STUDIES ON STAPHYLOCOCCAL CHLORAMPHENICOL ACETYLTRANSFERASES.

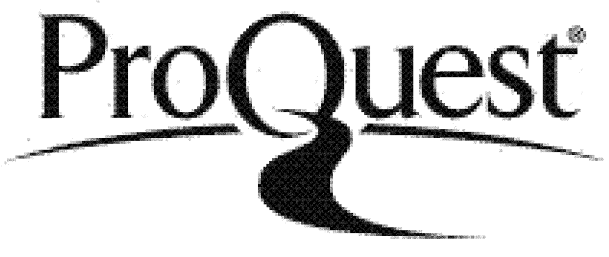
By

JOHN EDWARD FITTON, B.Sc.

A thesis submitted in partial fulfilment of the regulations governing the Ph.D. degree at the University of Leicester.

Department of Biochemistry,
Adrian Building,
University of Leicester,
University Road,
Leicester LEI 7RH.

October 1977.



ProQuest LLC
789 East Eisenhower Parkway
P.O. Box 1346
Ann Arbor, MI 48106-1346

To

whom it may concern.

ACKNOWLEDGEMENTS.

I wish to thank my supervisor, Professor W.V. Shaw, for all his help and encouragment over the last three years.

I am grateful to Dr. D. Peacock for his assistance in writing computer programmes and tuition in the use of the computer.

In addition I thank Ms. M. Wigney for her expert technical assistance, and the Medical Research Council for the award of a Research Studentship.

ABSTRACT.

I Four electrophoretic variants of chloramphenicol acetyltransferase (E.C.2.3.1.99) have been purified to homogeneity from cell-free extracts of staphylococci by affinity chromatography. All four enzymes show similar Km values for the substrates, acetyl Coenzyme A and D,threo chloramphenicol. Amino acid analyses and tryptic peptide maps of the four enzymes are similar.

II The N-terminal sequences of seven chloramphenicol acetyltransferase variants from both Gram positive and Gram negative bacteria have been determined by the method of Edman degradation of proteins covalently attached to solid phase supports. More than 90 percent of the primary sequence of one of the staphylococcal variants has been determined.

III Kinetic studies and the results of chemical modification experiments have implicated the importance of a histidine residue in the mechanism of catalysis, and a unique histidine residue in the native enzyme has been found to react with iodoacetamide with consequent formation of 3-amidocarboxymethylhistidine. There is no evidence of a covalent acyl-enzyme intermediate in the catalytic process.

IV The use of secondary structure prediction methods has allowed the comparison of both primary and predicted secondary structures of the N-termini of ten chloramphenicol acetyltransferase variants.

The results of this study are consistent with the view that the chloramphenicol acetyltransferase enzymes have evolved from a common ancestral protein and, although their primary sequences differ considerably in some cases, their secondary structures and catalytic mechanism are likely to be similar.

Abbreviations.

Diethylpyrocarbonate	DEPC.
1-Fluoro-2,4-dinitrobenzene	FDNB.
Bromoacetyl Coenzyme A	BrAcCoA.
N-ethyl maleimide	NEM.
5,5' dithiobis-(2-nitrobenzoic acid)	DTNB.
2-nitro-5-thiocyanobenzoic acid	NTCB.
2,2' dithiopyridine	DTDP.
Iodoacetic acid	IAA.
Iodoacetamide	INH_2.
p-Chloromercuribenzoic acid	PCMB.
Chloramphenicol Acetyltransferase	CAT.
Carboxymethyl Chloramphenicol Acetyltransferase	Cm CAT.
Chloramphenicol	CM.
D- threo-1-p-nitrophenyl (-2-dichloroacetamido) 1-hydroxypropane	3-deoxy CM.
D- threo-1-p-nitrophenyl(2-amino-1,3-propanediol)	CM base.
Phenylisothiocyanate	PITC.
Phenylthiohydantoin	PTH.
5-Dimethylamino-1-naphthalene Sulfonyl	dansyl

CONTENTS.

LIST OF FIGURES.

LIST OF TABLES.

CHAPTER I.

INTRODUCTION.

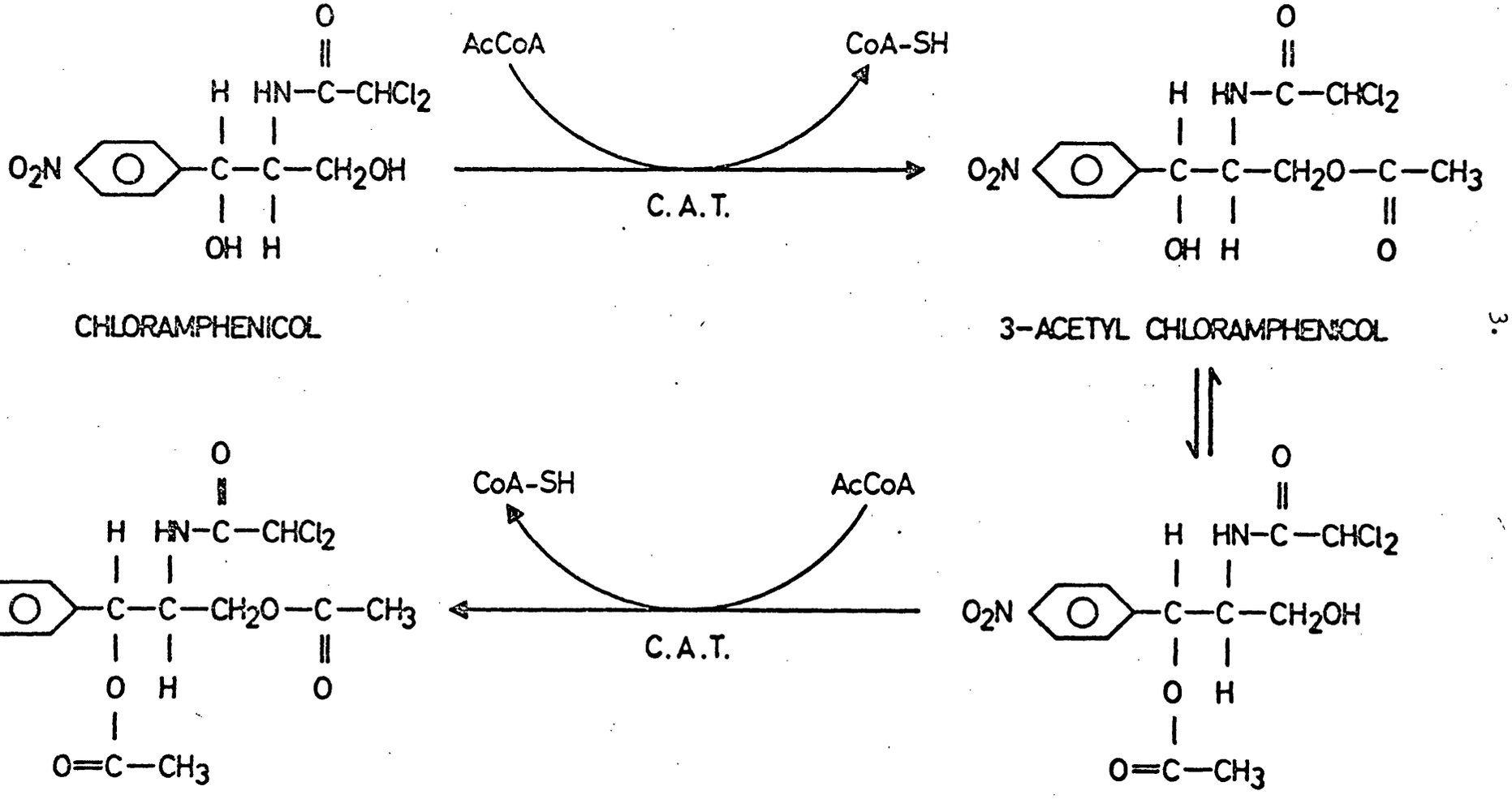
The discovery of chloramphenicol as the bacteriostatic agent in cultures of Streptomyces venezuelae was made independently by two laboratories in 1947 (1,2) and chloramphenicol was the first broad spectrum antibiotic introduced into medicinal use to inhibit the growth of bacteria (3). It was the first antibiotic to be chemically synthesised (4) and is still the only antibiotic industrially produced in this way.

Chloramphenicol (CM) has been shown to bind to the 50S subunit of prokaryotic ribosomes (5) and is thought to inhibit the peptidyl transferase reaction (6) so preventing protein synthesis and cell growth. Since the CM molecule has two asymmetric carbon atoms, four stereoisomers are possible. The two erythroisomers are biologically inactive and the L-threo-isomer is less than 0.5 percent. as active as the natural D-threo-isomer.

CM resistant strains had been isolated by 1954 (7) and resistance is now frequently encountered among many genera of bacteria. Apart from the intrinsic tolerance of certain micro-organisms (e.g. fungi, mycobacteria and Pseudomonas), virtually all examples of acquired CM resistance among eubacteria are due to enzymic acetylation and inactivation of the antibiotic by chloramphenicol acetyltransferase (E.C. 2.3.1.99) with acetyl CoA serving as the acyl donor (8). The primary product of the reaction is 3-acetyl CM, the 1,3 diacetyl product is only found after prolonged incubation of substrates with enzyme (9,10) and is probably formed by non-enzymic acyl transfer, followed by a second enzymic acylation reaction to give 1,3 diacetyl CM (figure 1). The O-acetoxy derivatives of CM do not bind to bacterial ribosomes, fail to inhibit protein synthesis and are consequently devoid of antibiotic activity (11). The synthesis of CAT is a constitutive property of CM resistant members of the Enterobacteriaceae whereas Gram positive species such as staphylococci or streptococci

Figure: 1.

Chloramphenicol acetyltransferase catalysed reaction. Acetylation of chloramphenicol, pH optimum 7.8.


AcCoA
CoA-SH
C.A.T.
CHLORAMPHENICOL
3-ACETYL CHLORAMPHENICOL
CoA-SH
AcCoA
C.A.T.
1,3 DIACETYL CHLORAMPHENICOL
1-ACETYL CHLORAMPHENICOL

synthesise CAT only after exposure to CM or closely related analogues possessing inducer activity (12). The CAT structural genes (and regulatory loci in the inducible systems) are associated with plasmids in most (but not all) of the cases examined to date (13). In the cases of the three (Types I, II and III) variants of CAT specified by R plasmids in enteric bacteria, the CM resistance determinant is usually only one of several resistance markers present. The R plasmids in question are usually large (more than 30×10^6 daltons) and their replication is regulated to yield only one copy per cell. The inducible CAT system in staphylococci is carried on small plasmids (3×10^6 daltons or less) which specify no known functions other than CM resistance and which exist as multi-copy replicons (52).

Most of the attention to date has focused on the three naturally occurring variants of CAT which are R-factor linked and are found in the Enterobacteriaceae (R-factor types I, II and III). Significant differences have been observed between CAT variants with respect to net charge of the native protein, substrate affinity, sensitivity to thiol inhibitors, heat denaturation and immunological reactivity (14, 8, 15, 16, 17, 18). More recent studies have revealed differences in elution behaviour on affinity and hydrophobic chromatographic supports (19). One variant of R-factor CAT (type II) has been shown to be especially sensitive to inhibition by a sulfhydryl-specific reagent (15).

All CAT variants studied in detail are tetrameric proteins composed of four identical subunits of molecular weight 24,000 daltons (96,000 daltons per tetramer) with the exception of three 'rogue' CAT types (R-factor type III , Agrobacterium tumefaciens and Streptomyces acrimycini) which have a subunit molecular weight of 25,000 daltons (20). Certain CAT variants have been shown to undergo subunit hybridization in vitro or, in some cases in vivo (16, 21).

Comparatively little is known of the mechanism of the enzyme reaction save for the observation that CAT is absolutely specific for the D-threo-isomer of CM and requires an acyl donor which is a short chain fatty acyl thioester of Coenzyme A (12). Both ATP and ADP are competitive inhibitors of CAT with respect to acetyl CoA, suggesting that the acyl donor is recognised via the adenine nucleotide moiety of the coenzyme (12, 22).

A large number of analogues of CM have been synthesised in the hope of finding a compound with antibiotic activity which is not a substrate for CAT. Although such an analogue has not yet been found the analogues have been useful in determining the necessary structural features for effective acylation by CAT. Previous studies have shown that the para substituent and the N-substituent are relatively unimportant whereas the steric configuration (D-threo) and the 1,3 propandiol substituent are of extreme importance in determining the ability of analogues to act as acyl acceptors (23).

Although the complete primary structure of an enteric CAT (type I, plasmid JR66) is known (W.V. Shaw, D.B. Burleigh, A. Dell, H. Morris and B.S. Hartley, in preparation) and X-ray crystalographic studies are in progress virtually nothing is known of the primary, secondary or tertiary structures of other CAT variants and this is particularly true of the CAT variants from Gram positive bacteria.

Immunological studies with antisera against type I CAT (plasmid JR66), type III CAT (plasmid R387) and type C staphylococcal CAT (plasmid C221) indicate that there is considerable homology within the staphylococcal variants but major differences may exist between the Gram positive and Gram negative variants (17, 15, 16.)

The present study was conceived with the following goals in mind:

a. to purify each of the four (types A,B,C and D) electrophoretically distinct variants found in staphylococci (16) and to characterise them by

kinetic analysis, chemical modification, amino acid analysis, peptide maps and N-terminal sequence determinations.

b. to determine as much of the primary structure of a staphylococcal CAT variant (type C, plasmid C221) as might be required to make a meaningful comparison with the known amino acid sequence of the plasmid JR66 (type I) enzyme.

c. and to determine the correlation, if any, between CAT variants, typed by electrophoretic mobility, immunological studies etc., with their N-terminal sequences.

It was hoped that such a study would bring the knowledge of staphylococcal CAT variants more in line with that of their R-factor counterparts, increase our understanding of the mechanism of catalysis and, by comparing CAT variants from a wider variety of organisms, determine if the CAT proteins are homologous in primary and (by the use of predictive models) secondary structure. If homology exists within the CAT proteins a systematic study at the primary sequence level may indicate the extent of extrachromosomal gene transfer in microbial systems.

CHAPTER II.

MATERIALS AND METHODS.

A. Materials.

Chemicals.

Dithiothreitol, β -mercaptoethanol, chloramphenicol and chloramphenicol base were purchased from the Sigma Chemical Co., U.S.A., diethylaminoethyl and carboxymethyl cellulose from the Whatman Corporation, DL-threonorpseudo ephedrine HCl from Phase Separations Ltd., dansyl amino acids, Whatman chromatography paper, polyamide thin layer chromatography sheets and 6-amino hexanoic acid from BDH Chemicals Ltd., [^{14}C] iodoacetic acid and [^{14}C] iodoacetamide from the Radiochemical Centre, Amersham, Bucks., ion exchange resins from Locarte Scientific Instruments Co., London, methyldichloroacetate from Aldrich Chemicals Co., Milwaukee, U.S.A., Sephadex gels and Sepharose 4 B from Pharmacia (GB) Ltd., London, 1-cyclohexyl-3-(2-morpholinoethyl)-carbodiimide metho-p-toluenesulphonate from Fluka AG Switzerland, dimethylformamide, N-methylmorpholine, trifluoroacetic acid, triethylamine, 0.4 M dimethylallylamine buffer, pH 9.5, acetonitrile, phenylisothiocyanate, methylisothiocyanate, aminopolystyrene resin, aminopropyl glass (75 Å porosity), guanidine hydrochloride and 5,5' dithiobis-(2-nitrobenzoic acid) from Pierce Chemical Co., U.S.A., methanol, dichloroethane and pyridine from Rathburn Chemical Co., Peebleshire, Scotland, alumina thin layer chromatography sheets (containing fluorescent indicator) from Merck, Germany, chlormethylated bio-beads S-X1 from Bio-Rad, U.S.A. and diisothiocyanate from Eastman Chemicals, U.S.A. DL-threo-2-dichloroacetamido-p-nitrophenyl-1-hydroxypropane (3-deoxy CM) was synthesised by the method of Rebstock (64) from DL-threo-norpseudo ephedrine hydrochloride. Methyl dichloroacetate and ethyl acetate were redistilled before use. Diisothiocyanate, iodoacetic acid and iodoacetamide were recrystallised before use. All other chemicals and reagents were of the highest purity available and were used without further purification.

Enzymes.

Carboxypeptidases A and B were obtained from Sigma Chemical Co., α-chymotrypsin and deoxyribonuclease from B.D.H. Chemicals Ltd., lysostaphin from Schwarz-Mann and staphylococcal protease (V8) from Miles (Research Division) Ltd. Elastase was a gift from Dr. Shotton. Commercially available trypsin (B.D.H.) contains substantial amounts of other proteases and was purified by affinity chromatography as described in 'Methods'. The CAT variant from Haemophilus parainfluenzae and that specified by plasmid R429 in E. coli were purified in this laboratory by Mr. J. Keyte and Mr. L. Packman respectively.

Bacterial strains.

Original Species.	Present Host Strain.	Plasmid (if known)	C.A.T. type.	Ref.
Klebsiella sp.	*E. coli* W677	JR66	I	(24)
Klebsiella sp.	*E. coli* J53	R429	I	(25)
Shigella sp.	*E. coli* J53	S-a	II	(26)
Shigella flexneri	*E. coli* J53	R387	III	(25)
Haemophilus Parainfluenzae	*E. coli* J53	'CT'	II	(27)
Streptomyces acrimycini	*S. acrimycini* 2236	?	?	(28)
'Flavobacterium'		?	?	(29)
S. aureus	*S. aureus*	? (curable)	A	(16)
S. epidermidis	*S. epidermidis* 39NC	? (curable)	B	(16)
S. aureus	*S. aureus* 8325	pC221	C	(17)
S. aureus	*S. aureus*	? (curable)	D	(16)

B. Methods.

2.1 Culture Methods.

In view of the high rate of loss of plasmid-linked chloramphenicol resistance (30), all strains were frequently subcultured on Penassay agar containing 50μg/ml chloramphenicol to maintain plasmid-rich stock cultures. Pilot experiments were carried out with bacteria grown on Penassay nutrient broth at 37°C on a rotary shaker. The large scale preparation of CAT involved the growth of S. aureus strain C221 in 300 litres of nutrient broth at the pilot plant facility of the Microbiological Research Establishment, Porton Down, under the supervision of Dr. Atkinson. Inducible strains were induced by the inclusion of 0.02mM D,L-threo-3-deoxy CM in the growth medium

2.2 Protein Determination.

Protein concentrations were determined by the method of Lowry et al (31). When amino acid analyses were available protein concentrations of pure enzymes were determined by amino acid analysis or extinction at 280 nM (1mg. C221 CAT/ml. 50mM Tris, pH 7.8, has an absorbance of 1.1 O.D. units).

2.3 Enzyme Detection. CAT Assay.

The rate of chloramphenicol acetylation was determined by the spectrophotometric method (17). This method is based on the stoichiometric liberation of 1 mole of reduced Coenzyme A per mole of chloramphenicol acetylated and by the coupled reaction of reduced CoA with 5,5'-dithio-bis-2-nitrobenzoic acid which is determined at 412 nM (molar extinction coefficient 13,600).

2.4 Chromatography.

Affinity chromatography - trypsin purification.

Commercial trypsin preparations contain significant amounts of chymotrypsin and other proteases which cleave polypeptides at sites other than lysine or arginine residues. In order to remove these proteolytic

enzymes trypsin was routinely purified by absorption on to a sepharose matrix containing covalently bound soya bean trypsin inhibitor (S.T.I.).

Preparation of S.T.I. - sepharose.

100ml (settled volume) of washed Sepharose 4B was stirred with 100ml of water adjusted to pH 11.0 with 2M NaOH. 4g of finely ground cyanogen bromide was added and the pH maintained at 11.0 by the addition of 2M NaOH. Temperature was maintained at 20°C by the addition of ice. After 15 minutes the mixture was poured on to a sintered glass filter and washed with 2,000ml of ice cold water followed by 2,000ml of cold 0.5M $NaHCO_3$. The activated gel was resuspended in 100ml of ice cold 0.5M $NaHCO_3$ containing 260mg soya bean trypsin inhibitor and stirred slowly overnight at 4°C. The mixture was then transferred to a column and washed with 500ml of:

1. 0.5M $NaHCO_3$.
2. 1.0M NaCl.
3. 0.1M glycine, pH 9.0.
4. 0.5M sodium acetate, pH 3.0.
5. Water.
6. Starting buffer, 0.1M Tris HCl, pH 7.1+20mM $CaCl_2$.

Capacity of column = 1mg trypsin bound/ml bed volume.

Elution buffers.

Buffer I	100mM Tris HCl (pH 7.1) containing	20mM $CaCl_2$		
Buffer II	5mM sodium acetate (pH 4.5)	"	"	"
Buffer III	3mM HCl	"	"	"

Trypsin was dissolved in buffer I (5mg/ml) and applied to the S.T.I. column and unbound proteins eluted with buffer I. Chymotrypsin is eluted by buffer II and finally pure trypsin eluted by buffer III. Eluted trypsin concentration is determined from: 1mg/ml trypsin has $O.D._{280}$ of 1.4 O.D. units and is stored at -20°C in the elution buffer.

Preparation of CAT specific affinity resins.

Sepharose 4 B (100g) was suspended in 400ml of water and activated by cyanogen bromide (10g), the pH being maintained at 11.0 by the addition of 5M NaOH. After 15 minutes the activated sepharose was washed with water (2 litres) and resuspended in 100mM $NaHCO_3$ (400ml) and 50g of 6-aminohexanoic acid added. This was slowly stirred at 4°C for 24 hours and resulted in a resin containing 10 μ equivalents of carboxyl groups per ml bed volume of gel. The hexanoic acid substituted sepharose was washed with 100mM $NaHCO_3$, 50mM NaOH, 100mM acetic acid and finally water. The substituted sepharose was suspended in water and the required amount (see Table I) of the free amine of chloramphenicol (D threo 1-p-nitrophenyl-2-amino 1,3 hydroxypropane) was added together with 10g of a water soluble carbodiimide, pH 4.5 (HCl). This was slowly stirred at room temperature for 24 hours and finally washed with 50mM Tris HCl buffer, pH 7.8, containing 1M NaCl (2 litres). The resin was then poured into a column of the required size and equilibrated with 50mM Tris HCl buffer, pH 7.8. By varying the amount of 'chloramphenicol base' added to the hexanoic acid derivatised sepharose and the number of times the coupling reaction between the hexanoic acid derivatised sepharose and the 'chloramphenicol base' is performed, a resin containing 2-10 μ equivalents of ligand can be obtained. It is important to note that adding more than 5g of 'chloramphenicol base' under the conditions stated does not increase the amount of bound ligand. It seems likely that the hydrolysis of the water soluble carbodiimide is the limiting factor and to obtain resins containing more than 4-5 μ equivalents ligand/ml gel a second or third round of coupling (using an additional 10g carbodiimide each time) must be performed.

Table I.

Weight 'CM base' coupled.	Number of times coupled.	μeq. ligand/ml bed volume gel (10μeq. COOH/ml).
1.2g.	1	2-3
5.0g.	1	4-5
5.0g.	2	7-8
5.0g.	3	9-10

Carboxymethyl cellulose.

Carboxymethyl cellulose ion exchange columns equilibrated with 8 M urea/20mM ammonium acetate, pH 5.0, were used in the separation of large, insoluble peptides generated by cyanogen bromide cleavage of proteins (Chapter IX). Whatman carboxymethyl cellulose (CM 52) was suspended in the urea buffer and poured into a 1x25cm glass column. Prior to use the column was fully equilibrated by washing with the urea/ammonium acetate buffer (500ml).

Sulphonated polystyrene.

Digests which resulted in a large number of small peptides (tryptic, chymotryptic and elastase digests) were initially separated by absorption on to the top of a jacketed column (bed volume 30ml) of ~~Dowex 50~~ sulphonic acid ion-exchange resin (32)* pre-equilibrated in the pyridinium form with 0.1 M pyridine/acetate buffer, pH 3.1. The products were fractionated at 50°C with 500ml of pyridine/acetate gradient devised so that the rate of increase in pH was as linear as possible rising from pH 3.1 to pH 5.0.

* Dowex 50 type resin supplied by Locarte Scientific Instruments - 'Fine' resin, 5% crosslinked.

The column was finally washed with pyridine/acetate buffer, pH 6.5. Flow rate was 40ml/hour and 3ml fractions were collected.

Gel filtration.

Sepharose gels (Sephadex-Pharmacia) and polyacrylamide gels (Bio-Rad P-resins) were pre-swollen according to the manufacturers' instructions in the required running buffer, poured into a column of the required size and washed with 3 column volumes of running buffer prior to use. Samples were dissolved in a minimum volume (<3 percent of the column volume) of the eluting buffer and carefully layered on to the top of the column. The samples were allowed to run into the top of the column, the remaining space at the top of the column filled with running buffer and the column was then connected to a reservoir of the running buffer.

2.5 Electrophoresis.

High voltage paper electrophoresis.

Peptide solutions were applied as 1cm bands to Whatman chromatography paper, dried in a current of air and sharpened to the origin by careful application of buffer before electrophoresis. Internal fluorescent markers (33), dyes and amino acid mixtures as side markers (34) were used. Electrophoresis was performed at pH 6.5, pH 3.5 or pH 2.1. (86).

Electrophoretic mobilities.

The electrophoretic mobilities of peptides at pH 6.5 were measured from the neutral band to allow for electroendosmotic flow and at pH 2.1 from the origin. Mobilities at pH 6.5 were calculated relative to free aspartic acid (basic peptides having a negative value) and mobilities at pH 2.1 being calculated relative to free serine. These mobilities were used to determine the correct allocation of amide groups in peptides containing asparagine and glutamine (35).

Polyacrylamide gels.

Purification of proteins and large peptides was monitored by S.D.S. polyacrylamide (11.5 percent) slab gel electrophoresis using the buffer system of Laemmli (36). Proteins and peptides were detected by staining with Coomassie blue. Polyacrylamide (7.5 percent) gels of native proteins were run in buffer, pH 8.8. as used for S.D.S. gels (36) without inclusion of S.D.S.

2.6 Detection of peptides.

a. Columns.

