STRUCTURAL STUDIES OF A BACTERIAL AMIDASE.

by

ANTHONY D. AUFFRET

A Thesis presented for the degree of Doctor of Philosophy of the University of Edinburgh.

Dept. of Molecular Biology,
Edinburgh.

1976

"Blessed are they who could not see, but being blind believed."

G. K. Chesterton.

ABSTRACT OF THESIS

The primary structure of the amidase from Pseudomonas aeruginosa strain PAC142, has been investigated. Some evidence is presented that the molecular weight of the amidase subunit is 40,000 to 42,000, higher than previously thought. Only partial sequence information has been obtained but some evidence has been found that a gene duplication may have occurred in the evolution of this amidase. The results are discussed with relation to other examples of protein evolution. An appendix, dealing with the reliability of molecular weight determination by amino acid analysis, is included.

Declaration

This thesis has been composed by myself and the work presented is my own unless otherwise acknowledged.

Acknowledgements

I wish to express my thanks to my supervisor Dr.R.P. Ambler, to my advisor Dr. A.F.W. Coulson, and to Professor M.R. Pollock F.R.S. for advice, encouragement and interest shown in this project. I am also grateful to Professor P.H. Clarke F.R.S. (University College of London) for encouragement and generous gifts of bacteria and protein.

Thanks are also given to Dr.A.Aitken, Dr.D.R.Thatcher, Dr.J.Moult, Mrs.H.E.Taylor and Mrs.M.Daniel for encouragement, criticism, technical advice and friendship.

I am grateful to the S.R.C. for a research studentship.

INDEX

LIST OF TABLES

LIST OF FIGURES

Appendix

ABBREVIATIONS USED

Abbreviations of accepted usage are not defined below.

Dansyl	1-dimethylamino naphthalene 5-sulphonyl.
D.M.A.A.	Dimethyl Allyl Amine.
g.l.c.	gas liquid chromatography.
K_m	Michaelis Constant.
V_{max}	Maximum velocity (at a given enzyme concentration).
O.F.N.	Oxygen Free Nitrogen.
PITC	Phenyl-isothiocyanate.
Quadrol	N,N,N',N'-tetrakis (2-hydroxypropyl)ethylene diamine.
S.D.S.	Sodium dodecyl sulphate.
T.C.A.	Trichloroacetic acid.
T.F.A.	Trifluoroacetic acid.
t.l.c.	thin layer chromatography.
Tris	2 amino-2(hydroxymethyl)-1,3-propanediol.
v/v	ratios by volume.
w/v	ratio of weight to final volume, (as gm per 100 ml.).

CHAPTER ONE

AN INTRODUCTION

a) General Introduction

The subject of this thesis is an investigation of the amino acid sequence of an amidase from Pseudomonas aeruginosa. This chapter presents the background information to this study. The biochemistry and genetics of the amidase system of P. aeruginosa are outlined, as are the physico-chemical properties of the enzyme. Amidases occurring in other micro-organisms, predominantly those in other pseudomonad species, are discussed with respect to their relationship to the P. aeruginosa enzyme. Amidase has been the subject of a laboratory evolution study and it is this which provides the major motivation for the present investigation. These experimental evolutionary studies are described, and the contribution of protein sequencing is discussed.

b) The Biochemistry and Genetics of the Amidase System.

A strain of P. aeruginosa which is capable of growth on acetamide or propionamide as the sole carbon and nitrogen source was shown (Kelly and Clarke 1960) to possess an inducible aliphatic amidase (an acylamino amidohydrolase,, E.C. 3.5.1.4.). As well as hydrolysing amides this enzyme catalyzes transfer of the acyl moiety of some amides and acids to hydroxylamine (Kelly and Kornberg 1962), and also catalyses the hydrolysis and acyl transfer of low molecular

weight esters (McFarlane et al 1965). The esters are poorer substrates than the corresponding amides. Transferase and hydrolase activities have different substrate specificity patterns, propionamide being a better hydrolase substrate than acetamide, but a poorer transferase substrate (Kelly and Kornberg 1962). Acetamide, propionamide, glycollamide ($HOCH_2CONH_2$) and acrylamide ($CH_2=CHCONH_2$) are good substrates for both reactions in the comparison of the specific activity of amidase with different substrates (Kelly and Kornberg 1964), formamide and lactamide are poor hydrolase substrates and glycinamide is poorer still. Transferase activity with formamide is low, and is only just detectable with lactamide and glycinamide. Kelly and Clarke (1962) have shown trace activity with butyramide and formamide as hydrolase substrates.

In the wild type strain amidase activity can be induced by a variety of amides (Kelly and Clarke 1962). N-methyl-, N-acetylacetamide, and lactamide are very good inducers but not good substrates. Acetamide, N-ethylacetamide, propionamide, and its N-methyl and N-ethyl derivatives, and methyl carbamate all induce activity. Glycollamide and N-ethyl formamide also do so, but poorly. An inducer concentration of 10^{-3}M. is effective with N-acetylacetamide and lactamide, but the effect of the other amides requires a concentration of 10^{-2}M. Butyramide and formamide are not inducers of amidase synthesis under these conditions.

Amidase synthesis is repressed (Kelly and Clarke 1962) by N-phenylacetamide, cyanoacetamide, glycinamide, sarcosine

amide, β-hydroxypropionamide, and thioacetamide, none of which are good substrates. Formamide is a weak inducer (Brammar et al 1967) but competes with other inducers to produce effective repression (Kelly and Clarke 1962). Brammar and Clarke (1964) showed that repression by cyanoacetamide, after induction, suggested competition for a single inducer/repressor binding site. The presence of a single constitutive permease for all amides (Brammar et al 1966) indicates that this site is not at the level of entry into the cell. Butyramide also represses amidase synthesis and has an affinity for the amide binding site similar to that of cyanoacetamide, and both of these compounds are capable of repressing synthesis in mutants constitutive for amidase production (Brown and Clarke 1970).

The genetic system appears to be relatively simple with an amidase structural gene closely linked to a regulator gene (Brammar et al 1967). Control properties are most easily explained by analogy to the negative control type found for the Escherichia coli lac operon (Jacob and Monod 1961), although as pointed out by Clarke (1974a) it is not yet possible to test this rigorously. The amidase genes have been assigned to the 65-67 minute region of the pseudomonad chromosome (Day et al 1975), but the current lack of markers in this region makes more detailed analysis impossible.

The synthesis of amidase is subject to catabolite repression by the hydrolysis products acetate and propionate, as well as by citrate, malate and pyruvate for organisms

grown in succinate medium (Brammar et al 1967). Increasing inducer concentration can reverse this repression (Brammar and Clarke 1964). The periodic oscillations of amidase synthesis in continuous culture (Boddy et al 1967) are probably due to a balance between induction and catabolite repression (Clarke et al 1968). Smyth and Clarke (1975a) have demonstrated catabolite repression by acetate, lactate, and glucose in pyruvate grown cultures, and showed that the repressing compound must be metabolized to exert its effect. This contrasts with the repression by amides and amide analogues, which are thought to act directly (Brown and Clarke 1970). Addition of cyclic-AMP gives partial relief of the mild repression by lactate in pyruvate medium, but not of the severe repression caused by succinate (Smyth and Clarke 1975b). Some mutants resistant to catabolite repression have been isolated (ibid.) and are probably promotor mutants, being closely linked with the amidase structural gene. One mutation, however, was not closely linked.

c) Physico-chemical Properties of Amidase.

Preliminary structural studies have been reported by Brown et al (1973). The molecular weight was found, by ultracentrifugation, to be 200,000. S.D.S.-gel electrophoresis, and gel filtration of the dissociated protein indicated a molecular weight of 33,000 - 35,000, suggesting 6 sub-units. Crosslinkage studies and peptide mapping supported the presence of 6 indistinguishable subunits.

Only one N-terminal amino acid, methionine, was found, and the C terminal sequence identified as glutamyl alanine.

Amidase activity can be inhibited non-competitively by urea and N-methylurea, and this inhibition can be reversed by hydroxylamine (Kelly and Kornberg 1964). Thiourea does not inhibit. Transferase activity is inhibited by iodoacetate and p-hydroxymercuribenzoate, reagents associated with cysteine modification. Addition of cysteine, indeed, reversed the inhibition caused by the latter compound. Both activities can be inhibited by low concentrations of fluoride, but only in the presence of hydroxylamine. The observation that diisopropylfluorophosphate (D.F.P.) does not inhibit precludes a mechanism analogous to that of the serine proteases. Similarly as substrate does not protect against 5,5'-dithiobis-(2-nitrobenzoic acid) inhibition, this would suggest that thiol groups are not involved in the mechanism (Woods and Orsi 1974).

Potent inhibition by acetaldehyde-ammonia has prompted the suggestion (Findlatter and Orsi 1973) that hydrolysis occurs via a sequential mechanism (both substrates bound before release of products) involving an acyl enzyme, a ternary complex, and release of ammonia as the first product. Using differential inhibition by iodoacetamide it was possible (Woods and Orsi 1974) to remove hydrolase activity and to study transferase action as a Bi-Bi (2 substrates, 2 products) system. In contrast to the earlier suggestion, characteristics associated with a

Ping-Pong mechanism (first substrate bound, first product released, second substrate bound, second product released) were found.

Studies of the dependence of activity on pH, and of chemical modification (Woods et al 1975) have implicated a lysine residue as necessary for amide binding, but as this is required in the protonated form it cannot be the nucleophile that is acylated in the reaction. The nature of that group is still unknown. A second lysine residue may be involved in transferase but not hydrolase activity.

d) The Occurrence of Amidases in Pseudomonads.

The taxonomic survey of Stanier et al (1966) indicates that acetamide can be utilized as a carbon source by most strains of P. aeruginosa, and by some strains of P. putida, P. acidivorans, and P.cepacia. It is, therefore, interesting to note that the original strain of P.aeruginosa obtained by Kelly and Clarke (1962) did not grow in a minimal salt medium containing acetamide.

Subsequently, strains of P.aeruginosa, P.putida (Biotype A), P.cepacia, P.acidivorans, and P.testosteroni have been tested for growth on aliphatic amides (Clarke 1972) and on phenylacetamide (Betz and Clarke 1973). All tested strains of P.aeruginosa, P.acidivorans, and P.cepacia grew on acetamide as did some strains of P.putida. Butyramide was also used as a growth substrate by some P.putida and P.acidivorans strains. Those strains which grew on acetamide

resembled the wild type P.aeruginosa (strain PAC1) in being inducible, and possessing hydrolase and transferase activities over similar substrate profiles. One strain of P.putida, and one of P.acidivorans were induced by butyramide, but in P.aeruginosa, and other P.putida and P.acidivorans strains butyramide repressed amidase synthesis. P.testosteroni did not grow on any of the aliphatic amides tested.

Partially purified amidases from P.aeruginosa, P.putida, P.acidivorans, and P.cepacia exhibited only slight electrophoretic mobility differences on starch gels. Cell free extracts were tested with antiserum prepared against purified P.aeruginosa (PAC 1) amidase; all P.aeruginosa extracts gave complete cross reaction, P.putida and P.acidivorans extracts gave partial cross reaction, whereas those from P.cepacia gave only weak cross reaction. The presence of similar antigenic determinants does therefore suggest that these amidases may have a common ancestral protein. This is supported by the other results cited above.

Growth of P.fluorescens on acetamide as sole carbon and nitrogen source (Jackoby and Fredericks 1964) appears to be due to an inducible aliphatic amidase which exhibits different inhibition properties (notably inhibition by D.F.P.) to the P.aeruginosa enzyme.

Phenylacetamide supported the growth of some strains of P.putida, P.acidivorans, and P.cepacia, but this appeared to be due to a distinct phenylacetamidase, separately inducible

from the aliphatic amidase. Two strains of P.putida, and one of P.cepacia which grew on phenylacetamide but not on acetamide are thought to possess only an aromatic amidase. The phenylacetamidase from one of these P.putida strains has been purified and gives no cross-reaction with antiserum prepared against P.aeruginosa aliphatic amidase (Clarke 1974 b) suggesting no close evolutionary relationship, although one cannot rule out a remote connection. The possibility that the naturally occurring acetamidases and phenylacetamidases may be related by an ancestral gene duplication has been noted by Betz and Clarke (1973).

An acetanilide hydrolysing enzyme has been found in P.acidivorans (Alt et al 1975b), and in P.striata (Hsuing et al 1975). The enzyme from P.acidivorans has been purified (Alt et al 1975a) and differs in its chemical and physical properties from the P.aeruginosa amidase. The P.striata enzyme is not well characterised but does not possess acyl transferase activity; an activity which has been demonstrated in the P.acidivorans acetanilidase.

e) Amidases in other micro-organisms.

Many molecules found in nature contain amide bonds and many enzymes are known which are capable of their hydrolysis. A comprehensive list of "amidase" activities would include enzymes for the hydrolysis of peptide bonds, antibiotics, nicotinamide, asparagine and glutamine. Some microbial

amidases which do appear to have some similarity with the pseudomonad amidases are described below.

Draper (1967) has described an inducible acylamino amido hydrolase in Mycobacterium smegmatis which possess acyl transferase activity. Crude suspensions of 6 mycobacterial strains (Myers et al 1957) have been shown to possess an enzyme exhibiting esterase and amidase activities. An ω-amido dicarboxylate amidohydrolase (E.C.3.5.1.3.) has been detected in Bacillus subtilis and in Thermus aquaticus which although having no activity on monocarboxylic acid amides is capable of acyl transfer to hydroxylamine (Fernald and Ramaley 1972).

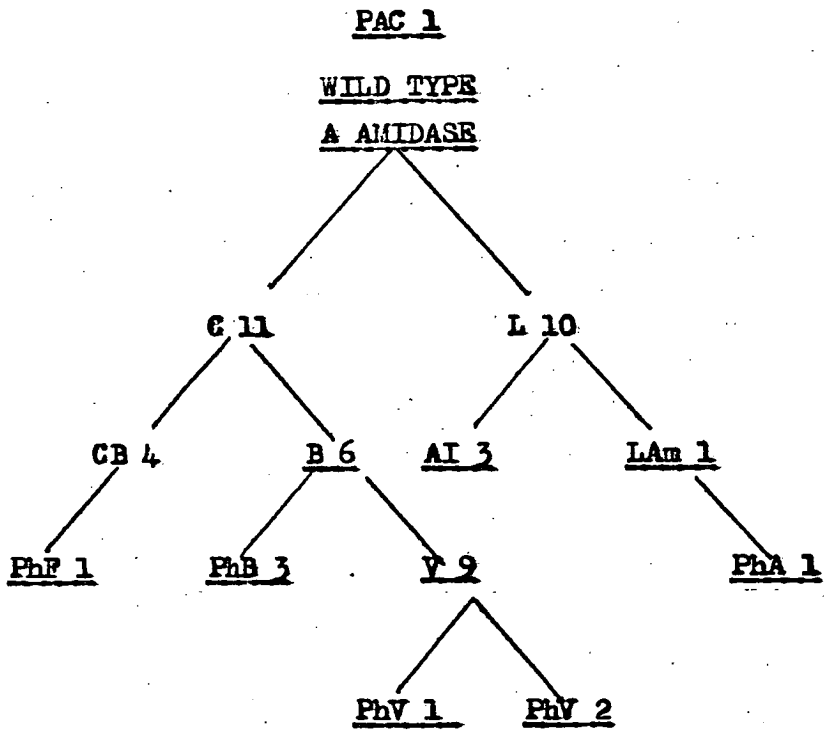
Dried extracts of an unidentified speciesof the yeast Torulopsis can hydrolyse acetamide, formamide, and lactamide, but hydrolysed butyramide and valeramide only poorly (Steiner 1959).

The fungus Aspergillus nidulans may possess 4 amidases (Hynes 1975), a formamidase, an acetamidase, a general aromatic amidase (e.g. active on benzamide and phenylacetamide) and a valeramide/hexanoamidase.

f) The Directed Evolution of P.aeruginosa Amidase.

Acetamide and propionamide are at the low molecular weight end of an homologous series of amides ($R\text{-}CONH_2$). Thus by changing the nature of the group R, or by using N-substituted amides it is possible to present a closely related series of novel substrates to the amidase system.

FIG 1 DERIVATION OF MUTANT AMIDASES.



All strains possess the genotype of the parent plus one additional mutation.

Strains producing an altered enzyme are underlined.

Selection for growth on, and hence utilisation of this series constitutes a laboratory simulation of evolution. Using such a series and by using specificity differences between the enzyme and its regulatory system , Clarke and her associates have isolated both regulatory and structural gene mutants of the P.aeruginosa amidase system. Some were isolated after mutagenesis, some are spontaneous mutants and all were isolated in single mutational steps. Fig. 1 outlines the derivation of some of these mutants.

The first compound of the homologous series, formamide, does not support the growth of the wild type strain (PAC1). Mutants capable of growth on formamide, as nitrogen source, have been isolated (Brammar et al 1967). These mutants are formamide inducible and as a result, can produce sufficient wild type amidase, A amidase, to allow growth on this amide.

Butyramide represses amidase synthesis even in some constitutive mutants which otherwise produce high levels of the enzyme (Brown and Clarke 1970). From one of these mutants, C11, two classes of strains capable of growth on butyramide have been isolated. One class, CB mutants, (Brown and Clarke ibid.) were resistant to butyramide repression and trace activity of the A amidase for this compound sufficed for growth. The other class, B mutants (Brown et al 1969) produced an enzyme with altered substrate specificity(B enzyme), which possessed enough butyramidase activity to allow growth despite continuing repression. B amidase has an altered electrophoretic mobility with respect to the A enzyme but is not detectably different by

immuno diffusion experiments. One of these mutants, B6, was used to isolate a heterogeneous class, V mutants, capable of growth on valeramide as sole carbon and nitrogen source. A second structural gene modification, detectable immunologically, seemed to be responsible for the extension of substrate range. Some V mutants had lost acetamidase activity.

Betz and Clarke (1972) have isolated 5 different phenylacetamide utilizing classes of mutants from various parents (see Fig 1). All are thought to produce further altered enzymes, and fall into two loose groups. Group 1 (PhB3,PhV1) hydrolyse butyramide,valeramide, and phenylacetamide, but not acetamide or formamide; Group 2 (PhV_2, PhA1, PhF) produce thermolabile enzymes which hydrolyse the same substrate range as group 1, but also have weak acetamidase activity. Group 1 and the V mutants lacking activity towards acetamide can be considered as "new" enzymes. PhA1 is a particularly interesting "revertant" as it is derived from a strain (LAm 1) producing an inactive enzyme.

Brown and Clarke (1972) were able to isolate an acetanilide utilising mutant (AI3) from strain L10 (a constitutive A amidase producer, resistant to butyramide and catabolite repression). A single amino acid substitution, isoleucine for threonine in the wild type sequence

Ser-Leu-Thr-Gly-Glu-Arg

is apparently responsible for the change in substrate specificity.

No amino acid changes have been characterised in any other of the altered amidases, but genetic analysis has been carried out (Betz et al 1974). Mutants V_2 and V_5 possess one common mutation and, as expected from their derivation, this is the same as the B6 mutation. Analysis also shows that V2 and V5 have different second sites of mutation. PhB mutants are also expected to have the B6 mutation, and genetic analysis suggests PhB3 and V2 may have a common second mutation site. Thus different mutations at the same site may confer different activities.

g) The Contribution of Protein Chemistry.

Immuno-diffusion experiments (Clarke 1972, Betz and Clarke 1973) have established that the naturally occurring pseudomonad acetamidases and phenylactemidases may have a common ancestor protein. One strain of P.putida possesses an acetamidase immunologically unrelated to P.aeruginosa amidase. Although the lack of immunodiffusion precipitates does not exclude evolutionary relatedness, Truffa-Bachi et al (1972) have detected immunological relatedness between aspartokinase homoserine dehydrogenases of E.coli where immunoprecipitation indicated none. (Kaminski et al 1969).

An understanding of the laboratory evolution of amidase requires an appreciation of how single mutational events affect the interaction of enzyme with substrate. As the amino acid sequence of a protein reflects the nucleotide base sequence of its structural gene it is a probe of both protein and

genetic fine structure. Location and identification of amino acid differences is essential for an understanding of the natural and directed evolution of pseudomonad amidases.

The effects of mutation on a single protein have been studied most extensively in human haemoglobin. Fasman (1976) lists 201 known single amino acid substitutions, in the α and β chains, which affect the functioning of the protein. As the tertiary structure is known to high resolution (Perutz et al 1968) it is possible to correlate molecular with functional changes. Hartley (1974) has briefly reviewed "Enzyme Families",in particular the serine proteases, where knowledge of the primary and tertiary structures has led to an understanding of the molecular basis of specificity differences.

Similar studies on pseudomonad amidases would be of great value. A knowledge of the sequence and 3-dimensional structure of the wild type A amidase and of the laboratory mutants, together with information from enzymological studies would explain the molecular basis of amidase specificity differences.

Comparison of amidase sequences between species is a second area where much can be learned. It would indicate how closely related the natural acetamidases, phenylacetamidases, and acetanilidases, are. Although as Clarke (1972) has warned, pseudomonads may have exchanged genes in relatively recent times by interspecies genetic transfer, and hence similarities in protein properties may not reflect long standing evolutionary relationships. Betz and Clarke

(1973) have noted the relatively low specific activity of both the naturally occurring and the laboratory evolved phenylacetamidases. The degree to which laboratory evolution has paralleled that of nature will be interesting to observe.

The limitations of sequence information must, however, be remembered. Li et al (1973) have estimated the nucleotide sequence divergence in the α-chain structural gene of tryptophan synthetase between 3 species of the enterobacteriacea by both protein sequencing and RNA-DNA hybridisation. These authors concluded that,"perhaps as many base differences do not alter the amino acid sequence as those that do."

Any conclusions drawn from the evolution of pseudomonad amidases must also be viewed in the light of other evolutionary studies. Results from comparative evolutionary studies range from the molecular to the behavioural level, and many form the basis of present taxonomic classifications. Experimental evolution has been studied on a number of characters in a number of organisms (see Hegeman and Rosenberg 1970, Clarke 1974a) and suggests several routes are available for protein evolution. The contribution of pseudomonas amidase evolution can only be fully assessed when its molecular basis is understood.

h) This Thesis.

This thesis describes studies of the P.aeruginosa amidase amino acid sequence and these are presented in the following order.The method of purification of amidase is given in chapter 2,together with the results from some experiments

carried out with the whole protein (which in this context refers to the sub-unit polypeptide chain, rather than the active hexameric enzyme). The methods used to investigate the amino acid sequence are given in chapter 3, and the results from these investigations are detailed in chapters 4 to 6. Chapter 7 summarises these results and shows to what extent the peptides can be overlapped. A discussion of the results and their relevance to protein evolution studies is also included.

An appendix describing the use of amino acid analysis in the calculation of molecular weights is included at the end, and gives some examples of the limitations of this method.

CHAPTER 2

THE PURIFICATION AND PROPERTIES OF AMIDASE

1) PURIFICATION TECHNIQUES

The amidase used in this investigation has been the wild type A amidase from *P.aeruginosa* strain PAC 142 (series no. L10, Brown and Clarke 1970).

a) The Purification of Amidase.

Amidase was purified by the method of Brown *et al* (1969) from frozen cells obtained from Prof. P.H. Clarke. The purification involved the following steps:

1) Thawed cells were suspended in ice cold buffer (0.1M tris, 0.15M KCl, 5mM 2-mercaptoethanol, 1mM E.D.T.A., pH 7.2) at 0.3 to 1.75 gm. cells (wet weight) per ml., and disrupted in a Manton-Gaulin homogeniser at a pressure of 500 kg. cm^{-2}.

2) Streptomycin sulphate (Glaxo, B.P.), 1 gm. per 50 gm. wet cells approximately, was added as an aqueous solution with mixing. Centrifugation (10,000g for 30 mins. at 4°C) removed cell debris and the precipitated nucleic acid.

3) 250 ml. portions of the supernatant were rapidly heated to 60°C, maintained at that temperature for 5 mins., and then cooled on ice and bulked. The precipitate was removed by centrifugation as before.

4) The supernatant was then fractionated with ammonium sulphate. With centrifugation of precipitates after 50, 55, 65 and 75% saturation (Brown *et al*. 1969) amidase activity

was equally distributed between the 50-55% and 55-65% cuts. Subsequently precipitates were collected after 40%, 50%, and 65% saturation. The bulk of the amidase activity was contained in the 50-65% fraction.

5) The amidase containing fraction was redissolved in buffer (minimum volume) and dialysed against buffer to remove ammonium sulphate.

6) The dialysate was loaded onto a DEAE-sephadex (A50) column equilibrated in buffer, and washed with buffer (for a minimum of 2 hrs. at a flow rate of 40 ml. per hr.) before eluting with a linear gradient of 0.15M to 0.35M KCl contained in 2 l. tris buffer. Effluent was monitored at 254nm (L.K.B. Uvicord) and amidase activity could be located by the test paper method (see below). Amidase activity was coincident with the major u.v. absorbing peak. Fractions possessing activity were assayed quantitatively (by the transferase method of Brammar and Clarke 1964) and samples tested for homogeneity by polyacrylamide gel electrophoresis (Davis 1964). Homogeneous fractions were pooled and stored as 80% saturated ammonium sulphate precipitates as difficulty was found in redissolving material extracted by the method of Brown et al (1969).

b) Enzyme Assays

The transferase method of Brammar and Clarke (1964) was used exclusively. 0.1 ml samples were mixed with 0.9 ml.

mixed substrates (purification buffer: 0.4M acetamide: 2M freshly neutralised hydroxylamine hydrochloride, 2:1:1 v/v) and incubated at 37°C. The reaction was terminated by the addition of 2ml. ferric chloride reagent (6% w/v $FeCl_3$, 2% w/v HCl). Optical densities were measured at 500 nm in a Unicam SP1800 spectrophotometer. The conversion factor, $3.5\mu\overset{\text{moles}}{\wedge}$ acethydroxamate per ml. has an E_{500} of 1.0 was used throughout (P.H. Clarke, pers.comm.).

