequencing. A trifluoroacetic acid-treated glass fiber filter disc, 12 mm in diameter, was inserted slightly folded into a 1.5-ml polypropylene microcentrifuge tube. The filter was wetted with 30 μl of a solution containing Polybrene (3 mg) and NaCl (0.2 mg) and dried. A solution of the protein to be tested was applied to the filter and dried. The filter was then saturated with 30 μl of anhydrous trifluoroacetic acid (sequencer reagent 3), and the tube was closed. After incubating for 4 min at 45°C, the tube was opened in the hood to allow most of the trifluoroacetic acid to evaporate. After 5 min at room temperature, the open tube was allowed to dry for another 10 min at 45°C. The tube was then closed and placed in an oven at 65°C for 16 hr or at 45°C for 3 days. Sequencing was then carried out according to the manufacturer's recommendations in an Applied Biosystems gas-phase sequencer model 470A equipped with an on-line phenylthiohydantoin (PTH) analyzer model 120A. Program O3RPTH was used.

RESULTS

It was found that several proteins that contain an N-acetylated serine residue, after treatment with anhydrous trifluoroacetic acid, undergo a time-dependent reaction in which they become deblocked and amenable to sequencing by the Edman method. Conditions that have been found to work well for a number of proteins are described under *Materials and Methods*. The protein is first treated with anhydrous trifluoroacetic acid for 4 min at 45°C. After removal of the acid, the treated protein is incubated at 65°C for 16 hr or for 3 days at 45°C. Initial yields vary, depending on the protein. Yields ranging from 3% to 40% have been obtained.

Table 1 shows the N-terminal sequence of rat parathymosin obtained by this method. The initial yield was about 7%. The N-terminal serine had previously been shown to be acetylated (11). An identical N-terminal sequence was obtained for bovine parathymosin by using the same method for deblocking the protein (14). The "background" in this se-

Abbreviation: PTH, phenylthiohydantoin.

*Present address: Department of Biochemistry, Post Graduate Institute of Basic Medical Sciences, Madras University, Madras 600113, India.

Table 1. N-terminal sequence of rat parathymosin

Position	1	2	3	4	5	6	7
Residue identified	Ser	Glu	Lys	Ser	Val	Glu	Ala
Amount, pmol	97*	115	55	87*	47	54	50
Yield, %	6.5	7.7	3.7	5.8	3.1	3.6	3.3

Parathymosin (1.5 nmol) was deblocked for 3 days at 45°C as described. The N-terminal residue was shown previously to be acetylated (11).
*Calculated from the sum of PTH-serine and the dithiothreitol adduct of PTH-dehydroalanine.

quencing run was low, and it was possible to sequence the first 20 residues. The same N-terminal sequence was also obtained for human parathymosin and subsequently confirmed by sequencing the cloned cDNA coding for this protein (15).

A number of other blocked proteins and peptides have been successfully sequenced by this method. These include bovine and rat prothymosin α, rat thymosin β_4, and synthetic *N*-acetylserine- and *N*-acetylthreonine-containing peptides.

The application of this procedure to *N*-acetylthreonine-containing peptides is shown in Table 2. An 11-residue peptide was sequenced in its entirety with an initial yield of about 6%. A separate sequencing run carried out without deblocking showed that at least 99.5% of the peptide was blocked.

In experiments with thymosin β_4, when the reaction time at 65°C was varied, it was found that the yield was optimal at about 16 hr (Table 3). Under these conditions, the initial yield was about 40%, substantially higher than with other proteins tested. The yields of two representative amino acids, lysine at position 3 and proline at position 4, are shown in Table 3. There was no increase in yield between 16 and 24 hr.

