Copyright is owned by the Author of the thesis. Permission is given for a copy to be downloaded by an individual for the purpose of research and private study only. The thesis may not be reproduced elsewhere without the permission of the Author.

THE SEQUENCES OF THE TRYPTIC PEPTIDES FROM ACTINIDIN

A thesis presented in partial fulfillment

for the degree of Doctor of Philosophy

at Massey University.

Alan Carne

December, 1976

ABSTRACT

Actinidin is a plant thiol protease which is isolated from the fruit of Actinidia chinensis, the chinese gooseberry. Determination of the primary amino acid sequence of actinidin was undertaken to extend the limited structural information available on this group of enzymes, and therefore enable a better understanding of their physical and chemical properties.

The order of arrangement of the 220 amino acid residues in the primary sequence of actinidin was determined from the sequences of the tryptic peptides. S-carboxy[$^{14}C_2$]methyl actinidin was digested with trypsin, and the twelve tryptic peptides produced were initially separated into seven fractions by gel chromatography on Sephadex G-50. The first four fractions contained tryptic peptides that were purified by DEAE-cellulose chromatography. The last three fractions contained peptides that were sufficiently small to enable purification by paper techniques, and these peptides were sequenced directly by the dansyl-Edman method.

Further degradation of the tryptic peptides purified on DEAE-cellulose with either chymotrypsin, thermolysin, pepsin or Staphylococcus aureus V8 protease was necessary to provide smaller peptides that could be sequenced by the dansyl-Edman method. S. aureus V8 protease was particularly useful in the determination of amide residues, because of the enzyme specificity for the carboxyl groups of glutamic acid.

The fourth tryptic peptide in the sequence of actinidin could not be located in the tryptic peptide elution profiles of either the Sephadex G-50 or DEAE-cellulose columns. The sequence of this peptide was determined from a tryptic peptide obtained by digestion of maleylated carboxymethyl actinidin.

The N-terminal of actinidin was determined by the dansyl Edman method, and the C-terminal by analysis of cyanogen bromide fragments, and by digestion with carboxypeptidase A.

Radioactively labelling the active site cysteine residue with iodo[$^{14}C_2$]acetic acid, and subsequent purification of the radioactive tryptic digest peptide, enabled the isolation of the tryptic peptide containing the active site cysteine residue. Further digestion of this peptide with chymotrypsin and determination of the sequence of the smaller radioactive peptide, provided the sequence about the active site cysteine residue.

Alignment of the tryptic peptides to reconstruct the primary sequence of actinidin was accomplished with information from cyanogen bromide fragments, information from tryptic peptides of maleylated carboxymethyl actinidin, and information from the three dimensional X-ray crystallographic structure of actinidin determined by Dr E.N. Baker.

The low proportion of basic residues and high proportion of acidic residues in actinidin are in agreement with the enzyme being an acidic protein. Colorimetric analysis of the tryptophan residue content, using 2-nitrophenylsulphenyl chloride, confirmed the presence of six tryptophan residues in the sequence of actinidin.

The amino acid sequences about the seven cysteine residues and the single histidine residue in actinidin were very similar to the analogous sequences in papain and other plant thiol proteases. Furthermore, comparison of the primary sequence of actinidin with that of papain, and the fragments of sequence available for other plant thiol proteases, indicated a considerable homology throughout the sequences of these proteins.

ACKNOWLEDGEMENTS

I wish to thank my supervisor Dr C.H. Moore for his continued interest in the project, and for the benefit of his experience in the form of helpful suggestions, throughout the course of this work.

The useful discussions with Dr G.G. Midwinter and Dr E.N. Baker during the project are much appreciated, as is the assistance of Mr P.R. Macdonald and Dr I.M. Morrison with proof reading of the thesis.

I wish also to acknowledge the encouragement given by my parents and Pauline during the course of this work.

CONTENTS

CONTENTS (contd.)

CONTENTS (contd.)

CONTENTS (contd.)

LIST OF FIGURES

LIST OF FIGURES (contd.)

LIST OF FIGURES (contd.)

LIST OF TABLES

LIST OF TABLES (contd.)

LIST OF TABLES (contd.)

ABBREVIATIONS

BAW	Butanol/glacial acetic acid/ distilled water.
BAWP	Butanol/glacial acetic acid/distilled water/pyridine.
N-α-CBZ-lys-pNP	N-α-carbobenzoxy-L-lysine-p-nitrophenyl ester.
Dansyl, DNS	1-dimethylaminonaphthalene-5-sulphonyl
⁷	Symbol used to indicate direct dansyl-Edman sequencing.
DEAE	Diethylaminoethyl.
DTT	Dithiothreitol.
E.U.	Enzyme unit (μmol substrate transformed.min^{-1}).
FDNB	1-fluoro-2,4-dinitrobenzene.
NpS	2-nitrophenylsulphenyl.
%	All percentages are weight/volume (w/v) unless otherwise stated.
PITC	Phenylisothiocyanate.
POPOP	1,4-di[2-(5-phenyloxazolyl)]benzene.
PPO	2,5-diphenyloxazole.
S. aureus	Staphylococcus aureus.
TFA	Trifluoroacetic acid.
TPCK	L-(1-tosylamido-2-phenyl)ethylchloromethyl ketone.
Tris	Tris(hydroxymethyl)amino methane.

Amino acid abbreviations:

Ala	Alanine
Arg	Arginine
Asn	Asparagine
Asp	Aspartic acid
Asx	Aspartyl or Asparaginyl
CMCys	Carboxymethyl cysteine
Cys	Cysteine
$CySO_3H$	Cysteic acid
Gln	Glutamine
Glu	Glutamic acid
Glx	Glutamyl or Glutaminyl
Gly	Glycine

ABBREVIATIONS (contd.)

His	Histidine
Hse	Homoserine
$\overset{\frown}{\text{Hse}}$	Homoserine lactone
Ile	Isoleucine
Leu	Leucine
Lys	Lysine
Met	Methionine
Phe	Phenylalanine
Pro	Proline
Ser	Serine
Thr	Threonine
Trp	Tryptophan
Tyr	Tyrosine
Val	Valine

Peptides obtained by enzymic digestion:

T	Tryptic
ChT	Chymotryptic
Th	Thermolytic
P	Peptic
GE	*S. aureus* V8 proteolytic
a	Peptide acidic charged at pH 6.5.
b	Peptide basic charged at pH 6.5.
N	Peptide neutral at pH 6.5

Enzyme EC numbers used are according to:

Recommendations (1972) of I.U.P.A.C. and the International Union of Biochemistry.

Elsevier Scientific Publishing Company, Holland (1975) And Supplement I (corrections and additions (1975)) published in *Biochim. Biophys. Acta*, **429**, 1-45 (1976).

General abbreviations:

Are used according to the "Policy of the Biochemical Journal" *Biochem. J.*, **153**, 1-21 (1976).