Amidase activity test paper was prepared by soaking Whatman 3MM chromatography paper in mixed substrates (cf. transferase assay) and drying. Small volumes of sample were spotted onto the paper and allowed to dry at room temperature. Activity was located by dipping the paper into a solution of ferric chloride (6% w/v) in acetone. Test papers could be used for 2-3 days after preparation, but thereafter background colouration masked the location of activity.

c) <u>The Estimation of protein</u>.

Protein concentrations were measured by the modified Folin-Lowry method of Miller (1959). Freshly prepared solutions of Bovine Serum Albumin were used to construct calibration graphs for each set of determinations.

d) <u>Polyacrylamide Gel Electrophoresis</u>.

The method of Davis (1964) was used and gels were stained with Amido black (1% w/v) in methanol:acetic acid: water (5:1:5 v/v). Bands were visualised by repeated washings

in the same solution minus the dye.

e) Reagents

Acetamide (B.D.H. laboratory reagent) was recrystallised from benzene:Ethyl acetate (10:3 v/v). Hydroxylamine hydrochloride (Hopkins and Williams laboratory reagent) was recrystallised from ethanol:water (3:1 v/v). Other reagents were "Analytical Reagent" grade, or of the highest purity available and were used as supplied.

f) Results

Four batches of enzyme were used in the investigation, and details are given below.

Batch 1 was obtained as purified protein (from P.H.Clarke). This protein was used in the first chymotryptic and tryptic digests. At a later (2 1/2 years)stage this material was found to contain 4 components (by S.D.S. polyacrylamide gel electrophoresis) of molecular weights 42,000, 40,500, 39,000 and 29,700. An attempt to purify these components by gel filtration after citraconylation failed owing to the small amount of protein then available. Amino acid analysis, sequenator analysis, and peptide maps (tryptic and chymotryptic) obtained at the time of use, however, are not distinguishable from those obtained from other batches of protein.

Batch 2 was prepared from frozen cells and rechromatography on DEAE-sephadex was necessary to obtain homogeneous material.

The specific activity of this batch was very low, 373 units per mg. although polyacrylamide gel electrophoresis, amino acid analysis, and peptide mapping suggested this material was homogeneous amidase. Protein from this batch was used for the second chymotryptic and tryptic digests.

Batch 3 amidase was purified from frozen cells and consisted of homogeneous amidase of specific activity 1725 units per mg. This material was used for the cyanogen bromide cleavage.

Batch 4 was obtained as partially purified material (from P.H.Clarke), as the 55-65% ammonium sulphate precipitate. The specific activity of the purified material was 1300 units per mg., and this protein was used for the Staphylococcus aureus protease digest.

2) Experiments with the Whole Protein

g) Molecular Weight Determination

Polyacrylamide gels containing S.D.S. were formulated according to King and Laemmli (1971). Slab gels, 15cmx15cm. were poured between glass plates, and gels containing a gradient of acrylamide concentration were prepared by mixing solutions of the initial and final acrylamide concentrations from a two chambered gradient maker. Samples, at a protein concentration of 1mg. per ml. were reduced in tris buffer (0.06M tris pH6.8 containing 2% w/v S.D.S., 5% v/v 2-mercaptoethanol, and 10% v/v glycerol) by heating at 100° for 15 mins. Samples were applied directly to the upper (stacking) gel, and run overnight at 10 mA.(at constant

TABLE 2A The Molecular Weight of Amidase

Protein sample	Gel A	Gel B	Gel C	Gel D
Batch 1 amidase	42,000 40,500 39,000 29,700 } 4 bands			
Batch 2 amidase (A.A.)	42,000	38,500		
Oxidized Batch 3 amidase (A-ox)		42,300	44,300	
Batch 4 amidase (A-M)			42,000	
A.Amidase (A-B)	42,000	diffuse band 40,000 - 43,000		
PhVl amidase (P.V.R.)	42,000	38,500		
PhVl amidase (P.V.A.)	40,750	40,000	a)43,000 i b)42,500 ii	a)40,000 (iii) b)39,750 (iv)

Sources of protein, and the composition of the gels are given in fig 2a.

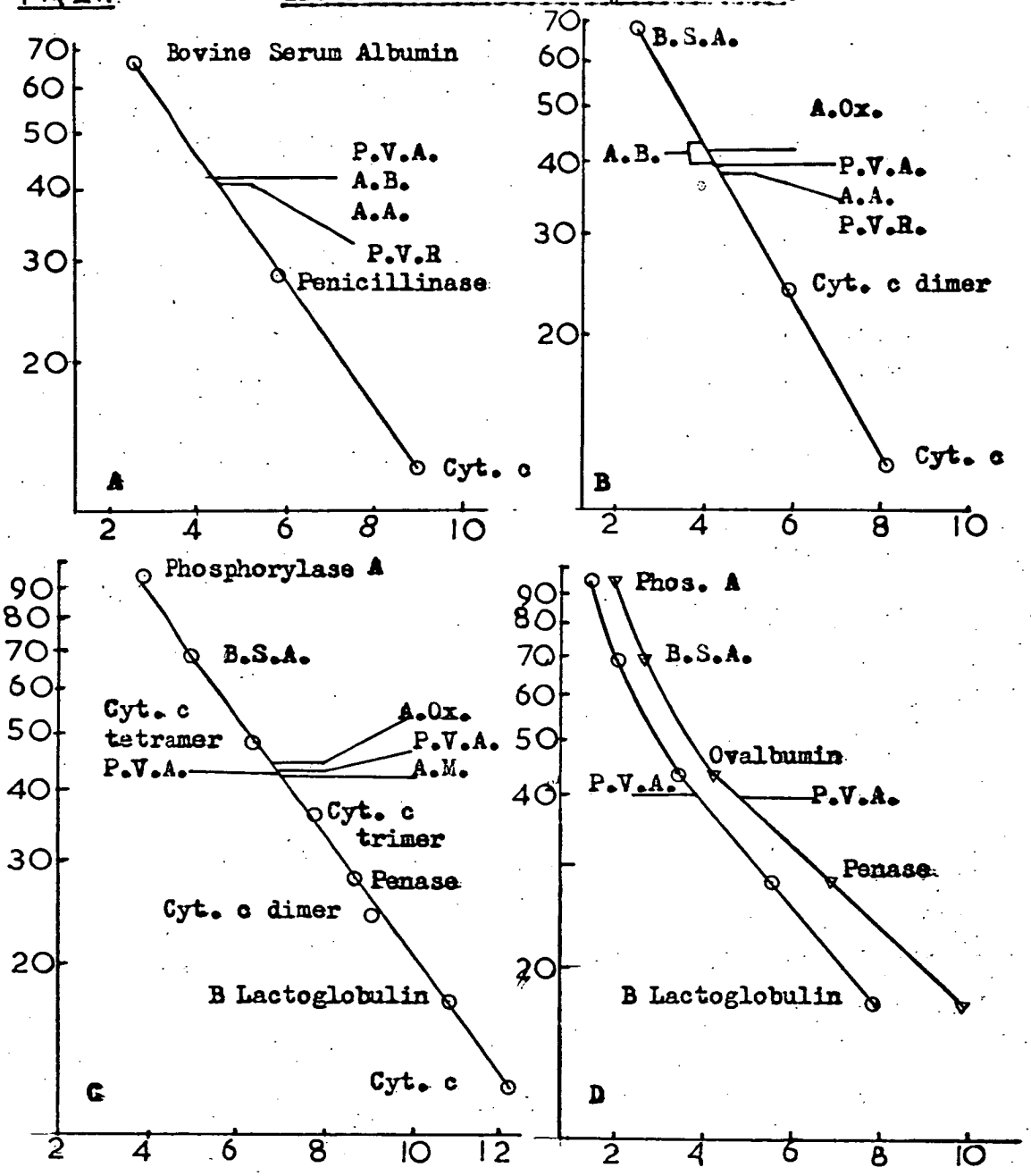
NOTE: Gel C (i) applied as single sample

(ii) applied to the same track as A-Ox

TABLE 2A - Contd.

Gel D - Both results are for amidases applied as sole sample and amidases applied to the same tracks as ovalbumin and pencillinase; as all amidases samples migrated identically (iii) measured after a few hours destaining (iv) measured after 1 week destaining.

FIG 2A. SDS GEL ELECTROPHORESIS OF AMIDASE.



Ordinate units are M.W. x 10^{-3}

Abscissa units are cm. migrated.

Gels A,B,andC contained a 10-30% gradient of acrylamide concentration.

Gel D contained 16% acrylamide and was measured after a few hours destaining (o——o) and after 1 weeks destaining (▽——▽).

AMIDASES
(Sources other than this laboratory)

A.A-Batch 3 amidase.
A.B.-Aamidase (P.R.Brown.)
A.Ox.-Oxidised A.A.
A.M.-Batch 4 amidase.
P.V.A.-PHV1 amidase(P.H.Clarke).
P.V.R.-PHV1 amidase(A.Paterson).

current.) The gels were washed in Methanol:Acetic acid: Water(5:1:5 v/v) to remove S.D.S.prior to staining with 1% Coomassie Brilliant Blue(in the same solvents)and destained by repeated washings in 10% v/v acetic acid.

I am grateful to Mrs.M.Daniel for assistance with the techniques and preparation of gels for electrophoresis.

Five different preparations of A amidase and two of PhVl amidase were tested in acrylamide gradient gels and in a constant acrylamide concentration gel. The results are shown in fig 2a. The results from these experiments are tabulated in table 2a and indicate a mean molecular weight of 40,900. Brown et al (1973) reported the subunit molecular weight of amidase to be 33,000-35,000,agreeing with a molecular weight of the hexameric enzyme of 200,000. The subunit molecular weight was determined by S.D.S. gel electrophoresis and by gel filtration of the dissociated enzyme. Brown et al chose the system of Weber and Osborn (1966)for S.D.S. gel electrophoresis using 10% (w/v) acrylamide gels in glass tubes, and phosphate buffers.

In this investigation slab gels and tris buffers were used, and both standard proteins and samples were run on the same gel under the same conditions. Amidase and close molecular weight markers have been applied to the same "track" in the gels, and the separation of native and oxidised amidase on the same track demonstrated the resolution of the system used. It is felt that such a method is more precise than the running of markers and samples on different gels under similar conditions with the

construction of calibration graphs by relative mobility to a single marker. The possibility that the discrepancy in apparent molecular weight may be due to behavioural diversity of the amidase in the different buffer systems cannot be overlooked.

h) The Calculation of Molecular Weight from Amino Acid Analyses

An attempt has been made to calculate the molecular weight of amidase from the amino acid frequencies by the method of Black and Hogness (1969). This method assumes that all amino acids are present in integral molar amounts and that at the correct molecular weight the calculated compositions should show minimal deviation from integral values (see the Appendix). A computer programme (supplied by Dr.J.Moult) which calculates a weighted sum of these deviations according to the equations of Black and Hogness has been used in this study.

Protein was hydrolysed with 0.4ml 6M HCl (B.D.H.Aristar) at 105°C in vacuo, and the hydrolysate analysed on a Beckman model 120C amino acid analyser. Values for tryptophan determined by the spectrophotometric method of Beaven and Holiday (1952), or after hydrolysis with 3M mercaptoethane sulphonic acid (Penke et al 1974; the reagent being obtained from Pierce Chemicals, Radford, Illinois, U.S.A.) are included for comparison but were not used in the calculation.

Data from this investigation, and the values of Brown et al(1973), has been used(Table 2B) to calculate the graphs in fig 2B. The amino acid frequencies of this investigation

TABLE 2B

Data for the Calculation of Molecular Weight from Amino Acid Analysis

Amino Acid	frequency[3]	Composition m.w.33,000	Composition m.w. 41,000	frequency[6]	Composition m.w.33,000	Composition m.w.41,000
Lys	0.0715	12.8	15.9	0.373	13.9	17.2
His	0.0338	6.04	7.5	0.166	6.2	7.7
Arg	0.0961	17.2	21.3	0.474	17.6	21.9
Trp (4)	-	5a 5.4 b 4.8	5a 6.9 b 6.0	0.144	5.2	7.05
Asx	0.1714	30.6	38.1	0.796	29.6	36.7
Thr (i)	0.0645	11.5	14.3	0.308	11.4	14.2
Ser (i)	0.081	14.5	18.0	0.372	13.8	17.2
Glx	0.2091	37.4	46.4	0.879	32.6	40.6
Pro	0.0832	14.9	18.5	0.380	14.2	17.5
Gly	0.1779	31.8	39.5	0.810	30.1	37.4
Ala	0.1593	28.5	35.4	0.769	28.6	35.5
Cys	-			0.208	7.73	9.6
Val (2)	0.1162	20.8	25.8	0.602	22.4	27.8
Met	0.0796	14.2	17.7	0.343	12.7	15.8

TABLE 2B - Contd.

Ile	0.1050	18.8	23.3	0.530	19.7	24.5
Leu (2)	0.1025	18.3	22.8	0.498	18.5	23.0
Tyr	0.0841	15.0	18.7	0.402	14.9	18.5
Phe	0.0457	8.2	10.2	0.199	7.4	9.2

(1) Thr. and Ser. values obtained by extrapolation to zero time.

(2) Ile and Val values obtained by extrapolation to infinite time.

(3) Values adjusted for differing amounts applied to analyzer columns.

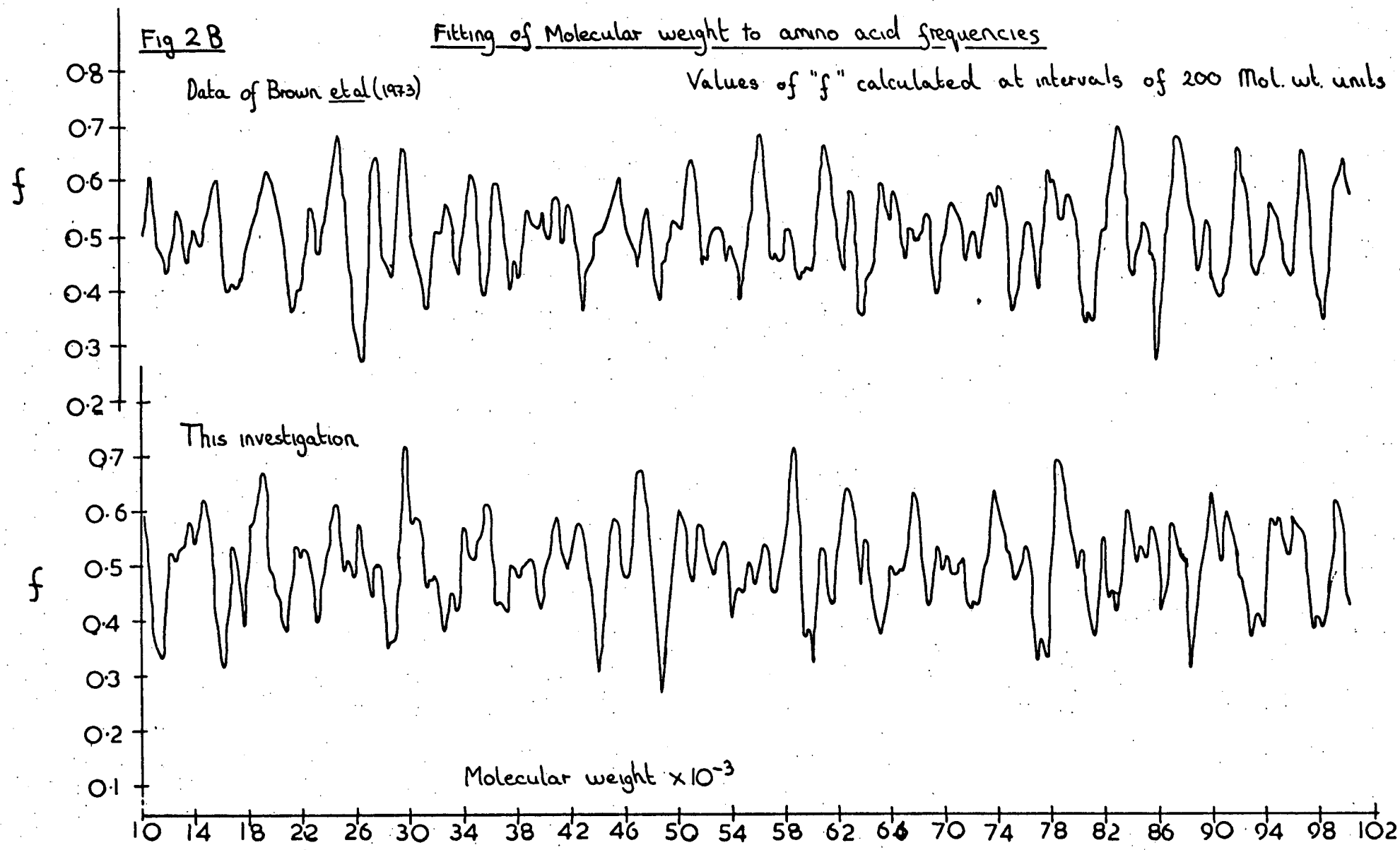
(4) Values not used in the calculation.

(5) Tryptophan values (a) estimated spectrophotometrically.
(b) after MESA hydrolysis.

(6) Data of Brown et al (1973).

Fig 2B

Fitting of Molecular weight to amino acid frequencies



show a minimum (f = 0.273) at a molecular weight of 48,600, whilst those of Brown et al fit a molecular weight of 26,400 (f = 0.273). The addition of excluded amino acids would increase these values by 1,000 to 2,000.

To test the validity of such a calculation amino acid frequencies have been calculated from the sequence of Rabbit Skeletal Actin (data from Dayhoff 1973). These frequencies have been used to calculate the molecular weight (see Appendix). The use of a computer programme which added random errors of known magnitude has shown that even 3% error is sufficient to predict an incorrect molecular weight for a protein of this size (M.W. 41,719). It is therefore believed that the errors in the data are sufficient to invalidate the result.

3) Sequence Studies on the Intact Protein

j) N-terminal Identification.

The dansyl procedure for proteins (Gray 1972a) has been used to confirm that Methionine is the only N-terminal amino acid (Brown et al 1973).

k) N terminal Sequence Studies

(i) Manual Edman Degradations.

The direct method of Edman and Henschen (1975) and the paper strip method of Frankel-Conrat (1954) succeeded only in identifying N-terminal methionine. The method of Weiner et al (1972) which uses an S.D.S. containing medium with dansylation

of the residual protein appeared to place an aspartic acid or asparagine as residue 5.

ii) Automatic Edman Degradation

Automatic Edman degradations were carried out in a Beckman model 890B sequenator operated by Mrs. M. Daniel. Residues were identified by thin-layer chromatography (t.l.c.) using the solvent system of Edman and Sjoquist (1956) and by gas liquid chromatography (g.l.c.) on SP4001 Chromasorb W. Residue 6 was initially identified as leucine but in one determination hydrolysis of the Phenyl thiohydantoin (P.T.H.) amino acid derivatives with hydriodic acid (Inglis et al 1971) and quantitative amino acid analysis established that this residue was in fact isoleucine. Subsequently care was taken to ensure clear identification of this residue.

Run 1. The degradation was carried out on oxidized batch 1 protein using a single coupling, double cleavage programme with quadrol buffer. The amino acid derivatives were identified by t.l.c. only and gave the result

Met- - -Gly-Asp-Leu-Asp-
→ ⇢ ⇢ → → → →

(an arrow → indicates each step, a broken arrow indicates substandard identification of the residue)

Run 2. was carried out as above on oxidized batch 1 amidase

Met- - -Gly-Asp-Leu
→ ⇢ ⇢ → → →

Run 3 was carried out on oxidized batch 2 protein using a double coupling single cleavage programme with dilute quadrol buffer. Residues were identified by t.l.c. and g.l.c. except for residue 3 (identified by a colour reaction on paper for histidine, Dent 1947) and residue 7 (g.l.c. only).

Met-_ →-His-Gly-Asp-Leu-Phe-

Run 4 degraded oxidized batch 3 protein, and quantitative amino acid analysis of the residues after hydriodic acid hydrolysis established the sequence

Met-Arg-His-Gly-Asp-Ile-_ →-_ →-_ →-Asx-

Run 5 used oxidized PhV1 amidase (obtained from A. Paterson) with a succinylated myoglobin "carrier" protein to improve the filming properties of the amidase. The programme used single coupling and single cleavage steps with D.M.A.A. buffer. Residues 1,4,5 were identified by t.l.c. and residue 1 confirmed by g.l.c.. Residues 2 and 3 were detected by reaction on paper; arginine by fluorescence with phenanthrenequinone (Yamada and Itano, 1963) and histidine as above.

Met-Arg-His-Gly-Asp

Run 6 was a repeat of run 5. Residues 1,4,5,6,9,10,11 were identified by t.l.c., and residues 6,9,10,12 by g.l.c. Residues 2 and 3 were identified as before.

Met→Arg→His→Gly→Asp→Ile→ _ →- _ →Ala→Gln→Asp→Ala→

It should be noted that in runs 1 to 3 residues 2 and 3 were not identified satisfactorily by paper spot tests. The overall sequence

Met-Arg-His-Gly-Asp-Ile-

is however, accepted. Residues 7,8, and 9 were poorly identified if at all, and the tentative identifications of runs 1 and 3 are rejected. It is thought that these residues may be those whose derivatives are easily destroyed, serine, threonine or tryptophan (destroyed on oxidation). Cysteine (as cysteic acid) was eliminated only by run 4. The appearance of a well identified alanine at position 9 in PhV1 amidase is interesting as Betz et al (1974) have indicated that one mutation in this strain may be close to one end of the amidase structural gene.

CHAPTER 3

METHODS OF PROTEIN SEQUENCE DETERMINATION

a) Introduction

Current methods for protein sequencing have been successfully applied to a diverse range of polypeptide chains extending in size up to that of calf collagen, 1052 amino acids (see Fasman 1976). Even so, as Brigden and Koch (1975) point out "...the elucidation of the complete sequence of a protein is a relatively rare event, especially if the protein is unrelated to other proteins of known sequence." The majority of available techniques are well documented (e.g. Hirs 1967, Hirs and Timasheff 1972, Needleman 1975) and the overall strategy remains constant. The protein is split at limited sites into its compound peptides, which are purified, analyzed for composition, and sequenced (as far as possible) by a variation on the PITC method for the sequential removal of N-terminal amino acids (Edman, 1956). Repetition with splitting at different sites generates overlapping sets of peptides which can be used to align the total sequence. The availability of automatic "sequenators" capable of carrying out sequential degradations at a high efficiency over extended lengths of polypeptide chains, coupled with the generation of large fragments of the protein can simplify the methodology. Unfortunately such a strategy is not always practicable; nor is it yet practicable to determine complete sequences by the automatic degradation of whole proteins.

The techniques used in this investigation are based on those developed for the study of bacterial electron transport proteins (e.g. Ambler and Wynn, 1973) and have been successfully applied to structural determinations of larger proteins (e.g. Ambler, 1975). Such techniques are, however, continuously changing, for example the Staphylococcus aureus protease used in this study was a barely recognised tool when this project began.

Unless otherwise stated the materials used were obtained from the sources given by Ambler and Wynn (1973).

b) Performic Acid Oxidation

Performic acid was prepared by incubating a mixture of anhydrous formic acid and 30% hydrogen peroxide (95:5 v/v) for two hours at room temperature (Hirs, 1956). In order to limit the halogenation of tyrosine, chloride free protein was prepared by ammonium sulphate precipitation and dialysis prior to oxidation. Amidase was dissolved in formic acid and a 12 fold excess (over sulphur containing groups) of performic acid added, the final protein concentration being 10-20 mg. per ml. Oxidation was carried out for 1-2hr. at 0°C, the longer times being required for the complete oxidation of methionine. The mixture was then diluted to give 5% v/v formic acid, and freeze dried. During oxidation the solution turned from a gelatinous mixture to a mobile liquid, this change being more apparent when a small volume (0.5 to 2ml) of methanol was included in the solution.

c) Enzymic Digestion

Oxidation was used to denature amidase prior to enzymic digestion. Chymotryptic, tryptic, and thermolytic digestions of protein or peptides were carried out in 0.2 M. ammonium acetate, pH 8.5, at 37°C using 1/75 by weight of enzyme to substrate. Oxidized amidase did not dissolve well in the solution used, and this buffer was usually prepared by dilution after initial dissolution of amidase in 2M. ammonia. Whole protein treated in such a manner was digested as a fine suspension which usually cleared within 1 hr.

Digestion with S.aureus protease was carried out in 0.125 M ammonia, pH 10.5 for 5 hr. at 37°C (enzyme : amidase ratio 1:100 w/w). The pH was then lowered to pH 9 with 1M acetic acid and digestion allowed to proceed for a further 5 hrs. prior to lyophilisation. Drapeau et al (1972) report the pH optimum of this enzyme as pH 8.5. The oxidised amidase used for this digestion precipitated below pH 10.5, and S. aureus protease digestion of fine suspension of proteins can be unsatisfactory (Dr. D.R. Thatcher, pers.comm.). The pH was lowered in mid digestion without significant precipitation.