Examination of the background peaks in the sequencing of thymosin β_4 and other proteins after deblocking suggested that some cleavage occurred at serine and threonine residues in the interior of the peptide chain. However, the extent of cleavage of these internal peptide bonds is considerably less than that of the N-terminal acetyl group. An estimate of the relative amount of internal and N-terminal cleavage may be obtained by measuring the yield of background PTH-glutamate at cycle 3 of the sequence. Thymosin β_4 is a 43-residue peptide containing three serine and three threonine residues in addition to the N-terminal acetylserine (16). The PTH-glutamate peak is relatively prominent in the background at cycle 3 because three of these hydroxyamino acid residues (one serine and two threonines) occur two positions before a glutamate residue. If, during the deblocking procedure, the peptide bond involving the amino group of one of these serine or threonine residues is cleaved, PTH-glutamate will appear at cycle 3. As shown in Table 3, the average background PTH-glutamate peak at position 3 corresponds to about 30% of the average PTH-lysine peak at position 3. If one assumes equal cleavage at all internal serine and threonine residues during deblocking, this would indicate that each internal cleavage corresponds to about 10% of the cleavage of the N-terminal acetyl group (Table 3, bottom line). Therefore, it may be concluded that the N-terminal acetyl linkage is considerably more labile than any internal peptide bond under the conditions used.

Table 3. Sequencing yields with thymosin β_4 after different reaction times at 65°C

Incubation time at 65°C, hr	1	4	10	16	24
Lys-3 yield, %	4.6	10.0	10.6	41	41
Pro-4 yield, %	5.1	10.1	10.0	37	40
Background Glu-3 yield, %	1.8	2.6	2.5	18	9.2
⅓ Glu-3 yield/Lys-3 yield*	0.13	0.09	0.08	0.15	0.07

Thymosin β_4 (2 nmol) was treated with anhydrous trifluoroacetic acid for 4 min at 45° as described. After drying, samples were placed in an oven at 65°C for the time indicated and sequenced.
*See text.

When horse heart cytochrome *c*, a protein with an *N*-acetylglycine N-terminal, was submitted to the deblocking procedure, no sequence was seen. A low background of PTH–amino acid derivatives was observed, suggesting that some cleavage had occurred at internal peptide bonds. This protein contains 10 threonine and no serine residues (17). Evidence for cleavage on the amino side of 9 of the 10 threonine residues was seen, but the yields were only 0.2–0.6%.

Attempts were also made to sequence *Drosophila* aldolase, a 360-residue polypeptide with an *N*-acetylthreonine terminal group (13). Although the data showed that unblocking had taken place and the protein was being sequenced from the N-terminal end, the background was too high for the sequence to be clearly read. This difficulty may be encountered when applying the method to large proteins, particularly those that are rich in serine and threonine. In such cases, it may be advantageous to isolate an N-terminal peptide prior to deblocking and sequencing.

DISCUSSION

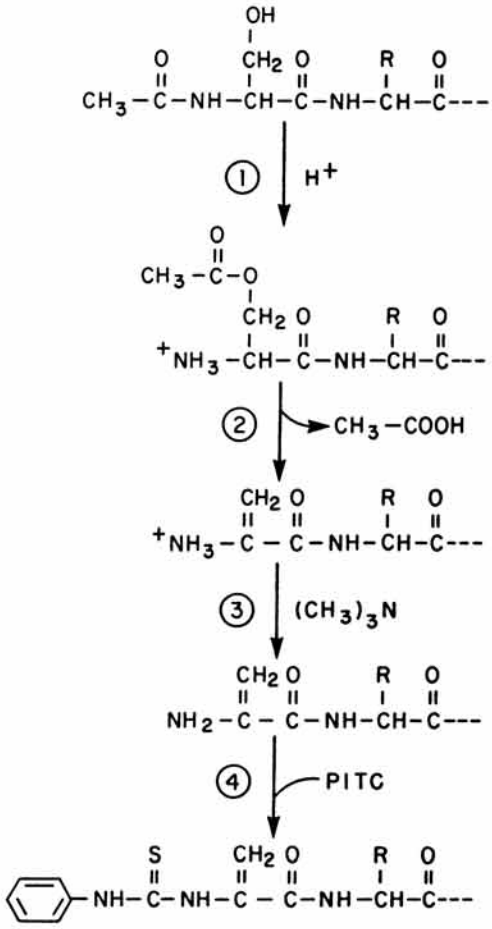
The method described here has been shown to be useful for deblocking proteins and peptides with an N-terminal acetylated serine or threonine. The data show that, after deblocking, the N-terminal sequence of several proteins and peptides could be successfully determined by Edman degradation.