Gel filtration and carboxymethyl cellulose columns were connected to a continuous flow recording spectrophotometer and the eluate from the column monitored at 280nM. If the eluting buffer had a low absorption at 230nM the collected fractions were manually monitored at 230nM. Samples (20 μl) from each fraction were taken for counting when digests were performed on radioactively derivatised proteins. Samples (50 μl) were removed from each fraction collected from ~~Dowex 50~~ sulphonic acid resin columns and applied as contiguous bands along the origin of sheets of Whatman No. 1 chromatography paper and fractionated by high voltage paper electrophoresis at pH 6.5 to give peptide 'fingerprints' from which the location of each peptide in the eluate could be seen. From the distribution of peptides a rational pooling of fractions could be made.

b. Paper.

After each electrophoretic run the paper was dried in a current of warm air and then examined under U.V. light to detect the positions of the fluorescent internal markers and fluorescent bands that accompany tryptophan containing peptides. The whole sheet or, in preparative runs, guide strips cut from the edge of the sheet, were then stained with cadmium-ninhydrin reagent and dried in a current of warm air until the initial colours of the

peptide spots developed. The papers were then stored overnight in polythene bags at room temperature and any changes in colour noted, the colours giving some indication of the N-terminal residue of the peptide (glycine or threonine - yellow, serine or asparagine - orange-yellow, valine or isoleucine - red, developing late.) The specific colour tests for peptides containing tryptophan (37) and arginine (38) were also used where appropriate, after initial staining with the fluorescent stain 4-phenylspiro [furan-2(3H)-1-phthalan]-3,3' dione (fluorescamine). Radioactively labelled peptides were detected by autoradiography overnight on Kodak Kodirex KD5T X-ray film.

2.7 Amino acid analyses.

Native enzyme was hydrolysed in 6N HCl containing 10 nmoles/ml phenol to prevent loss of tyrosine for 24, 48 and 72 hours at 105°C in evacuated Pyrex tubes. The chromatograms obtained were integrated manually and after normalisation of the three sets of data, a mean was taken for each of the amino acids except threonine, serine, valine and isoleucine. Threonine and serine values were extrapolated back to zero time and valine and isoleucine 72 hour hydrolysis values (which were consistently greater than the 24 and 48 hour values) used in calculating the amino acid composition of proteins. Performic acid oxidised and carboxymethylated enzymes were hydrolysed under identical conditions for 48 hours. Peptides purified from digests were also hydrolysed under identical conditions for 24 hours and are expressed uncorrected for any losses. Amino acid analyses were performed on a single column amino acid analyser obtained from Locarte Scientific Instruments Co. Tryptophan was determined by the method of Liu and Chang (39) using 4 M methane sulphonic acid.

2.8 Sequence analysis.

Endgroup determination.

Excluding automated methods, amino terminal amino acids were identified by the method of Gray and Hartley (40), the dansyl derivative being identified by chromatography on polyamide thin layer plates (41). C-terminal amino acids were determined by digestion with carboxypeptidases A and B followed by amino acid analysis of samples of the digest at set time intervals.

Dansyl-Edman.

Dansyl-Edman sequence determinations were performed as described by Gray and Hartley (40), the dansyl derivatives of amino acids being identified by the method of Woods and Wang (41). The criteria used in judging sequence results were those described by Ambler (42).

2.9 Automated sequence analysis.

a. Liquid phase.

The N-terminal sequence of CAT from a strain of S. aureus containing the plasmid C221 (type C CAT) was carried out on a Beckman 890 Sequencer by Dr. J. Bridgen at the M.R.C. Laboratory of Molecular Biology, Cambridge.

b. Solid phase.

All solid phase sequencing was performed on an Anachem APS 2400 solid phase sequencer in this laboratory.

Solid phase supports.

With the exception of triethylenetetramine polystyrene (TETA) resin all solid phase supports were purchased from Pierce Chemicals Ltd. since it was found that these products had a greater capacity and gave fewer non PTH contaminants than supports produced by the author. Various supports are recommended in the literature (43), the choice depending on the size of the protein or peptide to be sequenced and the functional groups through which

the protein or peptide is coupled to the solid support. In this study the following supports were used:

Table 2.

Support.	Coupling Method.	Ref.	Protein/Peptide.
Aminopropyl glass 75 Å porosity	Diisothiocyanate	(44)	CAT proteins. Large (>30 residues) lysine containing peptides.
N-(2-aminoethyl)-3-aminopropyl glass 75 Å porosity	via homoserine lactone	(45)	Large (>30 residues) CNBr peptides containing C-terminal homoserine.
TETA polystyrene	via homoserine lactone	(46)	Small (<30 residues) CNBr peptides containing C-terminal homoserine.
Amino polystyrene	Diisothiocyanate	(47)	Small (<30 residues) tryptic peptides containing C-terminal lysine.
Amino polystyrene	Carbodiimide	(48)	Small peptides (<30 residues) containing neither C-terminal lysine nor homoserine.

Coupling methods.

Peptide coupling to solid phase supports was achieved by published methods (43, 49).

CAT proteins (underivatised, carboxymethylated or performic acid oxidised) were coupled to aminopropyl glass by the following method:

Diisothiocyanate derivatised aminopropyl glass was prepared by

suspending 1g diisothiocyanate (DITC) in 5ml dimethylformamide (DMF). This was slowly stirred at room temperature and 100mg aliquots of aminopropyl glass were added at 5 minute intervals, the reaction vessel being flushed with nitrogen after each addition. After a total of 1g aminopropyl glass had been added the vessel was stirred at room temperature for an additional 75 minutes. The resulting derivatised glass was washed with 3x5ml DMF, 1x5ml methanol, 1x5ml water, 1x5ml methanol and finally dried under vacuum. Pure, salt-free, freeze dried CAT (2 to 5mg) was dissolved in 1 percent triethylamine and freeze dried overnight to remove all traces of ammonia. The resulting white powder was dissolved in 0.5ml of 10 percent triethylamine/trifluoroacetic acid, pH 9.5, saturated with guanidine hydrochloride and slowly stirred at 45°C. 50mg aliquots of DITC aminopropyl glass were added at 5 minute intervals, the reaction vessel being flushed with nitrogen after each addition until a total of 200mg of DITC aminopropyl glass had been added. The coupling reaction was allowed to proceed, with slow stirring, at 45°C for an additional 30 minutes. Excess DITC groups were blocked by the addition of 100μl ethanolamine and stirring continued at 45°C for a further 30 minutes. The resulting coupled protein/glass was washed with 2x5ml methanol and dried under vacuum.

Assay of solid phase supports.

2 percent by weight of the support was washed with 1ml trifluoroacetic acid to remove any non-covalently bound peptide or protein and sealed, under vacuum, in a thick-walled Pyrex hydroysis tube containing a 1:1 mixture (200μl) of propionic and concentrated hydrochloric acids. This was heated to 160°C for 20 minutes. The tube was then opened, the acids removed under vacuum and the dried residue loaded on to an amino acid analyser. This rapid hydrolysis procedure gives approximately 90 percent recovery of

the stable amino acids and so, by comparison with the known amino acid content of the peptide or protein, the total amount of material coupled can be easily calculated.

PTH identification.

Fractions from the solid phase sequencer were dried in a stream of nitrogen. 1 M HCl (200μl) was added to each tube and heated to 80°C for 10 minutes to convert the anilinothiazolines to the stable phenylthiodantoin amino acids. The fractions were then extracted with 1ml ethyl acetate and the upper organic phase evaporated under a stream of nitrogen in small vials. Using this procedure the PTH-derivatives of histidine, arginine and, where present, cysteic acid remain in the aqueous phase. These residues are identified by freeze drying the aqueous phase in a thick-walled pyrex tube, adding 200μl of 55 percent HI, sealing the tubes under vacuum and hydrolysing the PTH amino acids to the free amino acids overnight at 130°C which can then be quantitated on the amino acid analyser (50).

All other PTH amino acids, which are extracted by the ethyl acetate, are identified by chromatography on silica plates (50) using a 2 solvent chromatographic system (solvent 1, chloroform:ethanol; 98:2, solvent 2, chloroform:ethanol:methanol; 90:2:8, by volume).

2.10 Photo-oxidation.

Photo-oxidation was performed at 6°C with a 150 watt Crompton spot-light at a distance of 10cm from a water-jacketed conical tube containing the enzyme solution (potassium phosphate buffers, pH 5.7 to pH 7.0, and Tris HCl buffers, pH 7.3 to pH 9.0) and photosensitive dye (methylene blue). At intervals samples were withdrawn and immediately diluted ten-fold with 100mM Tris buffer, pH 8.0, in small glass tubes covered with aluminium foil to prevent further reaction. The decrease in enzyme activity was taken as a

measure of photo-oxidation. The dye, at a final concentration of 3μM, was approximately equimolar with the protein subunits.

2.11 Treatment with specific chemical reagents.

Diethylpyrocarbonate (DEPC).

Purified CAT (0.2mg/ml) in potassium phosphate buffer, pH 6.0, was made 0.1 mM with respect to DEPC and maintained at 0°C. Enzymes were assayed for activity after 10 minutes. Substrate protection was investigated by prior incubation of substrate (1 mM) with the enzymes for 10 minutes at room temperature.

Fluorodinitrobenzene (FDNB).

N-ethyl maleimide (NEM).

5,5' Dithiobis-(2-nitrobenzoic acid) (DTNB).

Nitrothiocyano benzoic acid (NTCB).

2,2' Dithiopyridine (DTDP).

p-Chloromercuribenzoic acid (PCMB).

CAT (0.2mg/ml) was incubated at 37°C with the test reagent (1 mM) in 50 mM Tris buffer, pH 7.8, and samples taken at intervals for assay. Substrate protection was investigated by prior incubation of substrate (1 mM) with the enzymes for 10 minutes at room temperature.

Bromoacetyl Coenzyme A (BrAcCoA).

BrAcCoA was prepared by the method of Chase and Tubbs (51). This preparation was found to inactivate pig heart citrate synthase and was considered to be contaminated with low molecular weight brominated precursors from the synthesis of the BrAcCoA and was further purified by gel filtration on a Sephadex G-10 column (1x20cm), BrAcCoA being eluted in the void volume of the column.

CAT (0.2mg/ml) was incubated at 37°C with BrAcCoA (0.05 mM) in 50 mM

Tris buffer, pH 7.8, and samples taken at intervals for assay. Substrate protection was investigated by prior incubation of substrate (1 mM) with the enzymes for 10 minutes at room temperature.

Iodoacetic acid (IAA).

Iodoacetamide (INH_2)

CAT (0.2mg/ml) was incubated at 37°C with IAA or INH_2 (5 mM) in 50 mM Tris buffer, pH 7.8, and samples taken at intervals for assay. Substrate protection was investigated by prior incubation of substrate (1 mM) with the enzymes for 10 minutes at room temperature.

In all inactivation experiments the appropriate enzyme control experiments (identical conditions minus test reagent) were performed. Modes of action of DTNB, NTCB, INH_2, IAA, DEPC and FDNB are illustrated in appendix B.

CHAPTER III.

ISOLATION AND CHARACTERISATION OF ENZYMES.

3.1 Purification of CAT enzymes.

Staphylococcal strains producing CAT types A, B, C and D were grown up and crude extracts prepared as follows:

15 litre batches of Penassay broth (containing 0.02 mM DL,deoxy CM) were inoculated with one litre of an overnight culture grown in the presence of CM (20µg/ml).The bacteria were grown on a rotary shaker at 37°C in 2 litre flasks. Cells were harvested in early stationary phase by centrifugation at 10,000 RPM for 20 minutes at 4°C. The cell paste was washed with 50 mM Tris HCl buffer, pH 7.8, containing 0.05 mM CM and resuspended in 500ml of the same buffer for cell disruption by lysostaphin (5 units/ml) and DNAase (50µg/ml) at 37°C. Lysis was complete in approximately 1½ hours as judged by the decrease in O.D. at 660 nM. The crude lysates were clarified by centrifugation and the supernatants removed and assayed for enzyme activity. Purification results for the four enzyme types are summarised in tables 3, 4, 5 and 6.

Initial studies on the type C enzyme (plasmid C221) had shown that the enzyme was stable to heating at 70°C for 10 minutes in the CM containing buffer. This stability was shared by the type A variant of CAT but appreciable losses occurred (approximately 30 percent) for the B and D enzymes and so the heat step was omitted in the purification of these variants.

Although several CAT enzymes have been purified by conventional methods (12, 17) R-factor CAT variants I, II and III have been successfully purified by affinity chromatography (19) and this technique was used for the final purification stage of the staphylococcal enzymes.

Preliminary experiments were performed using small (1ml bed volume) affinity columns containing resins of increasing substitution (see 'methods' 2.4) in order to determine the minimum ligand substitution necessary to bind

Table 3 Purification of type A staphylococcal CAT.

Preparation	Volume (ml)	Total units.	Total protein(mg)	Specific activity.	Percent. recovery.
Crude extract	430	4,480	10,148	0.44	100
Heat step	400	4,000	2,720	1.47	89
Affinity column	180	2,400	46.8	51	55

(from 46g wet weight cells).

Table 4 Purification of type B staphylococcal CAT.

Preparation	Volume (ml)	Total units.	Total protein(mg)	Specific activity.	Percent. recovery.
Crude extract	260	1,644	2,678	0.61	100
Affinity column	320	1,440	55	26	87

(from 24g wet weight cells).

Table 5 Purification of type C staphylococcal CAT.

Preparation	Volume (ml)	Total units.	Total protein(mg)	Specific activity.	Percent. recovery.
Crude extract	400	635	5,200	0.12	100
Heat step	380	600	2,500	0.24	95
Affinity column	72	420	14	30	67

(from 37g wet weight cells).

Table 6 Purification of type D staphylococcal CAT.

Preparation	Volume (ml)	Total units.	Total protein(mg)	Specific activity.	Percent. recovery.
Crude extract	435	1,535	1,226	0.13	100
Affinity column	66	1,010	26.4	38	66

(from 42g wet weight cells).

the CAT variant. When a resin capable of binding the CAT variant under study had been found the column was washed with buffers of increasing NaCl concentration $\pm$ CM (5mM) until an NaCl concentration was found which did not elute the CAT in the absence of the substrate CM but which eluted the CAT in a small volume when the substrate was present.

Considerable differences were found between the four staphylococcal variants as regards the extent of resin substitution and eluting buffer required to bind and elute the enzymes (table 7).

Table 7 Affinity resin substitution and elution buffer molarity required for staphylococcal CAT purification.

CAT type.	Resin substitution μeq. /ml.	NaCl concentration of eluting buffer (50 mM Tris HCl, pH 7.8, 5 mM CM).
A	7	0.6 M
B	9	0.6 M
C	2.5	0.3 M
D	9	0.6 M

This elution pattern was not due to the ionic concentration or composition of the crude extracts since both crude extracts and purified enzymes behaved in a similar manner after prior dialysis against 50 mM Tris HCl, pH 7.8.

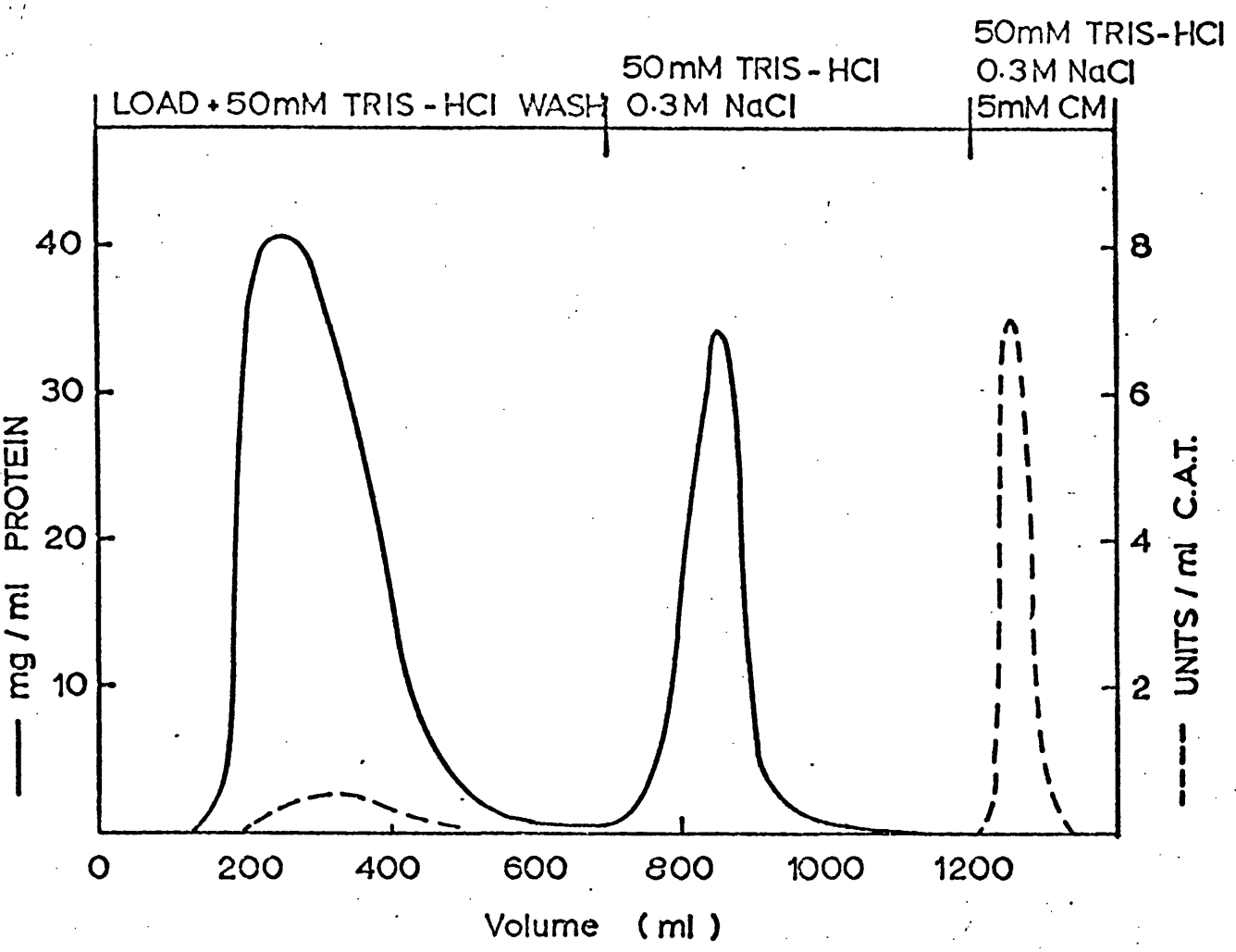
A typical elution profile from an affinity column is shown in figure 2. In all cases up to 20 percent of the enzyme loaded on to the column eluted in the Tris HCl buffer wash. This 'leakage' could be minimised by dialysing out the CM (0.05 mM) present in the crude extract but, since 10 percent of the enzyme activity was lost on dialysis against Tris HCl buffer and the eluted enzyme can be concentrated and added to stocks of crude enzyme awaiting purification, this alternative procedure was not generally followed. Indeed, by 'recycling' non-specifically eluted enzyme overall yields of 90 percent can be obtained.

All purification steps were monitored by S.D.S. gel electrophoresis (plate 1). Purified staphylococcal variants have identical mobilities on S.D.S. gel electrophoresis and this mobility is identical with that of the type I R-factor enzyme which has a known molecular weight of 24,000 (plate I).

Polyacrylamide gel electrophoresis of the native enzymes (plate 2) clearly shows the variation in mobility used in typing the staphylococcal enzymes, the type A variant being the slowest moving, through B and C to the type D enzyme which moves furthest towards the anode.

Storage of CAT enzymes at -20°C in the eluting buffer of the affinity column slowly inactivates the enzymes, clearly seen in the case of the R-factor type I enzyme run as a reference marker with the staphylococcal enzymes in plate 2 and for this reason purified enzymes were generally dialysed against 50 mM Tris HCl buffer containing 0.1 mM CM prior to freezing.

In addition to the purification of the four staphylococcal variants, 2mg of an R-factor type II enzyme (plasmid S-a) was purified from 11g (wet weight) of cells by the method of Zaidenzaig and Shaw (19) using an affinity resin of 2.5 μequ./ml substitution and an elution buffer containing



Figure(2).

Affinity column elution profile

Affinity chromatography of a heated extract of type C (C221) C.A.T. obtained from 37g (wet wgt.) cells. All buffers are pH 7.8, flow rate =200ml/hr. Affinity resin contained 2.5 μEq. CM bound/ml bed volume. Column size = 2.5 by 10cm.

Plate: 1.

S.D.S. polyacrylamide gel electrophoresis.

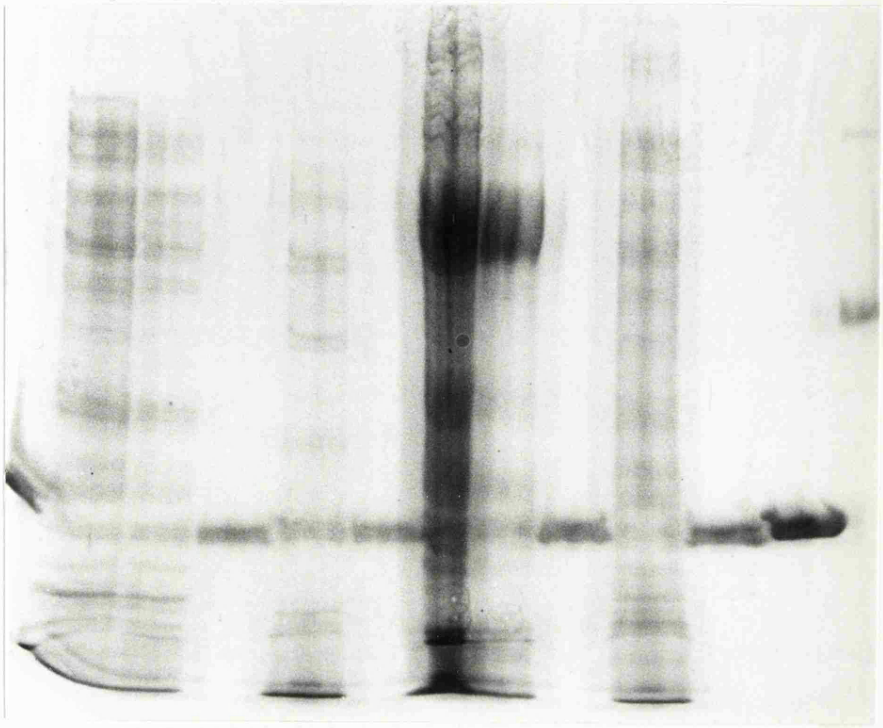
Purification of CAT variants A, B, C and D.

1. Crude extract type A.
2. Heat step type A.
3. Pure enzyme type A.
4. Crude extract type B.
5. Pure enzyme type B.
6. Crude extract type C.
7. Heat step type C.
8. Pure enzyme type C.
9. Crude extract type D.
10. Pure enzyme type D.
11. Pure R-factor type I enzyme (JR66).

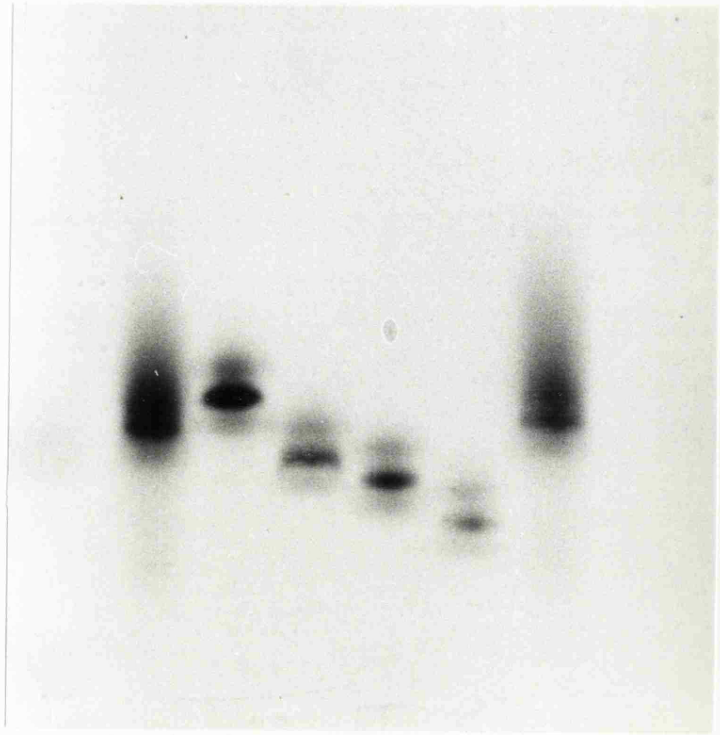
Plate: 2.

Polyacrylamide gel electrophoresis of native CAT enzymes.

1. Pure R-factor type I enzyme (JR66).
2. Pure type A enzyme.
3. Pure type B enzyme.
4. Pure type C enzyme.
5. Pure type D enzyme.
6. Pure R-factor type I enzyme (JR66).



1 2 3 4 5 6 7 8 9 10 11



1 2 3 4 5 6

0.6 M NaCl and 5 mM CM. This protein was used for the N-terminal sequence determination of a type II R-factor CAT (Chapter IV).

For the sequence studies of the type C variant 1.3 Kg. of cells were obtained from growth in a 400 L fermentor at the Microbiological Research Establishment, Porton. This was treated in exactly the same manner as the small scale purification of the type C variant to yield 10 litres of crude extract containing 38,000 units of enzyme (1.25g). This was purified in 200mg amounts using a 100ml (bed volume) affinity column as required.

3.2 Km determinations.

Michaelis constants for CAT variants A, B, C and D were determined using the spectrophotometric assay described in 'Methods' (2.3). The results are illustrated graphically in figures 3, 4, 5 and 6. Km values for the four enzymes under saturating (100μM CM, 200μM AcCoA) second substrate conditions are tabulated in table 8.

In addition to double reciprocal plots (Lineweaver-Burk) of velocity versus first substrate concentration under saturating conditions of second substrate, double reciprocal plots of velocity versus first substrate concentration, where the second substrate was present at a concentration near to the Km for that substrate (5μM in the case of CM and 50μM for AcCoA) were also plotted (figures 3, 4, 5 and 6). Intersecting lines were produced, the intercept being on the $\frac{1}{[S]}$ axis.

As seen with an enteric type I CAT (20) the staphylococcal variants are capable of deacylating AcCoA in the absence of the acceptor CM, the acyl acceptor probably being water to form acetate. The specific activity for this reaction is $<$0.1 percent (0.018μmoles/min./mg.) compared with the specific activity in the presence of CM.

figures: 3, 4, 5 and 6.