Enzymes were obtained from the following sources

Trypsin; Worthington Biochemical Corporation, Freehold, N.J., U.S.A., and was treated with diphenyl-carbamoyl chloride before use.

Chymotrypsin; Worthington Biochemical Corporation, Freehold, N.J., U.S.A., and was treated with soya bean tryptic inhibitor before use.

S. aureus protease; was prepared by Dr. D.R. Thatcher from the strain listed in Ambler (1975).

Thermolysin; Daiwa Kasei, K.K., Osaka, Japan.

d) Cyanogen Bromide Cleavage

Amidase (300 mg.) was dissolved in 4.5 ml. trifluoroacetic acid (T.F.A.) and diluted with 1.5 mls. water. Cyanogen Bromide (Eastman), 0.5g. in 0.5 ml 75% v/v T.F.A. (an approximate 40 fold excess over methionine) was added, and the mixture left for 24 hr, at room temperature with occasional swirling. The reactants were then diluted with a 10 fold excess of water and freeze dried.

e) Purification of peptides.

Gel filtration of peptides through Sephadex superfine G-25 was used as a preliminary purification step. Initially whole digests were fractionated this way, but later, larger "core" peptides were first removed by precipitation with 5% w/v (final concentration) trichloroacetic acid (T.C.A.). After standing for 30 minutes the precipitated core material was removed by centrifugation. Removal of T.C.A. from the supernatant was attempted using ether extraction. The treatment of core materials is dealt with in the relevant results sections. The residual soluble peptides were fractionated by gel filtration.

Effluent from the sephadex solumn was monitored at 254 nm. (L.K.B. Uvicord monitor) and collected in fractions. Portions

TABLE 3a. Electrophoresis Buffers and Chromatography Solvents

Electrophoresis at pH 6.5	Pyridine:Acetic acid;Water 25:1:225 (v/v) toluene coolant (Ryle et al, 1955)
Electrophoresis at pH 3.5	Pyridine:Acetic acid:Water 1:10:89(v/v), Paper wetted with half strength buffer. White spirit coolant, (Ryle et al, 1955)
Electrophoresis at pH 2	Acetic acid:formic acid:Water 4:1:45(v/v). White spirit coolant. (Ambler 1963)
Electrophoresis at pH 9	2% (w/v) Ammonium carbonate pH 9. White spirit coolant
Paper Chromatography	Butan-1-ol:Acetic acid:Water: Pyridine 15:3:12:10 (v/v) (Waley and Watson 1953)
T.l.c. of dansyl amino acids	solvent 1,Formic acid 5%(v/v) (Woods and Wang 1967)
	solvent 2, Toluene :Acetic acid 9:1 v/v (Woods and Wang 1967)
	solvent 3, Butyl Acetate: Methanol:Acetic acid 30:20:1 (Crowshaw et al 1967)
	solvent 4, 0.05m Na_3PO_4 in 25% (v/v) Ethanol (Hartley 1970)

of fractions were applied to Whatman 3M.M. chromatography paper in successive 10μl. applications at $\frac{1}{4}$ ins. intervals and subjected to electrophoresis at pH 6.5. After drying the paper was viewed in ultra-violet (U.V.) light and fluorescent peptides marked prior to staining with ninhydrin (0.2% w/v in acetone, containing 1% v/v acetic acid).. The paper was frequently counterstained with the Pauly reagent (Dent 1947) to locate histidine and tyrosine containing peptides. In one case the Ehrlich reagent (Dalglieshl952) was used to counterstain for tryptophan, and once the Sakaguchi reagent for arginine (Jepson and Smith 1953) was used prior to ninhydrin. The resultant map, a two-dimensional separation based on peptide size and electrophoretic mobility, was used to decide which fractions to pool for further purification.

f) Paper Electrophoresis and Chromatography

Further purification was carried on paper by combinations of electrophoresis at pH 3.5, 2.0, or 9.0, and descending chromatography. Buffers and solvents used are given in table 3a. Apparatus of the Michl (1951) type was used for high voltage paper electrophoresis. Whatman 3MM chromatography paper was routinely used with loadings up to 100nm. per cm., but Whatman No. 1 paper was used for analytic work and for separating small amounts of peptides.

Marker strips, of small amounts of material, run on both sides of the main band were cut off and the peptides

located by dipping these strips in a 0.2% w/v solution of ninhydrin in acetone. For pH 6.5, pH 3.5, pH 9.0 electrophoretograms and for chromatograms the ninhydrin reagent was acidified with acetic acid, 1% v/v final concentration (approx.). For pH 2 electrophoretograms impure collidine (Collidine, Fraction, Midland and Yorkshire Tar Distillers), 0.1% v/v final concentration was added to the acid ninhydrin reagent. Colours were developed at room temperature or at 105°C. Marker strips were occasionally counterstained with the Pauly reagent or with the starch/KI reagent for detecting peptides with weak ninhydrin colours (Rydon and Smith 1952). Low yielding peptides could be located with fluorescamine, papers were dipped in an N-ethyl morpholine acetate solution. (45 ml N-ethyl morpholine, 1.4ml acetic acid, 1l. acetone, D.R. Thatcher pers. comm.) and dried at room temperature prior to dipping in fluorescamine (0.2 mg per ml) in acetone.

After location the main bands were cut out and eluted with 0.1M ammonia. The eluate was desiccated and then taken up in 0.2 to 0.5 ml 0.1M ammonia for storage as frozen solutions.

g) Amino Acid Analysis

Peptides for qualitative amino acid analysis were hydrolysed in 6N HCl and the products separated by electophoresis at pH 2, at 5-6 Kv. Use of the ninhydrin-collidine dip and heating at 105°C for a short time gave rise to

marked colour differences between the amino acids. In some instances the hydrolysate was dansylated (see below) and examined by t.l.c.

Samples for quantitative analysis were hydrolysed in 6N HCl in vacuo, heated at 105°C for 24 to 96 hr. The amino acids released were analysed on an L.K.B. Biocal BC200 amino acid analyser by elution from a single column of resin. A Beckman 120 C analyser using 2 separating columns was also available. The former was more sensitive but the latter was more reliable and having a logarithmic scale more suited for whole protein analysis and for peptides having large differences in the amounts of the constituent amino acids. Peptide yields were calculated from the analysis but were not corrected for any losses (e.g. marker strips).

Cyanogen bromide peptides were heated at 105°C in pH 6.5 electrophoresis buffer in vacuo for 1 hr, to convert homoserine lactone to homoserine before analysis. Homoserine could be separated from glutamic acid on the analyser column by the use of pH 2.9 buffer (Ambler and Brown, 1967).

h) The Identification of N-terminal Groups

The method of Gray (1972a) was used. Samples, in 6 x 30mm. Durham tubes were dried in vacuo, 10μl. of sodium bicarbonate (0.1% w/v) added and the mixture dried in vacuo over NaOH pellets. Water, 10μl., and Dansyl chloride (1-dimethylamino naphthalene-5-sulphonyl chloride), 10μl. of a 2mg. per ml. solution in acetone, was added and the mixture incubated at

37° for 1 hr., or until the yellow colouration had been lost. After drying the dansyl peptide was hydrolysed overnight at 105° with 6N HCl. The hydrolysate was examined by t.l.c. on polyamide layer plates in the solvent systems given in table 3a. The unknown was applied to both sides of the t.l.c. plate and a mixture of known dansyl amino acids to one side only. Solvents 2,3, and 4 were run at right angles to solvent 1. Acid stable dansyl dipeptides were often identified. Dns-Arg, Dns-His, αDns-lys, and ℰDns-lys could be separated by solvent 4 but electrophoresis at pH 4.4 on a flat plate apparatus (Gray 1972a) was occasionally used as the t.l.c. separation was not always sufficient.

j) Dansyl Edman Degradations.

A modified method of Gray (1972 b) was used. Peptides (10nm. x no. of cycles, but less if not available) were dried down in 12 x 65mm screw cap tubes. Pyridine (50%, v/v, Rathburn Nonin Grade), 0.2ml. and phenylisothiocyanate (Pierce sequenal grade, 5% v/v in pyridine), 0.1ml were added, the tubes gassed with oxygen free nitrogen (O.F.N.), capped and incubated for 1 hr. at 37°C. Reagents were removed by drying in vacuo at 60°C. After cooling 0.2ml. T.F.A. was added, the tubes gassed with O.F.N. again, capped and inculated at 37°C for 30 mins. Occasionally incubation was for 5 mins. at 60°C. The T.F.A. was removed by drying in vacuo in a hot dessicator. Water, 0.25ml, was then added and the solution extracted 3 times with 2ml. butyl acetate. The phases were

separated by centrifugation and the upper, organic, layer discarded. The aqueous layer was dried overnight *in vacuo*, and the residue redissolved in 50% v/v pyridine. A sample was then removed, an equal volume of 50% pyridine added, and the next cycle begun. Newly exposed N-termini were identified by the dansyl procedure. Samples of degraded peptides were sometimes electrophoresed at pH 6.5 to locate amides. In one instance, samples were quantitatively analysed for amino acid composition.

k) Assignment of Amides

Amides were assigned on the basis of electrophoretic mobility at pH 6.5 using the equations of Offord (1966). Assignment of amides in the larger peptides and the histidine peptides has not usually been attempted.

l) The Use of Zeokarb

Beads of Zeokarb (Zk 225, S.R.C. 13)H^+ form, a sulphonated polystyrene resin was used to desalt peptides from T.C.A. contamination. The beads were added to an aqueous solution of the peptide until the pH had fallen to pH 2. Two volumes of distilled water were used to wash the beads, twice, and the peptides eluted with 2M ammonia (2x the bead volume, four times). It was found that larger peptides were not totally absorbed under these conditions.

m) Criteria of Purity

Ambler and Wynn (1973) proposed the following criteria

for satisfactory amino acid analyses,

1) no impurity amino acid should be present in an amount as great as 0.2 mol per mol.

2) The relative amounts of amino acids present, calculated on the basis that the average amount is integral should not fall outside the limits 0.8-1.2, 1.8-2.2, 2.7-3.3, or 3.7-4.3. Values as low as 0.6 (=1), however are considered satisfactory for tyrosine and homoserine. These criteria have been adopted as the basis for satisfactory analyses.

n) The Acceptability of Sequence Results

Dansyl Edman degradations could be considered unsatisfactory for the following reasons.

1) More than one dansyl amino acid present in significant amounts at any step. This was usually due to "ragged" degradation.

2) If no amino acid was identified at a particular step. The presence of tryptophan or lysine could lead to this, but where these residues were suspected identification of subsequent residues was usually satisfactory.

3) If the dansyl amino acid was not unambiguously identified. Poor t.l.c. separation between leucine and isoleucine derivatives, between the basic amino acid derivatives, or between serine, threonine, and methionine sulphone derivatives was the major cause of the failing.

p) The Recording of Sequence Data

Peptide sequences are listed in the text and amino acid analyses are grouped at the end of each chapter with a summary of the sequences. The following conventions have been used.

1) Amino acid analysis values outside the proposed criteria are marked thus +. Some unsatisfactory analyses are commented upon in the text. Unless otherwise stated values are after 24 hr. hydrolysis.

2) The three letter amino acid code has been used

3) An arrow, ——→ , underlining a residue indicates the amino acid was identified by the dansyl method. A dashed arrow, - - → , indicates unsatisfactory identification and the symbol, (—X), indicates that no residue was identified in that position. The symbol —⌒→ indicates that a dansyl dipeptide was identified. Sequences of arrows indicate results of dansyl Edman degradations.

4) Sequences are recorded by residues joined by hyphens, the N-terminal residue first. Residues enclosed in brackets are not ordered with respect to each other.

5) Peptide nomenclature consists of a letter (or letters) identifying the digest, and a number. Not all mumbers are assigned reflecting the fact that although other peptides were "purified" they are not included, usually owing to very unsatisfactory analysis. The ordering of the lists of compositions are on the basis of gel filtration poolings, the greatest elution volume pooling coming first. Within these poolings the peptides are ordered from the most acidic

to the most basic on the basis of pH 6.5 electrophoretic mobility.

6) Electrophoretic mobility (m) at pH 6.5 is relative to lysine (recorded as -) or relative to aspartic acid (recorded as +).

7) Purification methods (Purn.) are listed in the order of use; G = gel filtration; paper electrophoresis, 6 = pH 6.5, 3 = pH 3.5, 2 = pH 2, 9 = pH 9; B = paper chromatography; P indicates a pooling of peptide mixtures.

8) Yields are recorded as absolute yields, uncorrected for any losses.

9) Amino acids listed as cysteine were identified as cysteic acid. Similarly the analysis of methionine is as the oxidized form, summed with the native form if oxidation was not complete.

10) Relative analysis values below 0.1 mol per mol. are not recorded.

CHAPTER 4

CHYMOTRYPTIC DIGESTS

a) The First Chymotryptic Digest

1) Method

100 mg. of oxidised amidase (batch 1) were digested, with 1/75 by weight of chymotrypsin, in 0.2 M ammonium acetate pH 8.5, for 5 hr. at 37°C. Digestion was terminated by freeze drying. The residue was redissolved in 25% v/v formic acid and separated on a 90 x 1cm column of sephadex SG-25 eluted with 5% v/v formic acid. All peptides were isolated in poor yields. The oxidation of methionine was incomplete and analysis values are the sum of oxidized and unoxidized forms. Peptide compositions are given in table 4a, and a list of the peptides given in table 4c.

2) The Peptides

C1 The recovery of tyrosine from this peptide was low (0.5 = 1), and glycine contamination is unsatisfactory

Glu-Phe-Tyr

C2 Ser-Tyr-Phe

C3 Asp-Gly-Val-Tyr

C4 Heavy contamination with glycine is present (0.46 = 0), and the levels of serine and glutamic acid are high. Tyrosine has been placed on the basis of chymotrypsin's specificity

Ala-Asx-(Cys, His, Asx)Tyr

C5 This peptide also has serine and glutamic acid contamination

Ala-Ala-Gly-(Val, Asp, Gly, Tyr, Phe)

C6 The recovery of cysteine (as cysteic acid) and of tyrosine is low, and glycine contamination is evident. The calculated charge excess (0.75) indicates asparagine.

Ala-Asn-Cys-Tyr

C7 The recovery of isoleucine (0.5 =1) and valine (0.4 = 1) is low but the analysis is after 24hr hydrolysis and a resistant bond (Ile-Val) is present

Ile-Val-(Arg, Cys, Gln, Gly)Tyr

C8 The analysis is after 24 hr. hydrolysis only, and contaminants are present as shown in table 2a.

Ala-(Thr, Gln_2, Ala, Val, Leu_2)

C15 The value of alanine (0.4) is taken as 0, but as this value is after 24hr. hydrolysis only, the presence of alanine cannot be excluded.

Asx-Gly(Arg_2, Cys, Thr, Glx_6, Gly_3, Met, Ile, Leu, Tyr)

C16 The recovery of glutamic acid (2.6) has been taken as 3, which fits the mobility data. Placement of leucine within the sequence suggests that this may be the C-terminal peptide.

Glu-Gly-Leu-Glu-(Lys, Glu, Ala)

C17 A large aromatic peptide. Identification of the N-terminal 2 residues was good, although the analysis value of aspartic acid is unsatisfactory.

Ser-Gly-(Gly_3,Ala_2,Leu_2,Ser,Asx_{2-3},Glx_3,Phe_2,Tyr,
,Cys,Arg,Pro)

C18 Amino acid analysis showed an unidentified peak eluting immediately before the position of arginine. Qualitative analysis showed no basic residues were present and this is supported by the mobility. The similarity of composition to peptide CT7 suggests this peak may be a dipeptide (Val-Leu). An alternative explanation is that it is an artefact caused by the elution of ammonia in several positions within this region of the chromatogram.

Thr(Ala_2,Glu,Val,Leu)

C19 The presence of bonds resistant to hydrolysis may have contributed to the low recovery of amino acids after 24hr. hydrolysis. Contamination is quite heavy, and although o-Dns-tyrosine was recorded after dansylation, the analysis suggests none is present.

Val-(Ile,Asx,Glx,Gly,Ala,Leu,Lys)

C20 The Edman degradation was carried out on 2nm per cycle.

Val-Ile-Leu-Asx-(Lys,Gly,Val,Ile,Asx_2,Glx_2,Tyr)

C22 On analysis after 24 hr. hydrolysis proline gave a very small peak, not amenable to accurate quantitation. Qualitative analysis did not show the presence of proline.

Val-Ala-Val-(Gly_2,Ala_3,Asn,$Pro_{0.5}$)Phe

C23 There was a low recovery of tyrosine from this peptide (0.5 = 1). The neutral mobility indicates that two amides are present.

Met-Tyr-Pro(Ala,Lys,Asx,Glx_2,Val)

C24 Val-Ser-Glu-(Gly_2,Pro,Lys,Met)

C25 Lys-Ile-Leu

C27 This peptide was analysed after C18 and shows the same unidentified peak supporting the suggestion that it is an artefact. Heavy contamination by glycine (0.32 = 0) and alanine (0.37 = 0) is present.

Lys(Ile,Leu,Ser)

C31 This shows a similarity to C15 in mobility and composition. The discrepancy in the values for glutamic acid may be due to its high relative proportion, but there is also a difference in the arginine content.

Asx-Gly(Arg,Cys,Thr,Glx_4,Gly_2,Leu,Ile,Met,Tyr)

C32 Low yields of lysine (0.6 = 1) and valine (1.7 = 2) are recorded after 24 hr. hydrolysis. The N-terminal dipeptide was not identified and the mobility suggests two amides are present (i.e. a charge excess of 1).

Val-(Val,Arg,lys,Asx_3,Met,Glx_2,Ala_2,Leu)

C33 This was isolated as a U.V. fluorescent peptide which therefore may contain tryptophan. The N-terminus was identified as glutamic acid or glutamine but the quantitative analysis was lost owing to a machine fault.

C36 The N-terminus was identified by t.l.c. of the dansyl derivative as a basic residue, possibly arginine. The mobility suggests no amides are present, but the presence of

histidine makes this forecast unreliable.

$(Arg_2, His_2, Lys, Pro_2, Tyr, Glx_3, Thr, Ser, Leu, Ala, Gly)$

b) The Second Chymotryptic Digest

1) Method

Oxidized amidase (330mg) was digested under the same conditions as the previous chymotryptic digest. The amidase concentration was 25 mg. per ml. Insoluble material was removed by centrifugation and the digest freeze dried. By subsequent re-digestion this insoluble material was shown by peptide mapping to be undigested amidase which was then pooled with the bulk peptides.

The freeze dried mixture was taken up in 0.45M ammonium acetate pH 9.2 and the core peptides precipitated with 5% (w/v, final concentration) T.C.A. The pH after precipitation was pH 5.3. Core material was removed by centrifugation. T.C.A. was removed from the supernatant by ether extraction before freeze drying. The residue was then dissolved in 1.2mls of 65% (v/v) formic acid and fractionated on a 130 x 2.5cm column of sephadex SG-25, run in 5% v/v formic acid. The peptide map appeared similar to that of the first chymotryptic digest. Further extraction of some poolings was necessary for the removal of T.C.A. The pooling containing the largest peptides, and the neutral peptide fractions streaked very badly on paper electrophoresis and chromatography. These fractions were "desalted" on a short (20 x 1 cm) sephadex G-25 column but no improvement

was found. Ion exchange on Zeokarb, Zk225 SRC 13 (H^+ form) was also used to no avail.

The neutral fractions were then pooled and reseparated by gel filtration (60 x 1 cm column) and a peptide map run on pH 3.5 electrophoresis. It was then possible to purify some neutral peptides, but others were still not adequately pure after several further purification steps.

2) The Peptides

This digest is only partially characterised, and some redundant information has been obtained. Peptide compositions are given in table 4b and a summary of the information is in table 4d.

CT1 3 cycles of Edman degradation were carried out to ensure that only 3 residues were present.

Ser-Tyr-Phe
$\longrightarrow \longrightarrow \longrightarrow$

CT2 An extra Edman cycle was also used to establish the size of the tripeptide.

Glu-Phe-Tyr
$\longrightarrow \longrightarrow \longrightarrow$

CT3 Asp-Gly-Val-Tyr
$\longrightarrow \longrightarrow \longrightarrow \longrightarrow$

CT4 This peptide has heavy contamination with serine (0.36 = 0) and threonine (0.34 = 0). The mobility at pH 6.5 was not accurately measured but defines the excess charge limits $1.3 \leq \epsilon \leq 2.2$ suggesting at least one amide is present.

Gly-(Ala, Asx, Glx_2)
$\longrightarrow$

CT5 Even on 96 hr, hydrolysis the recovery of alanine (1.65 = 2) was low. Two impurities, valine (0.37) and methionine (0.45) were present only in the 96 hr. hydrolysate.

Ala-(Asx,Leu,Ile,Cys,Glx_3,Pro,Gly_2,Ala,Phe)

CT6 The 24 hr. hydrolysate suggested equimolar ratios of the constituent amino acids, but the listed composition was found after 96 hr. hydrolysis.

Ala-(Glx_2,Ala_2,Val,Leu_2)

CT7 Thr-Ala-Ala-Glu-Val-Leu

CT10 The analysis shows high glycine contamination but the Edman degradation was clean.

Val-Val-Phe-Pro-Glu-Tyr

CT11 Histidine appears to have been destroyed on hydrolysis as the recoveries were 24 hr. 0.72 = 1, 96 hr. 0.48 = 1.

Gly-(Ser,Gly,Ala,His,Ile)Phe

CT12 Recovery of tyrosine (0.51 = 1), lysine (0.56 = 1) and alanine (0.66 = 1) was low even on 96 hr. hydrolysis. The relative recovery of glutamic acid rose from 0.92 to 1.38 from 24 hr. to 96 hr. hydrolysis and the presence of 2 such residues cannot be excluded.

Lys-(Arg_2,Ser,Glx,Gly,Ala,Ile),Tyr

CT13 Lys-Ile-Leu

CT14 Edman degradation failed, but the mobility is in agreement with the presence of 1 amide (excess charge 4.99). Subdigestion with thermolysin revealed only a neutral peptide

band, and subdigestion with S.aureus protease showed only a slightly acidic peptide. Electrophoresis for only a short time did not reveal any very acid peptides that would normally have been lost. The products of subdigestion were not identified.

$\underrightarrow{Gly}(Gly_2, Cys, Met, Glu_4, Gln, Ile)Tyr$

CT15 No new N-terminus was identified after 4 dansyl-Edman steps but later residues were clearly identified. Lysine has therefore been placed from the composition. Such a sequence is in agreement with the carboxy peptidase digest results of Brown et al (1973) if this is the C-terminal peptide.

$\underrightarrow{Glu}\text{-}\underrightarrow{Gly}\text{-}\underrightarrow{Leu}\text{-}\underrightarrow{Glu}\text{-}\underrightarrow{Lys}\text{-}(\underrightarrow{x})\text{-}\underrightarrow{Glu}\text{-}\underrightarrow{Ala}$

CT16 No N-terminal amino acid was identified by t.l.c. or electrophoresis after the dansyl procedure. The recovery of aspartic acid fell after 96 hr. hydrolysis from 3.0 to 2.5.

$(Lys, Arg, Cys, Asx_3, Glx, Gly, Ala_2, Leu)$

CT17 This corresponds to peptide C23 with the addition of a methionine residue. This peptide is only slightly acidic, and probably contains 2 amides. The anodic mobility being due to the presence of N-terminal methionine sulphone. CT17 is resistant to S.aureus protease digestion suggesting two glutamine residues are present. Subdigestion with thermolysin revealed only neutral products. No residue was identified after 4 dansyl-Edman cycles but residue 6 was identified satisfactorily

$\underrightarrow{Met}\text{-}\underrightarrow{Tyr}\text{-}\underrightarrow{Pro}\text{-}\underrightarrow{Ala}\text{-}\underrightarrow{Lys}\text{-}(\underrightarrow{x})\text{-}\underrightarrow{Asx}\text{-}(Glx_2, Val, Met)$

Neutral peptides

CT18 Low recoveries of alanine and valine are recorded (0.5 = 1) after 24 hr. hydrolysis.

$\underrightarrow{\text{Ser}}$-(Arg,Thr,Ser,Glu,Gly,$\text{Leu}_2$,Val,Ala)

CT19 $\underrightarrow{\text{Gly}}$-Val-Phe

CT20 Low recoveries of valine (0.63 = 1), and glutamic acid (0.6 = 1) were recorded after 24 hr. hydrolysis.

$\underrightarrow{\text{Ser}}$-(Asn,Thr,$\text{Ser}_3$,Gln,Gly,Ala,Val,Leu,Tyr,Phe)

CT21 This was purified at all stages as a U.V. fluorescing peptide suggesting tryptophan was present in the original. The recovery of glutamic acid on 24 hr. hydrolysis is very low (0.4 = 1).

$\underrightarrow{\text{Ile}}$(Lys,Arg,$\text{Asx}_2$,Met,Glx,Gly,$\text{Ala}_2$,Val),Trp

CT22 Isoleucine (0.5 = 1) and glutamic acid (1.6 = 2) were recovered in low yield after 24 hr. hydrolysis.