Although the mechanism of the deblocking reaction is not known at this time, the finding that it occurs with proteins having *N*-acetylserine or *N*-acetylthreonine terminal residues but not with cytochrome *c*, which has an *N*-acetylglycine terminus, suggests that an N → O acyl shift may be involved (see Scheme I, reaction 1). Such acyl shifts occur readily under acid conditions and are rapidly reversed when the pH is increased above 7 (18). In addition to an N → O acyl shift, a second reaction must occur before sequencing is started. In the sequencing procedure, the protein or peptide is treated with trimethylamine vapors before and during the reaction with phenyl isothiocyanate. Under these conditions of high pH, the N → O acyl shift would be rapidly reversed, and the amino group would again be blocked. It is possible that the second reaction, which appears to be slower than the N → O transition, is a β-elimination reaction (Scheme I, reaction 2), yielding a protonated dehydroalanyl peptide. In the presence of trimethylamine and phenyl isothiocyanate (PITC in Scheme I), the corresponding phenylthiocarbamoyl derivative may then be formed (Scheme I, reactions 3 and 4). This proposed mechanism is supported by the finding that

Table 2. Sequence of *N*-acetylthreonine synthetic peptide

Position	1	2	3	4	5	6	7	8	9	10	11
Residue identified	Thr	Cys	Asp	Leu	Ala	Pro	Pro	Ala	Gly	Thr	Thr
Amount, nmol	0.79	*	0.74	1.07	1.35	0.90	1.02	1.10	0.66	0.41	0.31
Yield, %	4.0	—	3.7	5.4	6.7	4.5	5.1	5.5	3.3	2.1	1.6

A synthetic peptide (20 nmol) with the sequence *N*-acetyl-Thr-Cys-Asp-Leu-Ala-Pro-Pro-Ala-Gly-Thr-Thr was deblocked and sequenced as described. The incubation was at 65°C for 16 hr.
*The dithiothreitol adduct of PTH-dehydroalanine was seen at this cycle.



Scheme I

most of the PTH derivative observed at the first cycle is that of the dithiothreitol adduct of dehydroalanine. This is in contrast to the finding of a predominant PTH-serine peak usually seen, especially when serine occurs in the early part of a sequence. Although a PTH-serine peak is also seen in the first cycle of deblocked peptides, it may arise by hydration of PTH-dehydroalanine during the conversion step (when the anilinothiazolinone derivative is heated in aqueous acid to form the PTH derivative). Further investigation of this mechanism is needed.

The fact that so many proteins of interest to biochemists and molecular biologists are blocked at the N terminus is hampering efforts to elucidate their structures. Although many different approaches have been used to solve this problem (19), no generally applicable method for unblocking proteins has yet been devised. The present method is restricted to proteins where either serine or threonine is the acetylated N-terminal residue. Nevertheless, it may be applicable to a large number of proteins. A survey of N-acetylated proteins by Persson *et al.* (20) found that, of the known proteins in this class, about 41% have an N-terminal acetylated serine and another 2% an N-terminal acetylated threonine.

This investigation was supported by a Biomedical Research Support Grant and by Grant 1 S10 RR02855-01, both from the National Institutes of Health. Dr. Panneerselvam was the recipient of a Biotechnology Career Fellowship from the Rockefeller Foundation.