Double reciprocal plots of the dependence of S. aureus CAT (types A, B, C and D) activity on substrate concentration.

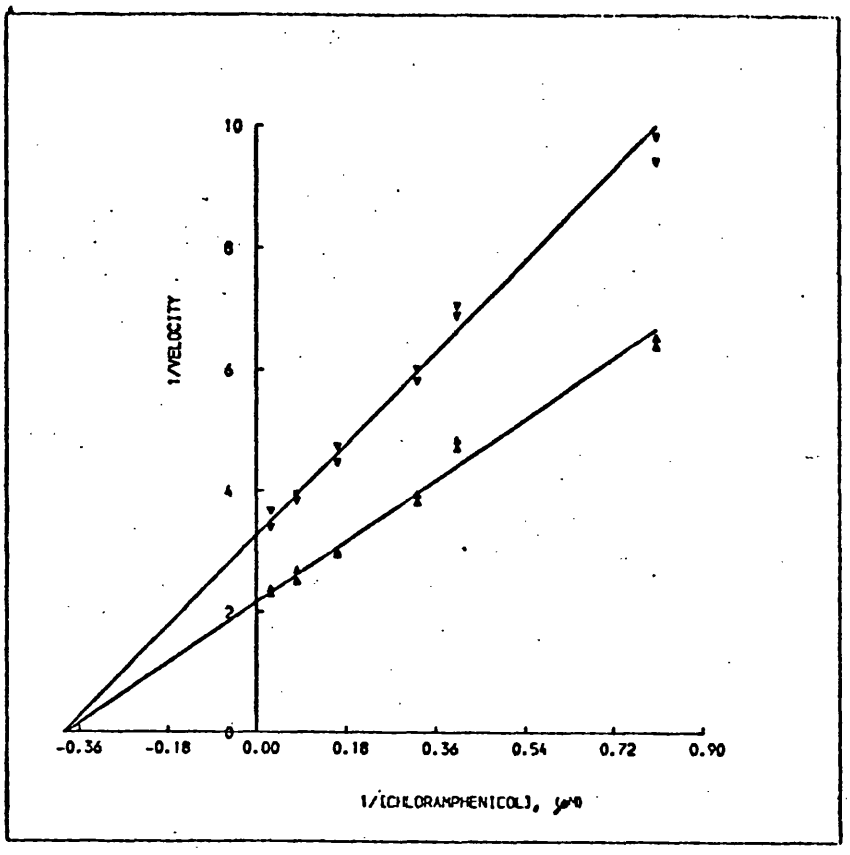
1/v expressed in arbitrary units.

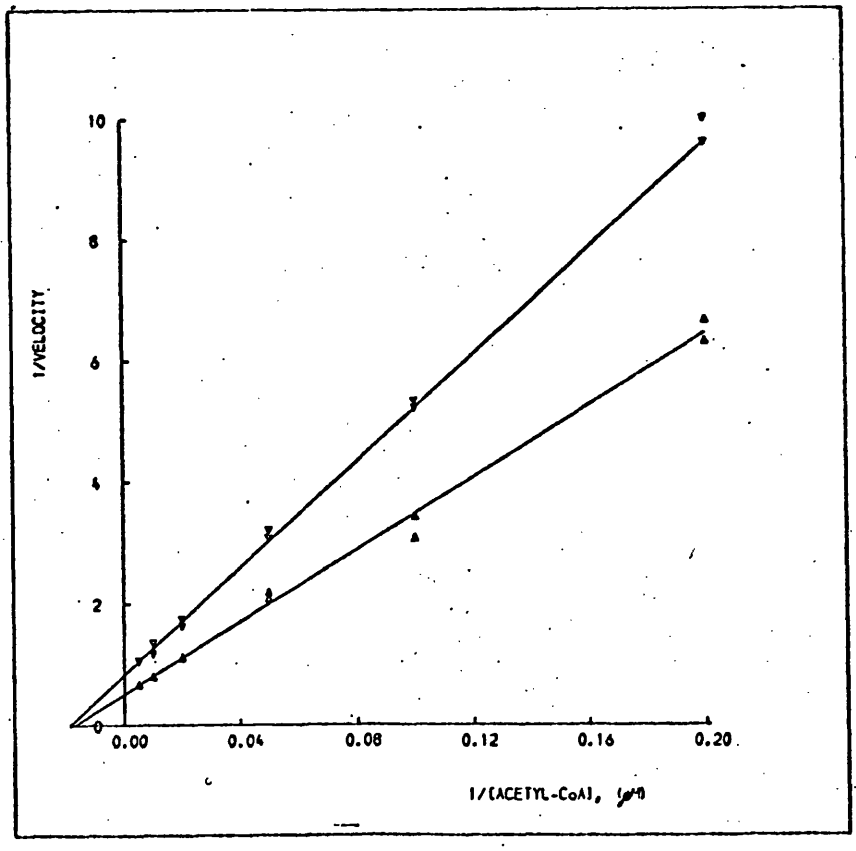
1/[substrate] expressed as μM^{-1}.

Δ = high second substrate concentration (CM = 100μM, AcCoA = 200μM).

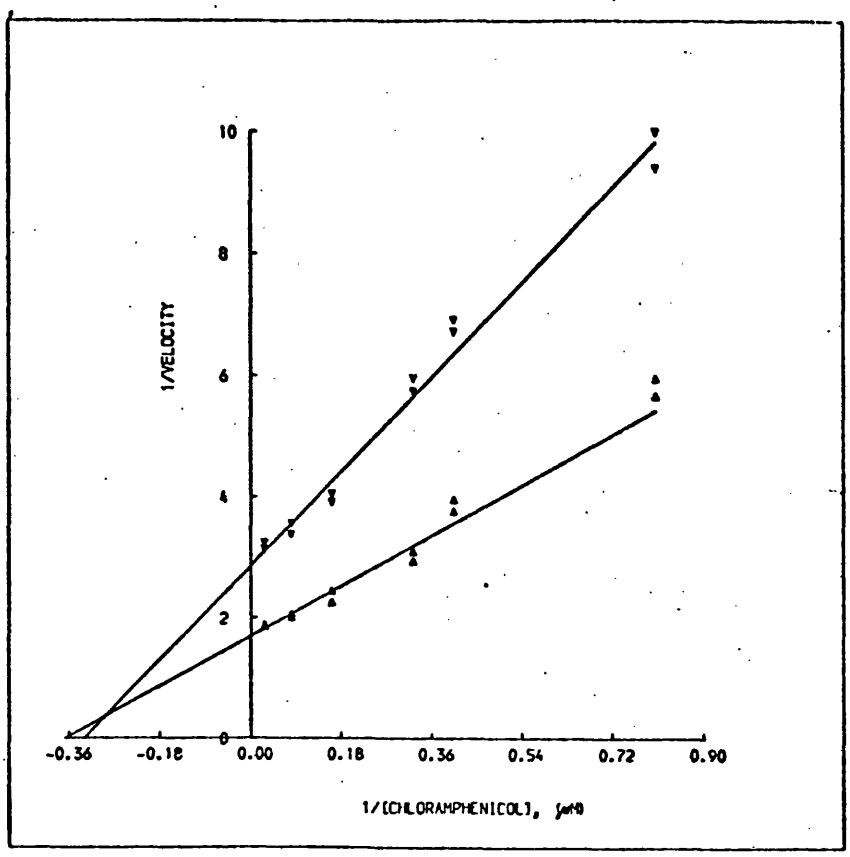
∇ = low second substrate concentration (CM = 5μM, AcCoA = 50μM).

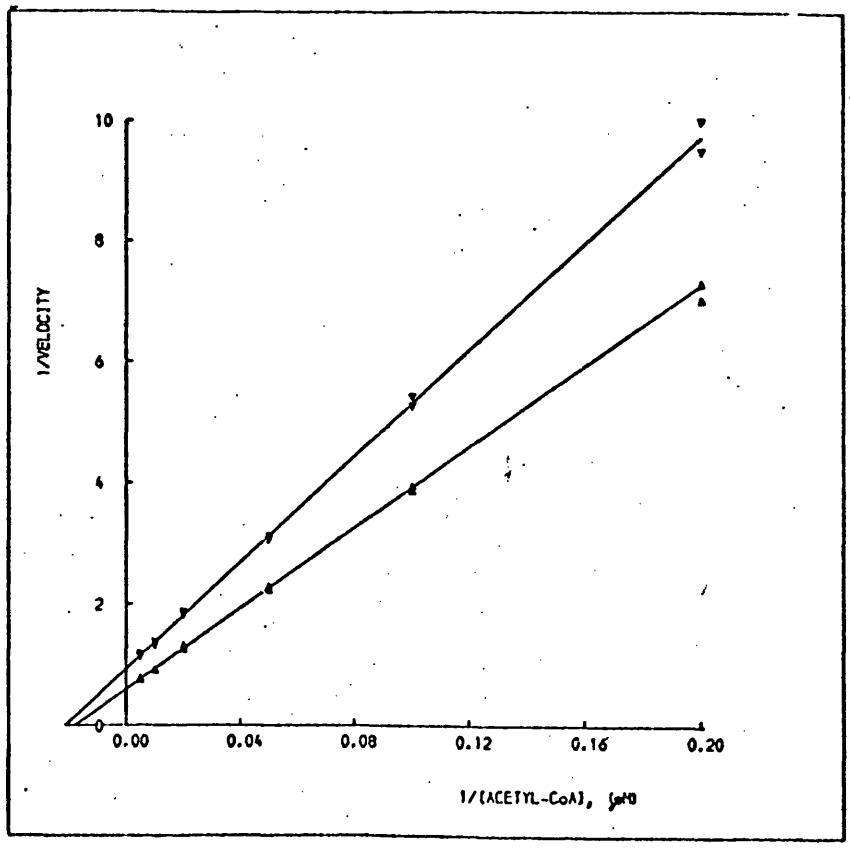
TYPE A C.A.T.


10
8
6
4
2
0
1/VELOCITY
-0.36
-0.18
0.00
0.18
0.36
0.54
0.72
0.90
1/[CHLORAMPHENICOL], μM⁻¹

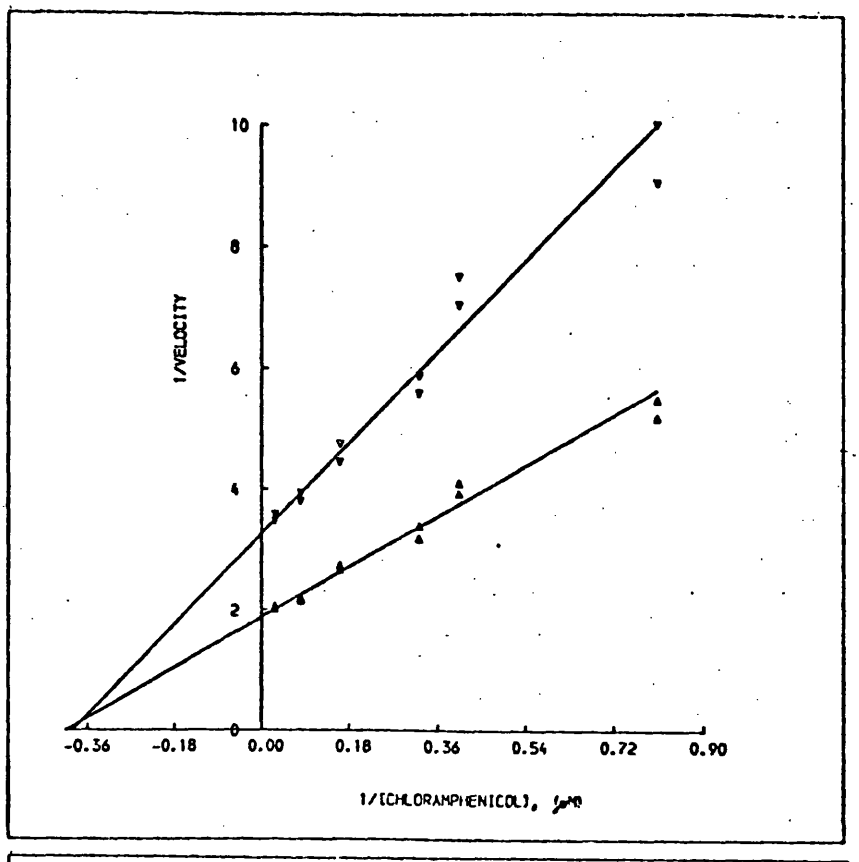

10
8
6
4
2
0
1/VELOCITY
0.00
0.04
0.08
0.12
0.16
0.20
1/[ACETYL-CoA], μM⁻¹

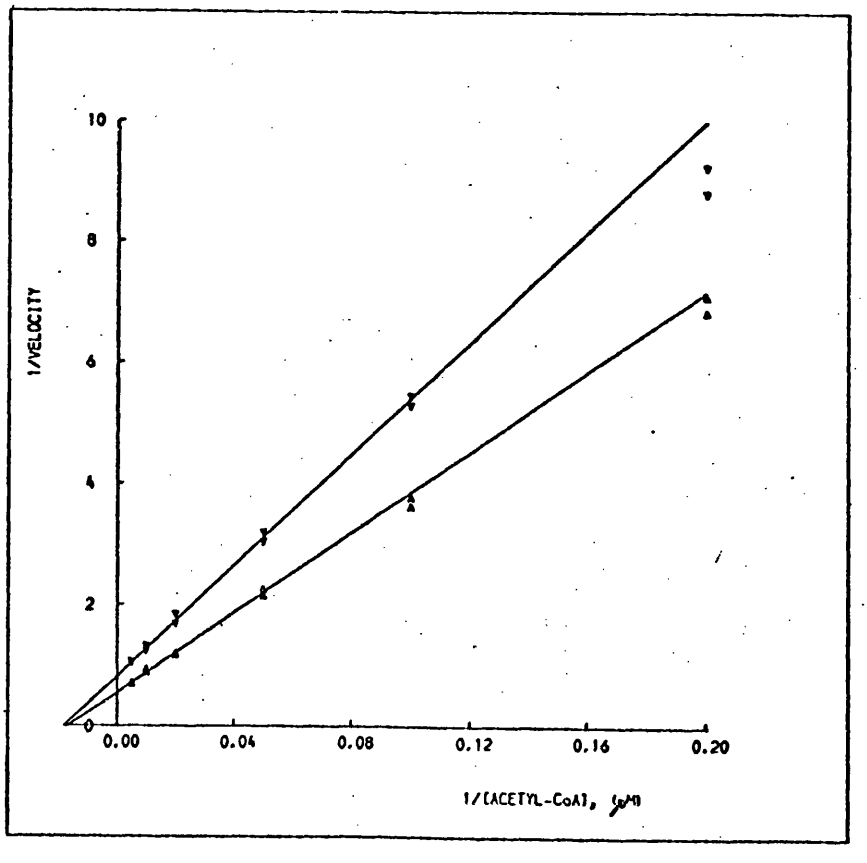
TYPE B C.A.T.


1/VELOCITY
10
8
6
4
2
0
-0.36
-0.18
0.00
0.18
0.36
0.54
0.72
0.90
1/[CHLORAMPHENICOL], μM⁻¹

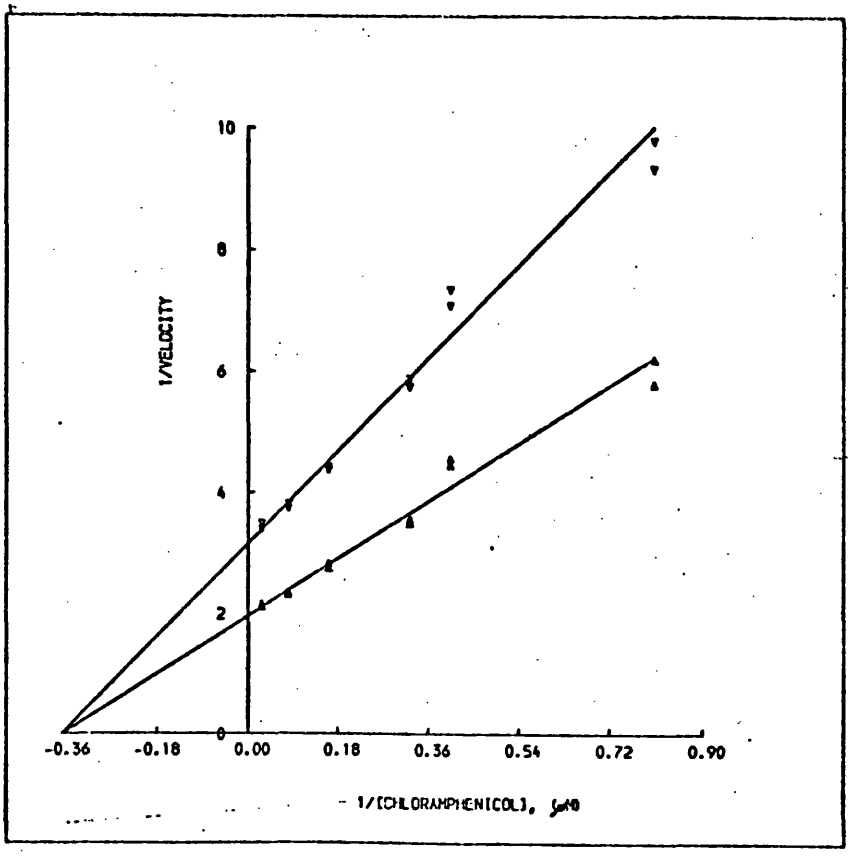

1/VELOCITY
10
8
6
4
2
0
0.00
0.04
0.08
0.12
0.16
0.20
1/[ACETYL-CoA], μM⁻¹

TYPE C C.A.T.


10
8
6
4
2
0
1/VELOCITY
-0.36
-0.18
0.00
0.18
0.36
0.54
0.72
0.90
1/[CHLORAMPHENICOL], μM


10
8
6
4
2
0
1/VELOCITY
0.00
0.04
0.08
0.12
0.16
0.20
1/[ACETYL-CoA], μM

TYPE D C.A.T.


10
8
6
4
2
0
1/VELOCITY
-0.36
-0.18
0.00
0.18
0.36
0.54
0.72
0.90
- 1/[CHLORAMPHENICOL], μM

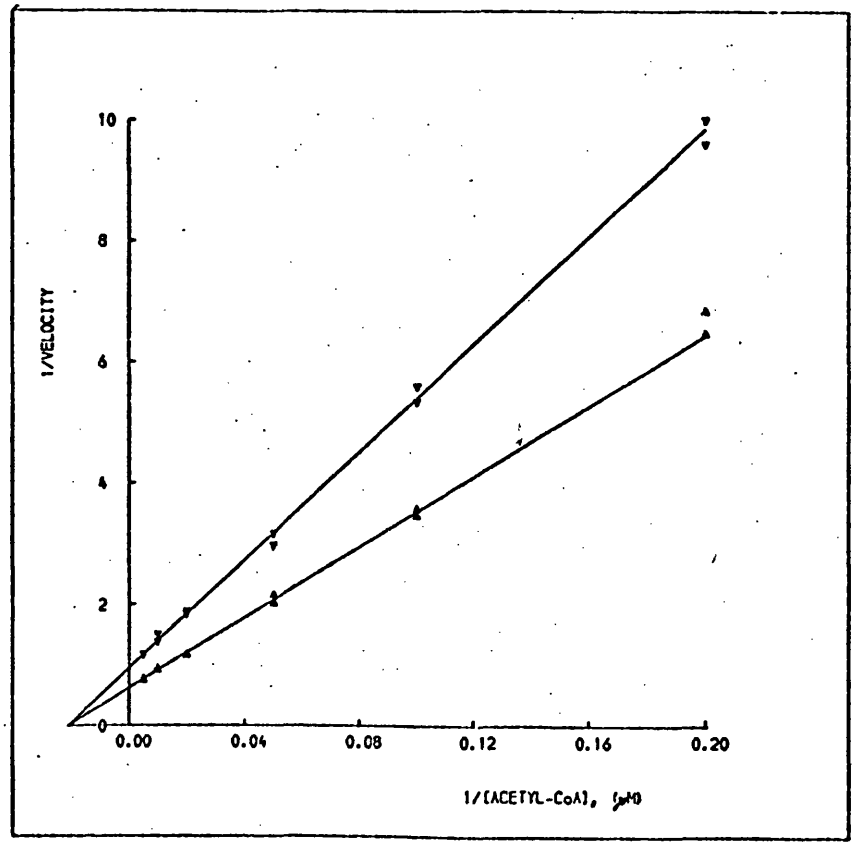

10
8
6
4
2
0
1/VELOCITY
0.00
0.04
0.08
0.12
0.16
0.20
1/[ACETYL-CoA], μM

Table 8 Km values determined for CAT variants A, B, C and D.

Enzyme variant.	Variable substrate.	Km (μM).	Standard error.
A	CM	2.58	± 0.16
	AcCoA	57.3	± 2.5
B	CM	2.72	± 0.22
	AcCoA	56.1	± 1.9
C	CM	2.47	± 0.17
	AcCoA	60.9	± 3.5
D	CM	2.70	± 0.17
	AcCoA	46.5	± 1.6

3.3 Modification studies.

Chemical.

In view of the sensitivity of R-factor type II CAT variants to the thiol reagent DTNB (15) serious consideration was given to the possibility that cysteine residue(s) may be involved in the catalytic mechanism. To determine if an analogous situation exists within the staphylococcal variants a number of potential thiol reagents were tested for their inactivating ability as outlined in 'Methods' (2.11).

Rate constants were derived by plotting loss of activity with time semilogarithmically (figure 7) and are tabulated in table 9.

Since all variants gave identical rates of inactivation ($\pm$ 3percent) mean values only are presented. Rate constants for the inactivation were calculated from the equation:

$E/Eo = -Kt$, where E/Eo is the fraction of the initial enzyme activity remaining after treatment with the inhibitor after time t (53).

Table 9 Mean rate constants for the inactivation of the four staphylococcal CAT variants.

Reagent	Concentration of reagent.	Rate constant min^{-1} $x10^{3}$.	Percent. protection by CM.	Percent. protection by AcCoA.
FDNB	1.0 mM	51	> 95	< 5
BrAcCoA	0.05 mM	240	80	85
NEM	1.0 mM	9.1	> 95	< 5
INH_2	5.0 mM	7.6	> 95	< 5
PCMB	1.0 mM	3.7	< 5	< 5
IAA	5.0 mM	< 0.3	-	-
DTNB	1.0 mM	< 0.3	-	-
NTCB	1.0 mM	< 0.3	-	-
DTDP	1.0 mM	< 0.3	-	-
Controls	-	< 0.3	-	-

Since DTNB, NTCB and DTDP react with thiol groups via disulphide exchange reactions these reagents might be expected to be more specific than the other reagents used, which are capable of modifying residues other than

cysteine (54), and it is interesting to note that the inactivation rates using these specific reagents were equal to the slow rate of inactivation found with the control samples.

It is also notable that iodoacetamide (INH_2) inactivated the CAT enzymes whereas the closely related compound iodoacetic acid (IAA) does not. Inactivation by INH_2 at pH values of 5.5 to 7.8 indicated an apparent pK of 6.3 for the inactivation process.

By far the most potent inactivating reagent was the substrate analogue bromoacetyl CoA and the CAT enzymes were protected from inactivation by both CM and acetyl CoA.

FDNB, originally used in the N-terminal determination of proteins, also proved to be an effective inactivating agent. FDNB not only reacts with α-amino groups but also thiol, imidazole and ε-amino groups of proteins. As FDNB is a relatively polar reagent and has a nitrated benzene ring, as has chloramphenicol, it is a potentially useful probe of the CAT active site.

With the exception of para-chloromercuribenzoate, where protection was not seen with either substrate, CM affords protection against inactivation whereas acetyl CoA generally has little effect (table 9).

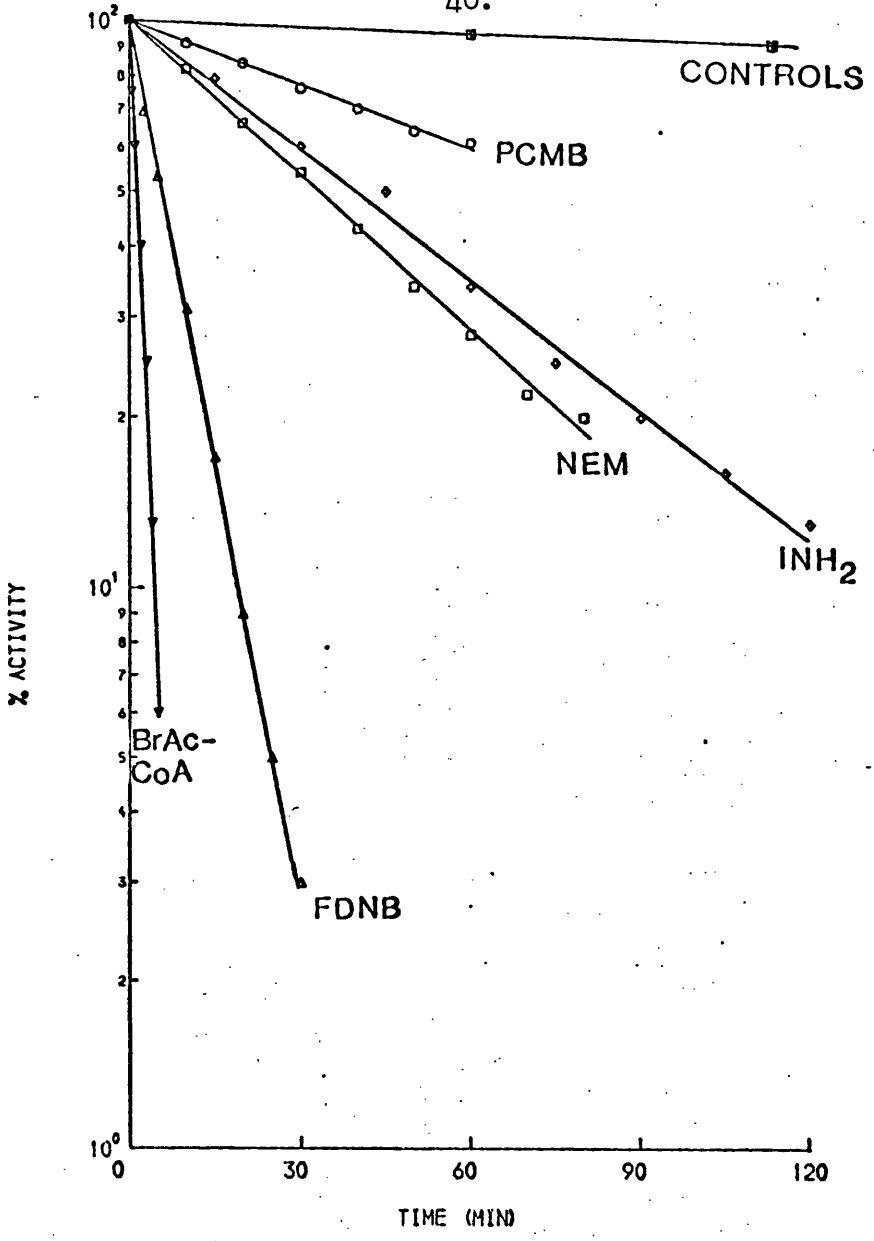
Diethylpyrocarbonate carboxyethylates amino acid residues of proteins, and at pH 6.0 histidine residues are specifically modified (55). Treatment of the four staphylococcal CAT variants with diethylpyrocarbonate as described in 'Methods' (2.11) resulted in the complete loss of activity of all four enzymes within 10 minutes. Inactivation was prevented by inclusion of CM but no protection was observed by the inclusion of AcCoA. Enzyme activity (50-70 percent) could be restored by treatment of the modified enzymes with 0.5 M hydroxylamine at pH 7.0 for 45 minutes, a procedure which removes carbethoxy groups from histidine but not from lysine or arginine (56).

figure: 7

Chemical modification of S. aureus CAT variants.
All variants were inactivated at the same rates ($\pm$ 3 percent) and mean values are plotted.

figure: 8

pH dependence of the rate of photo-oxidation of S. aureus CAT type C (C221).