$\underrightarrow{\text{Ser}}$(Arg,$\text{Asx}_2$,$\text{Glx}_2$,Ala,Ile)

TABLE 4A. CHYMOTRYPTIC PEPTIDES 1ST DIGEST

Peptide	m 6.5	V/Vo G-25	Purn	Yield %	Gly	Ala	Val	Leu	Ile	Ser	Thr	Asp + Asn	Glu + Gln	Phe	Tyr	Trp	Cys	Met	Pro	Lys	His	Arg	TOT	N–t	
C1	+0·46		G6	2·36	+ 0·22							0·14	1·08	0·91	+ 0·18								3	Glu	
C2	0		G63	5·38						0·96				1·03	0·84								3	Ser	
C3	+0·5		G63	3·75	1·05		0·98					0·97			0·69								4	Asp	
C4	+0·41		G63	0·68	+ 0·46	0·99				+ 0·28		2·01	+ 0·32		0·82		0·92				0·99		6	Ala	
C5	+0·29		G63	0·58	2·03	1·89	1·03			+ 0·24		1·19	+ 0·21	0·89	0·73								8	Ala	
C6	+0·29		G63	0·7	+ 0·3	0·97				0·18		1·03	+ 0·2		+ 0·47		+ 0·57						4	Ala	
C7	0		G63B3	2·27	1·03		+ 0·37		+ 0·46				0·97		0·79		+ 0·75					1·02	7	Ile I–V	24 hr hydrolysis
C8	0		G63B32	1·07	+ 0·23	+ 2·27	1·01	+ 1·73		+ 0·3	0·98	+ 0·24	2·02										8	Ala	
C15	+0·71		G63	0·99	4·2	+ 0·4		+ 1·24	1·13	+ 0·2	1·0	0·82	6·01		0·97		1·04	0·73		0·14		1·86	19–20	Asx	cf. C31
C16	+0·58		G63	2·23	1·17	1·04		1·03		+ 0·22			+ 2·6							1·15			7	Glu	
C17	+0·52		G63	1·19	3·94	2·0		1·98		2·18		+ 2·45	3·3	1·91	0·95		0·99		0·88			0·97	21–22	Ser	
C18	+0·37		G63	5·47		1·94	0·97	1·05			0·91	0·13	1·09										6	Thr	SEE TEXT
C19	+0·37		G63	0·5	0·82	+ 0·68	+ 0·65	+ 0·77	+ 0·58	+ 0·32	+ 0·25	1·17	1·19		+ 0·25					0·82			8	Val V–x	
C20	+0·21		G62	2·24	1·13		1·79	0·96	1·96			2·87	2·04		0·96					0·94			13	Val V–I	96 hr. hydrolysis

TABLE 4A CTD. CHYMOTRYPTIC PEPTIDES 1ST DIGEST.

Peptide	m 6.5	V/Vo G-25	Purn	Yield %	Gly	Ala	Val	Leu	Ile	Ser	Thr	Asp +Asn	Glu +Gln	Phe	Tyr	Trp	Cys	Met	Pro	Lys	His	Arg	TOT	N-t	
C22	0		G3B3	0.88	+ 1.72	4.14	+ 2.32	+ 0.23		0.11		1.03	+ 0.23	1.1	0.16				+ 0.45			+ 0.27	10-11	Val	
C23	0		G3B3	1.07	+ 0.26	1.01	+ 0.74			+ 0.24		1.06	1.93		+ 0.51			+ 1.23	0.99	0.82			9	Met	
C24	0		G3	12.7	1.93	0.14	1.03	0.11		0.94		+ 0.21	1.13					0.96	+ 1.22	0.85			8	Val	
C25	-0.51		G3	1.61	0.17	0.12		1.09	0.93	0.14			0.14							0.98			3	Lys	
C27	-0.47		G3	0.66	+ 0.32	+ 0.37	0.12	1.01	0.8	0.99		+ 0.23								0.99		0.25	4	Lys	SEE TEXT
C31	+0.72		G3	3.03	2.96	+ 0.28		0.85	0.97	0.11	0.89	1.0	4.3		0.71		+ 0.73	0.9				0.88	15	Asx	cf C15
C32	+0.19		G3	1.12		1.81	+ 1.67	0.92				3.09	2.05					1.07		+ 0.6		1.06	13	Val V-X	
C33	0		G33	1.55	Lost	1.99	2.2	1.14		←	LOST (+)	(+)	(+) →			+				0.87		2.8		Glx	(+) Present qualitatively
C36	-0.38		G3	2.55	1.05	1.03		1.0		0.83	0.95		3.13		0.76				2.12	0.92	2.02	2.06	16	Basic	

TABLE 4B CHYMOTRYPTIC PEPTIDES 2ND DIGEST

Peptide	m 6.5	V/Vo G-25	Purn	Yield %	Gly	Ala	Val	Leu	Ile	Ser	Thr	Asp +Asn	Glu +Gln	Phe	Tyr	Trp	Cys	Met	Pro	Lys	His	Arg	TOT	N-t	
CT1	0		G63	10·4						0·81		0·14		1·06	1·14								3	Ser	
CT2	+0·48		G63	7·43									1·16	0·96	0·88								3	Glu	
CT3	+0·51		G63	12·2	1·16		0·97			0·23 +		0·84			0·71								4	Asp	
CT4	–		G6P3	0·74	0·88	0·98				0·46 +	0·34 +	1·0	2·14										5	Gly	
CT5	–		G6P3	3·73	1·65 +	1·83		0·77 +	0·88		0·19	1·23 +	2·82	0·99			1·28 +		1·23 +				13	Ala	SEE TEXT / 96 hr values
CT6	–		G63P2	0·74		2·72	0·74 +	1·57 +					1·96										8	Ala	96 hr values
CT7	+0·42		G6P3	4·83		1·94	1·13	0·84			0·79		1·09										6	Thr	96 hr values
CT10	+0·31		G6P3	5·59	0·27 +	0·16	2·01	0·13		0·1		0·17	1·03	1·02	0·88				0·95				6	Val V-V	96 hr values
CT11	-0·11		G63	2·98	1·71 +	1·11	0·24 +		1·3 +	0·81	0·14	0·25 +	0·23 +	0·89							24 hr 0·72		7	Gly	96 hr value
CT12	–		G6PB3	1·24	0·89	0·7 +			1·09	0·98		0·22 +	1·38 +		0·51 +					0·56 +		1·98	9	Lys	96 hr value
CT13	–		G6PB3	1·37				1·0	1·15											0·8			3	Lys	
CT14	+0·91		G6	11·4	2·7				0·99			0·12	5·16		0·72		1·04	0·97					12	Gly	96 hr values
CT15	+0·61		G63	10·3	1·21 +	1·05		1·02					2·82							0·9			7	Glu	
CT16	+0·22 →0·29		G63	1·49	1·22 +	2·0	0·32 +	0·96				24 hr. 3·04	1·12				0·97			0·8		0·94	11	–	96 hr value

TABLE 4B CTD. CHYMOTRYPTIC PEPTIDES 2ND DIGEST.

Peptide	m 6.5	V/Vo G-25	Purn	Yield %	Gly	Ala	Val	Leu	Ile	Ser	Thr	Asp +Asn	Glu +Gln	Phe	Tyr	Trp	Cys	Met	Pro	Lys	His	Arg	TOT	N-t	
CT17	+0.1		G63	8.1		0.91	1.26 +					1.17	1.7 +		24 hr 0.74			1.85	1.19	0.9			10	Met	120 hr value
CT18	0		GGB39	0.24	1.2	0.54 +	0.45 +	2.13		1.79	0.95	0.22 +	1.06									0.78 +	10	Ser	
CT19	0		GG39	0.5	1.1		1.07							0.83									3	Gly	
CT20	0		GG39	0.5	1.16	0.78 +	0.63 +	0.96	0.14	4.0	1.22	1.08	0.6 +	0.8	0.61			0.18				0.25 +	13	Ser	
CT21	0		GG39	5.1	0.99	1.8	0.75 +	0.19	1.22	0.21 +	0.12	2.22 +	0.4 +			(+)		1.24 +	0.19	0.77 +		0.99	10 -12	Ile	
CT22	0		GG39	1.7		1.3 +	0.2 +		0.5 +	1.02		2.1	1.62 +							0.24 +		0.97	8	Ser	

TABLE 4C

Summary of Chymotryptic peptides: 1st digest.

C1 Glu-Phe-Tyr

C2 Ser-Tyr-Phe

C3 Asp-Gly-Val-Tyr

C4 Ala-Asx-(Cys,His,Asx),Tyr

C5 Ala-Ala-Gly-(Val,Asp,Gly,Tyr,Phe,)

C6 Ala-Asn-Cys-Tyr

C7 Ile-Val-(Arg,Cys,Gln,Gly),Tyr

C8 Ala-(Thr,Gln_2,Ala,Val,Leu_2)

C15 Asx-Gly-(Arg_2,Cys,Thr,Glx_6,Gly_3,Met,Ile,Leu,Tyr)

C16 Glu-Gly-Leu-Glu-(Lys,Glu,Ala)

C17 Ser-Gly-(Gly_3,Ala_2,Leu_2,Ser,Asx_{2-3},Glx_3,Phe_2,Cys,Pro, Arg,Tyr)

C18 Thr-(Ala_2,Glu,Val,Leu)

C19 Val-(Ile-Asx,Glx,Gly,Ala,Leu,Lys)

C20 Val-Ile-Leu-Asx-(Lys,Gly,Val,Ile,Asx_2,Glx_2,Tyr)

C22 Val-Ala-Val-(Gly_2,Ala_3,Asn,$Pro_{1/2}$),Phe

C23 Met-Tyr-Pro-(Ala,Lys,Asx,Glx_2,Val)

C24 Val-Ser-Glu-(Gly_2,Pro,Lys,Met)

C25 Lys-Ile-Leu

C27 Lys-(Ile,Leu,Ser)

C31 Asx-Gly-(Arg,Cys,Thr,Glx_4,Gly_2,Met,Ile,Leu,Tyr)

C32 Val-(Val,Arg,Lys,Asx_3,Met,Glx_2,Ala_2,Leu)

C33 Glx(Ala_2,Val_2,Leu,Trp,Lys,Arg_3...) see text

C36 (Arg_2,His_2,Lys,Pro_2,Tyr,Glx_3,Thr,Ser,Leu,Ala,Gly)

TABLE 4D

Chymotryptic Peptides - 2nd digest.

CT1 Ser-Tyr-Phe (=C2)

CT2 Glu-Phe-Tyr (=C1)

CT3 Asp-Gly-Val-Tyr (=C3)

CT4 Gly-(Ala,Asx,Glx_2)

CT5 Ala-(Asx,Leu,Ile,Cys,Glx_3,Pro,Gly_2,Ala,Phe)

CT6 Ala-(Glx_2,Ala_2,Val,Leu_2)

CT7 Thr-Ala-Ala-Glu-Val-Leu (=C18)

CT10 Val-Val-Phe-Pro-Glu-Tyr

CT11 Gly(Ser,Gly,Ala,His,Ile)Phe

CT12 Lys(Arg_2,Ser,Glx,Gly,Ala,Ile),Tyr

CT13 Lys-Ile-Leu (=C25)

CT14 Gly(Gly_2,Cys,Met,Glu_4,Gln,Ile),Tyr

CT15 Glu-Gly-Leu-Glu-Lys-Glu-Ala (=C16)

CT16 (Lys,Arg,Cys,Asx_3,Glx,Gly,Ala_2,Leu)

CT17 Met-Tyr-Pro-Ala-Lys-Asx-(Glx_2,Val,Met)

CT18 Ser(Arg,Thr,Ser,Glu,Gly,Leu_2,Val,Ala)

CT19 Gly-Val-Phe

CT20 Ser-(Asn,Thr,Ser_3,Gln,Gly,Ala,Val,Leu,Tyr,Phe)

CT21 Ile-(Lys,Arg,Asx_2,Met,Glx,Gly,Ala_2,Val)Trp.

CT22 Ser-(Arg,Asx_2,Glx_2,Ala,Ile)

CHAPTER 5

TRYPTIC DIGESTS

a) The First Tryptic Digest

1) Method

Oxidized amidase (150mg) was digested with 1/75 by weight of trypsin in 30 ml. of 0.2 m ammonium acetate pH 8.5 at 37°C. The protein did not dissolve well and some was lost when the flask was accidentally cracked. Digestion was allowed to proceed overnight before freeze drying. The residue was dissolved in formic acid (0.9 ml) and diluted to 1.8 mls. with water for gel filtration on a 1.5 x 120cm. column of sephadex sG-25 eluted with 5% v/v formic acid.

Peptide compositions are given in table 5A and a summary of results in table 5C. Many peptides appear to have arisen by chymotryptic type splitting.

2) The Peptides

Many free amino acids were isolated mostly in low yield but comparable to some of the peptides and are therefore included.

T1 Asp

T3 Two analyses were performed and only one showed heavy threonine contamination (0.4 = 0). This peptide has the sequence found for the protein C-terminus (Brown et al 1973)

Glu-Ala

T4 The recovery of glutamic acid is low (1.6 = 2) and owing to a machine fault the value for cysteic acid is a minimum value.

$\underrightarrow{\text{Gly}}$-(Cys,$\text{Glx}_2$,Ala,Leu,Tyr,$\text{Phe}_2$,Pro)Arg

T6 Ser

T7 Gly

T8 Ala

T9 Val

T10 Leu

T11 This was purified as a U.V. fluorescing peptide at each purification step. After the first Edman cycle no new terminal residue was identified but the identification of residue 3 was clear. Tryptophan has been tentatively positioned as residue 2 to explain these observations. On quantitative analysis the lysine peak position was overlapped by ammonia but qualitative analysis indicates that no lysine is present.

$\underrightarrow{\text{Ala}}$-$\underrightarrow{\text{Trp}}$-Val-(Gly,Val,Asn)
$\underrightarrow{(\rightarrow X)}$

T12 A large ammonia peak masked the position of lysine on the amino acid analysis, but qualitatively no lysine was found.

$\underrightarrow{\text{Gly}}$(Val,Phe)

T13 This peptide had a yellow grey ninhydrin colour. The presence of lysine is ruled out by qualitative analysis only, for the same reason as above.

$\underrightarrow{\text{Ser}}$(Gly,His,Tyr,Phe)

T14 An amide has been assigned from the mobility of this peptide.

Asn-Tyr-Lys
→

T15 Only a small amount of material was available for analysis, and the high aspartic acid value (0.39 = 0) is a maximum owing to difficult quantitation of a small peak.

Lys(Ala,Pro,Tyr)Lys.
→

T17 Tyr-Arg
→ - →

T18 The presence of lysine is ruled out qualitatively only owing to a large ammonia peak on the analyser trace.

Arg

T21 The value for aspartic acid (1.4 = 1) is high and glutamic acid contamination (0.35 = 0) is also high. The composition as shown has a charge excess of 1.1 and hence no amides are present.

Ala-(Met,Asp,Cys)Lys
→

T22 Lysine was not quantitated owing to masking by ammonia. Qualitatively lysine is present and its value is set at 1 from the mobility data.

Asp-(Glu,Met,Val)Lys
→

T23 The mobility indicates that one amide is present.

Asx-Glx-Val-Asx-Arg
→ → → →

T24 This was purified as a U.V. fluorescing peptide indicating the probable presence of tryptophan. The recoveries of alanine and isoleucine are low (24 hr. hydrolysis only) and contamination with threonine is high.

Ile-(Ile_2,Cys,Asx_4,Ser,Glx_2,Pro_2,Gly_3,Ala_2,Val,Leu,Tyr_2,Trp)Arg.

T25 A high recovery of glycine (1.38 = 1) is reported, and no amides can be present.

Leu-$(Glu_2$,Pro,Gly,Leu,Tyr)Lys

A chymotryptic subdigest was carried out but insufficient material was available for characterisation of the products.

T26 The recovery of serine (1.6 = 2) is low.

Gly-$(Asx_2$,Ser_2,Glx,Gly_3,Ala,Leu,Tyr,)Arg

A chymotryptic subdigestion released the following peptides.

T26C1 The recovery of arginine is low (0.6 = 1) and the serine and alanine values are higher than expected from the composition of T26. Only 10nm was available for analysis and these high values may be due to contamination.

Ala-$(Asx_2$,Ser_3,Glx,Gly,Ala,Arg)

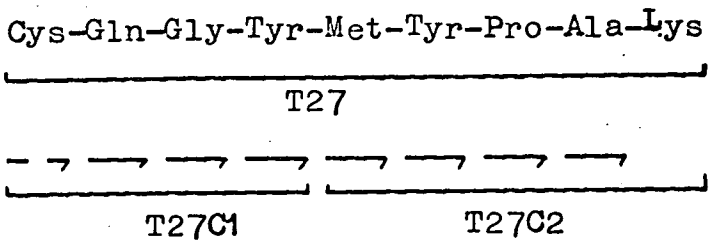
T26C2 a neutral peptide whose composition was determined qualitatively but insufficient material was available for further investigation.

Gly-Tyr

Another acidic peptide (m = 0.6 at pH 6.5, Nt. gly) and another neutral peptide (containing glycine and leucine) were also found but not in sufficient yield for further analysis. T26 therefore has the probable structure

Gly(Gly_2,Leu,Tyr)Ala(Asx_2,Ser_2,Glx,Gly)Arg

T27 Chymotryptic subdigestion (i.e. T27C1, T27C2) suggests the following sequence.



T27C2 was purified as 2 peptides, one having N terminal methionine, and one having N-terminal Methionine sulphone. They otherwise had the same composition and sequence.

T28 Ala-Cys-Arg

T29 Asp-Ala-Arg

T34 This peptide has the same sequence as that isolated by Brown and Clarke (1972) from a mixed chymotryptic and tryptic digest.

Ser-Leu-Thr-Gly-Glu-Arg

T35 This was isolated as a U.V. fluorescent peptide but Edman degradation suggests tryptophan is not present.

Gly-Ala-Glu-Leu-Ile-Val-Arg

T36 Lys-Ala-Pro-Tyr-Arg

T37 A low recovery of methionine (0.6 = 1) is recorded

Met-Pro-Arg

T38 Leu-Thr-Arg-Arg

T39 The recovery of lysine (0.74 = 1) is low

Gly-Met-Lys

T40 Ala-Arg

T41 The presence of lysine was shown by qualitative analysis,

but on quantitation this residue was masked by ammonia. The mobility indicates a charge excess of 2

$\underrightarrow{\text{Tyr-}}$(Arg,Lys)

T42 Lys

T43 On 96 hr. hydrolysis many residues gave low recoveries. The N-terminus was identified as valyl valine.

$\underrightarrow{\text{Val-}}$($Val_2$,Pro,Thr,Ser,$Glx_2$,Gly,Ala,Ile,Tyr,Phe,Arg,Lys)

T44 No basic residues were detected quantitatively or qualitatively. A charge excess of 4 is indicated by the mobility, thus 2 amides are present.

$\underrightarrow{\text{Thr-}}$(Cys,Met,$Glx_5$,Ile,Leu,$Gly_4$)

T45 A similar peptide to T44 except tyrosine is also present. The differences in the relatively high molar ratios of glutamic acid and glycine may not be significant. The mobility of T45 with the composition shown suggests only 1 amide is present.

$\underrightarrow{\text{Thr-Leu-Gly-}}$(Cys,Met,$Glx_4$,$Gly_2$Ile,Tyr)

T46 Only sufficient material for a 24 hr. hydrolysis was available and a low recovery of glycine (0.6 = 1) is recorded. The value for methionine does not differentiate between 1 and 2 mol. per mol.

$\underrightarrow{\text{Tyr-}}$(Asx,Thr,Gly,$Glx_2$,Pro,$Ala_2$,$Met_{1-2}$)

T47 Recoveries of glutamic acid and isoleucine are low even after 96 hr. hydrolysis. After 3 Edman cycles, when proline was identified, heavy contamination by isoleucine and glycine was recorded.

Val-Ala-Ile-Pro-Gly-(Thr,Ser,Glu_3,Ile,Phe)Arg

T50 As this peptide has a neutral mobility 2 amides must be present. The mobility of the residual peptides after 1,2 and 3 Edman cycles (-0.45,-0.49,-0.54 resp.) suggest a charge excess of 1.5. This may be due to the presence of methionine sulphone although this is usually only notable when this residue is N-terminal.

Asp-Gln-Gln-Val-Met-Met-Ala-Lys

T51 Val-Ser-Glu-Gly-Pro-Lys

T52 Even after 96 hr. hydrolysis the recovery of isoleucine (1.4 = 2) is very low.

Ile-Ala-Asp-Met-Ile-Val-(Met,Gly)Lys

T54 The presence of two histidines makes the placement of amides difficult.

His-Glx-Glx-His-Pro-Arg

T55 After 3 cycles of Edman degradation the new N-terminus was identified as a basic residue (ε-dansyl lysine was not recorded after any Edman cycles presumably due to the formation of ε-phenylthiocarbamyl lysine) but not unambiguously as histidine.

Ala-Asx-Asx-His-Glx-Asx-(Leu,Ser,Glx,Phe)Lys

A chymotryptic subdigest was carried out but lack of material led to inconclusive results.

T56 Ala-Gln-Ile-(Ser_2,Gln,Leu_2)Arg

T57 The recovery of leucine (0.5 = 1) was low even after 96 hr. hydrolysis, and the value for arginine rose from 0.2

(24 hr. value) to 0.5. N-terminal identification was clean, indicating isoleucyl isoleucine. This peptide was purified as being U.V. fluorescent and probably contains tryptophan. Edman degradation was unsatisfactory.

Ile-(Ile, Cys, Asx_2, Thr, Ser, Glx_3, Pro_4, Ala, Gly_4, Val_2, Tyr_2, Trp, Arg, Leu, Lys)

T58 Amino acid analysis showed an unidentified peak eluting after the position of histidine. This has been regarded as an ammonia artefact. Neutrality indicates that no amides are present, and Edman degradation was unsuccessful.

Asp(Glu_2, Val, Met_2, Ala, Lys_3)

T59 The N-terminus was identified as isoleucyl alanine. The amino acid composition bears similarities to that of T52, but with an apparently reduced isoleucine and methionine content. Both peptides T52 and T59 have identical mobilities after electrophoresis at pH 3,5 and pH 2, and the same chromatographic mobility. T59 therefore probably represents T52 with low recoveries of isoleucine and methionine. The composition listed is that found for T59.

Ile-(Ala, Asp, Met, Gly, Val)Lys

T60 After 6 cycles of Edman degradation valyl leucine was identified as the new N-terminus.

Leu-His-Thr-Ala-Ala-Glx-Val-(Leu, Ala, Asx_2)Arg

T61 It is thought that T61 represents T52 with an intact lysyl isoleucine N-terminus. This would suggest that recovery of isoleucine is low (1.0 = 2). The composition, below, is that

found for T61

Lys-Ile-Ala-Asp-(Met_2,Gly,Val)Lys

b) Second Tryptic digest

1) Method

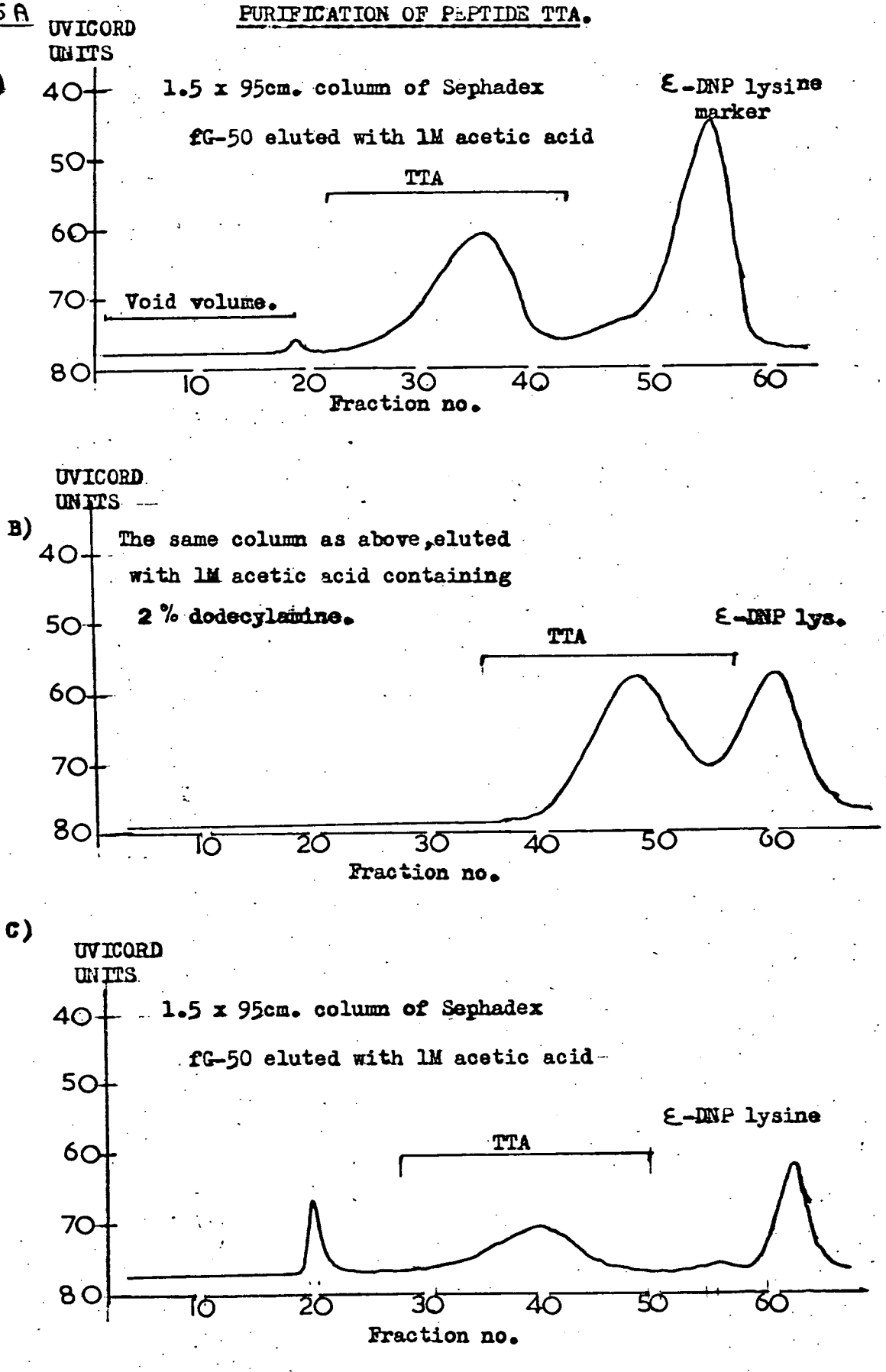
250 mg. of oxidized amidase were digested with trypsin (1/75 by weight) for 5 hr. at 37°C in 0.2 m ammonium acetate pH 8.5 (58 ml.). Insoluble material was removed by centrifugation before precipitation of core peptides with T.C.A. The core material was spun down, redissolved in formic acid and freeze dried. The soluble peptides were freeze dried and extracted with ether, both dry and in solution (5% v/v formic acid), and the aqueous phase freeze dried. This residue was again extracted with ether prior to separation on a 2.5 x 130cm. column of sephadex G-25 eluted with 5% v/v formic acid.