1. Narita, K. (1958) *Biochim. Biophys. Acta* **28,** 184–191.
2. Takahashi, N., Takahashi, Y. & Putnam, F. W. (1985) *Proc. Natl. Acad. Sci. USA* **82,** 73–77.
3. Press, E. M., Piggot, P. J. & Porter, R. R. (1966) *Biochem. J.* **99,** 356–366.
4. Brown, J. L. & Roberts, W. K. (1976) *J. Biol. Chem.* **251,** 1009–1014.
5. Nakamura, S., Iwanaga, S., Suzuki, T., Mikuni, Y. & Konishi, K. (1974) *Biochem. Biophys. Res. Commun.* **58,** 250–256.
6. Kobayashi, K. & Smith, J. A. (1987) *J. Biol. Chem.* **262,** 11435–11445.
7. Jones, W. M., Manning, L. R. & Manning, J. M. (1986) *Biochem. Biophys. Res. Commun.* **139,** 244–250.
8. Doolittle, R. F. & Armentrout, R. W. (1968) *Biochemistry* **7,** 516–521.
9. Fordyce, A. M., Midwinter, G. G. & Moore, C. H. (1979) *Biochem. Soc. Trans.* **7,** 721–723.
10. Chin, C. C. Q. & Wold, F. (1985) *Biosci. Rep.* **5,** 847–854.
11. Komiyama, T., Pan, L.-X., Haritos, A. A., Wideman, J. W., Pan, Y.-C. E., Chang, M., Rogers, I. & Horecker, B. L. (1986) *Proc. Natl. Acad. Sci. USA* **83,** 1242–1245.
12. Haritos, A. A., Caldarella, J. & Horecker, B. L. (1985) *Anal. Biochem.* **144,** 436–440.
13. Malek, A. A., Hy, M., Honegger, A., Rose, K. & Brenner-Holzach, O. (1988) *Arch. Biochem. Biophys.* **266,** 10–31.
14. Panneerselvam, C., Clinton, M., Wellner, D. & Horecker, B. L. (1988) *Biochem. Biophys. Res. Commun.* **155,** 539–545.
15. Clinton, M., Frangou-Lazaridis, M., Panneerselvam, C. & Horecker, B. L. (1989) *Biochem. Biophys. Res. Commun.* **158,** 855–862.
16. Low, T. L. K., Hu, S.-K. & Goldstein, A. L. (1981) *Proc. Natl. Acad. Sci. USA* **78,** 1162–1166.
17. Margoliash, E., Smith, E. L., Kreil, G. & Tuppy, H. (1961) *Nature (London)* **192,** 1125–1127.
18. Iwai, K. & Ando, T. (1967) *Methods Enzymol.* **11,** 263–282.
19. Chin, C. C. Q. & Wold, F. (1987) in *Methods in Protein Sequence Analysis–1986*, ed. Walsh, K. A. (Humana, Clifton, NJ), pp. 505–512.
20. Persson, B., Flinta, C., von Heijne, G. & Jörnvall, H. (1985) *Eur. J. Biochem.* **152,** 523–527.

ELSEVIER

Journal of Immunological Methods 314 (2006) 170 – 173

JIM
Journal of
Immunological Methods

www.elsevier.com/locate/jim

Technical note

Non-specific binding to protein A Sepharose and protein G Sepharose in insulin autoantibody assays may be reduced by pre-treatment with glycine or ethanolamine

Alistair J.K. Williams *,[1], Alastair J. Norcross [1], Kyla A. Chandler, Polly J. Bingley

Clinical Science at North Bristol, University of Bristol, UK

Received 13 January 2006; received in revised form 17 May 2006; accepted 7 June 2006
Available online 28 June 2006

Abstract

Insulin autoantibody (IAA) microassays are widely used for predicting type 1 diabetes. As levels of IAA are often low in type 1 diabetes, non-specific binding (NSB) needs to be minimised if assays are to achieve high analytical sensitivity. IAA microassays use protein A Sepharose (PAS) or protein G Sepharose (PGS) to isolate the antibody-bound label, but NSB by the gel can differ between commercially-produced batches. We investigated whether pre-incubation of gel with glycine or ethanolamine could overcome this problem. Batches of PAS/PGS shown to have high NSB (0.3–3.2%) were incubated with glycine or ethanolamine at various pHs between 8 and 10.6 for 2–18 h at 4 °C or room temperature. Treating PAS at pH 10.6 with 0.2 M glycine overnight at room temperature reduced NSB by >84%, with minimal reduction in specific binding (<5%). Treating PGS at pH 10.6 with 0.2 M ethanolamine overnight at 4 °C reduced background by >95%, with minimal reduction in specific binding by most sera. Treatment at high pH was critical in reducing NSB to both PAS and PGS, with slight reduction at pH 8, but a major reduction at pH 10.6. Pre-treatment with glycine or ethanolamine allows "poor" batches of PAS or PGS to be used in sensitive IAA assays, improving both consistency and performance.