CONTROLS
PCMB
NEM
INH2
BrAc-CoA
FDNB
% ACTIVITY
TIME (MIN)
0
30
60
90
120

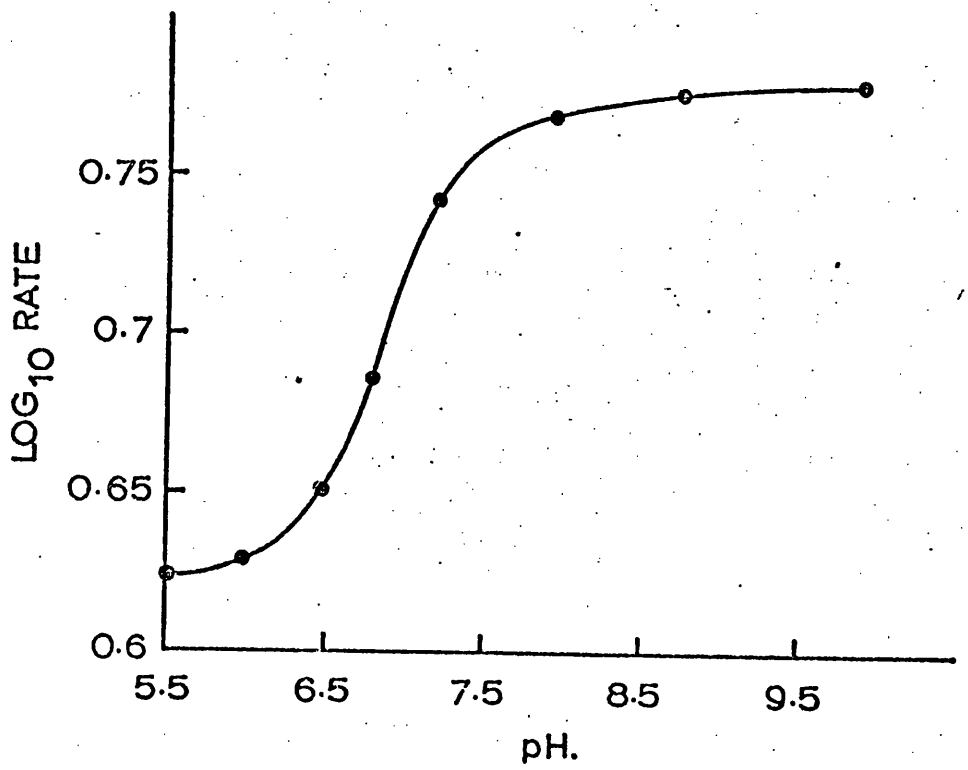

LOG10 RATE
0.75
0.7
0.65
0.6
5.5
6.5
7.5
8.5
9.5
pH.

Photo-oxidation.

Of the amino acids found in proteins only five are susceptible to photo-oxidation, histidine, tryptophan, tyrosine, methionine and cysteine (57). In studies on free amino acids, Weil (58) showed that the rate of photo-oxidation of tyrosine, tryptophan, histidine and methionine depended on the pH. For histidine the rate was entirely dependent on the ionisation of the imidazole ring, only the uncharged form being photo-oxidised.

The pH profile of photo-oxidation of type C CAT (figure 8) shows increased rates of loss of activity (and presumably photo-oxidation) with increasing pH, having an apparent pK of 6.9.

Inclusion of the substrate CM (1 mM) afforded complete protection against photo-oxidation.

3.4 Amino acid analyses.

Amino acid analyses of the four staphylococcal CAT variants were performed as described in 'Methods' (2.7). Cysteine was determined by amino acid analysis of 48 hour hydrolysates of carboxymethylated and performic acid oxidised type C (C221) CAT.

Amino acid contents of the four proteins are expressed as whole number of residues per enzyme monomer of molecular weight 24,000 daltons (table 10). Values for threonine and serine are zero time extrapolations. ('Methods' 2.7).

3.5 Tryptic peptide maps.

4mg of salt free, lyophilised enzymes (A, B, C and D) were reduced and carboxymethylated with iodoacetic acid in 6 M guanidine hydrochloride (see chapter V for methodology) and after extensive dialysis against water were freeze dried. The carboxymethylated proteins were then suspended in 1ml 50 mM ammonium bicarbonate, pH 8.0, and were digested by trypsin (1:50 molar ratio) for 4 hours at 37°C.

AMINO ACID ANALYSIS DATA FOR S. aureus TYPE C (pC 221) CAT.

DATA EXPRESSED AS μMOLES, NORMALYSED RELATIVE TO PHENYLALANINE (PHE =1 μMOLE).THREONINE & SERINE ARE ZERO TIME EXTRAPOLATIONS.

TIME OF HYDROLYSIS.(Hrs.).

	24	48	72	Best		Residue No.					
Carboxy-methyl cysteine											
Cysteic Acid		0.15		0.15		1.7					
Aspartic Acid	2.56	2.64	2.68	2.63		37.5					
Methionine Sulphone											
Threonine	0.72	0.72	0.74	0.73		10.4					
Serine	0.67	0.66	0.60	0.68		9.7					
Glutamic Acid	1.43	1.37	1.31	1.37		19.5					
Homoserine											
Proline	0.61	0.71	0.72	0.68		9.8					
Glycine	0.37	0.37	0.36	0.37		5.3					
Alanine	0.25	0.26	0.25	0.25		3.6					
Cystine											
Valine	0.38	0.45	0.46	0.46		6.5					
Methionine	0.28	0.28	0.29	0.28		4.0					
Iso-Leucine	1.21	1.33	1.35	1.35		19.3					
Leucine	0.88	0.90	0.85	0.87		12.4					
Internal Standard											
Tyrosine	0.88	0.86	0.87	0.87		12.4					
Phenylalanine	1.00	1.00	1.00	1.00		14.6					
Histidine	0.38	0.37	0.38			5.5					
Lysine	1.43	1.64	1.75	1.61		22.9					
(Tryptophan)											
Arginine	0.10	0.11	0.11	0.11		1.6					

Table 10 Amino acid content of staphylococcal variants A, B, C and D.

Amino acid.	Variant A.	Variant B.	Variant C.	Variant D.
Cysteine	N.D.	N.D.	2	N.D.
Aspartic acid	34	33	38	34
Threonine	11	10	11	13
Serine	14	14	10	10
Glutamic acid	20	18	20	20
Proline	8	8	8	8
Glycine	9	9	5	8
Alanine	4	5	4	4
Valine	12	10	7	7
Methionine	3	4	4	3
Isoleucine	16	16	20	21
Leucine	15	15	13	13
Tyrosine	15	13	13	13
Phenylalanine	14	15	15	14
Histidine	4	5	6	5
Lysine	17	18	23	20
Arginine	4	4	3	4
Tryptophan	N.D.	N.D.	4	N.D.
Total	(201)	(198)	201	(198)

N.D. = Not determined.

(Total) = types A, B and D assuming 4 tryptophans and 2 cysteines.

The digests were then freeze dried and the residues dissolved in pyridine-acetate buffer, pH 6.5.

After centrifugation the supernatents were spotted in 1cm squares on Whatman No. 1 chromatography paper and subjected to high voltage electrophoresis at pH 6.5 for 45 minutes. Four strips containing the partially separated tryptic peptides of the four enzymes were cut out and stitched to fresh sheets of paper and re-electrophoresed at pH 3.5 for 50 minutes. The sheets were then dried and stained as described in 'Methods' (2.6, b).

The resulting peptide maps are shown in figures 9, 10, 11 and 12. Staining characteristics of the peptides are denoted by the following abbreviations:

A - arginine stain positive.

N - ninhydrin stain positive.

R - stained red by ninhydrin.

O - stained orange by ninhydrin.

Y - stained yellow by ninhydrin.

IF - intrinsic fluorescence.

F - fluram stain positive.

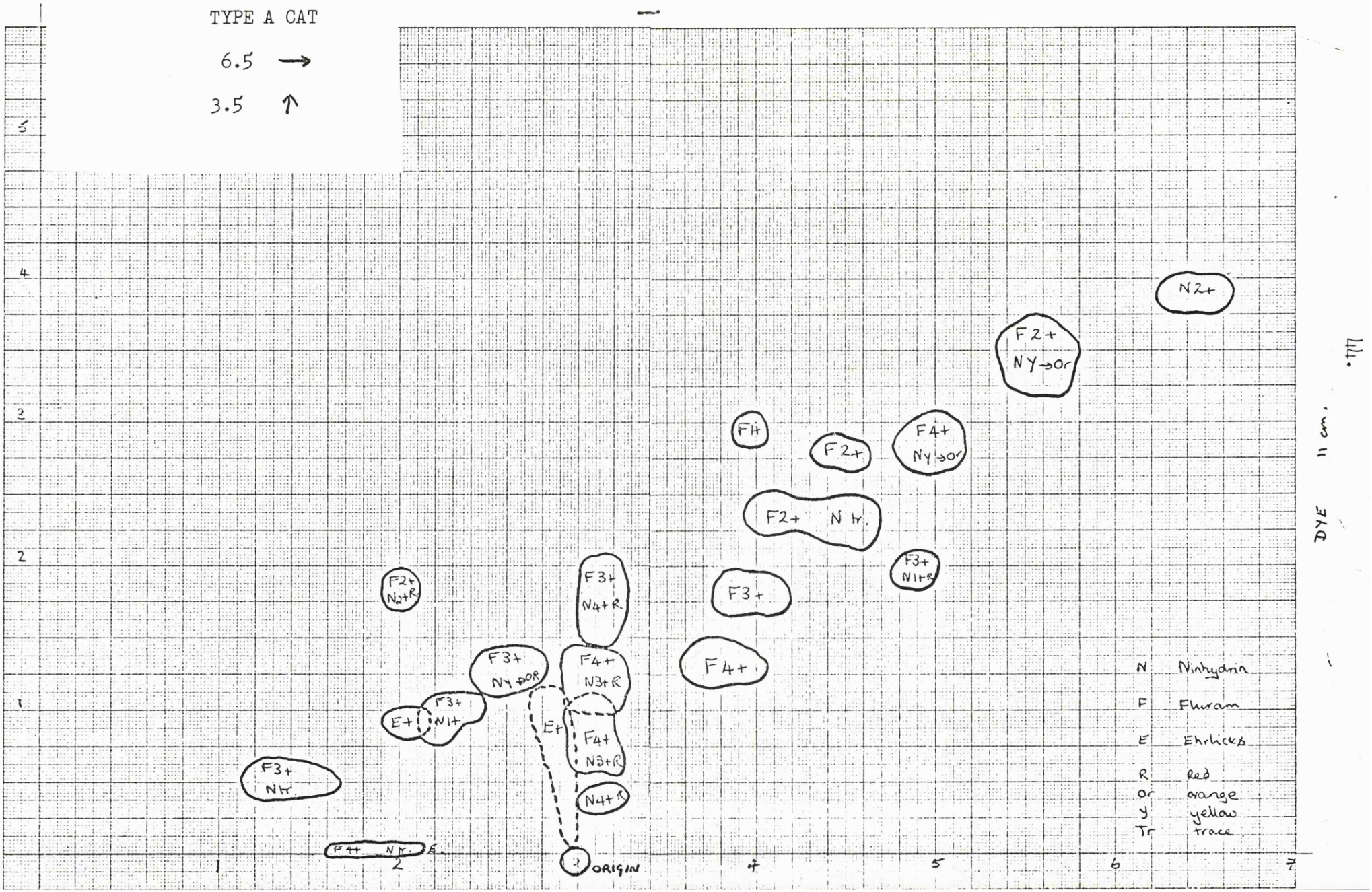
E - Ehrlich's tryptophan stain positive.

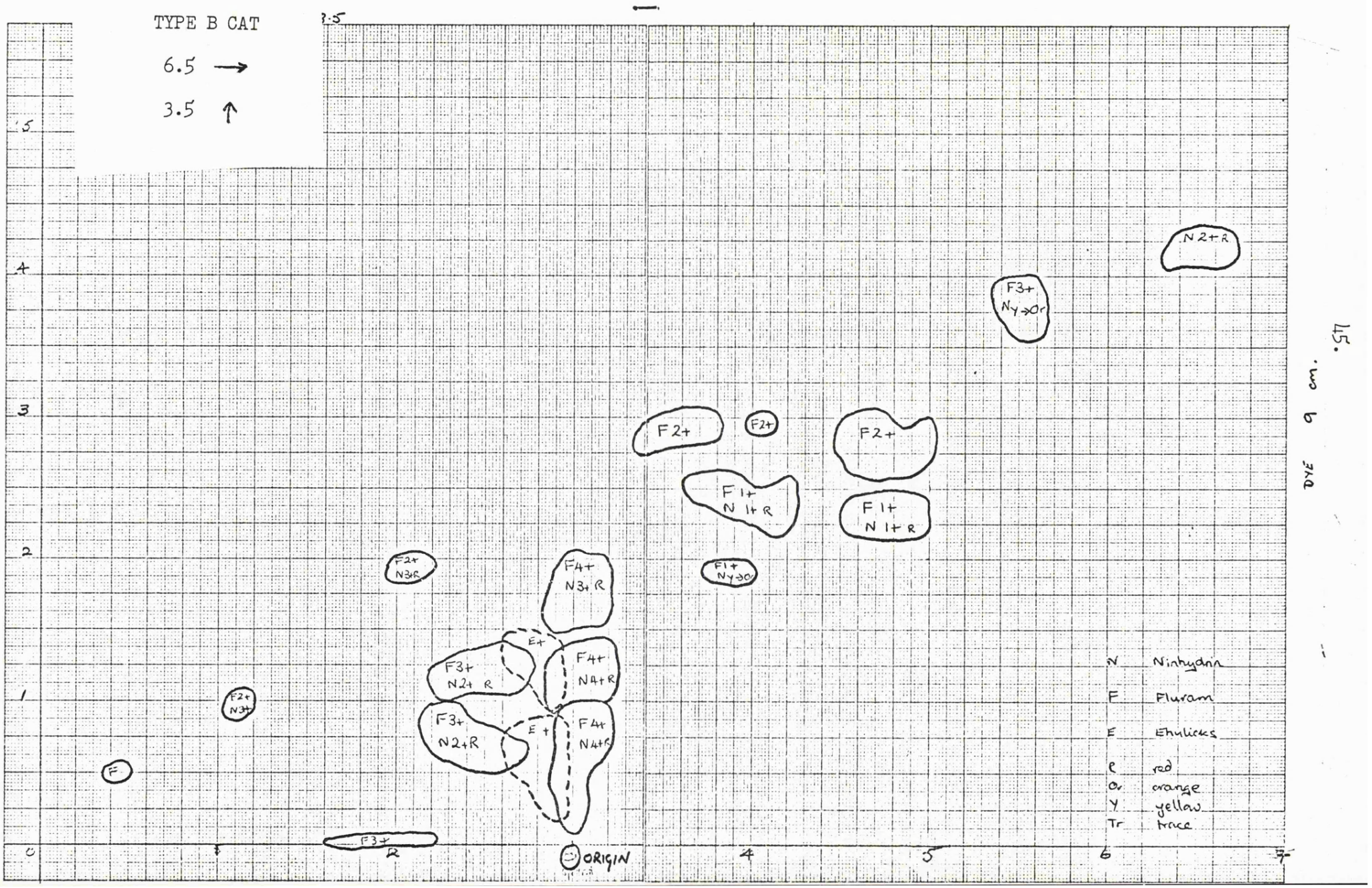
Tr - trace reaction with indicated stain.

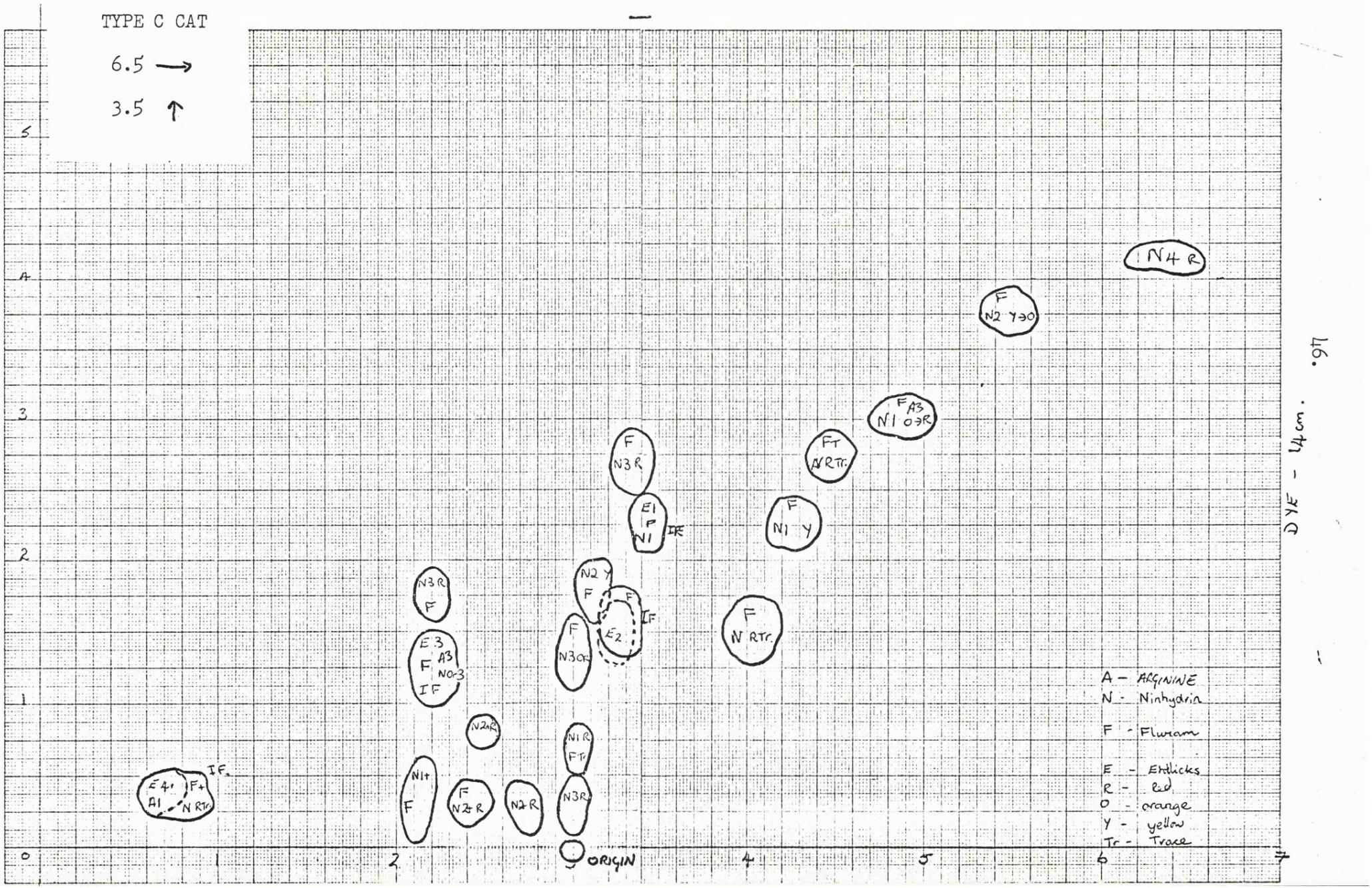
figures: 9, 10, 11 and 12.

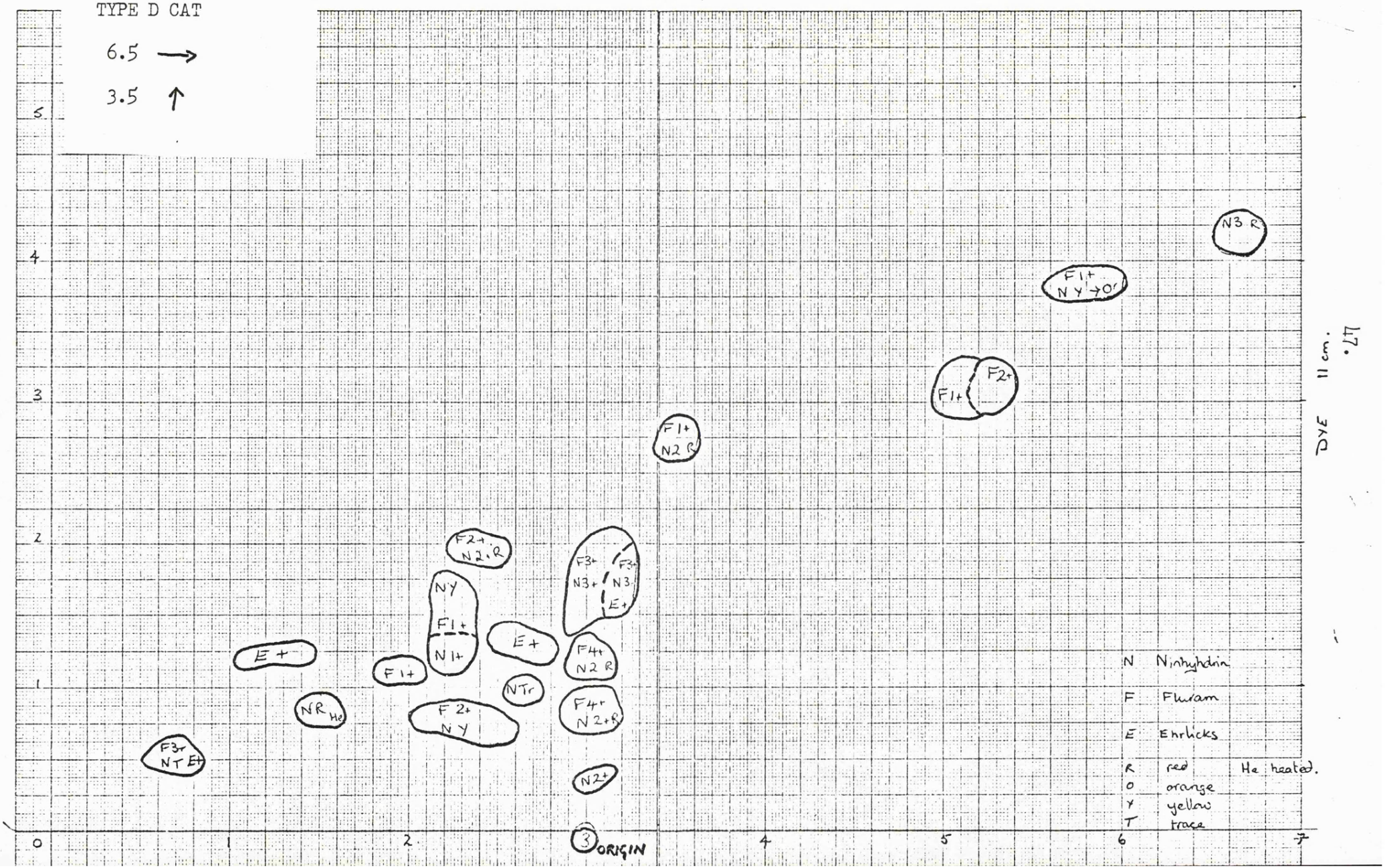
Tryptic peptide maps of S. aureus CAT types A, B, C and D.

High voltage paper electrophoresis at pH 6.5 and pH 3.5.


TYPE A CAT
6.5 →
3.5 ↑
N2+
F2+ NY→or
F1+
F2+
F4+ NY→or
F2+ N tr.
F3+ N1+R
F3+
F4+
F3+ N4+R
F2+ N2+R
F3+ NY→OR
F4+ N3+R
F3+ N1+
E+
E+
F4+ N3+R
N4+R
F3+ N1+
F4+ N1+
ORIGIN
N Ninhydrin
F Fluram
E Ehrlichs
R Red
or orange
y yellow
Tr trace
DYE 11 cm.


TYPE B CAT
6.5 →
3.5 ↑
F2+
F2+
F2+
F1+ N1+R
F1+ N1+R
F1+ Ny→Or
F3+ Ny→Or
N2+R
F2+ N3+R
F4+ N3+R
E+
F3+ N2+R
F4+ N4+R
E+
F3+ N2+R
F4+ N4+R
F2+ N3+
F
F3+
ORIGIN
N Ninhydrin
F Fluram
E Ehrlichs
R red
Or orange
Y yellow
Tr trace
DYE 9 cm.
145.