2) Attempted fractionation of core peptides.

Although the core material was soluble in formic acid it precipitated on dilution. Many solvent systems were tried and it was found that these peptides did not precipitate if dissolved in formic acid and adjusted to the solution formic acid : ethanol : water (1:2:2 v/v). Core material dissolved in 3 ml. of this solvent was applied to a 2.5 x 75cm. column of sephadex LH-20 equilibrated in the same solvent. Unfortunately this separation failed when the column became blocked probably

FIG 5A

PURIFICATION OF PEPTIDE TTA.



due to the precipitation of the core peptides. A small volume of pyridine was passed through the column to recover this material.

3) Purification of TTA

From the elution profile and peptide map of the soluble peptides fraction TTA appeared to be composed of soluble material too large to migrate on paper electrophoresis. A gel filtration purification using detergent, to cause separation on the basis of charge (Strid, 1973) was devised. After elution from a 1.5 x 95cm. column of sephadex fG-50, run in 1M acetic acid, the major U.V. absorbing component was freeze dried. The residue was taken up in 1M acetic acid containing 2% w/v dodecylamine, and applied to the same column equilibrated and eluted with the acetic acid dodecylamine solution. TTA eluted as a single peak with an altered elution volume. After extraction with n-heptane : butan-1-ol (5:1 v/v) TTA was eluted from a 1.5 x 95cm. column of sephadex fG-50 eluted with 1M acetic acid. This material eluted as a single peak with the original elution volume although as the U.V. absorption indicates there had been heavy extractive losses. The elution profiles are shown in fig 5A.

4) The Peptides

These peptides are all soluble peptides. Very few were characterised from this digest but TT4 is known to have been lost from the first tryptic digest owing to its very weak

colouration with ninhydrin. Compositions are listed in table 5B and a summary of this digest is given in table 5D.

TT1 This amino acid was not quantitated: Arg

TT2 Amino acid analysis gives the composition

$\underrightarrow{\text{Glu}}$-(Asp$_2$,Glu,Ala,Val)Arg

but Edman degradation suggested the part sequence

$\underrightarrow{\text{Glu-Ala-Ala-Ala-}}$

which indicates a rather larger composition (calculated from the same amino acid frequencies) with 2 amides i.e.:

Glx-Ala-Ala-Ala-(Val$_2$Glx$_4$,Asx$_5$Arg$_3$).

As no aromatic residues are present this would not have been expected to be greatly retarded on gel filtration and the former composition is more consistent with the elution volume.

TT4 Dansyl Edman degradation did not unambiguously define the sequence of this peptide. The sequence was deduced by amino acid analysis of the residual peptides after 1 to 3 Edman degradation cycles (i.e. TT4 ψ1 to TT4 ψ3 respectively as in the table of compositions).

Ile-Leu-His-Arg

TT5 The recovery of many amino acids is low after only 24 hr. hydrolysis. It cannot be overlooked that this apparently large peptide is perhaps a mixture of peptides, although only 1 N-terminal residue was found.

$\underrightarrow{\text{Lys}}$(Cys,Asx,Thr,Glx,Ser,Pro$_2$,Gly,Ala$_3$,Val,Ile Leu,Tyr$_2$,Lys,Arg$_2$)

TT6 Lys

TTA Amino acid analysis indicates that this peptide consists of about 60 amino acids. Automatic Edman degradation using a single coupling, single cleavage programme with dimethylbenzylamine buffer (the run was performed by Mrs. M. Daniel) established only the N-terminal sequence

Ala-Ile-Leu-Gln-

The residues were identified by both t.l.c. and g.l.c. but the presence of contaminating amino acids made the results sub-standard.

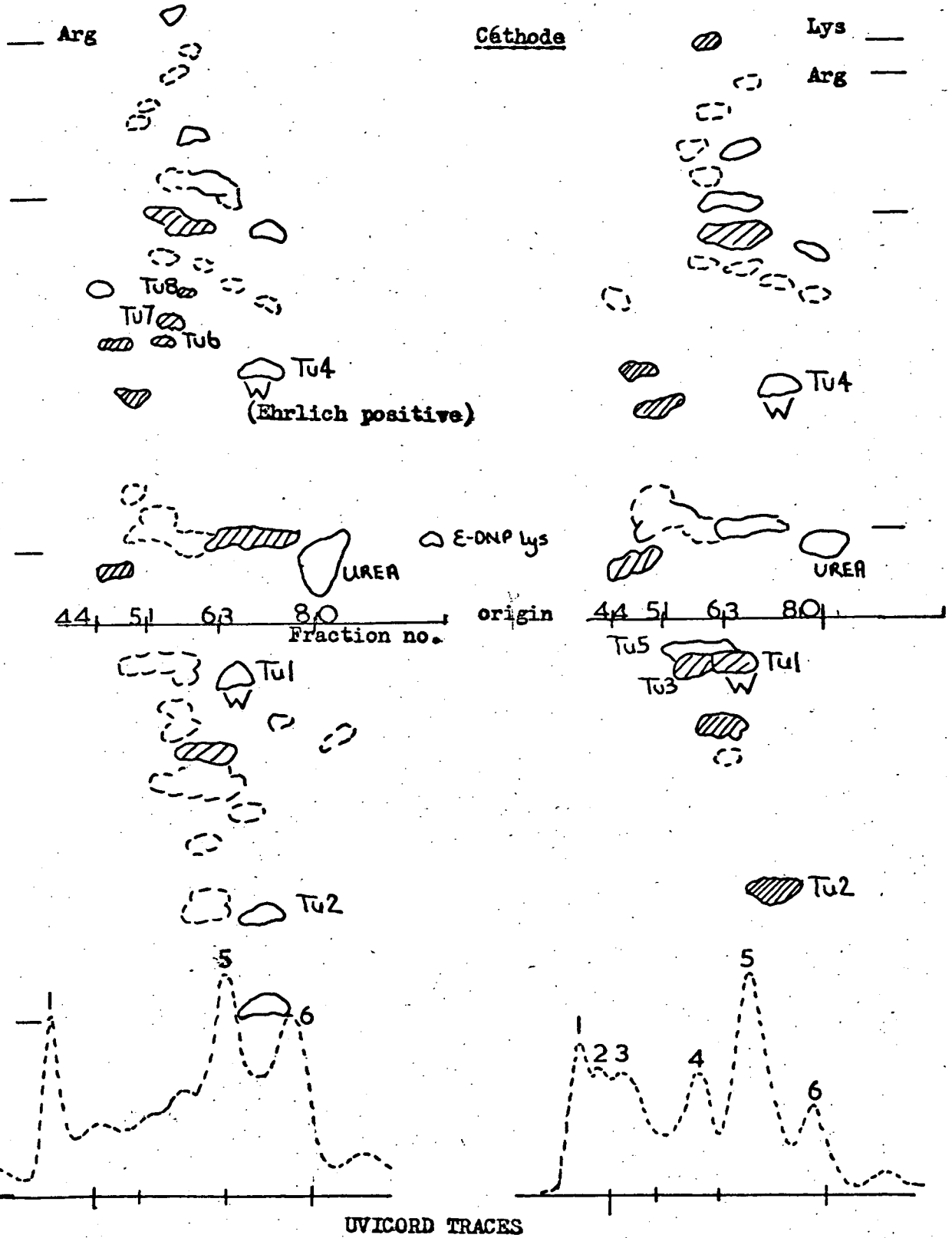
c) A Comparison of the Tryptic Peptide Maps of Wild Type and Ph Vl Amidases

Ph Vl amidase was donated by A. Paterson. Native amidases (15 mg each of A and Ph Vl) were dissolved in freshly deionized 8M urea (0.5 mls) and diluted to pH 8.5, 0.2M ammonium acetate, 2M in urea. Trypsin (0.2 mg) was added to each solution, which were then incubated at 37°C for 2 hr; a further 0.2 mg. trypsin was then added prior to another 2 hr incubation at 37°C before freeze drying. The residues were fractionated on the same 150 cm (3 connected 50 cm columns) x 1.0 cm column of sephadex sG-25 eluted with 5% v/v formic acid. The uvicord traces (given in fig 5B) show peaks 2,3,4 and 6 of greater intensity in Ph Vl. Peak 2 is part of the core material not separated by pH 6.5 electrophoresis. The peptide maps (fig 5B) had more than

FIG. 5B COMPARATIVE TRYPTIC PEPTIDE MAPS

A AMIDASE PHV1 AMIDASE

Degree of cross hatching indicates intensity of the ninhydrin colour. Dashed lines indicate faint spots.



the 3 differences expected. The most notable difference is the presence of many low yielding acidic peptides on the A amidase map, yet equal loadings of each digest were used. The most acidic spot seen on the A amidase map was probably run off the pH V1 map even though both maps were run on the same sheet of paper.

The Peptides For the peptides characterised bands were located directly with the fluorescamine reagent without the use of marker strips. Compositions are listed in table 5B and the digest summarized in table 5D as urea digest.

Tu 1 An Ehrlich positive peptide from both digests but of higher yield (2x) from the Ph V1 digest. It is assumed that only one mole of tryptophan is present per mole of peptide. This was not adequately purified in prior digests.

(Asp,Thr_2,Glu,Ala,Val,Trp)Lys

Tu 2 This is present on both maps but was only isolated from the Ph V1 digest.

No amides can be present and this corresponds to peptide T3.

(Glu,Ala)

Tu 3 This peptide is only evident on the Ph V1 map but compositional and mobility similarities to peptide T25 suggests the apparent absence from the wild type map is due to poor yield. As this is an acidic peptide no amides can be present.

(Glu_2,Pro,Gly,Leu_2,Tyr)Lys

Tu 4 An Ehrlich positive peptide not recovered in sufficient yield for analysis.

Tu 5 This was isolated from the Ph Vl digest but part of the amino acid analysis was lost owing to a machine fault. This peptide probably represents T26 with low recovery of serine (1.4 = 2). The apparent difference of 1 mole per mole of glycine may not be significant.

$(Tyr, Leu, Gly_3, Ala, Glx, Ser_2, \ldots)$

Tu 6/Tu 7/Tu 8 These peptides are apparent only on the wild type peptide map but were not recovered in sufficient amounts for accurate analysis. The most probable composition of Tu 6 is

$(Leu, His, Thr, Ala_2, Gly)$

although the expected peptide, in this position, T 13 is of a different composition.

Conclusion: It is suggested that the differences between the peptide maps are due primarily to differing yields. This is in agreement with the view (Harris and Hindley 1965) that "... mere visual inspection of peptide maps developed with the ninhydrin reagent can lead to highly misleading results".

TABLE 5A. TRYPTIC PEPTIDES 1ST DIGEST.

Peptide	m 6.5	V/Vo G-25	Purn	Yield %	Gly	Ala	Val	Leu	Ile	Ser	Thr	Asp +Asn	Glu +Gln	Phe	Tyr	Trp	Cys	Met	Pro	Lys	His	Arg	TOT	N-t
T1	+0·98		Gb3	1·19								1·0											1	Asp
T3	+0·72		Gb3	1·02	0·24+	0·97				0·17			1·02										2	Glu
T4	+0·47		Gb3	2·61	0·91	1·01		0·75+				0·23+	1·62+	1·94	1·14		0·53+		1·3+			1·0	11	Gly
T6	0		GbB3	1·68						1·0													1	Ser
T7	0		GbB3	0·87	1·0					0·24+	0·16	0·11	0·15										1	Gly
T8	0		GbB3	1·16	0·22+	1·0																	1	Ala
T9	0		GbB3	0·34			1·0																1	Val
T10	0		GbB3	1·09				1·0															1	Leu
T11	0		GbB3	4·7	1·21+	0·87	1·92			0·17		0·99		0·12		(+)							6	Ala
T12	0		GbB3	1·09	1·09	0·14	0·99	0·2+	0·13	0·22+		0·12	0·14	0·92	0·13			0·2+					3	Gly
T13	-0·35		Gb3	1·0	1·2					0·84				0·96	0·77						1·0		5	Ser
T14	-0·48		Gb3	0·87								0·86			0·6					1·14			3	Asn
T15	-0·48		Gb3	0·5	0·25+	0·93				0·18	0·1	0·39+	0·17		1·19				0·85	2·23+		0·18	5	Lys
T17	-0·61		Gb3	10·9											0·87							1·13	2	Tyr

TABLE 5A CTD. TRYPTIC PEPTIDES 1ST DIGEST.

Peptide	m 6.5	V/Vo G-25	Purn	Yield %	Gly	Ala	Val	Leu	Ile	Ser	Thr	Asp +Asn	Glu +Gln	Phe	Tyr	Trp	Cys	Met	Pro	Lys	His	Arg	TOT	N-t
T18	-0·9		G63	1·01																		1·0	1	basic residue
T21	+0·47		G63	9·2	0·23	0·99	0·17	0·1	0·14	0·11		1·4	0·35		0·12		0·8	1·02	0·15	1·09		0·15	5	Ala
T22	+0·47		G63	2·77	0·19	0·18	0·79			0·12		0·85	1·14				0·17	0·76		(+)			5	Asp
T23	+0·4		G63	4·84	0·13		1·06	0·11		0·1		1·81	1·08									1·05	5	Asx
T24	+0·4		G63	0·86	2·95	1·46	0·88	0·92	2·51	1·08	0·34	4·15	2·1	0·29	1·47	(+)	0·93	0·13	1·98			0·84	24	Ile I-x
T25	+0·29		G63	2·02	1·38			1·83					2·21		1·1				1·21	0·82			8	leu
T26	+0·29		G63	4·14	3·89	1·07	0·1	0·9		1·6	0·12	2·08	1·0		0·8							1·03	13	Gly
T26C1	+0·36		G6	—	1·06	2·05				3·12		2·12	1·13									0·6	"10"	Ala
T27	+0·13		G63	8·94	0·97	1·04							1·01		1·6		0·83	0·91	1·24	0·98			9	fail
T27C1	-0·35 -0·4		G6	—		0·96									0·95			0·92	1·12	1·05			5	Met
T27C2	+0·54		G6	—	1·18								0·95		0·9		0·96						4	Cys
T28	0		G63	6·58		0·89											0·99					1·12	3	Ala
T29	0		G63	10·7		0·97						0·99										1·03	3	Asp
T34	0		G63	5·86	0·97			0·99		1·12	0·83	0·28	1·1									1·0	6	Ser

TABLE 5A CTD. TRYPTIC PEPTIDES 1ST DIGEST.

Peptide	m 6.5	V/Vo G-25	Purn	Yield %	Gly	Ala	Val	Leu	Ile	Ser	Thr	Asp +Asn	Glu +Gln	Phe	Tyr	Trp	Cys	Met	Pro	Lys	His	Arg	TOT	N-t	
T35	0		G63	11·8	1·27 +	0·98	0·92	0·91	0·92	0·13		0·14	0·84	0·21 +								1·15	7	Gly	96 hr value
T36	-0·49		G63	4·35	0·21 +	1·03		0·22 +	0·17	0·18		0·1	0·22	0·11	0·93		0·13	0·26 +	1·21 +	1·07		0·7 +	5	Lys	
T37	-0·49		G632	3·48														0·61 +	1·08			0·92	3	Met	
T38	-0·55		G63	3·65	0·15	0·16		0·96		0·12	1·11							0·22 +		0·16		1·91	4	Leu	
T39	-0·65		G63	18·64	1·04					0·12								0·99		0·74		0·28 +	3	Gly	
T40	-0·75		G632	5·28		0·96																1·04	2	Ala	
T41	-0·86		G632	0·89											0·85					(+)		1·15	3	Tyr	
T42	-1·0		G63	23·4																1·0			1	Lys	
T43			G63PB3	0·71	1·08	0·74 +	2·48 +		1·1	0·93	0·7 +		1·7 +	0·99	0·91				0·92	0·48 +		1·02	15	Val V-V	96hr value
T44	+0·87		G63	2·98	3·85	0·15		1·08	1·05	0·25 +	0·94		5·05				1·02	0·9					14	Thr	96 hr value
T45	+0·81		G63	2·48	3·05			1·14	1·1	0·19	0·82	0·23 +	4·3		0·87		0·78 +	0·97					13	Thr	96 hr value
T46	+0·71		G63	0·65	0·67 +	1·8	0·12	0·22 +	0·23 +		0·99	1·26 +	2·2		0·63			1·4 +	1·0				10-11	Tyr	
T47	+0·37		G63	4·72	1·04	1·0	0·8		1·6 +	0·87	0·97	0·2	2·5 +	1·0					1·23 +			1·0	13	Val	96hr value
T50	0		G683	10·1		0·93	0·92					1·05	2·17					1·84		1·08			8	Asp	

TABLE 5A CTD. TRYPTIC PEPTIDES 1ST DIGEST.

Peptide	m 6.5	V/Vo G-25	Purn	Yield %	Gly	Ala	Val	Leu	Ile	Ser	Thr	Asp +Asn	Glu +Gln	Phe	Tyr	Trp	Cys	Met	Pro	Lys	His	Arg	TOT	N-t	
T51	0		G6B3	2·24	1·18		0·8			0·73+		0·11	0·97						1·14	1·19			6	Val	96 hr values
T52	0		G63	3·42	1·05	0·93	0·93		1·42+			1·08	0·1					1·7+		1·3+			9	Ile I-A	96 hr values
T54	-0·06		G63	2·29									2·2						1·1		1·91	0·82	6	His	
T55	-0·21		G63	5·09	0·13	0·9		0·91		0·89		3·02	2·13	0·83						1·2	1·1		11	Ala	
T56	-0·31		G63	1·8	0·2+	0·98		2·07	0·94	1·99			2·0									1·01	9	Ala	
T57	-0·25		G63	3·48	4·09	1·02	1·93	0·47+	2·15	1·05	1·01	1·86	2·98	0·23+	1·46	(t)	0·91	0·16	3·99	1·16		0·47+	27	Ile I-I	96 hr values
T58	0		G6B3	1·6	0·22+	1·19	1·07	0·1				0·96	1·85					1·74+		3·1			10	Asp	
T59	0		G6B3	2·36	0·83	1·12	0·76+	0·24+	0·98		0·16	1·1	0·2+					1·05		0·92	0·16	0·17	7	Ile I-A	96 hr values
T60	0		G6B3	2·61		3·2	1·04	1·77			1·02	1·8	1·2								1·0	0·9	12	leu	
T61	-0·35		G63	3·48	1·22+	1·08	1·07		1·0			1·08						1·9		1·92			9	lys	96 hr value except lys

TABLE 5B TRYPTIC PEPTIDES 2nd DIGEST AND UREA DIGEST.

Peptide	m 6.5	V/Vo G-25	Purn	Yield%	Gly	Ala	Val	Leu	Ile	Ser	Thr	Asp +Asn	Glu +Gln	Phe	Tyr	Trp	Cys	Met	Pro	Lys	His	Arg	TOT	N-t	
TT2	-0·78		G6	6·8	0·18	1·07	0·79					+ 1·77	+ 1·7									1·14	7	Glu	SEE TEXT
TT4	-0·74		G63	11·93				0·98	0·99												0·9	1·12	4	LEU OR ILE	96 hr value
TT4Y1	—		-	-				0·83	0·15												0·97	1·19	3	-	
TT4Y2	-		-	-	0·14			+ 0·24	0·12												0·89	1·1	2	-	
TT4Y3	-		-	-	+ 0·24			0·19		+ 0·26		0·14	0·13								+ 0·21	1·0	1	-	
TT5	-0·53		G63	0·26	1·18	+ 3·22	0·95	0·94	+ 0·42	+ 0·68	+ 0·79	1·12	0·8		1·5		0·96	+ 0·25	+ 1·62	2·08	+ 0·28	+ 2·25	20	Lys	
TT6	-1·0		G6	20·6																1·0			1	Lys	
TTA	—		SEE TEXT	2·93	7·31	(+) 6·47	+ 3·46	3·88	4·19	+ 3·53	+ 3·36	(+) 5·45	9·08	0·92	1·65		1·03	2·86	3·07	2·09		1·03		Ala	24 hr value
A. AMIOASE TU1	—		G63	7·83		0·96	1·06			+ 0·23	1·88	1·15	1·01			(+)				0·8	0·15		8	-	
PhVI Ami. TU1	—		G63	16·4		0·95	1·13				1·92	1·09	1·04			(+)				0·84			8	-	
PhVI TU2	—		G63	9·4		1·0							1·0										2	-	
PhVI TU3	—		G63	12·6	1·2			1·83					1·95		0·8				1·02	+ 0·77			8	-	
PhVI TU5	—		G63	16·8	3·14	1·06		0·82		+ 1·38		0·94	0·98		0·61					←	LOST	→			
A. Ami TU6	—		G63	2·08	+ 0·75	+ 2·29		+ 0·77		+ 0·35	1·15	+ 0·29	+ 0·34								1·04				

TABLE 5C

Summary of Tryptic peptides - 1st digest

This list excludes free amino acids given in the chapter

T3 Glu-Ala

T4 Gly(Cys,Glx_2,Ala,Leu,Tyr,Phe_2Pro)Arg

T11 Ala-Trp-Val(Gly,Val,Asn)

T12 Gly(Val,Phe)

T13 Ser(Gly,His,Tyr,Phe)

T14 Asn-Tyr-Lys

T15 Lys-(Ala,Pro,Tyr)Lys

T17 Tyr-Arg

T21 Ala-(Met,Asp,Cys)Lys

T22 Asp-(Glu,Met,Val)Lys

T23 Asx-Glx-Val-Asx-Arg

T24 Ile-(Ile_2,Cys,Asx_4,Ser,Glx_2,Pro_2,Gly_3,Ala_2,Val,Leu,Tyr_2,Trp)Arg

T25 Leu(Glu_2,Pro,Gly,Leu,Tyr)Lys

T26 Gly(Gly_2,Leu,Tyr)Ala(Asx_2,Ser_2,Glx,Gly)Arg

T27 Cys-Gln-Gly-Tyr-Met-Tyr-Pro-Ala-Lys

T28 Ala-Cys-Arg

T29 Asp-Ala-Arg

T34 Ser-Leu-Thr-Gly-Glu-Arg

T35 Gly-Ala-Glu-Leu-Ile-Val-Arg

T36 Lys-Ala-Pro-Tyr-Arg

T37 Met-Pro-Arg

T38 Leu-Thr-Arg-Arg

T39 Gly-Met-Lys

T40 Ala-Arg

TABLE 5C - Contd.

T41 Tyr-(Arg,Lys)

T43 Val-(Val_2,Thr,Ser,Glx_2,Pro,Gly,Ala,Ile,Tyr,Phe,Arg,Lys)

T44 Thr-(Cys,Met,Glx_5,Gly_4,Ile,Leu)

T45 Thr-Leu-Gly-(Cys,Met,Glx_4,Gly_2,Ile,Tyr)

T46 Tyr-(Asx,Thr,Gly,Glx_2,Pro,Ala_2,Met_{1-2})

T47 Val-Ala-Ile-Pro-Gly-(Thr,Ser,Glu_3,Ile,Phe)Arg

T50 Asp-Gln-Gln-Val-Met-Met-Ala-Lys

T51 Val-Ser-Glu-Gly-Pro-Lys

T52 Ile-Ala-Asp-Met-Ile-Val(Met,Gly)Lys

T54 His-Glx-Glx-His-Pro-Arg

T55 Ala-Asx-Asx-His-Glx-Asx(Leu,Ser,Glx,Phe)Lys

T56 Ala-Gln-Ile(Ser_2,Gln,Leu_2)Arg

T57 Ile-(Ile)(Cys,Asx_2,Thr,Ser,Glx_3,Pro_4,Ala,Gly_4,Val_2,Tyr_2, Leu,Arg,Lys,Trp)

T58 Asp-(Glu_2,Val,Met_2,Ala,Lys_3)

T59 Ile-(Ala)(Asp,Met,Gly,Val)Lys

T60 Leu-His-Thr-Ala-Ala-Glx-Val-(Leu)(Ala,Asx_2)Arg

T61 Lys-Ile-Ala-Asp(Met_2,Gly,Val)Lys

TABLE 5D

Summary of Tryptic Peptides; 2nd digest and Urea digest

Second tryptic digest

TT2 Glu(Asp_2,Glu,Ala,Val)Arg

TT4 Ile-Leu-His-Arg

TT5 Lys-(Cys,Asx,Thr,Glx,Ser,Pro_2,Gly,Ala_3,Val,Ile,Leu, Arg_2, Tyr_2, Lys)

TTA Ala-Ile-Leu-Gln-(...)