Keywords: Insulin autoantibody assays; Non-specific binding; Protein A Sepharose; Protein G Sepharose

Insulin autoantibodies (IAA) are early markers of beta cell destruction in type 1 diabetes and in combination with islet cell autoantibodies (ICA) and autoantibodies to glutamic acid decarboxylase (GAD) and the protein tyrosine phosphatase IA-2 form the basis of disease prediction in relatives (Bingley et al., 1999) and the general population (Bingley et al., 1997). Levels of IAA in patients with type 1 diabetes are often low and therefore non-specific binding (NSB) needs to be minimised if assays are to achieve high analytical sensitivity and optimise discrimination between health and disease. A microassay (Williams et al., 1997) that uses protein A Sepharose (PAS) and/or protein G Sepharose (PGS) to isolate the immune complexes has largely supplanted the large-volume IAA assays which use polyethylene glycol to precipitate serum immunoglobulins. Due to the much

Abbreviations: PAS, protein A Sepharose; PGS, protein G Sepharose; IAA, insulin autoantibodies; NSB, non-specific binding; GAD, glutamic acid decarboxylase.

* Corresponding author. Medical School Unit, Southmead Hospital, Bristol BS10 5NB, UK. Tel.: +44 117 959 5337; fax: +44 117 959 5336.
E-mail address: A.J.K.Williams@bristol.ac.uk (A.J.K. Williams).
[1] These authors made an equal contribution to the paper.

0022-1759/$ - see front matter © 2006 Elsevier B.V. All rights reserved.
doi:10.1016/j.jim.2006.06.003

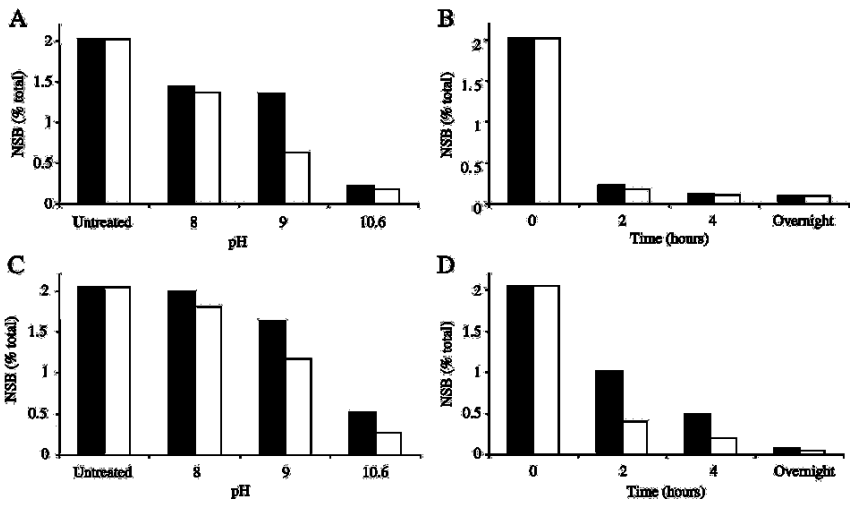


Fig. 1. Effect of incubation for 2 h at different pH (A and C) and at pH10.6 for different times (B and D) on mean non-specific binding of insulin label (NSB) after treating "poor" batches of PAS (A and B) and PGS (C and D) with 0.2 M ethanolamine at 4 °C (filled bars) or 0.2 M glycine at room temperature (open bars) in Tris buffer containing 1% Tween-20.

lower serum requirement, the microassay is the method of choice for measuring IAA in young children and in the most widely-used animal model of the disease, the NOD mouse (Yu et al., 2003).