TYPE C CAT
6.5 →
3.5 ↑
N4 R
F N2 Y-O
F A3 N1 O-R
F N3 R
F N R Tr.
E1 F N1 IF
F N1 Y
N2 Y F
F
E2
IF
F N R Tr.
N3 R F
E3 F A3 NO-3 IF
F N3 Or
N2 R
N1 R F Tr.
N1+ F
F N2 R
N2 R
N3 R
E4+ A1 F+ N R Tr. IF
ORIGIN
A - ARGININE
N - Ninhydrin
F - Fluram
E - Ehrlichs
R - Red
O - orange
Y - yellow
Tr - Trace
DYE - 14cm.
16.97


TYPE D CAT
6.5 →
3.5 ↑
F1+ N2 R
F2+ F1+
F1+ NY →O
N3 R
F2+ N2 R
F3+ N3+
F3+ N3
E+
NY
F1+
N1+
E+
E+
F1+
F4+ N2 R
NTr
NR He
F2+ NY
F4+ N2+R
F3+ NT E+
N2+
3 ORIGIN
0
1
2
4
5
6
7
N Ninhydrin
F Fluram
E Ehrlichs
R red
O orange
Y yellow
T trace
He heated.
DYE 11 cm.
47

CHAPTER IV.

STRUCTURAL STUDIES.

4.1 Introduction.

The ultimate aim of the protein chemist is the determination of the tertiary structure of the protein and the topography and composition of the functional areas (active sites in enzymes), enabling an insight into the evolution of the protein by comparison with analogous proteins and the postulation of a mechanism of catalysis. The initial step towards this goal is the determination of the primary sequence of the protein. Although the primary structure of a type I R-factor CAT (JR66) is known, no sequence data was available for the staphylococcal enzymes. In order to further comparisons between the R-factor and the Gram positive CAT variants the decision was made to initiate the determination of the primary structure of a type C staphylococcal CAT carried by plasmid C221 (52).

4.2 Strategy.

The now classical approach of studying the sequences of peptides obtained by digestion of the chemically modified protein with proteases of different specificity and cyanogen bromide cleavage in order to obtain the necessary sequence overlap information was used to determine the complete amino acid sequence of CAT from S. aureus (plasmid C221).

4.3 Peptide nomenclature.

Peptides produced by chemical or enzymic digestion of C221 CAT are identified by the primary letter code corresponding to that particular digest:

SP indicates digestion by staphylococcal protease (V8).

T indicates digestion by trypsin

C indicates digestion by chymotrypsin

E indicates digestion by elastase

CB indicates cleavage by cyanogen bromide

Freeze dried digests were vortexed with 3ml of the loading buffer used in the initial fractionation of the digest (either ion-exchange or gel filtration) and centrifuged. Peptides in the supernatent were termed 'soluble' and given the prefix S. Hence peptides soluble in pyridine acetate buffer, pH 3.1, derived from tryptic digestion of CAT (Chapter V) have the prefix TS. Peptides derived from insoluble 'core' material have the prefix C,thus, peptides insoluble in pyridine acetate buffer, pH 3.1, derived from a tryptic digest have the prefix TC. Peptides are also identified by the fraction numbers in which they are eluted from the column used in their purification. Hence peptide TS 76-77 is a tryptic peptide soluble in pyridine acetate buffer, pH 3.1, and was eluted from the ion exchange column used in their purification in fractions 76 and 77. Peptides purified further by high voltage paper electrophoresis are numbered alphabetically from the cathode eg. peptide TS 43-46 A is a soluble tryptic peptide which was the most basic peptide in fractions 43 to 46. Relative mobilities for the peptides are determined as described in 'Methods' (2.5). Although the various peptide staining tests (Chapter II, Detection of peptides 2.6 b.) were used extensively in these studies to detect and characterise peptides, the results of these tests are not reported unless they provide essential information not obtained by subsequent studies on the eluted peptides.

Amino acid compositions which have been quantitively established are indicated by the symbol: —— —— beneath the sequence of the peptide. Sequential analysis results are represented using the following symbols beneath the peptide:

→ represents the unequivocal identification of the PTH amino acid either from a 'spinning cup' or solid phase automatic sequencer.

⇢ represents the identification of the 'dansyl' derivative of the

indicated amino acid.

⟶ indicates strong evidence for the correct identification of the amino acid in the absence of 'dansyl' or automated sequence information either from the staining properties with cadmium ninhydrin reagent, or specific amino acid stains, or from the complete lack of reactivity of the N-terminal amino acid indicating cyclysation of a glutamic residue to form N-terminal pyrollidone carboxylic acid.

⟵ represents release of the indicated amino acid by digestion with carboxypeptidases A and B.

⟵ indicates strong evidence for the C-terminal amino acid from the known specificity of the method of peptide generation (eg. CNBr cleavage).

The one-letter abbreviations defined by the IUPAC-IUB Commission on Biochemical Nomenclature (59) are used for extensive sequences, all peptide sequence data are expressed as the three-letter abbreviations (60).

Dashes between three letter symbols denote regions of determined sequence, round brackets enclose regions where only the amino acid composition is known.

Protein N and C-terminal sequence determinations.

N-terminal determinations.

4.4 Liquid phase.

Quantitative N-terminal sequence determination of C221 CAT on a Beckman 890 Sequencer gave the N-terminal 30 residues (figure 13).

4.5 Solid phase.

Initial problems in coupling CAT proteins to solid phase supports were due to the insolubility of lyophilised CAT. The enzymes were totally insoluble in 50 percent. N-methyl morpholine, pH 9.5, or 0.4 M dimethyl-allylamine, pH 9.5, (both 'recommended' buffer systems for solid phase coupling).

The proteins were also insoluble in buffers containing up to 5 percent S.D.S. even after boiling.

Enzymes in their native state (ie.not lyophilised) did not couple to solid phase supports. Lyophilised CAT was found to be soluble in 0.4 M dimethylallylamine adjusted to pH 9.5 with trifluoroacetic acid and saturated with guanidine hydrochloride. Using this buffer system coupling yields of up to 50 percent were obtained. Yields were found to be greatest (70 percent) using the triethylamine/trifluoroacetic acid/saturated guanidine hydrochloride system described in 'Methods' (2.9, b.) The reason for increased coupling yields in triethylamine buffer is not yet known but may be related to the higher pK of triethylamine compared with N-methyl-morpholine and dimethylallylamine.

Underivatised, carboxymethylated and performic acid oxidised samples of enzymes were sequenced, however performic acid oxidised proteins were found to give a greater 'background' on TLC plates, probably due to acid cleavage of the polypeptide during the oxidation reaction and were therefore unsuitable for extended sequence runs.

Underivatised CAT gave the cleanest sequences and the longest N-terminal sequences, in spite of the possibility of cleavage at underivatised cysteine residues during the sequencing procedure (61). When no PTH amino acid was identified by TLC or back hydrolysis with hydriodic acid with both native and carboxymethylated protein, the residue was assumed to be lysine bound to the solid support in the case of DITC coupled proteins. Since all primary amino groups are covalently bound to the solid phase support using the DITC coupling procedure the N-terminal amino acid of each protein was determined by the 'dansyl' method (62).

N-terminal sequences obtained are illustrated in figure 13 and table 11. It can be seen that the length of sequence obtained is, to some extent, dependent on the amount of protein coupled although the maximum length of sequence obtained rarely exceeds 20 residues.

Table 11 N-terminal solid phase sequence of CAT proteins: number of residues obtained.

CAT Type.	Amount of CAT used in coupling.	Number of residues obtained.
Type I (RH29)	3.5mg.	18
Type II (S-a)	1.5mg.	13
H. parainfluenzae	2 mg.	13
S. aureus Type A	4 mg.	16
S. aureus Type B	4 mg.	16
S. aureus Type C	5 mg.	24
S. aureus Type D	4 mg.	19

Attempts to couple CAT proteins from Streptomyces acrimycini (28) and a Flavobacterium (29) (purified by Dr. S. Harford of this laboratory) failed. This is probably due to the low lysine content of these proteins (3 moles lysine per mole monomer), indeed, a definite correlation between lysine content and coupling efficiency was seen with the CAT proteins. It is interesting to compare the result obtained for the type C staphylococcal CAT by the liquid phase and the solid phase sequencers.

figure: 13

N-terminal sequences of CAT variants.

All sequences were determined on a solid phase sequencer with the exception of residues 25 - 30 of the type C *S. aureus* variant which were determined on a liquid phase sequencer. (see Chapter IV.)

	1 5 10 15 20 25 30
R-factor type I (R129)	M E K K I T G Y T T V D I S Q W H R
R-factor type II (S-a)	M N F T R I D L N T W N R
H. parainfluenzae	M N F T R I D L N T W N R
S. aureus type A	T F N I I N L E T W D R K E Y F
S. aureus type B	T F N I I K L E T W D R K E Y F
S. aureus type C	T F N I I K L E N W D R K E Y F E H Y F N Q Q T T Y S I T K
S. aureus type D	T F N I I E L E N W D R K E Y F E H ? F

The N-terminal 30 residues were identified by the liquid phase (spinning cup) technique using a total of 20mg of protein whereas the solid phase gave a 24 residue sequence using only 5mg protein. Although the solid phase technique cannot equal the liquid phase in length of sequence obtained it offers advantages where limited quantities of protein are available, especially if the sequencer is adapted to take advantage of radioactive micro-sequencing methods (63).

4.6 C-terminal sequence determination.

0.6mg (25 n Moles) of lyophilised S. aureus C221 CAT was dissolved in 0.25ml 50 mM N-methylmorpholine buffer, pH 8.5, which contained urea at a final concentration of 4 M and digested with a 1:50 molar ratio carboxypeptidases A and B. 50μl samples were taken at 0, 1, 2, 4 and 24 hours, freeze dried and loaded on to the amino acid analyser. Peaks corresponding to isoleucine and serine (or asparagine or glutamine since these amino acids co-elute with serine) were observed. Control samples containing CAT only and carboxypeptidases A and B only gave blank amino acid analyses in each case.

The rates of release of the free amino acids were:

Isoleucine = 2.4 n Moles/hour.

'Serine' =1.3 n Moles/hour.

hence the C-terminus of C221 CAT was concluded to be:

$$-\begin{pmatrix}\text{Ser}\\ \text{Asn}\\ \text{Gln}\end{pmatrix}-\text{Ile}$$

Although Ser,Asn and Gln are not resolved by amino acid analysis,they could be differentiated by high voltage paper electrophoresis and future work will utilise this method of identification.

CHAPTER V.

PRODUCTION, FRACTIONATION AND SEQUENCE OF TRYPTIC PEPTIDES.

5.1 Preparation of Cm CAT.

In general, native proteins are not good substrates for extensive proteolytic degradation although digestion of native proteins has been used for the production of large 'core' fragments which may retain functional activity (65). Peptide bonds on the surface of the protein in contact with the external environment are accessible and may be susceptible to enzymic attack, but peptide bonds buried in the protein matrix are not readily hydrolysed. Since the main objective in structural investigations is to cleave all potentially sensitive bonds quantitively, prior denaturation of of the substrate is necessary for proteolytic degradation.

Since the CAT specified by plasmid C221 contains cysteine these residues must be derivatised to avoid disulphide exchange reactions and to enable the identification of the labile cysteine residues. Carboxymethylation in 6 M guanidine hydrochloride not only carboxymethylates cysteine but also effectively denatures the protein providing a suitable substrate for enzymic degradation and was performed as follows:

Lyophilised, salt free C221 CAT (130mg) was dissolved in degassed 6 M guanidine HCl/100 mM Tris HCl, pH 8.5, (40ml). Dithiothreitol (100μMoles) was added and the solution stirred under nitrogen for 2 hours at room temperature. [^{14}C] iodoacetic acid (400μMoles, 0.125μC/μMole) was then added and the reaction vessel covered with aluminium foil to prevent the formation of iodine which may react with tyrosine, tryptophan and histidine residues. Stirring under nitrogen was continued for a further 2 hours, excess mercaptoethanol (500μMoles) was added to quench further reaction and the alkylated protein dialysed exhaustively against water.

Amino acid analysis of the alkylated protein indicated 1.8 moles carboxymethyl cysteine/mole monomer and counting an aliquot of the alkylated

protein (knowning the specific activity of the [^{14}C] iodoacetic acid to be 0.125μC/μMole) gave a value of 2.15 moles iodoacetic acid/mole monomer incorporated. The carboxymethyl CAT was then freeze dried.

5.2 Tryptic digestion of C221 Cm CAT.

Pilot digestion.

Pilot digests on 1mg aliquots of Cm CAT indicated that a 1:50 molar ratio trypsin to Cm CAT for 4 hours at 37°C gave the best yields of peptides as judged by high voltage paper electrophoresis at pH 6.5 and 3.5 followed by staining with fluorescamine and cadmium-ninhydrin reagent. 5mg Cm CAT was then digested under these conditions followed by electrophoresis and staining to give a peptide map of the protein digest. Of the theoretical 27 peptides (23 lysine residues and 3 arginine residues per monomer) obtainable from a tryptic digest of Cm CAT, 20 peptides were distinguishable on the peptide map (figure 11).

Autoradiography of the peptide map resulted in only one radioactive region. This was at the origin and indicates that the cysteine containing sequences are contained in peptides which do not run under the conditions employed for the production of the tryptic peptide map.

Preparative digestion.

Cm CAT (120mg) was suspended in 50ml of 20 mM ammonium bicarbonate, pH 8.0, and sonicated for 30 seconds to give a fine suspension. Trypsin (purified as described in 'Methods' 2.4) was added at a molar ratio of 1:50 and the suspension slowly stirred at 37°C for 5 hours. As the digestion progressed the insoluble material decreased in quantity but even after 5 hours the solution was still turbid. The digest was then quickly shell-frozen and lyophilised.

5.3 Fractionation of the tryptic peptides of Cm CAT.

The preparative tryptic digest of Cm CAT (digest T) was fractionated by ion-exchange chromatography on ~~Dowex 50~~ a sulphonic acid resin column using the method described in 'Methods' (2.4). The freeze dried tryptic peptides were vortexed with 5ml of the starting buffer (0.2 M pyridine/acetate, pH 3.1,) and centrifuged. The insoluble 'core' material was stored at -20°C for further study.

The supernatent was loaded on to the ion-exchange column and separated with a pH gradient pH 3.1 to 5.0 and the column finally washed with pyridine/acetate, pH 6.5. Samples from each fraction collected were fingerprinted by high voltage paper electrophoresis ('Methods' 2.6) and pooled on the basis of the fingerprint pattern. The pooled fractions were rotary evaporated at 40°C to 1ml for further purification by high voltage paper electrophoresis ('Methods' 2.5) as required.

The 'core' material (insoluble in pyridine/acetate, pH 3.1) proved to be rather difficult to separate being insoluble in 8 M urea, 50 percent acetic acid or 6 M guanidine HCl. It was eventually dissolved in a saturated aqueous solution of guanidine HCl and fractionated on a G-75 Sephadex column equilibrated with saturated guanidine HCl solution. Four distinct peaks were obtained (figure 14). The pooled fractions were desalted by dialysis against 50 mM ammonium bicarbonate, pH 8.0, using boiled Spectropore 3 dialysis membrane (exclusion limit approximately 800 daltons).

5.4 Amino acid analyses, characteristics and sequence of tryptic peptides.

Peptide TS 5 - 10 B.

Analysis : Glu (1.9), Thr (0.9), Lys (1.0).

Mobility : pH 6.5 (+0.45), pH 2.1 (0.90). Net charge -1

Yield : 550 n moles. Purif. HVPE pH 6.5.

All attempts to sequence this peptide, both manually and on the solid phase

sequencer, failed. This indicates a blocked N-terminus, probably an N-terminal pyroglutamyl residue. Mass spectrometry (performed by Dr. A. Dell at Imperial College, London) gave the N-terminal two residues: Glu-Thr- and the total sequence of this peptide was concluded to be:

Glu-Thr-Glu-Lys.

Peptide TS 18 - 22.

Analysis : Asp (1.2), Ser (2.1), Glu (1.1), Lys (1.0).

Mobility : pH 6.5 (0.0), pH 2.1 (1.05). Net charge zero.

Yield : 300 n moles.

Sequence : Ser-Gln-(Asp,Ser)-Lys.

Although several attempts were made to sequence this peptide only the N-terminal 2 residues were obtained. From the pH 6.5 mobility it is apparent that an amide is present in this peptide (either Asn or Gln) and it is likely that the inability to obtain the sequence of more than 2 residues is due to the cyclisation of a glutamine residue at position 2 during the sequencing procedure.

Peptide TS 37 - 41.

Analysis : Asp (1.9), Thr (0.9), Pro (0.95), Val (1.1), Leu (1.5), Tyr (0.9), Phe (0.95), Lys (1.0).

Mobility : pH 6.5 (-0.25), pH 2.1 (1.0). Net charge +1.

Yield : 600 n moles.

Sequence : Leu-Asn-Pro-Leu-Tyr-Thr-Val-Phe-Asn-Lys.

This peptide was sequenced on the solid phase sequencer, the asparagine residues being identified as such by this method and the sequence (2 asparagine residues) agrees with the electrophoretic mobility indicating a net charge of +1.

Peptide TS 43 - 46 A.

Analysis : Asp (1.9), Thr (0.9), Ser (0.9), Glu(1.1), Gly (1.1), Ile (1.1), Lys (1.0).

Mobility : pH 6.5 (0.0), pH 2.1 (1.2). Net charge zero.

Yield : 750 n moles. Purif. HVPE pH 3.5

Sequence : Asn-Gly-Ile-Glu-Ser-Asn-Thr-Lys.

The N-terminal residue (which is not identified on the automatic sequencer due to the DITC coupling procedure, 'Methods' 2.9 b) was identified as Asx by the dansyl N-terminal method and was assumed to be Asn to give the peptide which was neutral at pH 6.5.

Peptide TS 43 - 46 D.

Analysis : Asp (2.1), Ser (0.4), Val (1.1), Ile (0.9), Trp +ve.

Mobility : pH 6.5 (+0.95), pH 2.1 (0.7). Net charge -2.

Yield : 750 n moles.

Sequence : Val-Asp-Asp-Trp-Ile.

This peptide was the only high yield peptide which did not contain lysine or arginine and was considered a likely candidate for the C-terminal tryptic peptide. No serine was detected during the sequencing procedure and the low serine content (0.4) in the analysis was considered to be a contaminant.

An analogous peptide has been isolated from a tryptic digest which was purified by gel filtration (G-25, in 50 mM ammonium bicarbonate) and paper electrophoresis. This peptide had the analysis:

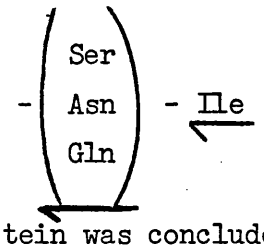
Asp (2.0), Val (0.9), Ile (1.0), Trp +ve.

and had a net charge of -1 (indicating one Asp and one Asn).

Sequence : Val-Asx-Asx-Trp-Ile.

If the sequence data of these peptides is compared with the C-terminal

determination of the C221 enzyme (Chapter IV, 4.6) :



the C-terminus of the protein was concluded to be :

Val-Asp-Asn-Trp-Ile.

Peptide TS 43 - 46 D had become deamidated during purification (hence net charge -2) whereas the analogous peptide purified by gel filtration had not (hence net charge -1).

Amino acid analyses of samples taken for the C-terminal determination of the protein using carboxypeptidases would not have detected tryptophan since this amino acid is eluted with ammonia using our normal analysis programme. Amino acid release would have stopped when the aspartic acid residue had been reached since carboxypeptidases only release aspartic acid slowly (62).

Peptide TS 59 - 63 C.

Analysis : Asp (1.1), Thr (2.4), Glu (4.1), Tyr (2.7), Phe (2.1), His (1.0), [redacted].

Mobility : pH 6.5 (+0.28), pH 2.1 (0.70). Net charge -1.5.

Yield : 100 n moles. Purif HVPE pH 6.5

Sequence :

Glx-Tyr-Phe-Glx-His-Tyr-Phe-Asx-Glx-(Glu, Thr_2, Tyr)[redacted].

This peptide corresponds to the N-terminal residues 14 to 26 (see N-terminal sequence determinations, figure 13).

Peptide TS 59 - 63 D.

Analysis : Asp (1.9), Glu (1.0), Leu (1.8), Tyr (0.9), Lys (1.0).

Mobility : pH 6.5 (+0.35), pH 2.1 (1.2). Net charge -1.

Yield : 550 n moles. Purif. HVPE pH 6.5.

Sequence : Asn-Asp-Leu-Leu-Glu-Tyr-Lys.

The N-terminal residue was identified as Asx by the dansyl method and was concluded to be the amide due to its staining characteristics with ninhydrin (red - orange) and the expected net charge of -1 from the electrophoretic data.

Peptide TS 66 - 68.

Analysis : Asp (1.1), Glu (0.9), His (1.1), Lys (1.0).

Mobility : pH 6.5 (+0.25), pH 2.1 (1.45). Net charge $-\frac{1}{2}$.

Yield : 500 n moles. Purif. HVPE pH 3.5.

Sequence : Asp-His-Glu-Lys.

In order to have a net charge of $-\frac{1}{2}$ both Asp and Glu must be free acids rather than the amides.

Peptide TS 69 - 71.

Analysis : Asp (1.0), Met (0.95), Ile (0.85), Lys (1.0).

Mobility : pH 6.5 (0.0), pH 2.1 (1.5). Net charge zero.

Yield : 1,200 n moles.

Sequence : Asp-Met-Ile-Lys.

Peptide TS 72 - 75 B.

Analysis : Asp (1.1), Thr (0.9), Ile (1.2), Phe (1.1), Lys (1.0).

Mobility : pH 6.5 (-0.3), pH 2.1 (1.3). Net charge +1.

Yield : 600 n moles. Purif. HVPE pH 6.5

Sequence : Thr-Phe-Asn-Ile-Ile-Lys.

In order to have a net charge of +1 the third residue must be Asn rather than Asp. The low Ile content from the amino acid analysis is due to the slow rate

of hydrolysis of Ile-Ile bonds. This peptide corresponds to the N-terminal residues 1 - 6 (see figure 13).

Peptide TS 72 - 75 C.

Analysis : Asp (0.9), Glu (1.9), Pro (0.9), Met (0.9), Phe (0.9), Lys (1.9).

Mobility : pH 6.5 (+0.25), pH 2.1 (1.4). Net charge -1.

Yield : 950 n moles. Purif. HVPE pH 6.5

Sequence : Asp-Lys-Glu-Glu-Met-Phe-Pro-Lys.

→ → → → → → → →

It is worthy of note that this peptide contains an internal lysine residue which was not cleaved by trypsin, probably due to the three adjacent acidic residues (as determined by the net charge of the peptide).

Peptide TS 76 - 77.

Analysis : Asp (2.1), Ser (0.9), Glu (1.0), Tyr (0.9), Phe (1.0), Lys (1.0).

Mobility : pH 6.5 (0.0), pH 2.1 (1.2). Net charge zero.

Yield : 700 n moles.

Sequence : Phe-Tyr-Ser-Glx-Asx-Asx-Lys.

→ → → → → → →

This peptide contains 2 amides (from the net charge of the peptide). The exact location of the amides is not yet known.

Peptide TS 82 - 84.

Analysis : Thr (0.8), Ser (0.9), Ile (0.9), Lys (1.0)

Mobility : pH 6.5 (-0.45), pH 2.1 (1.6). Net charge +1.

Yield : 500 n moles.

Sequence : Ser-Ile-Thr-Lys.

→ → → →

This sequence is identical to residues 27 - 30, however residue 26 is a tyrosine rather than the normal cleavage point of lysine or arginine so the

position of this peptide in the sequence is not certain.

Peptide TS 87 - 89 B.

Analysis : Asp (2.1), Glu (0.9), Leu (1.0), Arg (1.1), Trp +ve, ninhydrin stain, red - orange.

Mobility : pH 6.5 (+0.3), pH 2.1 (1.2). Net charge -1.

Yield : 450 n moles. Purif. HVPE pH 6.5.

Sequence : Leu-Glu-Asn-Trp-Asp-Arg.

This peptide corresponds to residues 7 - 12 of the protein N-terminus.

Peptide TS 108 - 104.

Analysis : Asp (1.0), Gly (1.2), Leu (1.0), Tyr (0.9), Lys (1.0), Trp +ve.

Mobility : pH 6.5 (0.0), pH 2.1 (1.2). Net charge zero.

Yield : 500 n moles.

Sequence : Leu-Gly-Tyr-Trp-Asp-Lys.

Peptide TS 116 - 121.

Analysis : Gly (1.0), Val (1.2), His (1.1), Arg (1.0).

Mobility : pH 6.5 (-0.6), pH 2.1 (2.3). Net charge +1½.

Yield : 150 n moles.

Sequence : His-Val-Gly-Val-Arg.

Peptide TS 102 - 104.

Analysis : Asp (1.0), Lys (1.0).

Mobility : pH 6.5 (-0.7), pH 2.1 (2.3). Net charge -1.

Yield : 1,500 n moles.

Sequence : Asn-Lys.

The N-terminal residue was judged to be Asn from the ninhydrin stain (red - orange) and the net charge +1.

Peptide TS 127 - 129.

Analysis : Leu (0.8), Arg (1.1).

Mobility : pH 6.5 (-0.65), pH 2.1 (2.1). Net charge +1.

Yield : 200 n moles.

Sequence : Leu-Arg.

Peptide TS 137 - 140.

Analysis : Met (1.0), Phe (0.9), Arg (1.1).

Mobility : pH 6.5 (-0.45), pH 2.1 (1.7). Net charge +1.

Yield : 1,200 n moles.

Sequence : Met-Phe-Arg.

Peptide TS 146 - 148.

Analysis : Gly (1.1), Val (1.0), His (0.9), Arg (1.0).

Mobility : pH 6.5 (-0.6), pH 2.1 (2.3). Net charge +1½.

Yield : 450 n moles.

Sequence : His-Val-Gly-Arg.

Peptide TS 149 - 151 B.

Analysis : Phe (1.0), Arg (1.0).

Mobility : pH 6.5 (-0.55), pH 2.1 (1.85). Net charge +1.

Yield : 180 n moles. Purif HVPE pH 2.1.

Sequence : Phe-Arg.

This peptide was obtained in relatively low yield and was considered to be a partial cleavage product of peptide TS 137 - 140.

Peptide TC 1.

This was the first peak eluted from the G-75 column (figure 14) used to fractionate the insoluble 'core' material. From its amino acid analysis,

yield (40 n Moles) and mobility on S.D.S. gel electrophoresis it was judged to be undigested CAT protein and no further studies were undertaken.