Tryptic digest in the presence of urea

PhVl peptides

Tu1 (Asp,Thr_2,Glu,Ala,Val,Trp)Lys

Tu2 (Glu,Ala) (= T3)

Tu3 (Gly,Glu_2,Pro,Leu_2,Tyr)Lys (= T25)

Tu5 (Tyr,Leu,Gly_3,Ala,Glx,Ser_2,...) (= T26)

"A" amidase Peptides

Tu1 (Asp,Thr_2,Glu,Ala,Val,Trp)Lys

Tu6 (Leu,His,Thr,Ala_2,Gly) see text.

CHAPTER 6

CYANOGEN BROMIDE CLEAVAGE AND THE STAPHYLOCOCCUS AUREUS PROTEASE DIGEST

a) Cyanogen Bromide Cleavage

i) Method

Preliminary experiments suggested cyanogen bromide cleavage would produce fragments of differing molecular weight amenable to purification by gel filtration. Amidase (300 mg) was treated with a 40 fold excess of cyanogen bromide as described in chap. 3. On dilution prior to freeze drying a heavy precipitate formed. The freeze dried residue was taken up in formic acid (3 mls), samples were removed for amino acid analysis and gel electrophoresis, before diluting to 5% v/v formic acid and adjusting to pH 4 with ammonia. Only a very slight precipitate developed on standing and this was spun down. Larger fragments were then precipitated by the addition of a T.C.A. solution, to a final concentration of 5% (w/v), and were removed by centrifugation.

Amino acid analysis revealed that 75% of the methionine residues had been converted to homoserine/homoserine lactone. The sample taken for gel electrophoresis appeared not to dissolve in the S.D.S.-gel electrophoresis sample buffer (as in chap 2) and no bands could be detected on the gels. (This gel was prepared and run by Dr. D.R. Thatcher).

2) Separation of the T.C.A. Precipitated Material.

An attempt to purify the precipitated fragments by gel filtration was not successful. Elution from sephadex sG-75 suggested that the bulk of the material was excluded from the gel. Oxidation, to reduce cysteine interactions, decreased the solubility of the fragments and gel filtration on sephadex sG-50 gave little separation. Small elution volumes again suggested an anomolously high molecular weight.

3) Purification of the Soluble Fragments.

Initial separation of this material was by gel filtration on sephadex sG-25 and by pH 6.5 electrophoresis. Two forms of each peptide are expected corresponding to C-terminal homoserine or homoserine lactone. Prior to preparative pH 6.5 electrophoresis, pooled fractions were treated with pH 6.5 electrophoresis buffer *in vacuo* for 1 hr, heated at 105°C, to convert homoserine lactone to homoserine. This method is routinely used prior to quantitative amino acid analysis (Ambler and Brown 1967), without apparent significant modification of other amino acids. Before subsequent electrophoreses the gentler method of overnight incubation at 37°C in 0.1 M ammonia (after Ambler and Brown 1967) was used to open the homoserine lactone ring.

4. The Peptides.

Residues listed as methionine were identified as homoserine or homoserone lactone. A list of compositions is given in table 6A and table 6B summarizes the peptides which were characterised.

X1 This was identified by dansylation and qualitative analysis as homoserine

Met

X2 The acidic mobility of this peptide indicates no amides are present. Edman degradation was unsatisfactory as the peptide became ragged after only one cycle.

Tyr-(Pro,Asp,Ala,Glu)Met

X3 A low recovery of isoleucine (0.7=1) was recorded even after 96 hr. hydrolysis.

The N-terminus was identified as isoleucyl valine

Ile-Val-Gly-Met

X4 Ala-Lys-Met

X5 The mobility indicates 1 amide is present

Asx-Leu-Val-Val-Phe-Pro-Glx-Tyr-(Ser,Glx,Gly,Leu,Ile)Met

X6 The mobility of X6 suggests 2 amides are present. No residue was identified in position 2 by the dansyl Edman method, but derivatives of residues 3 to 5 were identified clearly. Lysine has been placed in position 2 in order to explain this

Ala-Lys-Asx-Glx-Glx-(Gly,Val)Met

X7 The mobility indicates that 1 amide is present

Lys-(Gln,Gly_2,Pro,Leu)Met

b) Staphylococcus aureus Protease Digest

1) Method

250 mg. of oxidized amidase were digested as described in chap 3. After freeze drying the residue was redissolved in formic acid (final conc. 6% v/v) and core material precipitated with T.C.A. (5% w/v final concentration). The heavy precipitate was removed by centrifugation and a small portion examined by S.D.S.-polyacrylamide slab gel electrophoresis (performed by Dr. D.R. Thatcher). No undigested material remained and the bulk of these peptides were in the approximate molecular weight range 2,000 to 3,000.

2) Attempted Purification of Core Material

After freeze drying the core peptides were separated on a 2.5 x 110 cm column of sephadex sG-25 equilibrated and run in 40% (v/v) formic acid. Whilst loading the column these peptides, dissolved in 50% formic acid, appeared to form a gel which reduced but did not stop the flow. As the peptides became diluted on the column the flow rate quickly recovered. From the gel electrophoresis result it was hoped some peptides would be suitable for purification by paper electrophoresis. A pH 6.5 map of the column effluent, however, showed that none of the core peptides migrated under the conditions used.

3) Purification of the Soluble Peptides

The soluble peptides were fractionated on a 2.5 x 110cm. column of sephadex sG-25 eluted with 1M acetic acid. The two smallest molecular weight range poolings were then extracted with ether to remove T.C.A. One such pooling appeared devoid of peptides from the map but preparative pH 6.5 electrophoresis at a higher loading revealed the presence of many low yield and poorly separated peptides.

4) The Peptides.

This digest was only partially characterized. Peptide compositions are given in table 6A and the digest results are summarised in table 6B.

F1 The recorded mobility is not in good agreement with the composition, suggesting an excess charge of 2.5. Thus the presence of 1 amide cannot be excluded.

Thr-Ala-Val-Ala-Ile-Pro-Gly-Glx-(Glx,Thr)Glu.

F2 The mobility of this peptide suggests a charge excess of 2.7(≈3) and therefore 1 amide is probably present.

Tyr-(Asx,Met_2,Ser,Glx_2,Pro,Gly,Ala,Ile,Leu,Tyr)Glu

F3 After 3 cycles of Edman degradation no new N-terminus was identified, although subsequent residues were identified. Lysine has been placed in position 4, from the composition. The sequence suggests this is the C-terminal peptide, and the mobility indicates that no amides are present. It is noted that two glutamyl

bonds are present, which have not been hydrolysed by the protease

Gly-Leu-Glu-Lys-Glu-Ala
—→ —→ —→ (—X) —→ —→

F4 As this is a neutral peptide 1 amide must be present. The mobility of the residual peptide after Edman degradations (3 cycles residual m = + 0.74, 4 cycles residual m = + 0.55) indicates charges of 2.0 and 1.2 respectively. These values do not fit the composition very well. Good values are not obtained if one assumes that an extra acidic/amide residue is present and deamidation has occurred. The mobilities suggested that arginine has been modified but this would not be expected.

Lys-Ala-Asx-Asx-(Arg,Val)Glu
—→ —→ —→ —→

TABLE 6A

CYANOGEN BROMIDE AND S. AUREUS PROTEASE PEPTIDES.

Peptide	m 6.5	V/Vo G-25	Purn	Yield%	Gly	Ala	Val	Leu	Ile	Ser	Thr	Asp +Asn	Glu +Gln	Phe	Tyr	Trp	Cys	Met	Pro	Lys	His	Arg	TOT	N-t	
X1	0		G63	—														(t)					1	Met	
X2	+0.64		G63	1.6		0.97						1.17	0.9		0.87			0.87	0.96				6	Tyr	
X3	0		G63	4.1	1.08	0.25+	0.92		0.7+	0.11		0.17	0.26+					1.01					4	Ile I-V	96 hr value
X4	-0.61		G63	1.74		1.13												0.8		0.87			3	Ala	
X5	+0.39		G63	2.8	1.07		1.95	2.0	0.97	0.82		1.03	2.05	0.88	0.68			0.89	1.09				14	Asx	96 hr value
X6	0		G63	1.3	0.74+	1.05	1.26+					1.22	1.84					1.1		24hr+ 1.22			8	Ala	96 hr value
X7	-0.42		G63	3.1	2.24+			1.01					0.98					0.77	0.99	0.77+			7	Lys	
F1	+0.62		G63	11.0	1.04	2.01	1.1		1.0		1.76+		3.11						1.02				11	Thr	96 hr value
F2	+0.5		G63	10.9	1.06	1.12		0.99	1.0	24 hr. 0.84		1.07	3.06		24 hr. 1.82		0.15	2.07	0.9				14	Tyr	96 hr value
F3	+0.33		G63	2.0	0.9	0.95		0.98					2.09							1.08	0.2+		6	Gly	
F4	0		G6B3	1.97		0.98	0.6+					1.87	0.84							1.03		1.28+	7	Lys	

TABLE 6B

Summary of Cyanogen Bromide and S.aureus Protease Peptides

Soluble Cyanogen Bromide fragments

X1 Met

X2 Tyr(Pro,Asp,Ala,Glu)Met

X3 Ile-Val-Gly-Met

X4 Ala-Lys-Met

X5 Asx-Leu-Val-Val-Phe-Pro-Glx-Tyr-(Ser,Glx,Gly,Leu,Ile),Met

X6 Ala-Lys-Asx-Glx-Glx-(Gly,Val),Met

X7 Lys(Gln,Gly_2,Pro,Leu),Met

S.aureus protease digest, soluble peptides

F1 Thr-Ala-Val-Ala-Ile-Pro-Gly-Glx-(Glx,Thr),Glu

F2 Tyr-(Asx,Met_2,Ser,Glx_2,Pro,Gly,Ala,Ile,Leu,Tyr),Glu

F3 Gly-Leu-Glu-Lys-Glu-Ala

F4 Lys-Ala-Asx-Asx-(Arg,Val),Glu

CHAPTER 7

SEQUENCE DEDUCTION AND DISCUSSION.

a) The Purity of the Protein

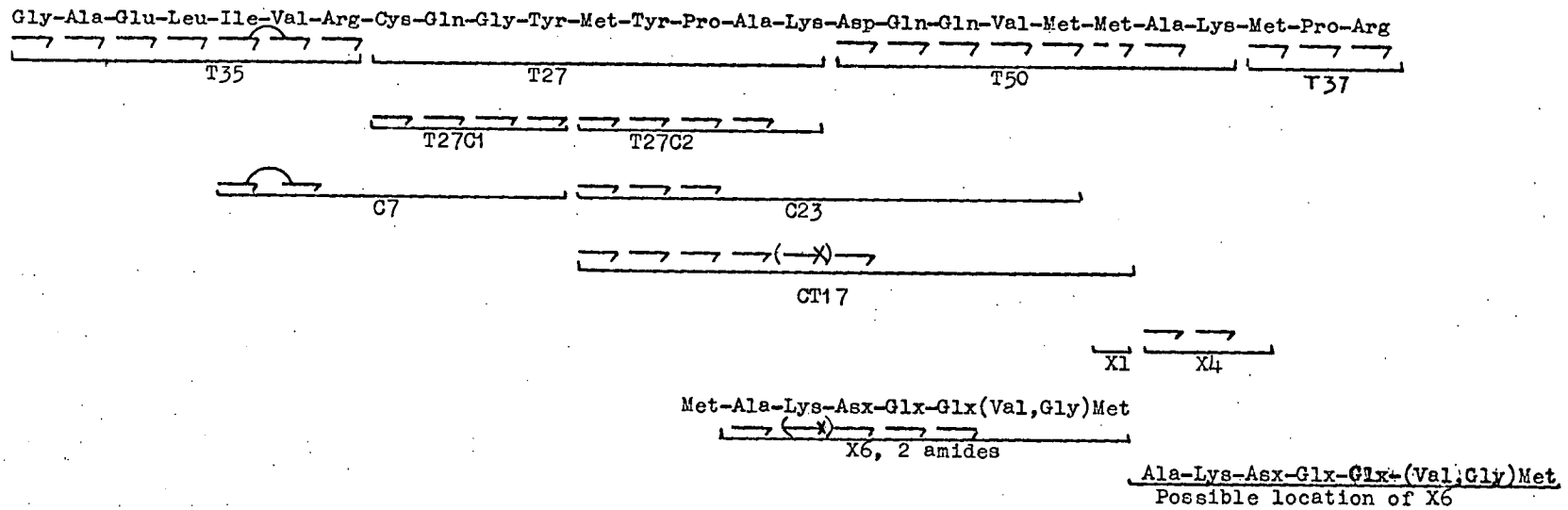
The validity of some of the results presented in the preceding chapters may be questioned on the grounds that the amidase used was not sufficiently pure. Batch 1 protein has subsequently been found to contain 4 components, although it was stored for 2 1/2 years before that experiment was carried out. Experiments carried out at the time of use did not indicate heterogeneity and it is possible that proteolysis may have occurred during storage. Similarly batch 2 protein had a very low specific activity, yet the available evidence suggested that this was sufficiently pure amidase to be used in this study. Those parts of the sequence which do overlap tend to support these views.

b) Sequence Deduction

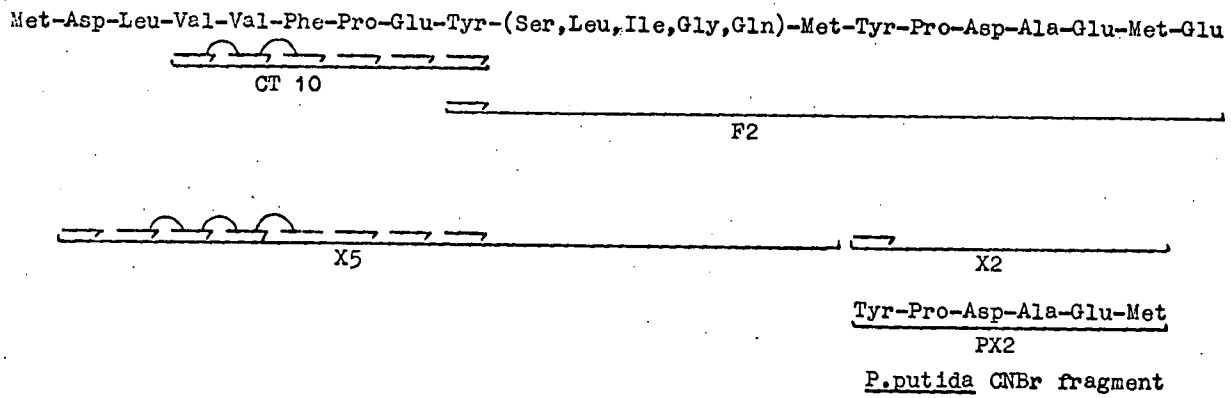
From the peptide structures presented in the preceding chapters the partial sequences shown in fig 7A have been compiled. (N.B. table 7A does not include all peptides for which only limited sequence information is known). These regions are discussed below. It has not been possible to determine the total primary structure of amidase from the available data. Some sequence information on a P.putida amidase sequence has been included in table 7A, and this

FIG 7A REGIONS OF AMIDASE PARTIAL STRUCTURE

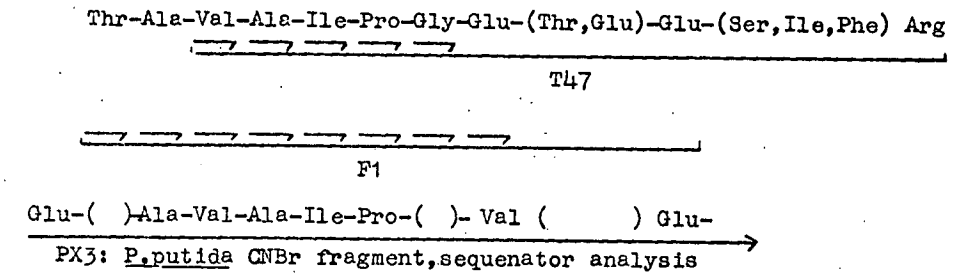
Region 1



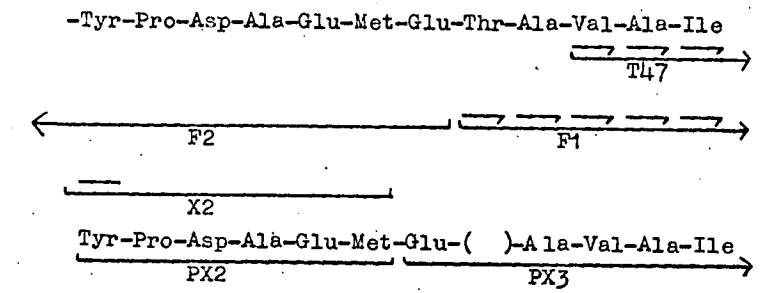
Region 2A



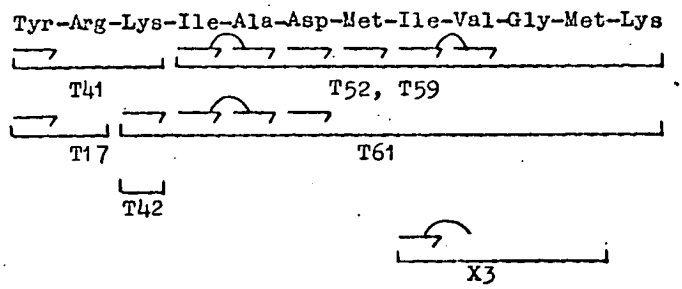
Region 2B



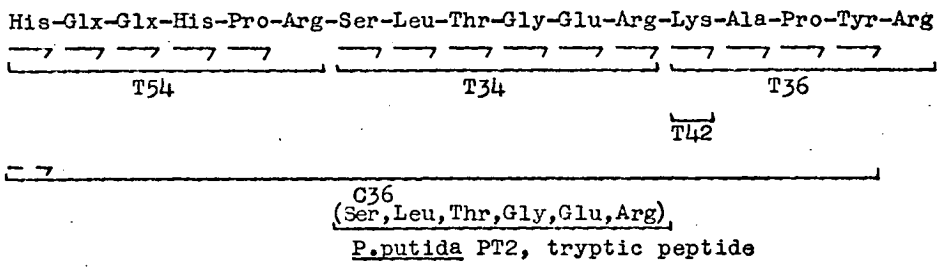
Overlapping of 2A and 2B



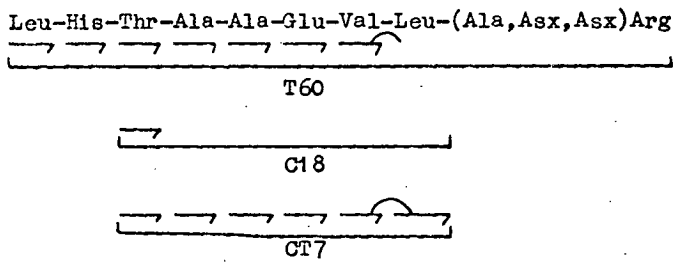
Region 3



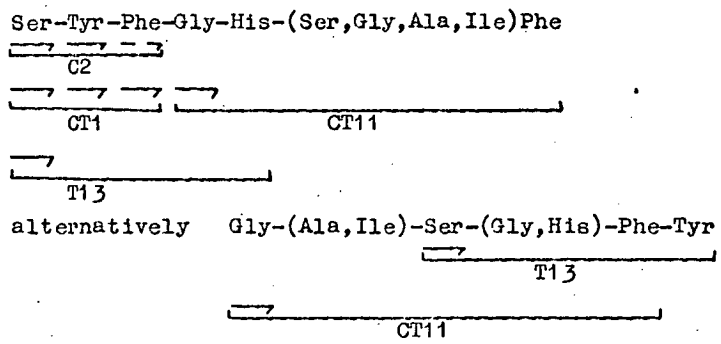
Region 4



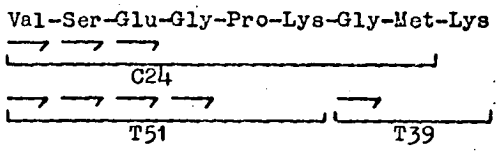
Region 5



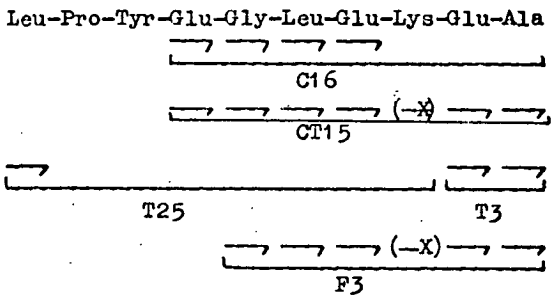
Region 6



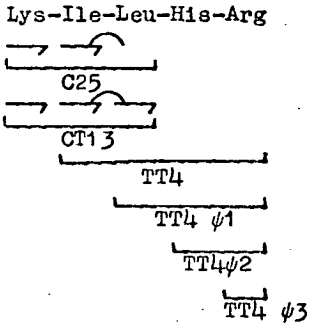
Region 7



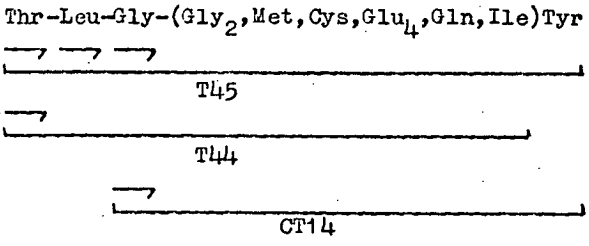
Region 8



Region 9



Region 10



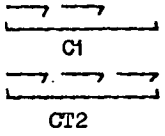
see text — mobility gives charge = 4

mobility gives charge = 5

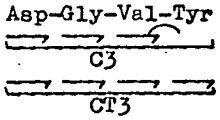
TABLE 7A CTD.

Sequences not overlapped

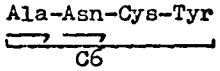
1) Glu-Phe-Tyr



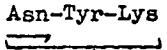
2) Asp-Gly-Val-Tyr



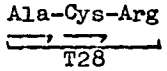
3) Ala-Asn-Cys-Tyr



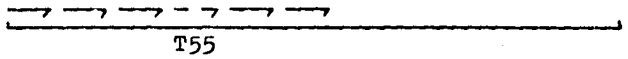
4) Asn-Tyr-Lys



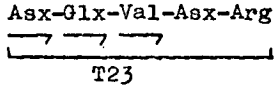
5) Ala-Cys-Arg



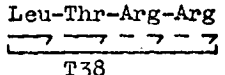
6) Ala-Asx-Asx-His-Glx-Asx-(Leu,Ser,Glx,Phe)Lys



7) Asx-Glx-Val-Asx-Arg



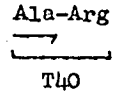
8) Leu-Thr-Arg-Arg



(Leu, Thr, His, Arg)

P. putida PT2
(tryptic peptide)

9) Ala-Arg



10) Met-Lys-Gln-Gly-Leu-Pro-Gly-Met

X7

Lys-Gln-Gly-Leu-Pro-Gly-Met

PX4, P.putida cyanogen bromide fragment.

11) Lys-Ala-Asx-Asx-(Arg,Val)Glu

F4

12) Asp-Ala-Arg

T29

13) Gly-Val-Phe

CT19

T12

may not represent the same sequence

14) Met-Arg-His-Gly-Asp-Ile — sequenator analysis of amidase

15) Ala-Ile-Leu-Gln — TTA, sequenator analysis.

has been made available by Dr. R.P. Ambler from his investigations of the structure of the acetamidase from P.putida strain A87C1 (Clarke 1972). This data has been included for comparison only.

Region 1.

Tryptic peptides covering the whole of this 27 residue sequence have been found. The C terminal 3 residues are only weakly overlapped, but peptides T37 is the only tryptic peptide isolated with the predicted N-terminal methionine. Peptide T27 was aligned from its subdigest products on the basis of enzymic specificity and this is confirmed by the overlapping shown. This region is rich in methionine but X6 is not the predicted cyanogen bromide (CNBr) fragment. The failure to identify residue 2 of X6, strictly renders the sequencing results unsatisfactory, but the proposed sequence is very similar to that of T50. The assignment of amides in X6 is unknown but 2 are present. It is possible that X6 is located at the C-terminus of region 1 as shown.

On the basis of composition the P.putida enzyme appears to have the N-terminal tryptic peptide of this region, and the CNBr fragment PX1 suggests a leucine for methionine substitution (a single minimum base change in the genes).

Region 2B

The sequences of T47 and F1 shows homology, but T47

is apparently derived from an alanyl-valine bond split by trypsin. The common contaminant activities of trypsin (Smith 1967) would lead to chymotryptic type splits. Even so an alanyl-valine cleavage would not be expected and this suggests either heterogeneity in F1 (it would require a minimum of 2 base changes for the substitution of lysine or arginine for alanine), or the location of T47 elsewhere in the sequence.

The N-terminal sequence of the P.putida amidase CNBr fragment PX3 aligns with region 2B, and suggests a valine for glutamic acid substitution (a single minimum base change), although non identification of some residues renders the P.putida sequence unsatisfactory.