Like previous assays, specificity of binding in the IAA microassay is confirmed by competitive displacement with excess unlabelled insulin. In our assay the threshold for positivity usually represents less than 0.1% specific binding of the ^{125}I-insulin label added, but we have recently found non-specific binding of label by some batches of commercially available PAS and PGS to be as high as 3.2%, rendering them unsuitable for use. This may partly explain the wide variation in performance of IAA assays from different laboratories demonstrated in international workshops (Bingley et al., 2003), since we found NSB of label by four of six batches of PGS from one manufacturer to be greater than 0.2%. These gels are prepared by coupling the affinity ligands to cyanogen bromide (CNBr) activated Sepharose (Axèn et al., 1967; Kohn and Wilchek, 1984). Since ethanolamine and glycine have previously been used to block unreacted sites of CNBr-activated beads after coupling of proteins (Harlow and Lane, 1988) we investigated whether further treatment of these gels with glycine or ethanolamine could reduce the high levels of NSB.

The effect of different blocking protocols on specific insulin binding with PAS was tested using an insulin antibody-positive serum from a patient with long-standing type 1 diabetes diluted with an antibody-negative human serum and two sera from healthy volunteers. As our mouse IAA assay uses PGS rather than PAS to isolate

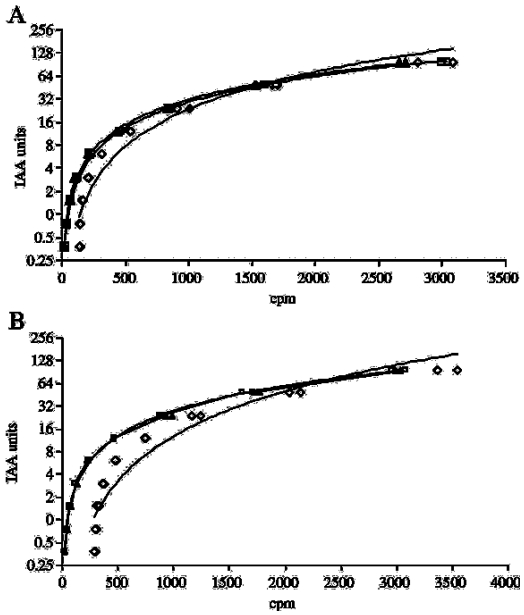


Fig. 2. Effect of pre-treating (A) PAS with 0.2 M glycine and (B) PGS with ethanolamine overnight at pH 10.6 on insulin-binding assay standard curves. Standard curves comprised dilutions of a serum from an antibody-positive long-standing patient with antibody-negative human serum. Untreated "poor" gels (diamonds) show increased background binding with poor logarithmic curve fit (R^2 values of 0.94 for untreated PAS, 0.90 for untreated PGS). After treatment (squares) "poor" batches of gel showed reduced background with quality of curve fit similar to those obtained with an untreated "good" batch (triangles) of gel ($R^2 > 0.99$ for both the "good" and treated gels).

the insulin antibody complexes, the effect of blocking on specific insulin binding with PGS was tested using pooled normal mouse serum (Sigma, Poole, UK), a mouse IgG1 anti-human insulin monoclonal antibody (Hu018, Novo Nordisk, Bagsvaerd, Denmark) (Andersen et al., 1993) diluted in normal mouse serum and an affinity purified polyclonal mouse anti-insulin serum diluted in normal mouse serum.

All samples were assayed for IAA as previously described (Williams et al., 1997). 5 μl serum was incubated with 15,000 cpm ^{125}I-labelled human insulin (Amersham, Little Chalfont, UK) for 72 h at 4 °C and immune complexes isolated using PAS or PGS (Pharmacia, Uppsala, Sweden; Zymed, San Francisco, USA; and Sigma). Results were expressed as percentage binding of label added. Specific binding was calculated by subtraction of counts for serum incubated in the presence of excess unlabelled insulin (Humulin, Lilly, Basingstoke, UK) from counts obtained after incubation with label alone. The assay achieved a laboratory-defined sensitivity of 36% with 100% specificity in the First Diabetes Antibody Standardization Program (Bingley et al., 2003) and 69% sensitivity with 83% specificity in the second murine workshop (Yu et al., 2003).