Peptide TC 2.

Analysis : Cm Cys (1), Asp (15), Thr (4), Ser (10), Glu (10), Gly (5), Ala (4), Pro (3), Val (4), Met (3), Ile (10), Leu (5), Tyr (4), Phe (5), Lys (10), Arg (1).

Yield : 50 n moles.

Sequence : No sequence obtained by dansyl-Edman or solid phase sequencing.

In view of the large size and number of lysine and arginine residues this peptide was considered to be a partial cleavage product.

Peptide TC 3.

Analysis : Asp (1.9), Thr (0.4), Ser (2.0), Glu (1.9), Gly (3.1), Val (0.9), Ile (0.8), Leu (1.2), Tyr (0.9), Phe (0.8), Lys (1.0), Arg (1.1).

Yield : 500 n moles.

Sequence : No sequence obtained by dansyl-Edman or solid phase sequencing.

Since this peptide contains one lysine and one arginine residue it can be considered a partial cleavage product although obtained in high yield.

Peptide TC 4.

Analysis : Asp (2.0), Thr (1.0), Ser (1.1), Glu (1.0), Val (1.0), Ile (1.9), Leu (0.8), Tyr (1.7), Phe (1.0), Lys (1.0).

Yield : 600 n moles.

Sequence : No sequence obtained by dansyl-Edman or solid phase sequencing.

No sequence information was obtained for any of the tryptic 'core' peptides. This could be due to N-terminal blocking by carbamylation

figure : 14.

Elution profile of tryptic core peptides from a Sephadex G-75 column (1 x 80cm) equilibrated and run in saturated guanidine HCl.

Flow rate = 5ml/hour.

2.5ml fractions collected.

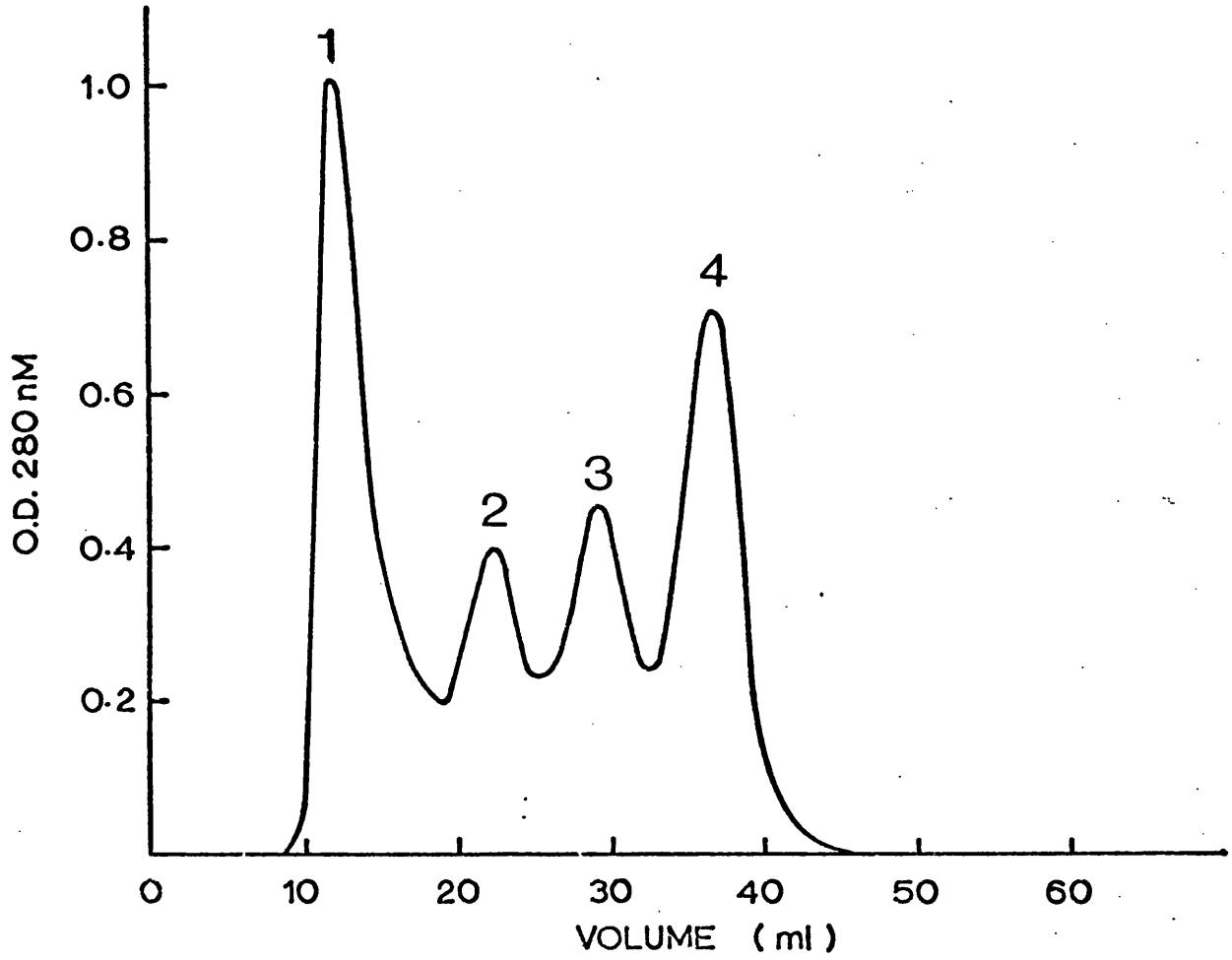
figure : 15.

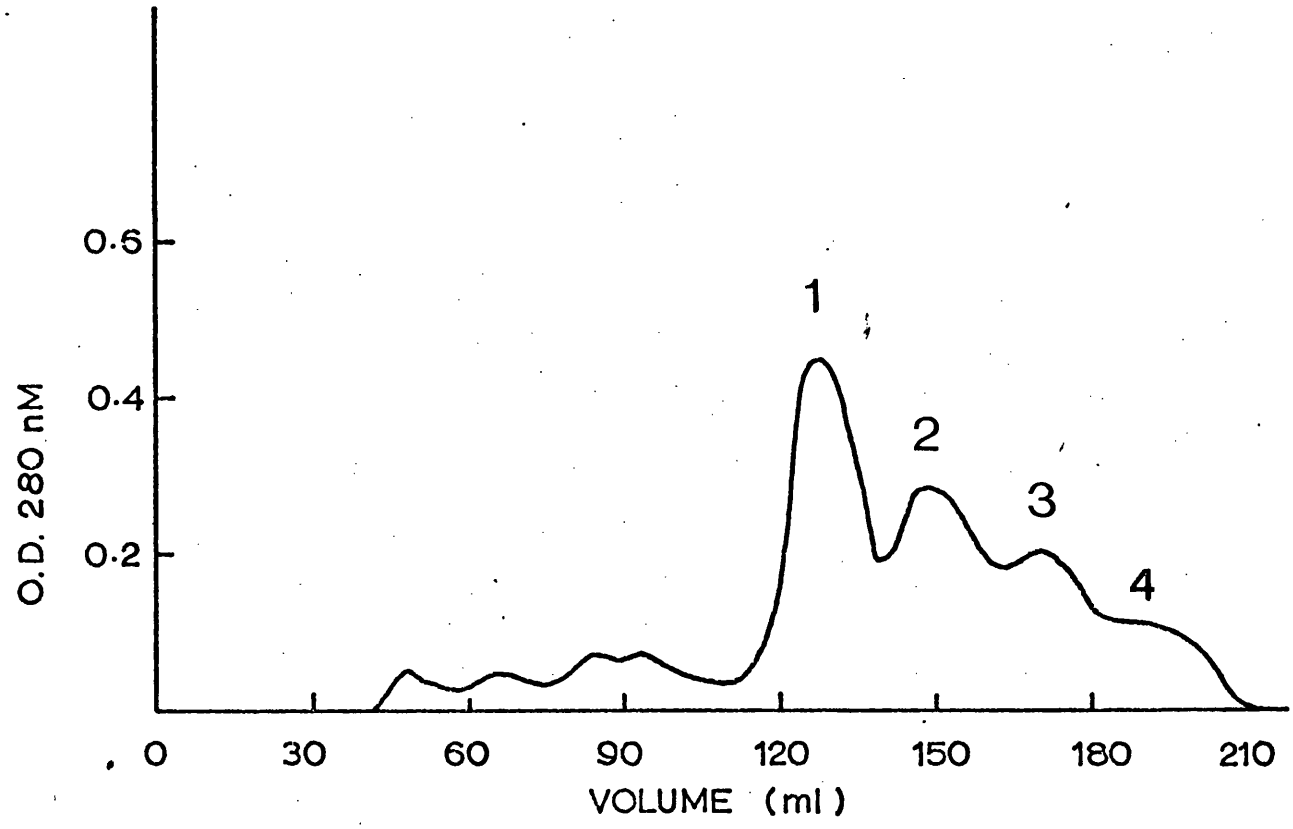
Elution profile of tryptic digest of citraconylated Cm CAT.

Column = Bio-gel P60 (2.5 x 80cm) equilibrated and run in 50 mM ammonium bicarbonate (pH 8.0).

Flow rate = 10ml/hour.

2.0ml fractions collected.


1
2
3
4
1.0
0.8
0.6
0.4
0.2
O.D. 280 nM
0
10
20
30
40
50
60
VOLUME (ml)


1
2
3
4
0.6
0.4
0.2
O.D. 280 nM
0
30
60
90
120
150
180
210
VOLUME (ml)

since attempts to solubilise the core material had involved the use of urea solutions. Alternatively the extreme insolubility of the peptides may have caused the sequencing problems.

5.5 Citraconylation of Cm CAT.

Reversible masking of the ε-amino groups of lysine restricts tryptic hydrolysis to arginyl residues in the polypeptide chain. This may be accomplished by a number of methods, including carbamylation, dinitrophenylation, succinylation, trifluoroacetylation, maleylation and citraconylation. A comparison of the specificity, ease of reversal and homogeneity of the masked and unmasked derivatives (66) has shown that citraconylation is the most satisfactory method.

Cm CAT (70mg) was dissolved in ice cold 6 M guanidine HCl/100 mM N-methylmorpholine acetate buffer, pH 8.3, (15ml). Aliquots (20μl) of citraconic anhydride were added at 10 minute intervals, the pH being maintained at 8.3 by the addition of 2 M NaOH. The total amount of citraconic anhydride added (200μl) was a 20 fold excess over lysine residues and the total reaction time was $2\frac{1}{2}$ hours. The resulting solution was dialysed extensively against 30 mM ammonium bicarbonate. The citraconylated Cm CAT precipitated and then redissolved as dialysis progressed, finally resulting in a clear solution.

5.6 Tryptic digestion of citraconyl Cm CAT.

The derivatised Cm CAT was digested with trypsin under identical conditions to digest T. (37°C, 1:50 ratio trypsin, 5 hours) and freeze dried.

5.7 Fractionation of limited tryptic peptides.

The lyophilised peptides were dissolved in 3ml of 50 mM ammonium bicarbonate, pH 8.0, applied to a Bio-gel P-60 column (2.5 x 80cm) and the column was developed with the same solvent. Four major peaks were obtained (figure 15).

Although the peaks were insufficiently resolved to obtain pure samples of each peptide, reasonable amino acid analyses, and, in one case, an N-terminal sequence, were obtained by taking samples from the fractions at the centre of each peak. Citraconyl blocking groups were removed by incubating the peptides at 40°C in dilute acetic acid, pH 4.0, for 3 hours.

Peak 1.

Analysis : Asp (16), Thr (5), Ser (4), Glu (6), Pro (2), Gly (3), Ala (2), Val (4), Met (3), Ile (7), Leu (5), Tyr (3), Phe (5), Lys (10), His (1), Arg (1).

Estimated yield : 600 n moles.

N-terminal sequence : Ala-Ile-Met-Glu-Val-Val-Asn-
→ → → → → → →

Peak 2.

Analysis : Asp (10), Thr (5), Ser (3), Glu (6), Pro (1), Gly (1), Ala (1), Val (2), Met (1), Ile (4), Leu (4), Tyr (3), Phe (5), Lys (6), His (1), Arg (1).

Estimated yield : 500 n moles.

Peak 3.

Analysis : Asp (5), Thr (4), Ser (2), Glu (9), Gly (1), Ala (1), Val (1), Ile (1), Leu (1), Tyr (7), Phe (4), Lys (4), His (3).

Estimated yield : 500 n moles.

Peak 4.

The peptides contained in these fractions were purified further by high voltage paper electrophoresis. A large number of lysine containing peptides were obtained in low yield, probably produced by slow unmasking of lysine residues at pH 8.0 during the tryptic digest. Only one peptide was isolated

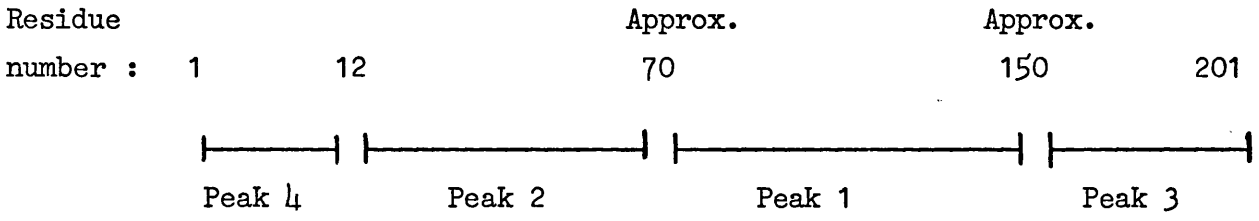
in high yield :

Analysis : Asp (3.0), Thr (0.9), Glu (1.0), Ile (1.3), Leu (1.0), Phe (1.0), Lys (1.0), Arg (1.0).

Estimated yield : 400 n moles.

This peptide corresponds to the N-terminal 12 residues of the C221 protein.

Since the peptides obtained from a digest of the citraconylated Cm CAT were only partially purified, little sequence information was obtained. Peak 3 however contains very little arginine and must contain the C-terminal fragment of the protein. Peak 4 contains the expected 12 residue N-terminal peptide, and since the N-terminal sequence of Peak 1 was determined, Peak 2 must contain the fragment starting at residue 13 of the C221 protein. This limited study of the peptides derived from tryptic digestion of citraconylated Cm CAT, although producing little sequence information, gives some indication of the distribution of amino acids throughout the C221 protein since the order of the fragments produced can be deduced :



Summation of the amino acid analysis of Peaks 1 to 4 accounts for >90 percent of the amino acid residues of the C221 CAT protein. This value is <100 percent since hydrolyses were performed for 24 hours and hydrophobic residues are not quantitively hydrolysed under these conditions.

Tryptic digestion of citraconylated C221 CAT is a particularly useful cleavage method since it specifically cleaves the polypeptide chain at arginyl residues to produce four large peptides which are ideally suited to

sequencing on the solid phase sequencer. Since the peptides produced are strongly anionic, solubility problems are minimised. The peptides could easily be purified to homogeneity by judicious pooling of the collected fractions followed by rechromatography using the P-60 gel filtration resin.

CHAPTER VI.

PRODUCTION, FRACTIONATION AND SEQUENCE OF CHYMOTRYPTIC PEPTIDES.

6.1 Chymotryptic digestion of Cm CAT.

Cm CAT (80mg) was prepared as described in Chapter V. All chymotryptic digests were performed under identical conditions to the tryptic digest described in Chapter V (50 mM ammonium bicarbonate, pH 8.0, 1:50 ratio chymotrypsin to substrate, 37°C for 5 hours). Preliminary digests using 2mg Cm CAT followed by high voltage paper electrophoresis at pH 6.5 and pH 3.5 were performed to give peptide maps. Because of the broad specificity of chymotrypsin these maps were complex and no precise estimation could be made of the number of unique peptides generated. The preparative digest was done under identical conditions using a total of 75mg Cm CAT. After 5 hours the digest was quickly frozen and lyophilised.

6.2 Fractionation of chymotryptic peptides.

The chymotryptic peptides were purified in an identical manner to the tryptic peptides. The freeze dried digest was vortexed with 5ml of the ion-exchange starting buffer (0.2 M pyridine/acetate, pH 3.1) and centrifuged. The supernatent was loaded on to a ~~Dowex 50~~ sulphonic acid ion-exchange column and eluted with a linear pH gradient (pH 3.1 to pH 5.0). Samples from each fraction collected were 'fingerprinted' by high voltage paper electrophoresis at pH 6.5 ('Methods' 2.6). At this stage it was noticed that the yield of peptides was low since the 'fingerprint' showed rather faint staining reactions with both fluorescamine and cadmium-ninhydrin reagents. The pooled fractions from the ion-exchange column were rotary evaporated and further purified by high voltage paper electrophoresis.

6.3 Amino acid analyses, characteristics and sequence of chymotryptic peptides.

When the amino acid analyses of the purified chymotryptic peptides became available it was apparent that all the peptides were present in very

low yield. Further work indicated that the chymotrypsin used for the digest had very little proteolytic activity, despite the apparent success of the pilot digest. Because of the poor digestion of the Cm CAT very little sequence information has been obtained from this digest although a large number of amino acid analyses are available.

Peptide C2 - 5 A.

Analysis : Asp (1.4), Glu (0.6), Ile (1.0), Leu (1.0), Phe (1.0).

Yield : 80 n moles.

Sequence : Asx-Leu-Ile-Phe.

Peptide C6 - 11 B.

Analysis : Pro (0.9), Ala (1.0), Val (1.0), Ile (0.9), Leu (1.0).

Yield : 100 n moles.

Sequence : Ile-Pro-Val-Ala-Leu.

Peptide C6 - 11 C a.

Analysis : Asx (1.4), Thr (2.0), Glu (2.1), Tyr (0.9), Phe (0.9).

Yield : 100 n moles.

Sequence : Phe-Asx-Glx-(Thr_2,Glu)-Tyr.

This peptide corresponds to residues 20 - 26 of the N-terminus of the protein

Peptide C6 - 11 C b.

Analysis : Ser (0.9), Glu (0.9), Pro (0.9), Ile (1.9), Leu (1.0), Tyr (0.9), Lys (1.9).

Yield : 100 n moles.

Sequence : Glx-Ile-Tyr-(Ser,Pro,Ile,Leu).

Peptide C6 - 11 D.

Analysis : Asp (2.0), Thr (1.0), Ser (1.0), Glu (0.9), Ile (1.0), Tyr (0.9), Phe (0.9), Trp +ve.

Yield : 80 n moles.

Sequence : Thr-Asx-Ile-(Phe,Trp,Asp,Ser,Glu)-Tyr.

Peptide C34 - 36 B.

Analysis : Asp (1.0), Ile (1.0), Phe (1.0).

Yield : 80 n moles.

Sequence : Ile-Asx-Phe.

Peptide C39 - 44 B.

Analysis : Asp (0.6), Thr (0.9), Ser (0.3), Val (1.0), Phe (1.2).

Yield : 100 n moles.

Sequence : Thr- Val-Phe.

Although quite contaminated the sequence obtained was unambiguous.

Peptide C47 - 51 B.

Analysis : Asp (2.0), Pro (0.9), Leu (2.0), Tyr (0.7), Lys (1.0).

Yield : 80 n moles.

Sequence : Asx-Lys-Leu-Asx-Pro-Leu-Tyr.

Peptide C53 - 57 B.

Analysis : Asp (1.2), Glu (0.9), Ala (0.9), Ile (1.0), Val (1.4), Met (0.7), Phe (0.9), Lys (1.1).

Yield : 100 n moles.

Sequence : Ala- Ile-Met-(Asp,Glu,Val_2,Lys)-Phe.

Peptide C53 - 57 D.

Analysis : Asp (0.8), Ser (1.1), Thr (1.0), Ile (0.6), Tyr (1.0), Phe (1.0).

Yield : 150 n moles.

Sequence : Asx-Ser-Phe-(Thr,Ile)-Tyr.

Peptide C58 - 61 B.

Analysis : Asp (2.0), Glu (1.0), Leu (1.9), Tyr (0.9), Lys (1.0).

Yield : 100 n moles.

Sequence : Lys-Asx-Asx-Leu-Leu-Glu-Tyr.

Peptide C87 - 89 A.

Analysis : Asp (1.0), Val (1.1), Phe (1.0), Lys (1.0).

Yield : 35 n moles.

Sequence : Asx-Val-(Lys)-Phe.

Peptide C87 - 89 B.

Analysis : Ser (1.0), Ala (1.0), Leu (1.0), Phe (0.9), His (1.0).

Yield : 50 n moles.

Sequence : His-Ala-(Ser,Leu)-Phe.

Peptide C96 - 101 A.

Analysis : Asp (1.0), Glu (1.0), Tyr (0.7), Lys (1.0), Arg (1.0).

Yield : 100 n moles.

Sequence : Asx-Arg-Lys-Glx-Tyr.

This peptide corresponds to residues 11 to 15 of the C221 protein sequence.

In addition to the chymotryptic peptides described approximately fifty peptides were isolated with yields less than 25 n moles for which no sequence information is available. The insoluble 'core' material has not yet been studied.

CHAPTER VII.

PRODUCTION, FRACTIONATION AND SEQUENCE OF ELASTASE PEPTIDES.

7.1 Derivatisation, digestion and fractionation of elastase peptides.

In order to gain information concerning the amino acid residue(s) involved in the catalytic mechanism of the CAT enzymes the protein used for the trial elastase digest was inactivated with [^{14}C] iodoacetamide prior to carboxymethylation. Iodoacetamide was chosen as an active site probe since it is readily available in radioactive form, it is a small molecule and should cause no problems in isolating peptides containing modified residues, and because both its cysteine and histidine derivatives are stable to acid treatment (save for the conversion of the amide to the free acid yielding the carboxymethyl derivatives of cysteine and histidine on amino acid analysis rather than the amidocarboxymethyl derivatives.)

Trial digest.

C221 CAT (14mg) was dialysed extensively against 50 mM Tris HCl buffer, pH 7.8, and inactivated with 5 mM [^{14}C] iodoacetamide (0.2μC/μmole) for 4 hours at 37°C. After 4 hours less than 15 percent enzyme activity remained. Excess mercaptoethanol was then added to prevent further reaction, the inactivated protein was dialysed extensively against water and lyophilised. The derivatised protein was then carboxymethylated (Chapter V, 5.1) with non-radioactive iodoacetic acid to block unreacted cysteine residues and then dialysed extensively against 50 mM ammonium bicarbonate, pH 8.0. Using this method, amino acid residues which reacted with iodoacetamide in the native protein should contain [^{14}C] label and any unreacted cysteine residues should be present as non-radioactive carboxymethyl cysteine.

The derivatised CAT protein was digested with elastase (1:100 molar ratio) for 3½ hours at 37°C and then lyophilised. The dried peptides were dissolved in 50 mM ammonium bicarbonate, pH8.0, (3ml) and fractionated using a Sephadex G-25 column (2.5 x 80cm) equilibrated with the same buffer. Aliquots (20μl)

were taken from each fraction for scintillation counting. A single major radioactive peak was obtained eluting in fractions 95 - 104. The fractions were pooled, lyophilised and the peptides further purified by high voltage paper electrophoresis, the radioactive peptides being identified by autoradiography. A single radioactive peptide (peptide 95 - 104) was obtained.

Peptide 95 - 104.

Analysis : Cm Cys (0.8), Asp (1.1), Gly (1.1), Ala (1.8), Val (1.0), Tyr (0.9), His (1.8), 3-Cm His (0.8)

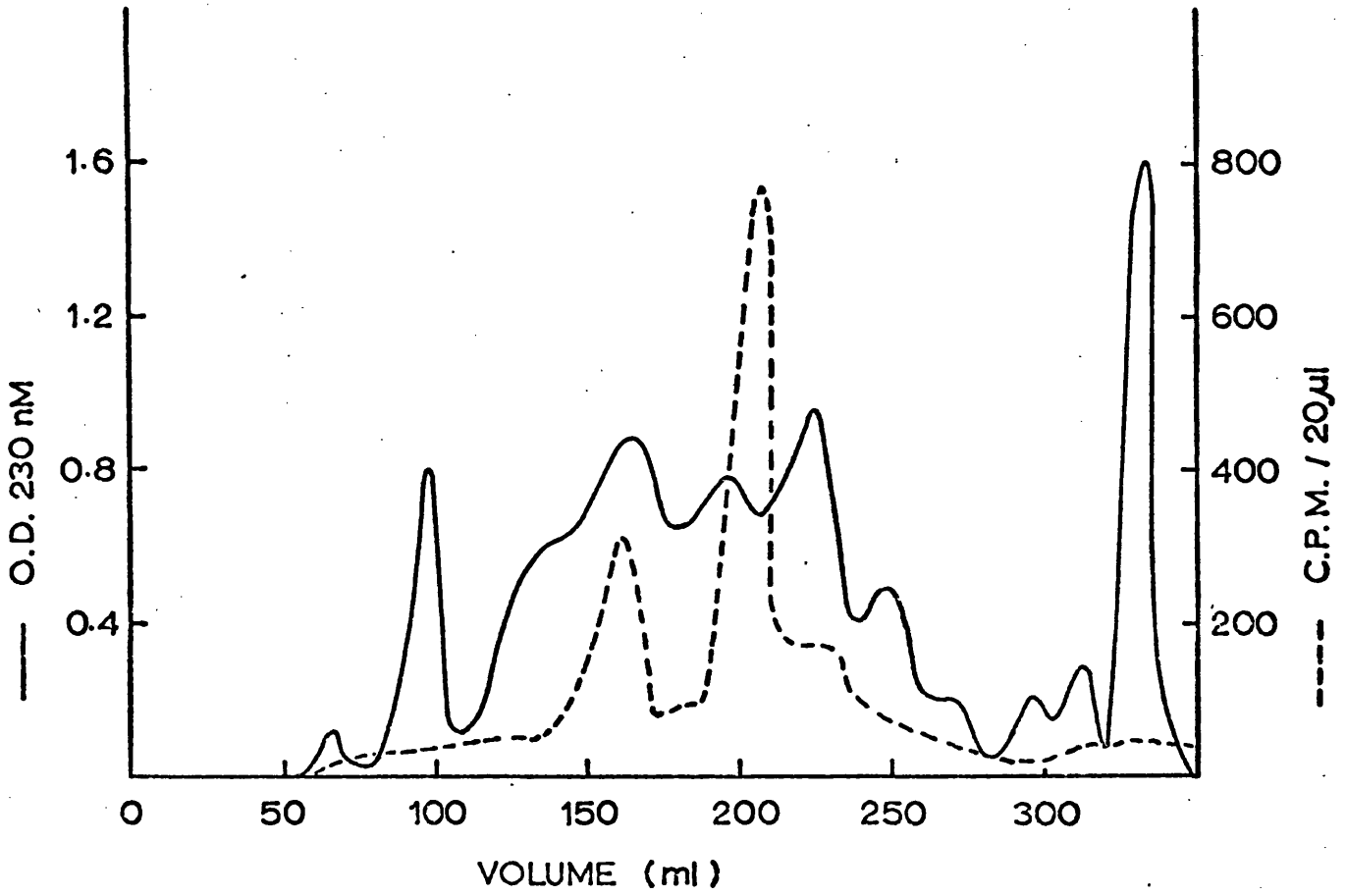
Mobility : pH 6.5 (0.0). Purif. HVPE pH 6·5, 2·1, 3·5.