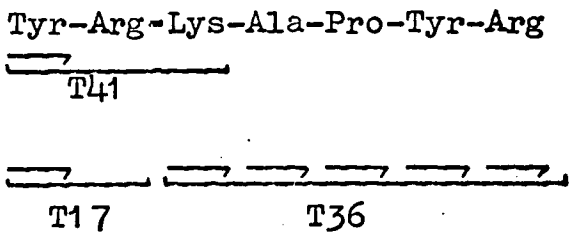
Overlapping of Regions 2A and 2B

Region 2A may immediately precede 2B in the total sequence but the evidence for the overlap is weak. It can be postulated on the basis of homology of X2 and PX2, and on the specificity of the S.aureus protease. As both the cyanogen bromide and S.aureus protease digests were only partially characterised other alternatives may be available.

Region 3.

Most of this 12 residue stretch has been aligned by overlapping tryptic peptides, with the ordering of the C-terminus by peptide X3. The peptides from the N-terminal

region are consistent with 2 sites of tryptic cleavage at the arginyl lysine sequence. The evidence for the postulated Arg-Lys-Ile sequence is not strong. An alternative overlapping of T41 with peptide T36 (at present located in region 4) is given below.



Region 4

The alignment of peptides T54, T34, and T36, is solely dependent on the composition of peptide C36. Dansylation of C36 indicated a basic residue at the N-terminus but this was not unambiguously identified. It was not possible to purify this peptide (C36) from the second chymotryptic digest as this was contained in one of the poolings which smeared badly on electrophoresis. Both C36 and T54 appear to be derived from cleavage at the same site. The composition of other tryptic peptides suggests that they have arisen by chymotryptic splitting and this may be the case for T54. An alternative location of T36 is given above, and the composition of peptide CT18 suggests another possible location of T34, which would then have arisen by chymotryptic type cleavage. Peptide T34 is the "difference peptide" described by Brown and Clarke (1972) and appears to be located in a region which

is rich in basic residues. On the basis of amino acid composition peptide T15 could replace T.36.
The composition of the neutral tryptic peptide PT2 from P.putida acetamidase is consistent with the T34 sequence being conserved in this protein.

Region 5

Chymotryptic splitting at histidine is evident from this sequence.

Region 6

The composition of T13 and the small number of histidine residues present in amidase suggest this overlap. Peptide T13 appears to have arisen solely by chymotryptic like splitting. An alternative overlapping of T13 and CT11 suggests a less satisfactory origin for T13; the evidence for both alternatives is based mainly on compositions.

Region 7

T51 is another tryptic peptide that may have arisen through a chymotryptic like cleavage.

Region 8

This is postulated as the C-terminal sequence of amidase, in agreement with the results of Brown et al

(1973). Lysine has been placed indirectly as it was not identified in dansyl Edman degradations, but its position is consistent in CT15 and in F3, and predicts peptide T3. Evidence for the N-terminal 3 residues is based on the composition of T25 (present in both A and PhV1 amidases, as is T3) and the specificity of chymotrypsin.

Region 9

This is a short basic region which could proceed from any of several tryptic peptides on the basis of enzymic specificity and peptide compositions (e.g. T21, T22, T25, T52, T55).

Region 10

The compositional similarity of CT14 with T45 suggests this possible overlap. Peptide T44 which is similar to T45 has been included, and slight differences in compositions of the 3 peptides and therefore in amide assignments weakens the overlap. The composition of peptide CT14 has been used as the basis for the structure shown.

Sequences not overlapped

Of these peptides the sequence is weakest for X7 (No. 10 in list) which has been ordered by homology with the corresponding P.putida amidase fragment PX4. The N-terminal methionine is postulated from the cyanogen

bromide cleavage site.

Peptides CT19 and T12 (see no 13 in the list) may be identical although neither have been fully sequenced.

The composition of the P.putida amidase tryptic peptide PT2 would indicate homology with T38 (listed as no 8), with a histidine for arginine substitution (a single minimum base difference).

c) Discussion of the Sequence

The sequences presented above account for just over 200 residues of the amidase sequence. This total does not include some peptides identified by composition only.

The most notable omission is that of the N-terminal region. Not even the tryptic dipeptide Met-Arg, predicted by the automatic sequenator results, has been isolated. No digest, however, was fully characterised as most effort has been concentrated on the soluble peptides. As these may be expected to be more amenable to the flexible techniques of paper electrophoresis and chromatography

their purification should be more easily possible. A comparison of tryptic or chymotryptic peptide maps with and without removal of core material showed that the only obvious difference was the absence of the larger peptides which did not migrate on paper electrophoresis. If these were purified in a comparably low yield to the soluble peptides the amount of sequence information obtainable may not have been great.

Even so full characterisation of the soluble peptides has not been possible. The low yield of all peptides has limited the information available. It was not uncommon to find that after several purification steps, when only sufficient material was available for amino acid analysis, the peptide was still not adequately pure. Despite the 75% conversion of methionine to homoserine and its lactone, even the cyanogen bromide fragments were recovered in low yield.

It is relatively easy to look back and see where time may have been more profitably spent, or indeed to say what might have been possible if more time were available. This investigation has proceeded along the established route of overlapping proteolytic and cyanogen bromide peptides. The first tryptic digestion proved to be the most informative and several peptides seem to have arisen by chymotryptic like splitting. This probably reflects the long digestion time rather than the degree of contamination of the trypsin preparation used. The

chymotryptic digests, especially the neutral peptides, proved to be the most difficult to purify. The repeat serine protease digests, and the S.aureus protease digests which were only partially characterised would have yielded more information had more time and effort been available for them. Similarly information could have been obtained from the core materials.

It was interesting to note that in the comparative mapping experiment the yields of PhV1 peptides were greater than those of the corresponding A amidase peptides. This suggests that the mutant protein is more susceptible to proteolysis, possibly being more easily denatured. It was this possibility that led to the sequenator analysis of PhV1 amidase, which established the N terminal protein sequence more satisfactorily than analysis of the A amidase.

Many of the peptides listed do not conform to the proposed criteria of purity (Ambler and Wynn, 1973). Recoveries of serine and theorine uncorrected for losses, low recovery of valine and isoleucine after 24 hr. hydrolysis (when insufficient material was available for 96 hr. hydrolysis), and contamination with other peptides or the common "paper impurities" (e.g. alanine, glycine, serine) are all contributing factors to the unsatisfactory analyses. Those peptides which are presented in this thesis do not represent the total amount of peptide

material purified. Many other peptides were "purified" but have been rejected on the basis of having unsatisfactory amino acid analysis, and often insufficient material remained for further purification.

There is some evidence from the sequence that a gene duplication, with subsequent modification, may have occured during the evolution of P.aeruginosa amidase. Fragment X6 shows homology with region 1, but must have arisen from cleavage of a methionyl-alanine bond, and there is also a plausible substitution of glycine for methionine in X6. This substitution probably occurs in the same position as the leucine substitution in P.putida amidase seen in PXI. Although the leucine for methionine substitution may be due to a single base change, a glycine substitution would require a minimum 2 base changes.

Peptide T47 has the same partial sequence as peptide FI, yet overlapping suggests it has arisen by tryptic cleavage of an alanyl-valine bond. An alternative explanation is that T47 arises from a region similar to FI but located elsewhere in the sequence.

The composition of peptide T15 is similar to that of T36 (whose sequence is known) except that it contains 2 lysines whereas T36 has one lysine and one arginine. Two other tryptic peptides, T50 and T58, have a compositional similarity although only the sequence of T50 is known. The multiple lysines of T58 could be

explained on the basis of incomplete cleavage of lysyl lysine sequences, but the mobility data requires that deamidation must also have occurred if it is derived from the same region as T50.

Peptides T44 and T45 may represent the same part of the sequence but have arisen by incomplete splitting, with differences in amino acid contents explained by insensitivity of analysis. Similarly C15 and C31 may represent the same sequence but with poor analysis suggesting different peptides. The relationships between T61, T59, and T52, can also be explained by incomplete bond cleavage and artefactual differences in the analyses. Nevertheless one cannot rule out the possibility of similar regions giving rise to these sets of peptides.

Assuming a duplication has occurred these results do not differentiate between an intra chain repeat and heterogeneity of subunits. Biochemically it has not been possible to differentiate amidase subunits (Brown et al 1973) and the enzyme is thought to be coded for by a single amidase structural gene (Betz et al 1974). The presence of different subunits has not been rigorously excluded.

Other explanations for these observations could be that the protein has been purified from a mixed population of bacteria or that different protein batches were not prepared from identical parent strains. As

both pure protein and frozen cells were obtained from another laboratory it is not possible to make an informed estimate of the probability of these occurences. The number of these putative changes and the fact that some may have required at least 2 base changes would suggest that the former is unlikely.

d) Comparison of P.aeruginosa and P.putida Amidases

P.putida strain A87 (Clarke 1972) possess distinct acetamidase and phenylacetamidase enzymes. The acetamidase has a similar substrate specificity to the P.aeruginosa A amidase, and gives partial cross reaction with antiserum prepared against the purified P.aeruginosa enzyme. An investigation of the sequence of this P.putida enzyme (Dr. R.P. Ambler, this laboratory) is in progress. Although only partial sequence information is available for the P.aeruginosa amidase and only very limited information is available for the P.putida sequence, the structural homology found supports a common evolutionary origin. Three apparent substitutions have been found so far, all of which can be explained by single base changes within the structural gene. Only one of the substitutions involved in the P.aeruginosa directed evolution is known (Brown and Clarke, 1972), an isoleucine for threonine change conferring acetanilidase activity upon the mutant enzyme. The P.aeruginosa wild type amidase sequence involved appears

to be present in the P.putida acetamidase. One can speculate that the substitutions occuring in the P.putida enzyme are not concerned with the maintenance of enzymic activity.

e) Protein Evolution

The partial sequence presented here does not, per se, contain much information about amidase evolution. It does, however, (as would a completed sequence) form a basis for comparing amidases. Virtually nothing is known about the substitutions occuring in the laboratory mutants, although an investigation is in progress (A. Paterson, pers.comm.) A similar situation exists with respect to the naturally evolving amidases.

It was suggested earlier in this report that the conclusions drawn from amidase evolution could contribute to our understanding of protein evolution as a whole. Before general conclusions can be drawn, it will be necessary to have studied the evolution of many other proteins. We do not know the path by which pseudomonad amidases have evolved (although it might be possible to retrace it), and the laboratory mutants have each changed under a peculiar selection pressure.

Nevertheless the observation that a gene duplication may have occured, in the evolution of amidase, is not without interest. Gene duplication has been discussed as an evolutionary mechanism for at

least 25 years (Lewis 1951). Duplication without fusion of the gene copies produced would lead to a redundancy of information such that one daughter would be free to evolve. Koch (1972) has calculated that haploid evolution could occur more quickly through the sequence 1) gene duplication. 2) inactivation of 1 copy by loss of transcription or translation, 3) multiple changes in the silent copy, 4) reversion to expression, than via multiple changes in a single gene copy.

Laboratory studies show that mechanisms are available for such events. Chemostat growth of Klebsiella aerogenes on Xylitol (Rigby et al, 1974) has led to duplication of a gene coding for ribitol dehydrogenase, a poor xylitol dehydrogenase. Duplication of a glycyl-tR.N.A. synthetase gene has been found as a remedy for a similar case of poor enzymic activity in Escherichia coli (Folk and Berg 1971). Jackson and Yanofsky (1973) have shown a duplication event, with translocation of products, in the tryptophan operon of E.coli. Their observation that such duplications were unstable and were rapidly lost from the population, together with the calculation (Koch, 1972) that 2 copies of a gene would confer a slight selective disadvantage, suggests that inactivation of 1 copy may be necessary for duplication to have evolutionary significance in the long term. Khan and Hayes (1972), however, have shown that Saccharomyces

cerevisiae has 2 loci, for α-methylgalactosidase activity, producing very similar enzymes without apparent disadvantage.

Such duplication events would also have the potential for subjecting genes to different regulatory controls, and of creating new genes by fusion of products (Jackson and Yanofsky 1973). An elegant demonstration of gene fusion (Yourno et al, 1970) produced a single polypeptide chain possessing 2 enzymic activities concerned with histidine biosynthesis in Salmonella typhimurium. Crawford (1975) has concluded that fusion events have occured at various times during the evolution of the tryptophan pathway in microorganisms. Many globular proteins, in fact, may have evolved by such a route. Crystallographic studies show (Rossman and Liljas, 1974) that the peptide chains of many globular proteins are folded into 2 or more spatially separate domains, and these domains often have unrelated tertiary structures. Sequence homology between proteins of unrelated functions (Ycas 1976) supports this view and may indicate that present day proteins are evolved from relatively few ancestors.

Gene duplication with fusion of products is another process which may have occured in evolution. Some amino acyl-tR.N.A. synthetases from Bacillus stearothermophilus (Koch et al, 1974) and from E.coli (Kalousek and Konisberg, 1976) seem to have arisen by

duplication and fusion of single ancestral genes. Repeated sequences within other protein structures (Sjodahl, 1976 and examples cited therein) could be further examples of this process.

The laboratory evolution of amidase, however, demonstrates that evolution can occur, without duplication, by modification of regulatory and structural genes to use or improve existing side specificities. The utilisation of "unnatural" pentoses and penitols by K.aerogenes is due mainly to regulatory changes and exploitation of side specificities of the enzymes for utilisation of the "natural" pentoses and penitols (see Wood, 1966). Growth on only one of the 6 unnatural compounds appears due to the appearance of a new activity. The regulatory changes were all for constituitive enzyme production, but in E.coli growth on D-arabinose (an unnatural pentose) occurs through a change in inducer specificity (Leblanc and Mortlock, 1971).

Structural gene modification was also the response of S.cerevisiae when grown at low pH to reduce acid phosphatase activity (Francis and Hansche, 1972). A mutant enzyme regained 40% of the lost activity by having an altered pH optimum. S.cerevisiae has also been shown to produce an altered alcohol dehydrogenase when grown on allyl alcohol (Wills, 1976). Modification in this case resulted in an impaired enzyme such that production of toxic acrolein was limited.

The work of Campbell *et al* (1973) adopted a different approach. Using *E.coli* strains with a deleted β-galactosidase (lac z) gene, selection was for growth on lactose. β-galactosidase activity from a different locus (called egb A) was detected. The new enzyme was induced by lactose and the original gene product had β-galactosidase activity but lactose was not a substrate. (Hall and Hartl, 1974a and b). Presumably, therefore gene duplication of egb A would not allow growth. The same locus (egb A) has subsequently been evolved (Hall, 1976) to methylgalactosidase activity. This again suggests modification of specificities and although the relationship between the lac z and egb A loci is not known their products are immunologically distinct (Hall and Hartl, 1974a). The conclusion being that they are not closely related enzymes.

A similar question was asked of *S.typhimurium* isopropylmalate isomerase (iPMI), an enzyme of the leucine biosynthetic pathway normally coded for by 2 genes, leu C and leuD. Leucine auxotrophs carrying a deletion in leu D reverted by a mutation at a locus sup Q (normal function unknown). Sup Q and an additional gene, new D, are thought (Kemper, 1974) to code for 2 subunits of a multimeric enzyme and that new D protein (without change) can complex with leu C protein to give a functional iPMI. It was proposed that a modification in sup Q decreased

the affinity of sup Q subunit for new D subunit, such that sufficient new D protein was available for complexing with leu C product. Again the evolutionary relationships between these three locii are not known but evolution occured by modification. It would be interesting if the affinity of sup Q protein for new D protein is sufficiently high that a duplication of new D (assuming the amount of protein produced is related to the number of gene copies) would not lead to enough subunit being available for complexing with the leu C protein.

In both these cases the existance of 2 related loci may have been brought about by duplication of an ancestor gene with subsequent modification. Ycas (1976) has proposed that before the present there has been a period of extensively occuring duplication and deletions. If some present day enzymes were created by fusion of parts of ancestral genes and the component parts of the fusion event retained their original function (as in the event described by Yourno et al, 1970) then evolution may have produced single enzymes with 2 functions. An example of this could be cod muscle pyruvate kinase which also is responsible for oxaloacetate decarboxylase activity (Creighton and Rose, 1976). It should be noted that possession of different activities does not preclude narrow specificity with respect to an individual activity. Kirschner and Bisswanger (1976) have discussed "Multifunctional Proteins", and also conclude that gene

fusions may have played a part in the evolution of such polypeptides.

It is difficult to tell how widespread the possession of more than one activity is. Most enzymes have probably never been tested for activities different to that originally assigned to them. P.aeruginosa amidase is also a transferase and an esterase and it remains to be seen if these activities are mechanistically related as a single ancestral enzyme or by fusion of different ancestral polypeptides.

f. Conclusion

Only one aspect of evolution has been considered, how can new enzymic activities arise? I have not discussed how activities both old and new can be evolved together to produce metabolic pathways, how regulatory systems may have evolved, or how new characters may become fixed in a population. Structural studies of P.aeruginosa amidase cannot answer these questions. The partial sequence determined in the investigation forms a basis for understanding the evolution of one particular enzyme. Eventually the results of studies of the evolution of many enzyme systems may provide sufficient clues for us to comprehend the courses of protein evolution and to design "useful" enzymes ourselves. From the examples given in section e it is evident that Nature has a variety of evolutionary routes which can be used in complimentary combinations.

Fusions, duplications, translocations and modifications are all possible. When these processes are understood it may be possible to predict the evolution of a particular protein. It has earlier been speculated that the S. typhimurium sup Q locus may have evolved by "Modification", as a gene duplication may not have led to a functional enzyme. Similarly one can envisage a situation such that the rate of utilisation of a substrate is limited by the rate of entry of that substrate into the cell. If such a limitation brought a selection pressure to bear then, of all the enzymes concerned with the use of that compound, only improvement of the "permease" would confer a selective advantage. If sufficient is known about the biochemistry of a particular enzyme to predict its evolution (under given circumstances) then this will provide a test of protein evolution theories, as well as giving us the potential of evolving enzymes for specific purposes.

Two thoughts, however, occur in concluding this thesis

(1) As each laboratory evolutionary study occurs under peculiar circumstances then such studies, amidase included, will have most value (for the understanding of protein evolution) when taken at the comparative level.

(2) As evolution is a continual process then not even the products of directed evolution can be regarded as "permanent".

REFERENCES CITED

1) Alt, J., Heymann, E., and Krisch, K. (1975a) Eur.J. Biochem. 53, 357-369.

2) Alt, J., Krisch, K. and Hirsch, P. (1975b) J.gen. Microbiol. 87, 260-272.

3) Ambler, R.P. (1963) Biochem. J. 89, 349-378.

4) Ambler, R.P. (1975) Biochem. J. 151, 197-218.

5) Ambler, R.P. and Brown, L.H. (1967) Biochem.J. 104, 784-825.

6) Ambler, R.P. and Wynn, M. (1973) Biochem. J. 131, 485-498.

7) Beaven, G.H. and Holiday, E.R. (1952) Adv.Protein Chem. 7, 320-386. Academic Press.

8) Betz, J.L., Brown, J.E., Clarke, P.H., and Day, M. (1974) Genet.Res.Camb. 23, 335-359.

9) Betz, J.L., and Clarke, P.H. (1972) J.gen.Microbiol. 73, 161-174.

10) Betz, J.L. and Clarke, P.H. (1973) J.gen.Microbiol. 75, 167-177.

11) Black, L.W. and Hogness, D.S. (1968) J.Biol.Chem. 244, 1976-1981.

12) Boddy, A., Clarke, P.H., Houldsworth, M.R. and Lilly, M.D. (1967) J.gen.Microbiol. 48, 137-145.

13) Brammar, W.J., and Clarke, P.H. (1964) J.gen.Microbiol. 37, 307-319.

14) Brammar, W.J., Clarke, P.H. and Skinner, A.J. (1967) J.gen.Microbiol. 47, 87-102.

15) Brammar, W.J., McFarlane, N.D. and Clarke, P.H. (1966) J.gen.Microbiol. 44, 303-309.

16) Bridgen, J. and Kock, G.L.E. (1975) in Amino acids, Peptides and Proteins 6, p.118. Chemical Society,London.

17) Brown, J.E., Brown, P.R. and Clarke, P.H. (1969) J.gen.Microbiol. 57, 273-285.

18) Brown, J.E. and Clarke, P.H. (1970) J.gen.Microbiol. 64, 329-342.

19) Brown, P.R. and Clarke, P.H. (1972) J.gen.Microbiol. 70, 287-298.

20) Brown, P.R., Smyth, M.J., Clarke, P.H. and Rosemeyer, M.A. (1973) Eur.J.Biochem. 34, 177-187.

21) Campbell, J.H., Lengyel, J.A. and Langridge, J. (1973) Proc. Nat. Acad. Sci. U.S.A. 70, 1841-1845.

22) Clarke, P.H. (1972) J.gen.Microbiol. 71, 241-257.

23) Clarke, P.H. (1974a) Symp.Soc.Gen.Microbiol. 24, 183-217, Cambridge University Press.

24) Clarke, P.H. (1974b) Biochem. Soc. Trans. 2, 831-834.

25) Clarke, P.H., Houldsworth, M.A. and Lilly, M.D. (1968) J.gen.Microbiol. 51, 225-234.

26) Crawford, I.P. (1975) Bacteriol.Revs. 39, 87-120.

27) Creighton, D.J. and Rose, I.A. (1976) J.Biol.Chem. 251, 69-72.

28) Crowshaw, K., Jessup, S. and Ramwell, P.W. (1967) Biochem.J. 103, 79-85.

29) Dalgleish, C.E. (1952) Biochem.J. 52, 3-14.

30) Davis, B.J. (1964) Ann.N.Y.Acad.Sci. 121, 404-427.

31) Day, M., Potts, J.R. and Clarke, P.H. (1975) Genet. Res.Camb. 25, 71-78.

32) Dayhoff, M. (1973) Atlas of Protein Sequence and Structure, 5 Supplement 1, Nat.Biomedical Research Foundation, Washington, U.S.A.

33) Dent, C.E. (1974) Biochem.J. 41, 240-253.

34) Drapeau, G.R., Boily, Y. and Houmard, J. (1972) J.Biol.Chem. 247, 6720-6726.

35) Draper, P. (1967) J.gen.Microbiol. 46, 111-123.

36) Edman, P. (1956) Acta.Chem.Scand. 10, 761-768.

37) Edman, P. and Henschen, A. (1975) in Molecular Biology, Biochemistry and Biophysics 8, Protein Sequence Determination, 232-279. (Ed. Needleman,S.B.) Springer-Verlag, Berlin.

38) Edman, P. and Sjoquist, J. (1956) Acta.Chem.Scand. 10, 1507-1509.

39) Fasman, G.D. (1976) Handbook of Biochemistry and Molecular Biology 1-3 Chemical Rubber Company Press.

40) Fernald, N.J. and Ramaley, R.F. (1972) Arch.Biochem. Biophys. 153, 95-104.

41) Findlatter, J.D. and Orsi, B.A. (1973) FEBS Lett. 35, 109-111.

42) Folk, W.R. and Berg, P. (1971) J.Mol.Biol. 58, 595-610.

43) Francis, J.C. and Hansche, P.E. (1972) Genetics 70, 50-73.

44) Frankel-Conrat,H. (1954) J.Am.Chem.Soc. 76, 3606-3607.

45) Gray, W.R. (1972a) Methods Enzymol. 25, 121-143. Academic Press.

46) Gray, W.R. (1972b) Methods Enzymol. 25, 333-344. Academic Press.

47) Hall, B.G. (1976) J. Bacteriol. 126, 536-538.

48) Hall, B.G. and Hartl, D.L. (1974a) Genetics 76, 391-400.

49) Hall, B.G. and Hartl, D.L. (1974b) Nature (London) 248, 152-153.

50) Harris, J.I. and Hindley, J. (1965) J.Mol.Biol. 13, 894-913.

51) Hartley, B.S. (1970) Biochem.J. 119, 805-822.

52) Hartley, B.S. (1974) Symp.Soc.Gen.Microbiol. 24, 151-182. Cambridge University Press.

53) Hegeman, G.D. and Rosenberg, S.L. (1970) Annu.Rev. Microbiol. 24, 429-462.

54) Hirs, C.H.W. (1956) J.Biol.Chem. 219, 611-621.

55) Hirs, C.H.W. (1967) Methods Enzymol. 11. Academic Press.

56) Hirs, C.H.W. and Timasheff, S.N. (1972) Methods Enzymol. 25, Academic Press.

57) Hsuing, K.P., Kuan, S.S. and Guilbaut, G.G. (1975) Biochem. Biophys. Res. Commun. 66, 1225-1230.

58) Hynes, M.J. (1975) J.gen.Microbiol. 91, 99-101.

59) Inglis, A.S., Nichols, P.W. and Roxburgh, C.M. (1971) Aust.J.Biol.Sci. 24, 1247-1250.

60) Jackoby, W.B. and Fredericks, J. (1964) J.Biol. Chem. 239, 1978-1982.

61) Jackson, E.N. and Yanofsky, C. (1973) J.Bacteriol. 116, 33-40.

62) Jacob, F. and Monod, J. (1961) J.Mol.Biol. 3, 318-356.

63) Jepson, J.B. and Smith, J. (1953) Nature (London) 172, 1100-1101.

64) Kalousek, K. and Konisberg, W. (1976) FEBS.Lett. 61, 151-153.

65) Kaminski, M., Falcoz-Kelly, F., Truffa-Bachi, P., Patte, J-C. and Cohen, G.N. (1969) Eur.J.Biochem. 11, 278-282.

66) Kelly, M. and Clarke, P.H. (1960) Biochem.J. 74, 21P.

67) Kelly, M. and Clarke, P.H. (1962) J.gen.Microbiol. 27, 305-316.

68) Kelly, M. and Kornberg, H.L. (1962) Biochim.Biophys. Acta. 64, 190-191.