The PAS/PGS was washed by centrifugation 3–4 times in 150 mmol NaCl to remove the preservative and incubated with 0.2 M glycine (BDH, Poole, UK) or 0.2 M ethanolamine (Sigma) in 50 mmol/l Tris buffer containing 1% Tween-20, at a ratio of 25:1 blocking buffer/gel. Buffer pH was adjusted by addition of 1 M NaOH. The reactions were incubated with mixing at room temperature or at 4 °C. After blocking, the gel was washed in 150 mmol NaCl, made up as a 50% slurry with 20% ethanol in sterile 150 mmol NaCl and stored at 4 °C until use.

Non-specific binding with standards and negative controls by "good" batches of PAS or PGS is usually less than 15 cpm (<0.1% binding of label). Blocking experiments were performed on "poor" batches of PAS ($n=3$) and PGS ($n=1$) with mean NSBs > 0.2% binding of label (range 0.21 to 1.8% and 2.2% respectively). Pre-treatment of all 3 batches of PAS with 0.2 M glycine overnight at pH 10.6 at room temperature reduced the mean NSB to less than 0.12% (range 0.07% to 0.11%), while specific binding by human sera was similar to that found for untreated "good" batches of PAS. Pre-treatment of the PGS batch with 0.2 M ethanolamine overnight at pH 10.6 at 4 °C was similarly effective, reducing the mean NSB from 1.7% to 0.08%, although mean specific binding by the mouse monoclonal antibody Hu018 (diluted 1/4000 with normal mouse serum) was reduced by 51%. The NSB reduced progressively as the pH was increased from pH 8 to pH 9 to pH 10.6 and time-course experiments showed that with both treatments maximal reduction in NSB was achieved after overnight incubation at pH10.6

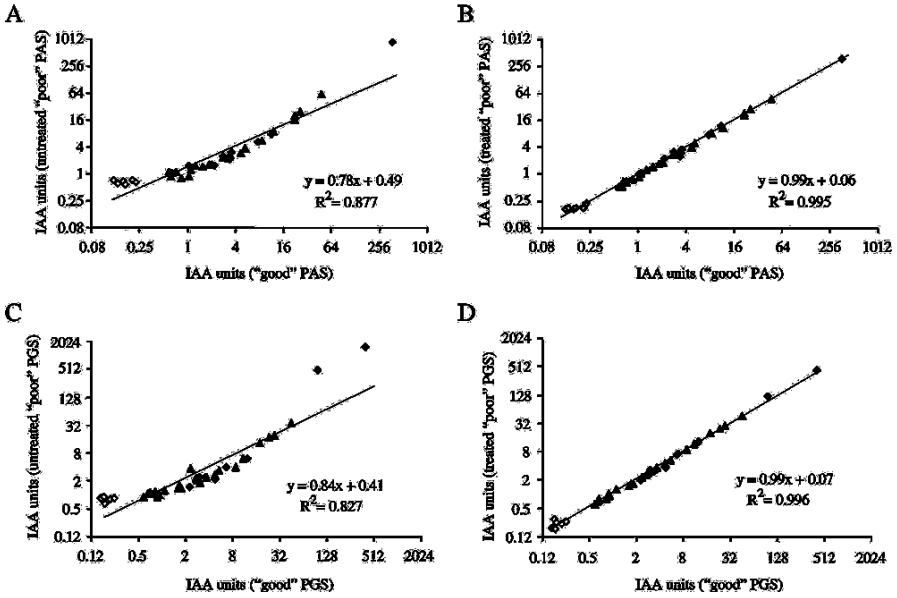


Fig. 3. Insulin binding results obtained with untreated (A) and treated (B) "poor" batches of PAS and untreated (C) and treated (D) "poor" batches of PGS in comparison to those obtained with untreated "good" batches for sera from 6 negative controls (open diamonds), 10 positive controls (filled diamonds) and 24 IAA positive patients with newly-diagnosed type 1 diabetes (filled triangles). Pre-treating "poor" batches of gel greatly improved the correlation of results with those obtained using "good" batches of gel, yielding R^2 values > 0.99.

(Fig. 1). Pre-incubation of PAS or PGS overnight at high pH alone also reduced NSB, although the reduction was greater in the presence of blocking agent (data not shown).