Yield : 25 n moles.

Sequence : His-3-Cm His-Ala-Val-Cm Cys-Asx-Gly-Tyr-His-Ala.

At each round of the dansyl Edman degradation of this peptide the butyl acetate extractions (containing the anilinothiazoline and PTH amino acid) were counted in the scintillation counter and the polyamide plates auto-radiographed. By this method the radioactive label was found to be incorporated into the second residue of the peptide and, on the standard thin layer chromatographic system used in the identification of dansyl amino acids, runs slightly slower in solvent I than dansyl histidine. No radioactivity was found in residue five which was identified as Cm cysteine.

Carboxymethyl derivatives of histidine were prepared by reacting iodoacetic acid with histidine at pH 7.5 and purifying the products (1-Cm histidine, 3-Cm histidine and 1,3 Cm histidine) by high voltage paper electrophoresis. On amino acid analysis the 1,3 dicarboxymethyl derivative elutes prior to aspartic acid, the 1-carboxymethyl derivative after proline



Figure(16).

Elution profile of an elastase digest of type C (C221) C.A.T. from a Sephadex G-25 column (2.5 by 80cm). Eluting buffer was 50mM ammonium bicarbonate pH 8.0. Flow rate = 40ml/hr.

and the 3-carboxymethyl derivative after alanine, in agreement with the results of Crestfield, Stein and Moore (74). Both dansyl 1-Cm histidine and dansyl 3-Cm histidine chromatograph just behind histidine using the standard chromatographic system for the identification of dansyl amino acids (41).

An identical digest (digest E) was performed using 70mg CAT (figure 16). Because of the extensive use of paper electrophoresis required to purify the peptides a large number of peptides were obtained in low yield and peptides derived from this digest are only tabulated if the analogous peptides were not isolated from digest EI.

Digest EI.

In an attempt to reduce the number of high voltage paper electrophoresis steps in the purification of the elastase generated peptides the peptides were initially fractionated using a ~~Dowex 50~~ sulphonic acid. ion-exchange column rather than gel filtration.

Cm CAT (70mg) was digested with elastase (1:100 molar ratio) in 50 mM ammonium bicarbonate, pH 8.0, for 4 hours at 37°C. Digestion was terminated by lyophilisation and the peptides purified by ion-exchange chromatography and paper electrophoresis using the system devised for the purification of the tryptic peptides (Chapter V, 5.3).

7.2 Amino acid analyses, characteristics and sequence of elastase peptides.

Peptide EI 4 - 10 A.

Analysis : Ile (0.8), Leu (0.9), Phe (0.9).

Mobility : pH 6.5 (0.0), pH 2.1 (0.32) Net charge zero.

Yield : 40 n moles.

Sequence : Ile-Phe-Leu.
$\longrightarrow \longrightarrow \longrightarrow$

Peptide EI 11 - 14 A.

Analysis : Glu (1.1), Ala (1.0), Leu (2.0).

Mobility : pH 6.5 (0.0), pH 2.1 (0.67). Net charge zero.

Yield : 100 n moles.

Sequence : Ala-Leu-Gln-Leu.

Peptide EI 11 - 14 B.

Analysis : Ser (0.8), Glu (1.0), Ile (1.0), Tyr (0.9).

Mobility : pH 6.5 (+0.3), pH 2.1 (0.54). Net charge -1.

Yield : 100 n moles.

Sequence : Tyr-(Glu,Ser)-Ile.

Peptide EI 15 - 17 A.

Analysis : Asp (1.0), Glu (0.9), Val (1.2), Met (0.8), Ile (0.9).

Mobility : pH 6.5 (+0.38), pH 2.1 (0.67). Net charge -1.

Yield : 30 n moles.

Sequence : Ile-Met-Glu-Val-Val-Asn.

Peptide EI 27 - 32 B.

Analysis : Glu (1.0), Val (1.2), Met (1.0).

Mobility : pH 6.5 (+0.5), pH 2.1(0.77). Net charge -1.

Yield : 30 n moles.

Sequence : Met-Glu-Val-Val.

Peptide E 67 - 70 F.

Analysis : Glu (1.0), Val (1.0), Met (0.9).

Mobility : pH 6.5 (+0.65). Net charge -1.

Yield : 125 n moles.

Sequence : Met-Glu-Val.

Peptide EI 36 - 37 A.

Analysis : Glu (1.0), Val (1.0).

Mobility : pH 6.5 (+0.5), pH 2.1 (0.83). Net charge -1.

Yield : 90 n moles.

Sequence : Glu-Val.

Peptide EI 15 - 17 B.

Analysis : Asp (0.9), Thr (1.3), Ile (1.9), Phe (0.9).

Mobility : pH 6.5 (0.0), pH 2.1 (0.56). Net charge zero.

Yield : 80 n moles.

Sequence : Thr-Phe-Asn-Ile-Ile.

This peptide corresponds to the known N-terminal 5 residues of the protein.

Peptide EI 23 - 26 C.

Analysis : Pro (0.9), Val (1.0), Ile (1.0), Tyr (0.8).

Mobility : pH 6.5 (0.0), pH 2.1 (0.55). Net charge zero.

Yield : 100 n moles.

Sequence : Tyr-Ile-Pro-Val.

Peptide EI 27 - 32 C.

Analysis : Thr (0.9), Ser (1.0), Ile (1.0), Tyr (0.8).

Mobility : pH 6.5 (0.0), pH 2.1 (0.68). Net charge zero.

Yield : 80 n moles.

Sequence : Tyr-Phe-Ile-(Ser,Thr).

Peptide EI 36 - 37 C.

Analysis : Asp (2.0), Ile (1.0), Trp +ve.

Mobility : pH 6.5 (+0.77), pH 2.1 (0.53). Net charge -2.

Yield : 70 n moles.

Sequence : Asp-Asp-Trp-Ile.

This peptide corresponds to the C-terminal 4 residues of the protein.

Peptide EI 98 - 100 B.

Analysis : Asp (1.9), Val (1.0), Ile (0.9), His (0.8), Lys (0.9), Trp +ve.

Mobility : pH 6.5 (-0.1), pH 2.1 (1.2). Net charge $+\frac{1}{2}$.

Yield : 90 n moles.

Sequence : His-Lys-Val-Asp-Asn-Trp-Ile.

This peptide corresponds to the C-terminal 7 residues of the protein.

Peptide E 45 - 48.

This peptide, isolated from digest E, was the first peptide to be eluted from the G-25 column used in the purification of peptides from digest E and did not require further purification.

Analysis : Asp (2.8), Glu (3.2), Thr (0.7), Pro (2.0), Met (0.9), Ile (2.8), Tyr (1.0), Phe (0.9).

Mobility : remains at origin.

Yield : 400 n moles.

Sequence : Glu-Tyr-Lys-Asp-Lys-Glu-Glu-Met-Phe-Pro-Lys-(Asp_2,Glu, Thr,Pro,Ile_3,Lys).

A tryptic digest of this peptide produced three secondary peptides :

Peptide E 45 - 48 TA.

Analysis : Glu (0.9), Tyr (0.9), Lys (1.1).

Yield : 50 n moles.

Sequence : Glu-Tyr-Lys.

Peptide E 45 - 48 TB.

Analysis : Asp (1.0), Glu (2.0), Pro (0.8), Met (0.9), Phe (0.8), Lys (2.1).

Yield : 100 n moles.

Sequence : Asp-Lys-Glu-Glu-Met-Phe-Pro-Lys.

Peptide E 45 - 48 TC.

Analysis : Asp (1.7), Thr (0.8), Glu (1.1), Pro (1.3), Ile (2.0), Lys (1.1).

Yield : 80 n moles.

Sequence : Asp-Ile-Ile-Pro-(Glu,Asx,Thr)-Lys.

Peptide EI 46 - 50 A.

Analysis : Asp (2.0), Ser (1.0), Glu (1.0), Gly (1.0), Ile (1.0), Leu (1.0), Lys (1.0).

Mobility : pH 6.5 (-0.1), pH 2.1 (0.85). Net charge zero.

Yield : 200 n moles.

Sequence : Gly-Ile-Glx-Ser-(Asx_2,Lys)-Leu.

Peptide EI 46 - 50 C.

Analysis : Thr (0.6), Val (1.0), Tyr (0.8).

Mobility : pH 6.5 (0.0), pH 2.1 (0.63). Net charge zero.

Yield : 140 n moles.

Sequence : Tyr-Thr-Val.

Peptide EI 46 - 50 E.

Analysis : Cm Cys (0.9), Asp (1.1), Ser (0.8), Gly (1.1), Ala (1.0), Tyr (0.9), His (1.0).

Mobility : pH 6.5 (+0.5), pH 2.1 (0.95). Net charge -1.

Yield : 80 n moles.

Sequence : Cys-Asp-Gly-Tyr-His-Ala-Ser.

Peptide EI 51 - 54 B.

Analysis : Ile (1.0), Leu (1.0).

Mobility : pH 6.5 (0.0), pH 2.1 (0.89). Net charge zero.

Yield : 400 n moles.

Sequence : Leu-Ile.

Peptide EI 56 - 59 B.

Analysis : Asp (1.0), Ile (1.0).

Mobility : pH 6.5 (+0.58), pH 2.1 (0.73). Net charge -1.

Yield : 30 n moles.

Sequence : Asp-Ile.

Peptide EI 61 - 65 D.

Analysis : Asp (2.0), Leu (1.0), Tyr (0.7), Lys (1.0).

Mobility : pH 6.5 (0.0), pH 2.1 (0.96). Net charge zero.

Yield : 90 n moles.

Sequence : Tyr-Asx-(Asp,Lys)-Leu.

Peptide EI 61 - 65 E.

Analysis : Asp (2.0), Leu (1.0), Tyr (0.9), Phe (0.6), Lys (1.0).

Mobility : pH 6.5 (0.0), pH 2.1 (0.78). Net charge zero.

Yield : 250 n moles.

Sequence : Phe-Tyr-Asx-Asx-(Lys)-Leu.

Peptide E 61 - 66 D.

Analysis : Asp (2.0), Thr (1.0), Ser (0.8), Glu (0.9), Ile (0.8), Phe (0.9), Trp +ve.

Mobility : pH 6.5 (+0.52), pH 2.1 (0.45). Net charge -2.

Yield : 250 n moles.

Sequence : Asx-Ile-Trp-Thr-(Asp,Glu,Ser,Phe).

Peptide E 67 - 70 H.

Analysis : Asp (1.2), Thr (0.7), Ile (1.0), Trp +ve.

Mobility : pH 6.5 (+0.5),Net charge -1.

Yield : 200 n moles.

Sequence : Asp-Ile-Trp-Thr.

Peptide EI 66 A.

Analysis : Asp(1.3), Val (0.9), Lys (0.9).

Mobility : pH 6.5 (-0.5), pH 2.1 (1.31). Net charge +1.

Yield : 80 n moles.

Sequence : Val-Asn-(Lys).

Peptide EI 67 - 68 A.

Analysis : Asp (1.8), Val (1.0), Lys (0.9).

Mobility : pH 6.5 (-0.55). Net charge +1.

Yield : 200 n moles.

Sequence : Val-Asn-Lys-Asn.

Peptide EI 69 - 75 B.

Analysis : Asp (0.8), Thr (1.0), Glu (1.0), Val (0.7), Phe (1.0), Lys (1.0).

Mobility : pH 6.5 (-0.29), pH 2.1 (1.12). Net charge -1.

Yield : 30 n moles.

Sequence : Val-Phe-(Asp,Thr,Glu,Lys).

Peptide EI 70 - 73 B.

Analysis : Asp (1.0), Thr (1.9), Glu (2.0), Val (0.9), Phe (1.1), Lys (1.9).

Mobility : pH 6.5 (-0.3). Net charge -1.

Yield : 200 n moles.

Sequence : Val-Phe-(Asp,Thr_2,Glu_2,Lys_2).

Peptide EI 70 - 73 A.

Analysis : Thr (0.9), Glu (1.0), Ile (1.1), Lys (1.1).

Mobility : pH 6.5 (-0.1). Net charge zero.

Yield : 60 n moles.

Sequence : Thr-(Lys)-Glu-Ile.

Peptide EI 81 - 90 A.

Analysis : Glu (0.9), Ile (1.0), Lys (1.0).

Mobility : pH 6.5 (-0.1), pH 2.1 (1.33). Net charge zero.

Yield : 70 n moles.

Sequence : Lys-(Glu)-Ile.

Peptide E 57 - 60 B.

Analysis : Asp (1.0), Thr (0.9), Glu (1.0), Ile (2.0), Lys (0.9).

Mobility : pH 6.5 (+0.39), pH 2.1 (1.1). Net charge -1.

Yield : 100 n moles.

Sequence : Lys-Glu-Ile- (Asp,Ile,Thr).

Peptide EI 70-73 C.

Analysis : Asp (2.1), Pro (0.8), Gly (1.0), Leu (2.1), Tyr (0.9), Lys (1.1), Trp +ve.

Mobility : pH 6.5 (0.0). Net charge zero.

Yield : 100 n moles.

Sequence : Gly-Tyr-Trp-Asx-(Asp,Pro,Leu_2,Lys).

Peptide EI 98-100 C.

Analysis : Asp (1.1), Gly (0.8), Leu (1.0), Tyr (0.9), Lys (1.0), Trp +ve.

Mobility : pH 6.5 (0.0), pH 2.1 (0.96). Net charge zero.

Yield : 150 n moles.

Sequence : Gly-Tyr-(Asp,Trp,Leu,Lys).

Peptide EI 76 - 79 D.

Analysis : Thr (1.0), Tyr (0.9).

Mobility : pH 6.5 (0.0), pH 2.1 (0.8). Net charge zero.

Yield : 200 n moles.

Sequence : Tyr-Thr.

Peptide EI 76 - 79 F.

Analysis : Asp(1.0), Ser (0.8), Tyr (1.0), Phe (0.9).

Mobility : pH 6.5 (0.0), pH 2.1 (0.62). Net charge zero.

Yield : 30 n moles.

Sequence : Ser-Phe-Tyr-Asn.

Peptide E 85 - 93 B.

Analysis : Thr (1.0), Phe (1.0), Arg (1.0).

Mobility : pH 6.5 (-0.25), pH 2.1 (1.35). Net charge +1.

Yield : 250 n moles.

Sequence : Phe-Arg-Thr.

Peptide EI 98 - 100 A.

Analysis : Val (1.3), Lys (1.0).

Mobility : pH 6.5 (-0.57), pH 2.1 (1.66). Net charge +1.

Yield : 150 n moles.

Sequence : Lys-Val.

Peptide EI 101 - 127 A.

Analysis : Ala (1.0), His (2.0).

Mobility : pH 6.5 (-0.35). Net charge +1.

Yield : 40 n moles.

Sequence : His-His-Ala.

CHAPTER VIII.

PRODUCTION, FRACTIONATION AND SEQUENCE OF STAPHYLOCOCCAL PROTEASE GENERATED PEPTIDES.

8.1 Staphylococcal protease digestion of citraconylated Cm CAT.

The purification and properties of a proteolytic enzyme from culture filtrates of Staphylococcal aureus, strain V8 has been reported (67). This enzyme specifically cleaves polypeptides at glutamoyl bonds when digestion is carried out in ammonium bicarbonate buffer. Trial experiments on 1mg aliquots indicated both native and carboxymethylated CAT were poor substrates for the protease. Citraconylated Cm CAT, which is both denatured and soluble, appeared to be readily digested by the enzyme (as judged by the number and intensity of peptides stained after high voltage paper electrophoresis). Optimum conditions were found to be digestion by a 1:30 molar ratio protease : citraconylated Cm CAT, at 37°C for 20 hours in 50 mM ammonium bicarbonate buffer (pH 8.0).

Preparative digest.

Native C221 CAT (60mg) was dialysed extensively against 50 mM Tris HCl buffer, pH 7.8, to remove CM and inactivated with [^{14}C]iodoacetamide (0.2μC/μmole) as described in Chapter VII (7.1). After exhaustive dialysis against water to remove non-covalently bound iodoacetamide the protein was lyophilised, carboxymethylated (as described in Chapter V, 5.1) and finally citraconylated (Chapter V, 5.5). Thus the protein used for the s.protease digest had been labelled with [^{14}C] iodoacetamide in its native conformation, denatured, reactive groups derivatised with 'cold' iodoacetic acid and finally citraconylated to provide a suitable substrate for the proteolytic enzyme. The citraconylated Cm CAT (4mg/ml) was dialysed against 50 mM ammonium bicarbonate, pH 8.0, and digested with a 1:30 molar ratio s.protease at 37°C for 20 hours. The digest was then lyophilised and the citraconyl masking groups removed by incubating the digest with dilute acetic acid, pH 4.0, at 40°C for 4 hours.

8.2 Fractionation of s.protease peptides.

The deblocked peptide mixture was lyophilised, dissolved in 3ml of 50 mM ammonium bicarbonate, pH 8.0, and loaded on to a Sephadex G-25 column (2.5 x 80cm) previously equilibrated with 50 mM ammonium bicarbonate (figure 17). Samples (100μl) were taken from each fraction collected and 'fingerprinted' by high voltage paper electrophoresis at pH 6.5 ('Methods' 2.6). From the optical densities at 230 nM and 280 nM and the 'fingerprints', the collected fractions were pooled judiciously and then lyophilised. The peptides were further purified by high voltage paper electrophoresis ('Methods' 2.5).

8.3 Amino acid analyses, characteristics and sequence of s.protease peptides.

Peptide SP 60 - 67 A.

Analysis : Asp (2.8), Thr (0.9), Ser (1.0), Glu (0.9), Gly (1.2), Val (2.0), Ile (1.8), Phe (0.9), Lys (3.0), Arg (1.1).

Yield : 400 n moles.

Sequence : Val-Val-Asx-(Asp,Thr,Ser,Gly,Ile_2,Lys_3,Arg)-Glu.

This peptide remained at the origin during high voltage paper electrophoresis at pH 6.5 and was purified by running at pH 2.1.

Peptide SP 60 - 67 B.

Analysis : Cm Cys (0.9), Asp (4.0), Thr (1.9), Ser (1.9), Glu (2.1), Pro (0.9), Gly (2.1), Ala (1.6), Val (2.0), Ile (2.1), Leu (2.0), Tyr (0.8), Phe (1.0), Lys (2.0), His (3.1), Arg (1.1), 3-Cm His (0.8).

Yield : 70 n moles.

Sequence : Thr-Asx-Lys-(Asp_2,Thr,Ser_2,Glu,Pro,Val,Gly,Ile_2,Leu_2,Phe, His,Lys,Arg[His,3-Cm His,Ala,Val,Cm Cys,Asp,Gly,Tyr,His, Ala])-Glu

Peptide SP 60 - 67 B remained at the origin during high voltage paper electrophoresis at pH 6.5 and was purified at pH 2.1. This peptide was the primary [^{14}C] containing peptide obtained from the s.protease digest and contained 0.8 moles 3-Cm histidine (assuming a ninhydrin colour factor equivalent to that of glycine) per mole of peptide. Dansylation of a hydrolysed sample of the peptide followed by polyamide thin layer chromatography and autoradiography of the plate indicated that the [^{14}C] label was incorporated into Cm-histidine and not Cm cysteine. This peptide probably contains the sequence obtained for the elastase peptide 95 - 104 (Chapter VII), which had the sequence :

His-3 Cm His-Ala-Val-Cm Cys-Asp-Gly-Tyr-His-Ala.

Peptide SP 68 - 76 A a.

Analysis : Asp (2.1), Thr (0.8), Ser (1.0), Glu (1.9), Pro (1.9), Gly (1.2), Val (1.8), Met (1.4), Ile (2.0), Phe (1.8), Lys (3.0), Arg (1.0).

Yield : 400 n moles. Purif. HVPE pH 2·1

Sequence : Met-Phe-Pro-Lys-Asx-Ile-Ile-(Asp,Thr,Ser,Glu,Pro,Met,Gly, Val_2,Phe,Lys_2,Arg)-Glu.

Peptide SP 77 - 86 B.

Analysis : Asp (1.0), Glu (1.2), Tyr (0.8), Lys (1.9).

Yield : 300 n moles. Purif. HVPE pH 2·1.

Sequence : Tyr-Lys-Asx-Lys-Glu.

Peptide SP 77 - 86 F.

Analysis : Ser (1.0), Glu (1.0), Pro (0.9), Ala (0.9), Met (1.0), Ile (1.9), Leu (0.9), Tyr (1.1), Arg (1.1).

Yield : 300 n moles. Purif. HVPE pH 2·1.

Sequence : Ile-Tyr-Pro-Ser-Leu-Ile-Arg-Ala-Ile-Met-Glu

Peptide SP 89 - 97.

Analysis : Asp (1.0), Thr (1.8), Glu (1.0), Ile (0.9), Phe (0.9), Trp +ve.

Yield : 250 n moles.

Sequence : Lys-Phe-Thr-Asx-Ile-Trp-Thr-Glu.

Peptide SP 99 - 107.

Analysis : Asp (1.9), Glu (1.0), Lys (1.0), Arg (0.9), Trp +ve.

Yield : 500 n moles.

Sequence : Asn-Trp-Asp-Arg-Lys-Glu.

This peptide corresponds to residues 9 - 14 of the N-terminal sequence of the protein.

Peptide SP 108 - 115.

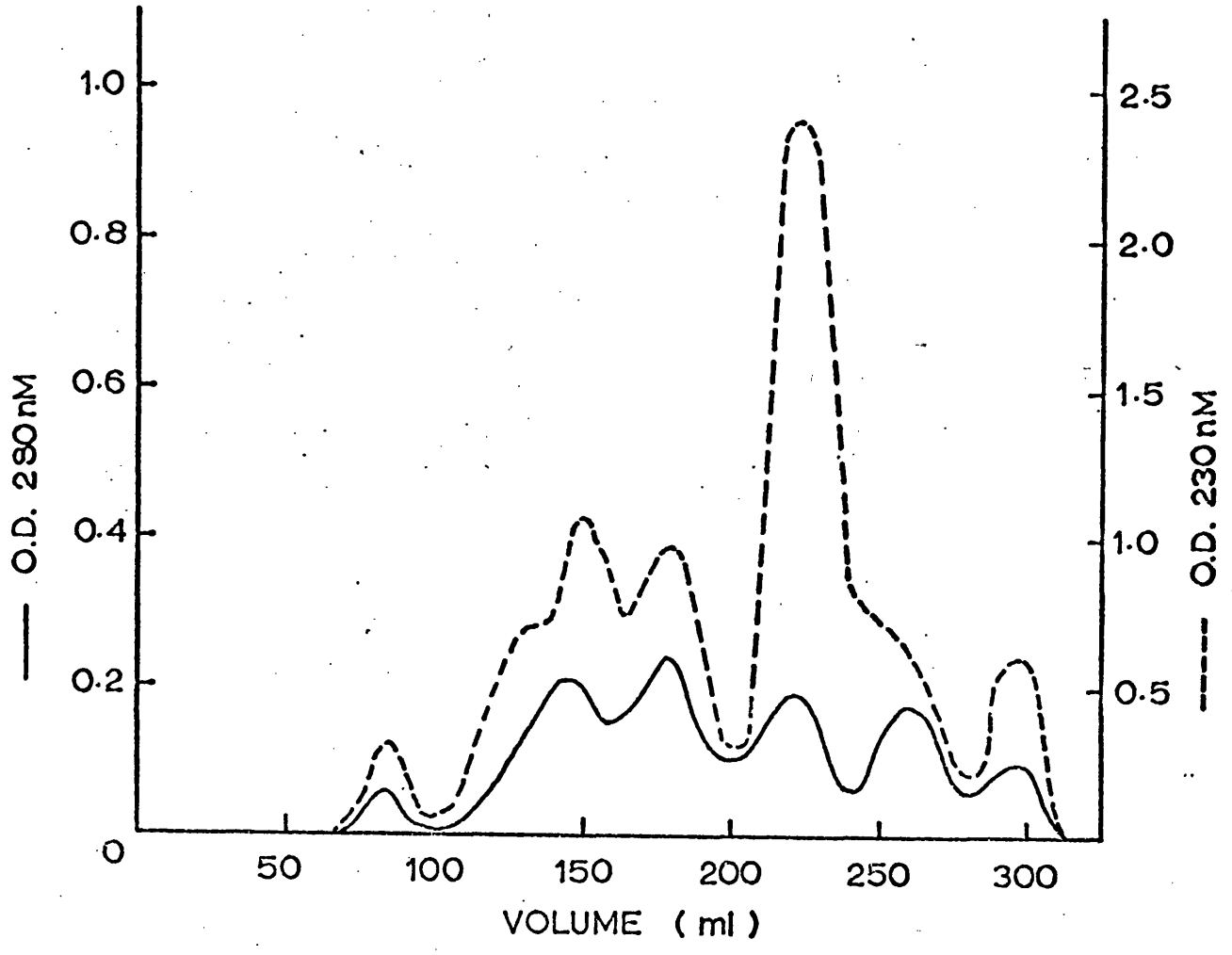
Analysis : Glu (1.0), Tyr (0.9), Phe (0.9).

Yield : 400 n moles. Purif. HVPE pH 3.5.

Sequence : Tyr-Phe-Glu.

This peptide corresponds to residues 15 - 17 of the protein sequence.