69) Kelly, M. and Kornberg, H.L. (1964) Biochem.J. 93, 557-566.

70) Kemper, J. (1974) J.Bacteriol. 120, 1176-1185.

71) Khan, N.A. and Hayes, R.H. (1972) Molec.Gen.Genetics 118, 279-285.

72) King, J. and Laemmli, U.K. (1971) J.Mol.Biol. 62, 465-477.

73) Kirschner, K., and Bisswanger, H. (1976) Annu.Rev. Biochem. 45, 143-166.

74) Koch, A.L. (1972) Genetics 72, 297-316.

75) Koch, G.L.E., Boulanger, Y. and Hartley, B.S. (1974) Nature (London) 249, 316-320.

76) Leblanc, D.J. and Mortlock, R.P. (1971) J.Bacteriol. 106, 90-96.

77) Lewis, E.B. (1951) Cold Spring Harbour Symp. Quant. Biol. 16, 159-174.

78) Li, S.L., Denney, R.M. and Yanofsky, C. (1973) Proc. Nat.Acad.Sci. U.S.A. 70, 1112-1116.

79) McFarlane, N.D., Brammar, W.J. and Clarke, P.H. (1965) J.gen.Microbiol. 44, 303-309.

80) Michl, H. (1951) Monatsh.Chem. 82, 489-493.

81) Miller, G.I. (1959) Anal.Chem. 31, p964.

82) Myers, D.K., Tol, J.W. and De Jonge, M.H.T. (1957) Biochem.J. 65, 223-232.

83) Needleman, S.B. (1975) Ed. Molecular Biology, Biochemistry and Biophysics 8 Protein Sequence Determination. Springer-Verlag, Berlin.

84) Offord, R.E. (1966) Nature (London) 211, 591-593.

85) Penke, B., Frenczi, R. and Kovacs, K. (1974) Anal. Biochem. 60, 45-50.

86) Perutz, M.F., Muirhead, H., Cox, J.M. and Goaman, L.C.G. (1968) Nature (London) 219, 131-139.

87) Rigby, P.W.J., Burleigh, B.D. jun. and Hartley, B.S. (1974) Nature (London) 251, 200-204.

88) Rossman, M.G. and Liljas, A. (1974) J.Mol.Biol. 85, 177-181.

89) Rydon, H.N. and Smith, P.W.G. (1952) Nature (London) 169, 922-923.

90) Ryle, A.P., Sanger, F., Smith, L.F. and Kitai,R. (1955) Biochem. J. 60, 541-556.

91) Sjodahl, J. (1976) FEBS. Lett. 67, 62-67.

92) Smith, D.G. (1967) Methods Enzymol. 11, 214-231. Academic Press.

93) Smyth, P.F. and Clarke, P.H. (1975a) J.gen.Microbiol. 90, 81-90.

94) Smyth, P.F. and Clarke, P.H. (1975b) J.gen.Microboil 90, 91-99.

95) Stanier, R.Y., Palleroni, N.J. and Doudoroff, M. (1966) J.gen.Microbiol. 43, 159-271.

96) Steiner, M. (1959) Symp.Soc.Exp.Biol. 13, 177-192.

97) Strid, L. (1973) FEBS. Lett. 33, 192-196.

98) Truffa-Bachi, P., Guiso, N., Cohen, G.N., Theze, J. and Burr, B. (1975) Proc.Nat.Acad.Sci. U.S.A. 72, 1268-1271.

99) Waley, S.G. and Watson, J. (1953) Biochem.J. 57, 529-538.

100) Weiner, A.M., Platt, T. and Weber, K. (1972) J.Biol. Chem. 247, 3242-3251.

101) Weber, K. and Osborn, M. (1966) J.Biol.Chem. 244, 4406-4412.

102) Wills, C. (1976) Nature (London) 261, 26-29.

103) Wood, W.A. (1966) Annu.Rev.Biochem. 35, 521-528.

104) Woods, K.R. and Wang, K.T. (1967) Biochim.Biophys. Acta. 133, 369-370.

105) Woods, M.J., Edgeworth, M.A. and Orsi, B.A. (1975) Biochem.Soc.Trans. 3, 1216-1219.

106) Woods, M.J. and Orsi, B.A. (1974) Biochem.Soc. Trans. 2, 1344-1345.

107) Yamada, S. and Itano, H.A. (1963) Biochim.Biophys. Acta 130, 538-544.

108) Ycas, M. (1976) J.Molec.Evoln. 7, 215-244.

109) Yourno, J., Kohno, T. and Roth, R.J. (1970) Nature (London) 228, 820-824.

APPENDIX

The Reliability of Molecular Weight Determination by Amino acid analysis.

1) Methods

Hydrolysis of a protein followed by amino acid analysis is routinely used to determine the relative molar frequencies of the constituent amino acids. Calculation of the number of moles of each amino acid in one mole of protein can be achieved in several ways, and the qualitative criteria for acceptance are usually

(i) that the number of moles of each amino acid per mole of protein should be nearly integral

(ii) the composite should yield a molecular weight in agreement with values calculated from independent experiments (e.g. S.D.S. gel electrophoresis)

As Nyman and Lindskog (1964) state, ".. at the true molecular weight, the sum of the deviations from integral values for the different amino acids should be minimized." This leads to a mathematical model for calculating molecular weight from the analysed amino acid frequencies. The method proposed by Nyman and Lindskog (1964) assumes that amino acids occurring with a smaller frequency give a more reliable estimate of the molecular weight and should be given a greater contribution. They state, in rather unclear English, that, ".. this weighting has been achieved by multiplying the deviation from an integral

number for each amino acid, by the minimum molecular weight for the amino acid... A curve showing this sum should give a minimum at the best estimate of molecular weight." I have interpreted "minimum molecular weight for the amino acid", as the protein molecular weight if only one residue of the amino acid concerned is present. A plot of

$$\left(\text{min.mol.wt.} \times \left|\text{number of residues} - \text{nearest integral number}\right|\right)$$

versus protein molecular weight should then show a minimum at the true molecular weight. An example of such a calculation is given below

(i) Analysis yields n_i relative moles of amino acid i per mole of protein, n_j moles of amino acid j and so on.

(ii) Choose protein molecular weight M_a and calculate the matrix

amino acid	calculated relative residue no./ mole protein	nearest integer	difference	weighted difference
i	n_i	I_i	$\lvert n_i - I_i \rvert$	$\lvert n_i - I_i \rvert \times \frac{Ma}{I_i}$
j	n_j	I_j	$\lvert n_j - I_j \rvert$	$\lvert n_j - I_j \rvert \times \frac{Ma}{I_j}$
z	n_z	I_z	$\lvert n_z - I_z \rvert$	$\lvert n_z - I_z \rvert \times \frac{Ma}{I_z}$

(iii) Sum the last column to give

$$\sum_{i=i}^{z} \left|\frac{n_i - I_i}{I_i}\right| \times Ma$$

(iv) Repeat for molecular weight M_b etc. and plot the sums against Ma, Mb etc. The minimum point indicates the molecular weight.

It can be shown that the sum

$$\sum_{i=i}^{z}\frac{|n_i-I_i|}{I_i} \times M_a = \frac{|n_i-I_i|}{I_i} M_a + \frac{|n_j-I_j|}{I_j} M_a + \ldots + \frac{|n_z-I_z|}{I_z} M_a$$

$$= M_a\left(\frac{|n_i-I_i|}{I_i} + \frac{|n_j-I_j|}{I_j} + \ldots + \frac{|n_z-I_z|}{I_z} \right)$$

$$= Ma \sum_{i=i}^{z} \frac{|n_i-I_i|}{I_i} \qquad \text{Eqn 1}$$

The sum of the weighted difference $\sum\frac{|n_i-I_i|}{I_i}$, although weighting in favour of deviations for small values of I_i, unfortunately decreases as I increases. The inclusion of Ma counteracts the decrease preventing artefactual minima at high values of molecular weight.

Black and Hogness (1969) have modified the method in a more rational way, as given below.

(i) Analysis yields frequencies $\emptyset_i, \emptyset_j, \ldots \emptyset_z$ for amino acids i, j,...z, respectively

(ii) Choose protein molecular weight M and calculate n_i, the number of moles of amino acid i from the relationship

$$n_i = \phi i \ \frac{(M-18)}{\Sigma(\phi_{iRi})} \qquad \text{Eqn 2.}$$

where R_i is the residue molecular weight of amino acid i

(iii) Repeat for all amino acids j to z

(iv) Calculate the sum $\sum_{i=i}^{z} \frac{|n_i - I_i|}{I_i}$ Eqn 3

which provide a weighting..."to account for the fact that the smaller the residue number the more significant is its deviation from an integral value." (Black and Hogness, 1969).

(v) Now the maximum value such a sum can take is 0.5 times the sum of the reciprocals of the integers. (This is easily verified by generating a set of data such that, $|n_i - I_i| = 0.5$, for all i). Now calculate

$$f = \frac{(\Delta_i / I_i)}{0.5(\Sigma^1 / I_i)} \qquad \text{Eqn 4}$$

where $\Delta i = |ni - Ii|$

Repeat for different values of M, and plot f versus Molecularweight M.

This quotient (Eqn. 4) allows "..easy comparison to that obtained if random errors in the measurement of the frequencies become dominant... Values of this ratio near 0.5 are expected if random errors dominate the result." (Black and Hogness, 1909). It also allows easy comparison between different proteins over diff-

erent molecular weight ranges. With the previous method the values of the ordinate would depend on the molecular weight range taken.

Using a computer programme (supplied by Dr.J.Moult, Dept. of Physics, Edinburgh) I have used the equations of Black and Hogness to determine molecular weights from amino acid calculations. This programme was also capable of adding normally distributed random errors, of defined size, into the data. Glutamine has been counted as glutamic acid, and asparagine as aspartic acid as amino acid analysis methodology does not distinguish these amides from the acids. I have also assumed that in equation 2 M-18=M for the high values of molecular weight considered. Values of f were calculated at intervals of 200-250 molecular weight units.

2) Calculations Using Ideal Data.

Amino acid frequencies ϕ_i were taken from the sequences of the following proteins, so that all ϕ_i were integers.

Potato Chymotryptic Inhibitor 1 (Richardson, 1974, M.W. 9332), Anabaena variabilis plastocyanin (Aitken 1975, M.W. 11,184, P.fluorescens azurin (Ambler and Brown 1967 M.W. 13,946), B.licheniformis penicillinase (Meadway 1969 M.W. 29,503).

Staphylococcus aureus penicillinase (Ambler 1975, M.W.28,794), Rabbit actin (Dayhoff 1973, M.W.41,719).

TABLE A1 The Calculation of Molecular weight from amino acid analyses.

Azurin data with 3% error Range 5000-25000 (at 200 unit intervals)

Run No.	Posn. of min	value of f	f value of next lowest trough.
1	14,000	0.16	0.398
2	14,400	0.177	0.358
3	14,000	0.232	0.354
4	14,000	0.14	0.338
5	14,200	0.144	0.369
6	13,800	0.164	0.308
7	14,200	0.133	0.363
8	14,000	0.14	0.372
Azurin data with 5% error			
1	14,600	0.253	0.335
2	13,800	0.216	0.387
3	13,200	0.240	0.391
4	14,200	0.173	0.429
5	13,800	0.223	0.411
6	13,600	0.181	0.284
7	14,000	0.159	0.314
8	14,600	0.193	
Azurin data with 10% error			
1	14,000	0.281	0.389
2	14,800	0.213	0.343
3	14,000 and 14,200	0.235	0.434
4	14,800	0.253	0.413

Molecular weights calculated from these ideal data all, except for Actin, gave unambiguous minima (f 0.05) at a calculated molecular weight within one molecular weight interval (i.e. 200 to 250) of the known value. The data for actin gave an unambiguous minima (f = 0.02) at a molecular weight of 40,600 (200 unit intervals). In no case was the next lowest trough (excluding values at integral multiples of the molecular weight) lower than f=0.33.

3) Calculations Using Ideal Data with Added Errors.

Ideal amino acid frequencies (i.e. taken from known protein sequences) and the computer programme which added normally distributed random errors of a defined size to these frequencies was used in this simulation. The size of these errors is quoted as a percentage, and errors of this standard deviation (% of ideal value) were added or subtracted from each amino acid frequency. Some results are given in tables A1 and A2 and by fig. A1 (intervals of 200 m.w. units were used). Eight runs were carried out at each error level.

Adding 3% error to ideal azurin data (see table A1) the calculated molecular weight was in the range 13,800 to 14,200. With 5% error this range increased to 13,200 to 14,600. At 10% error the minima ranged from 12,800 to 14,800, and in one case (out of 8) a trough at a

TABLE A2 The Calculation of Molecular Weight from amino acid analysis.

S.aureus Penicillinase range 10,000 to 40,000 in steps of 200.

Data with 3% error

Run No.	Posn. of minimum.	Value of f	f value at next lowest trough.
1	29,400	0.192	0.361
2	28,800	0.251	0.338
3	28,600	0.238	0.348
4	28,000	0.271	0.35
5	28,800	0.151	0.376
6	28,600 29,400 29,600	0.277	0.362
7	29,000	0.24	0.377
8	28,200	0.224	0.375

Data with 5% error

1	30,400	0.301	0.332
2	15,600	0.300	0.328 (at 31,000)
3	27,400	0.239	0.362
4	29,800	0.245	0.36
5	39,000	0.345	0.347 (at 31,400)
6	28,800	0.247	0.376
7	29,400	0.281	0.312
8	28,800	0.311	0.33

Data with 10% error

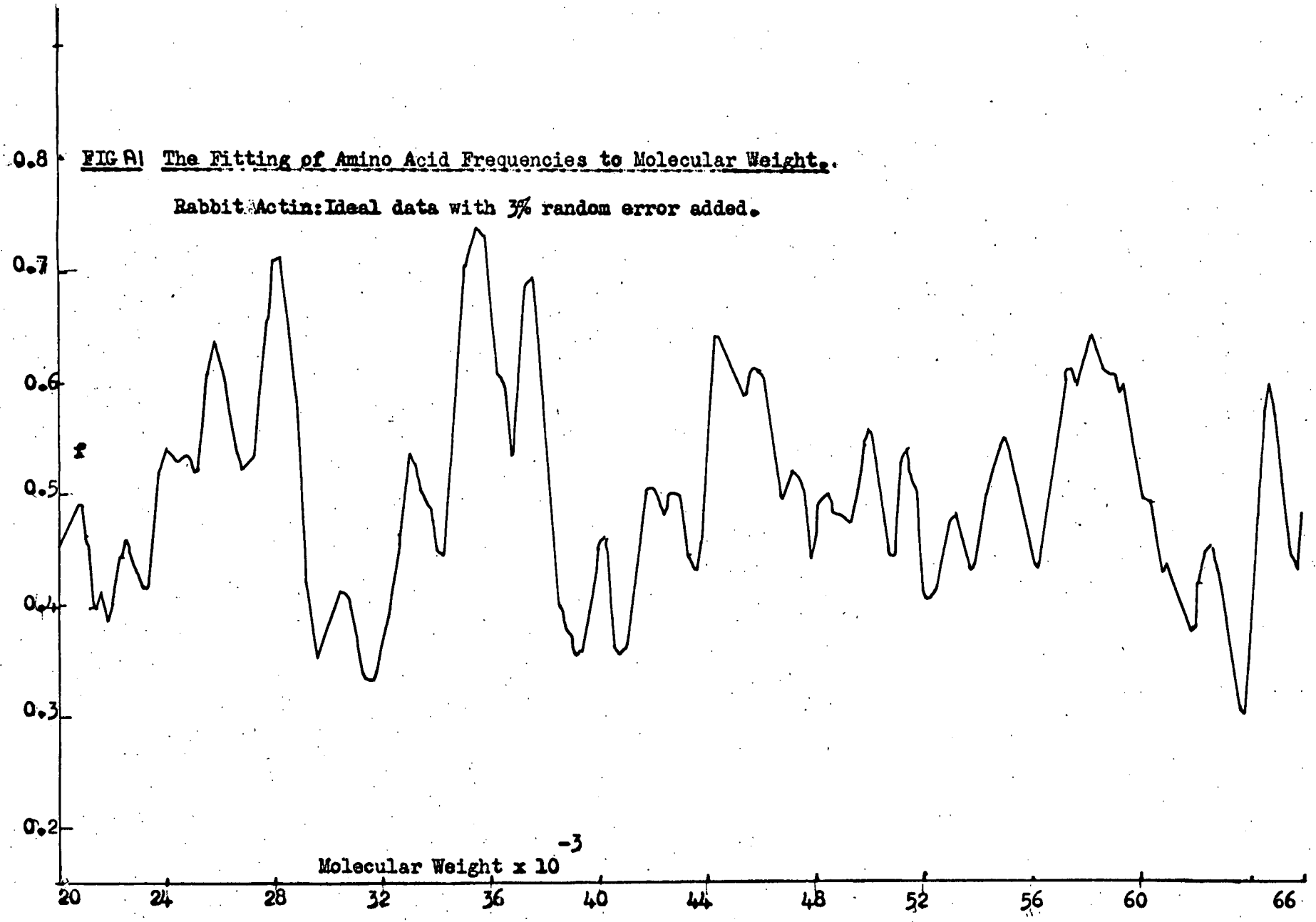
1	24,400	0.314	0.317 (at 12,800)

TABLE A2 - Contd.

2	32,800	0.255	0.361
3	27,400	0.312	0.32 (at 38,200)
4	18,400	0.294	0.37 (at 34,400)
5	16,000	0.299	0.300 (at 31,200)
6	12,400	0.246	0.381 (at 28,000)
7	27,800	0.253	0.363
8	12,400	0.312	0.372 (at 36,400)

FIG A1 The Fitting of Amino Acid Frequencies to Molecular Weight.

Rabbit Actin: Ideal data with 3% random error added.



different position had the lowest value predicting an "incorrect" molecular weight.

The S.aureus penicillinase data (table A2) shows similar trends. At 3% error it is possible to predict the correct molecular weight to within 1,000 units. With 5% error, however, it can be seen that the trough with the lowest f value does not always predict the correct (approximate) value. With 10% error the minimum value of f would have predicted an incorrect molecular weight 5 times and an approximately correct value 3 times.

The data for actin values with known error is presented in a different form (fig A1). This figure represents one run chosen, randomly out of the eight carried out at 3% error, and indicates that meaningful interpretation is not possible. The other 7 runs are in agreement with this statement. With 5% and 10% errors (results not presented here) one cannot assign any meaningful molecular weight as each calculation showed multiple minima within close ranges of f.

4) Conclusions

The results show that for the one example considered of molecular weight 41,719 even small errors can lead to incorrect results. Where the molecular weight does not exceed 30,000, however, it should be possible to predict protein molecular weights from amino acid analyses.

TABLE A3

Amino acid frequencies of B.licheniformis Penicillinase

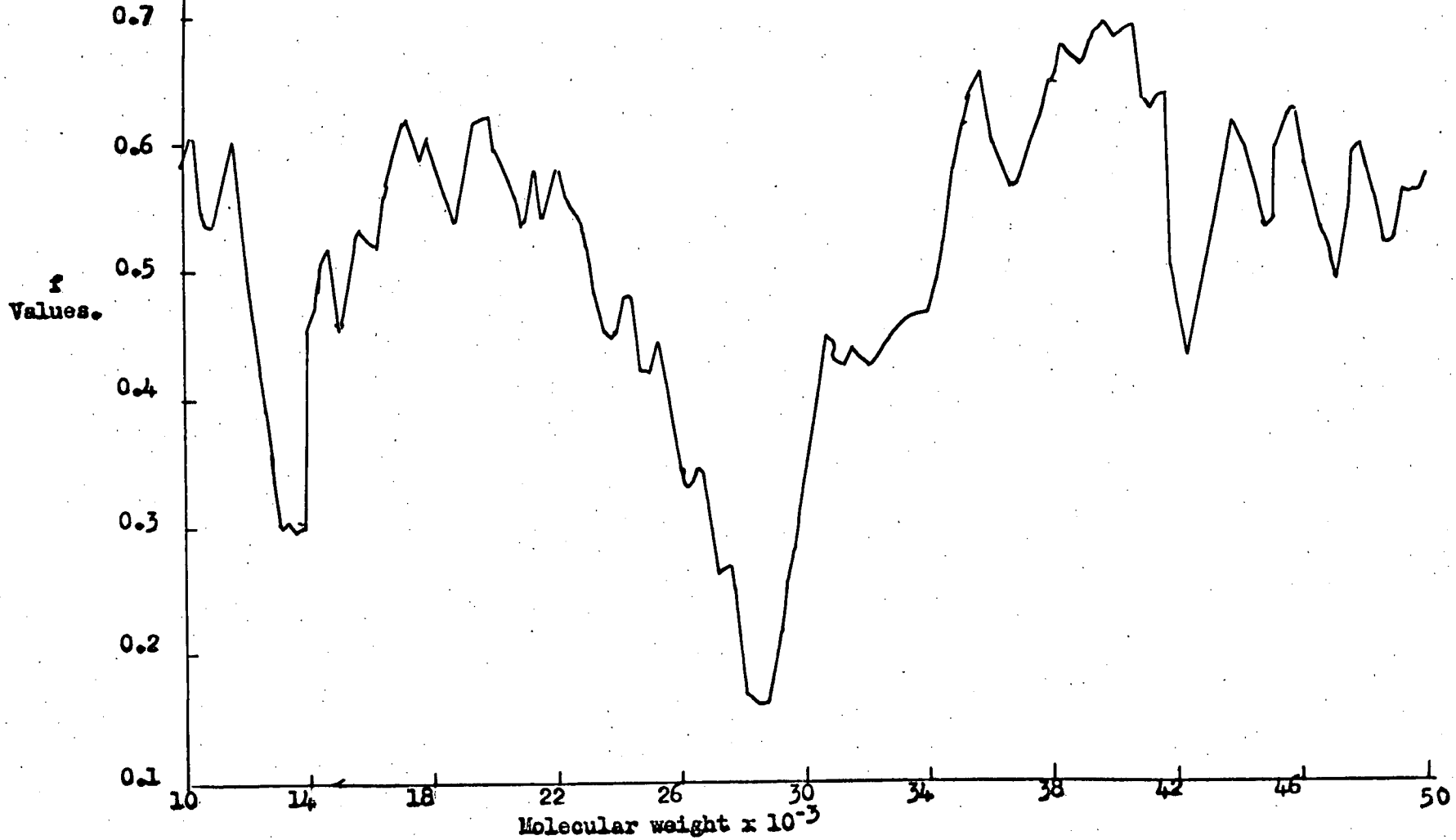
Amino acid	Residues in sequence(1)	Amino acid frequencies(2)	Calculated analysis at MW 28400(3)
Lys	24	.0897	22.8
His	1	.0040	1.02
Arg	15	.0526	13.4
Trp	3	.0078	1.98
Asx	37	.1394	35.4
Thr	21	.0820	20.8
Ser	11	.0427	10.8
Glx	27	.1112	28.2
Pro	11	.0422	10.7
Gly	15	.0575	14.6
Ala	26	.0974	24.1
Cys	0	0	0
Val	15	.0593	15.1
Met	5	.0160	4.06
Ile	14	.0548	13.9
Leu	24	.1033	26.24
Tyr	6	.0237	6.0
Phe	7	.0269	6.8

(1) Meadway P.J. 1969.

(2) Data of R.P. Ambler (pers.comm), a mean of 3 values determined by analysis after acid hydrolysis except for tryptophan (estimated spectrophotometrically)

(3) 28400 is the predicted m.w., from fig A2, using this table of frequencies.

FIG 42 The Fitting of Amino Acid Frequencies to Molecular Weight for B. licheniformis Penicillinase



To test this hypothesis an amino acid analysis of B.licheniformis penicillinase (data supplied by Dr.R.P. Ambler, as a mean of three determinations, see table A3) has been used to calculate this proteins molecular weight. It can be seen in fig A2 that from these results it is possible to assign a molecular weight of 28,400. This agrees quite well with the value of 29,503 calculated from the amino acid sequence (Meadway 1969).

References Cited in Appendix.

Aitken, A. (1975) Biochem.J. 149, 675-683.

Ambler, R.P. (1975) Biochem.J. 151, 197-218.

Ambler, R.P. and Brown, L.H. (1967) Biochem.J. 104, 784-825.

Black, L.W. and Hogness, D.S. (1964) J.Biol.Chem. 244, 1976-1981.

Dayhoff, M. (1973) Atlas of Protein Sequence and Structure 5 Supplement 1 National Biomedical Research Foundation Washington, U.S.A.

Meadway, R.J. (1969) Ph.D.Thesis, Univ.Edinburgh.

Nyman, P.O. and Lindskog, S. (1964) Biochim.Biophys.Acta. 85, 141-151.

Richardson, M. (1974) Biochem.J. 137, 101-112.

ABSTRACT OF THESIS

Name of Candidate ANTHONY DAVID AUFFRET

Address c/o University Chemical Laboratories, Lensfield Road, Cambridge.

Degree Doctor of Philosophy *Date* 7/12/76

Title of Thesis STRUCTURAL STUDIES OF A BACTERIAL AMIDASE.

The primary structure of the amidase from Pseudomonas aeruginosa strain PAC 142, has been investigated. Some evidence is presented that the molecular weight of the amidase subunit is 40,000 to 42,000, higher than previously thought. Only partial sequence information has been obtained but some evidence has been found that a gene duplication may have occurred in the evolution of this amidase. The results are discussed with relation to other examples of protein evolution. An appendix, dealing with the reliability of molecular weight determination by amino acid analysis, is included.

Use other side if necessary.