The effect of pre-treatment on insulin binding was demonstrated in our routine screening assay (Williams et al., 1997) using positive and negative control sera (from 2 patients with long standing type 1 diabetes, 9 relatives of patients with type 1 diabetes and 4 healthy volunteers) and sera from 24 IAA positive patients with newly-diagnosed type 1 diabetes. Pre-treatment of PAS with glycine or PGS with ethanolamine overnight at pH10.6 improved the fit of the standard curve from 0.93 to 0.99 and from 0.9 to 0.99 respectively with parameters that agreed much more closely with those obtained using "good" batches of gel (Fig. 2). The regression R^2 of results obtained using "poor" batches of PAS or PGS against those obtained using good batches improved after treatment from 0.87 to 0.99 and from 0.82 to 0.99 respectively (Fig. 3). This pre-treatment had little effect on insulin binding of the human sera when compared to a "good" batch of PAS or PGS. One serum did, however, show 45% higher binding with untreated "poor" PGS and also demonstrated much higher binding with PGS than PAS, suggestive of high levels of IgG3 autoantibodies. Sensitivity of PGS to treatment is therefore likely to be isotype specific and blocking conditions should be optimised for each application.

Blocking can also improve the performance of other assays which use affinity gels to isolate immune complexes, although the blocking method is dependent on the mode of gel activation. Currently we are using PAS prepared by the epoxide method (Sundberg and Porath, 1974) which shows little NSB of insulin, but high NSB in our GAD and IA-2 autoantibody assays (Bingley et al., 1999) when compared to batches prepared by the cyanogen bromide method. Pre-incubation of this gel with 1% BSA in Tris buffer at pH 7.5 for 4 days at room temperature reduced background binding by up to 60% (data not shown).

High NSB of insulin by some batches of commercially available PAS or PGS can seriously limit the ability of IAA microassays to detect low levels of antibodies. Pre-treatment with glycine or ethanolamine at high pH can greatly reduce NSB of some gels and allows the use of even relatively "poor" batches of PAS of PGS in sensitive assays.

References

Andersen, L., Dinesen, B., Jørgensen, P.N., Poulsen, F., Røder, M.E., 1993. Enzyme immunoassay for intact human insulin in serum or plasma. Clin. Chem. 39, 578.

Axèn, R., Porath, J., Ernback, S., 1967. Chemical coupling of peptides to polysaccharides by means of cyanogens halides. Nature 214, 1302.

Bingley, P.J., Bonifacio, E., Williams, A.J.K., Genovese, S., Bottazzo, G.F., Gale, E.A.M., 1997. Prediction of IDDM in the general population. Strategies based on combinations of autoantibody markers. Diabetes 46, 1701.

Bingley, P.J., Williams, A.J.K., Gale, E.A.M., 1999. Optimized autoantibody-based risk assessment in family members. Diabetes Care 22, 1796.

Bingley, P.J., Bonifacio, E., Mueller, P.W., participating laboratories, 2003. Diabetes Antibody Standardization Program: first assay proficiency evaluation. Diabetes 52, 1128.

Harlow, E., Lane, D., 1988. Chapter 13: Coupling with activated beads. In: Harlow, E., Lane, D. (Eds.), Antibodies: A Laboratory Manual. Cold Spring Harbor Laboratory, New York.

Kohn, J., Wilchek, M., 1984. The use of cyanogens bromide and other novel cyanylating agents for the activation of polysaccharide resins. Appl. Biochem. Biotechnol. 9, 285.

Sundberg, L., Porath, J., 1974. Preparation of adsorbents for biospecific affinity chromatography. 1. Attachment of group-containing ligands to insoluble polymers by means of bifunctional oxiranes. J. Chromatogr. 90, 87.

Williams, A.J.K., Bingley, P.J., Bonifacio, E., Palmer, J.P., Gale, E.A.M., 1997. A novel micro-assay for insulin autoantibodies. J. Autoimmun. 10, 473.

Yu, L., Eisenbarth, G., Bonifacio, E., Thomas, J., Atkinson, M., Wasserfall, C., 2003. The second murine autoantibody workshop: remarkable interlaboratory concordance for radiobinding assays to identify insulin autoantibodies in nonobese diabetic mice. Ann. N. Y. Acad. Sci. 1005, 1.