Although the peptides described above were obtained in relatively good yield a large number of peptides (not yet studied) were present in low yield. Whether this was due to the protease cleaving the polypeptide at aspartic acid residues (as is said to occur when digestion is carried out in phosphate buffers) or whether the citraconyl groups, which carry a negative charge, interfered with the 'normal' specificity of the enzyme, is not yet known. If the latter possibility is the case, the use of s. protease may be limited to cleavage of proteins which are substrates when in their native conformation or which are soluble in ammonium bicarbonate buffers after denaturation.



Figure(17).

Elution profile of Staph. protease (V8) digest of type C (C221) C.A.T. from a Sephadex G-25 column (2.5 by 80cm). eluting buffer was 50mM ammonium bicarbonate pH 8.0. flow rate = 40ml/hr.

CHAPTER IX.

PRODUCTION, FRACTIONATION AND SEQUENCE OF PEPTIDES GENERATED BY CYANOGEN BROMIDE CLEAVAGE.

9.1 Cyanogen bromide cleavage of Cm CAT.

Cyanogen bromide reacts with the thioether side chain of methionine and, when methionine is in a peptide linkage, the bond involving the carboxyl group of methionine is cleaved. All peptides produced by the action of cyanogen bromide will contain C-terminal homoserine or its lactone except for the C-terminal peptide of the protein.

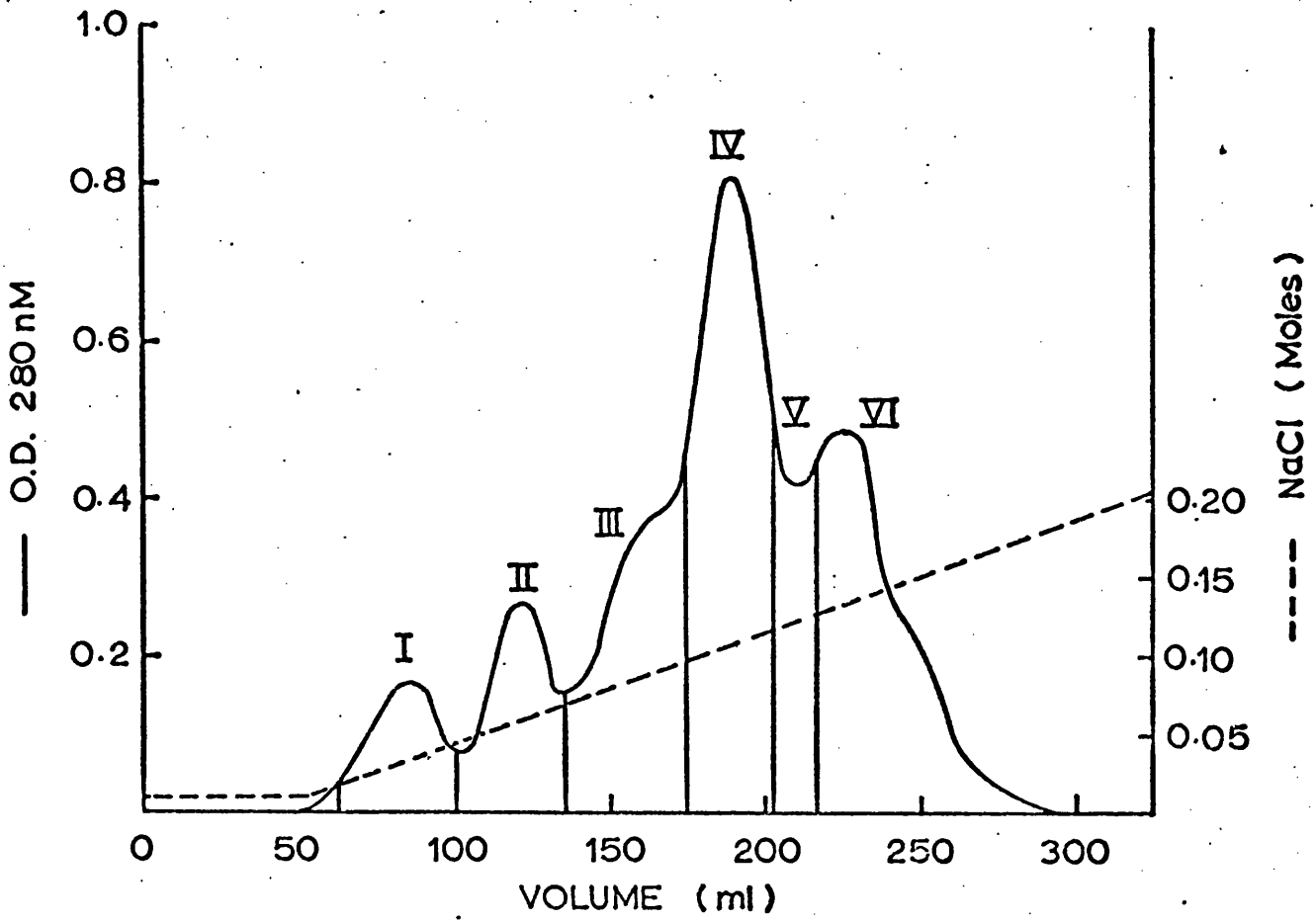
Trial digests were performed on 1mg aliquots of Cm CAT to determine the optimal cleavage conditions. Cleavage was assessed by S.D.S. gel electrophoresis. From the results of these preliminary experiments the best cleavage conditions were found to be protein concentration >5mg/ml, cleavage time 18 hours and a 1:50 molar ratio of methionine (5 methionyl residues per monomer) : cyanogen bromide. 70 percent formic acid was the solvent used in all trial experiments.

Preparative digest.

90mg of Cm CAT (prepared as described in Chapter V, 5.1) was dissolved in 70 percent formic acid (8ml) and 100mg CNBr added. The flask was flushed with nitrogen, sealed and kept in the dark at room temperature for 18 hours. Water (100ml) was then added and the digest lyophilised.

9.2 Fractionation of CNBr peptides.

Initial problems encountered in the fractionation of the CNBr peptides were due to the extreme insolubility of the peptides and their tendency to aggregate. The only solvent capable of dissolving the lyophilised digest was 98 percent formic acid. However, when this solution was diluted prior to fractionation by gel filtration (column buffer 30 percent acetic acid, 30 percent formic acid) the peptides precipitated. To date, the following purification procedure has been the only method capable of fractionating the CNBr peptides.



Figure(18).

Elution profile of a cyanogen bromide digest of type C (C221) C.A.T. from a carboxymethyl-cellulose column (1.0 by 20cm) equilibrated with 8M urea/20mM ammonium acetate pH 5.0. Flow rate = 12ml/hr.

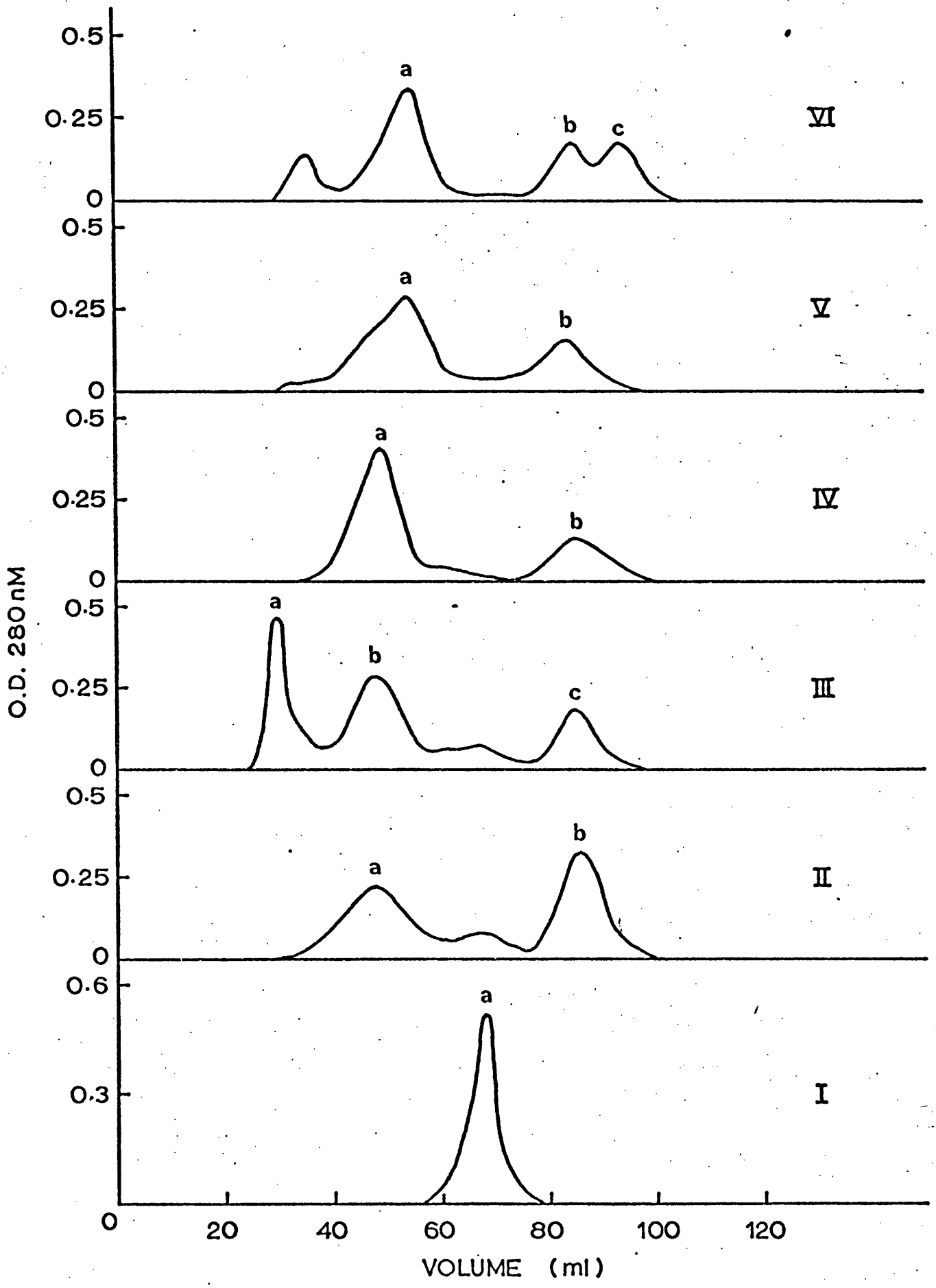
figure : 19

Elution profile of cyanogen bromide pooled fractions from Cm - cellulose ion-exchange column (I to VI).

Column = Sephadex G-75 (1 x 80cm) equilibrated and run in 5 percent acetic acid.

Flow rate = 10ml/hour.

1.5ml fractions collected.


0.5
0.25
0
a
b
c
VI
0.5
0.25
0
a
b
V
0.5
0.25
0
a
b
IV
0.5
0.25
0
a
b
c
III
0.5
0.25
0
a
b
II
0.6
0.3
0
a
I
O.D. 280nM
0
20
40
60
80
100
120
VOLUME (ml)

The lyophilised CNBr peptides were dissolved in 1ml formic acid (98 per cent). This was made up to 5ml with 8 M urea/20 mM ammonium acetate, pH 5.0, and dialysed against the same for 8 hours to remove the formic acid. The slight precipitate was removed by centrifugation and the supernatant loaded on to a carboxymethyl cellulose ion-exchange column (1 x 20cm) equilibrated with 8 M urea in 20 mM ammonium acetate, pH 5.0. The CNBr peptides were eluted by a linear gradient of NaCl (0 - 0.25 M). The eluted fractions were pooled as shown (figure 18) and the urea removed by dialysis against acetic acid (5 percent) using boiled Spectropore 3 dialysis membrane. The dialysed fractions were then lyophilised and redissolved in 1ml of 5 percent acetic acid. The partially purified CNBr peptides were further fractionated by gel filtration using Sephadex G-75 (1 x 80cm) equilibrated with 5 percent acetic acid (figure 19).

From the elution position from the G-75 column and N-terminal determination ('dansyl' Edman method) peaks II a, III b and IV a were judged to contain the same peptide and were pooled together. Peak V a was thought to consist of a mixture of IV a and VI a and has not been studied further. The peaks which eluted near the Vt of the column (peaks II b, III c, IV b, V b, VI b and VI c) appeared to consist of a large number of low yield peptides and/or free amino acids, possibly formed by acid cleavage of the polypeptide during the cleavage reaction, and have not been studied in detail.

9.3 Amino acid analyses, characterisation and partial sequence of CNBr peptides.

Peak I a.

Analysis : Asp (2.0), Thr (0.7), Ser (0.8), Glu (1.2), Pro (1.7), Ile (2.2), Phe(0.9), Hse (0.9), Lys (1.7).

Yield : 400 n moles.

Sequence : Phe-Pro-Lys-Asn-Ile-Ile-Pro-Glu-Ser-Asn-Thr-Lys-Hse.
→ → → → → → → → → → → → ←

As yet this sequence is tentative since more than one PTH amino acid was identifiable at positions 5,7,8 and 9 of this peptide and the peptide has a considerable absorbance at 280 nm although the sequence contains no aromatic amino acid.

Peaks II a, III b, IV a.

Analysis : Asp (7.0), Thr (3.1), Ser (2.0), Glu (4.1), Pro (0.9), Gly (1.2), Val (1.1), Ile (2.0), Leu (2.0), Tyr (2.0), Phe (1.9), Lys (4.1), Arg (1.0), Hse (1.0).

Yield : 400 n moles.

Sequence : Glu-Val-Val-Asn-
→ → → →

The N-terminal sequence of this peptide is also found in peptide SP60 - 67 A and in 'Peak I' derived from the tryptic digest of citraconylated Cm CAT (Chapter V, 5.7). In all cases no sequence could be obtained beyond the asparagine residue.

Peak III a.

Analysis : Asp (8.1), Thr (3.9), Ser (2.0), Glu (5.0), Pro (1.0), Gly (1.2), Val (1.0), Ile (4.1), Leu (3.0), Tyr (2.9), Phe (3.1), Lys (6.0), Arg (1.0), Hse (1.0).

Yield : 200 n moles.

Sequence : Thr-Phe-Asn-Ile-Ile-
→ → → → →

This peptide was concluded to be the N-terminal CNBr peptide of the C221 protein.

Peak VI a.

Analysis : Asp (5.9), Thr (2.8), Ser (2.1), Glu (3.1), Gly (1.2), Val (1.0), Ile (2.6), Leu (2.2), Tyr (2.0), Phe (1.8), Lys (4.2), Arg (0.8), Hse (1.0).

Yield : 300 n moles.

Sequence : Ile-Lys-Asn-Lys-Gly-Tyr-Glu-
→ → → → → → →

The following peptides have been isolated from a tryptic digest of this peptide :

Peptide VI a TA.

Analysis : Asp (1.0), Lys (1.0).

Yield : 50 n moles.

Sequence : Asn-Lys.
→ →

This peptide had a positive net charge on electrophoresis at pH 6.5 and must therefore contain asparagine.

Peptide VI a TB.

Analysis : Thr (0.9), Glu (1.1), Val (1.0), Phe (1.1), Arg (1.0)

Yield : 20 n moles.

Peptide VI a TC b.

Analysis : Asp (1.0), Glu (2.1), Lys (1.1), Hse (1.0).

Yield : 20 n moles.

From the analysis this appeared to be the C-terminal fragment of the CNBr peptide and is included in the tryptic peptide TS 72 - 75 C as follows :

Asp-Lys-Glu-Glu-Met-Phe-Pro-Lys.

As was noted previously, cleavage on the C-terminal side of lysine did not take place, probably as a result of the density of negative charge from the three adjacent acidic amino acids.

Peptide VI a TD a.

Analysis : Asp (2.0), Thr (0.9), Ser (1.0), Glu (1.0), Gly (1.0), Ile (1.1), Lys (1.1).

Yield : 25 n moles.

This peptide corresponds to peptide TS 43 - 46 A.

Asn-Gly-Ile-Glu-Ser-Asn-Thr-Lys.

Peptide VI a TD c.

Analysis : Asp (2.1), Glu (1.0), Leu (2.1), Tyr (0.8), Lys (1.0).

Yield : 20 n moles.

This peptide corresponds to TS 59 - 63 D.

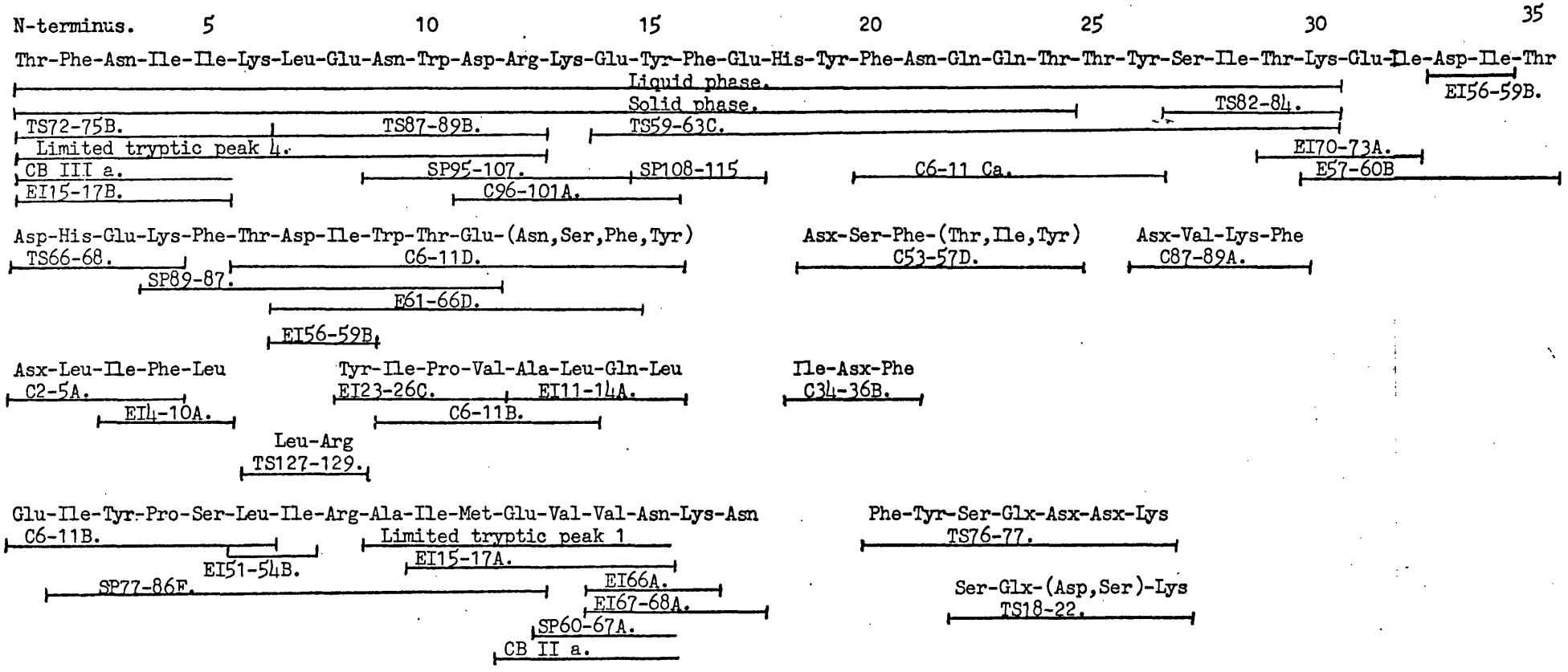
Asn-Asp-Leu-Leu-Glu-Tyr-Lys.

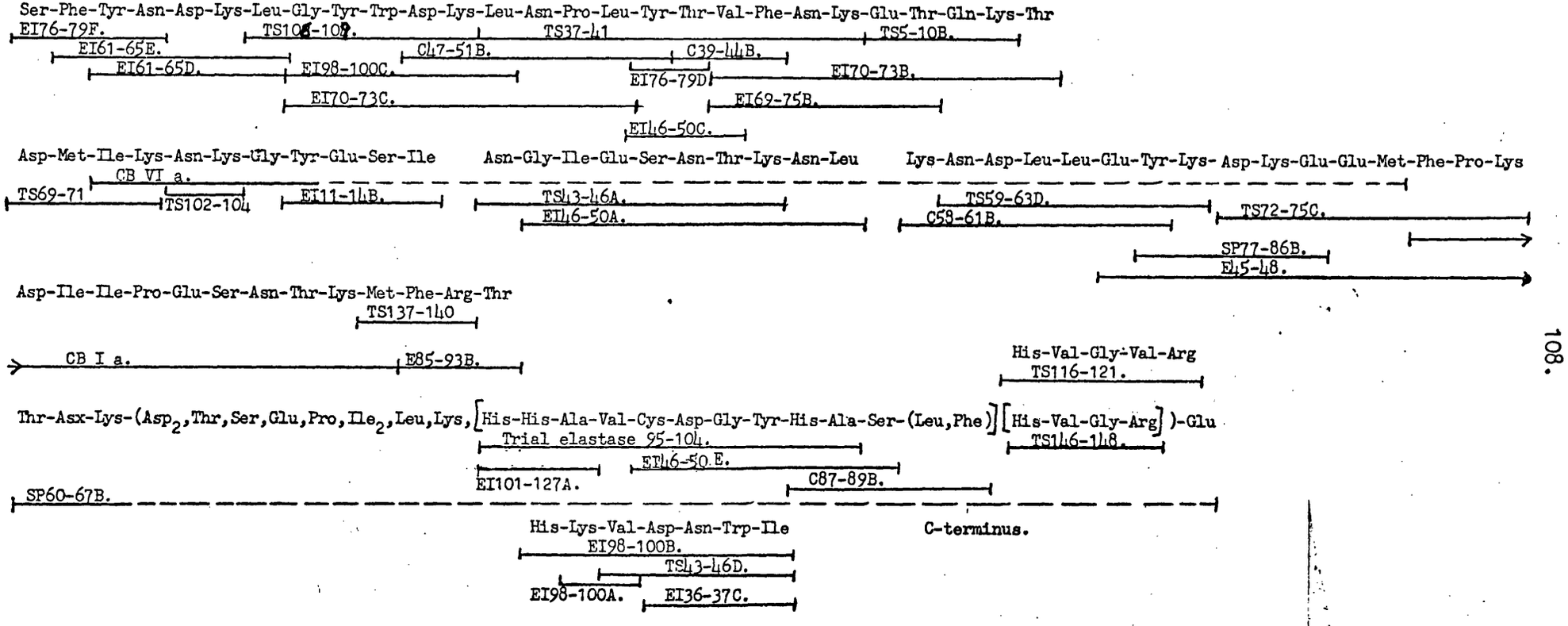
CHAPTER X.

THE PARTIAL PRIMARY STRUCTURE OF C221 CAT AND ITS COMPARISON WITH OTHER CAT SEQUENCES.

figure :. 20.

Summary of the covalent sequence of type C (C221) staphylococcal CAT determined in this study.


N-terminus.
5
10
15
20
25
30
35
Thr-Phe-Asn-Ile-Ile-Lys-Leu-Glu-Asn-Trp-Asp-Arg-Lys-Glu-Tyr-Phe-Glu-His-Tyr-Phe-Asn-Gln-Gln-Thr-Thr-Tyr-Ser-Ile-Thr-Lys-Glu-Ile-Asp-Ile-Thr
Liquid phase.
EI56-59B.
Solid phase.
TS82-84.
TS72-75B.
TS87-89B.
TS59-63C.
Limited tryptic peak 4.
EI70-73A.
CB III a.
SP95-107.
SP108-115
C6-11 Ca.
E57-60B
EI15-17B.
C96-101A.
Asp-His-Glu-Lys-Phe-Thr-Asp-Ile-Trp-Thr-Glu-(Asn,Ser,Phe,Tyr)
Asx-Ser-Phe-(Thr,Ile,Tyr)
Asx-Val-Lys-Phe
TS66-68.
C6-11D.
C53-57D.
C87-89A.
SP89-87.
E61-66D.
EI56-59B.
Asx-Leu-Ile-Phe-Leu
Tyr-Ile-Pro-Val-Ala-Leu-Gln-Leu
Ile-Asx-Phe
C2-5A.
EI23-26C.
EI11-14A.
C34-36B.
EI4-10A.
C6-11B.
Leu-Arg
TS127-129.
Glu-Ile-Tyr-Pro-Ser-Leu-Ile-Arg-Ala-Ile-Met-Glu-Val-Val-Asn-Lys-Asn
Phe-Tyr-Ser-Glx-Asx-Asx-Lys
C6-11B.
Limited tryptic peak 1
TS76-77.
EI51-54B.
EI15-17A.
EI66A.
SP77-86F.
Ser-Glx-(Asp,Ser)-Lys
EI67-68A.
TS18-22.
SP60-67A.
CB II a.


Ser-Phe-Tyr-Asn-Asp-Lys-Leu-Gly-Tyr-Trp-Asp-Lys-Leu-Asn-Pro-Leu-Tyr-Thr-Val-Phe-Asn-Lys-Glu-Thr-Gln-Lys-Thr
EI76-79F.
TS10[illegible]-10[illegible].
TS37-41
TS5-10B.
EI61-65E.
C47-51B.
C39-44B.
EI61-65D.
EI98-100C.
EI76-79D
EI70-73B.
EI70-73C.
EI69-75B.
EI46-50C.
Asp-Met-Ile-Lys-Asn-Lys-Gly-Tyr-Glu-Ser-Ile
Asn-Gly-Ile-Glu-Ser-Asn-Thr-Lys-Asn-Leu
Lys-Asn-Asp-Leu-Leu-Glu-Tyr-Lys- Asp-Lys-Glu-Glu-Met-Phe-Pro-Lys
CB VI a.
TS69-71
TS102-104
EI11-14B.
TS43-46A.
EI46-50A.
TS59-63D.
C58-61B.
TS72-75C.
SP77-86B.
E45-48.
Asp-Ile-Ile-Pro-Glu-Ser-Asn-Thr-Lys-Met-Phe-Arg-Thr
TS137-140
CB I a.
E85-93B.
His-Val-Gly-Val-Arg
TS116-121.
Thr-Asx-Lys-(Asp2,Thr,Ser,Glu,Pro,Ile2,Leu,Lys,[His-His-Ala-Val-Cys-Asp-Gly-Tyr-His-Ala-Ser-(Leu,Phe)][His-Val-Gly-Arg])-Glu
Trial elastase 95-104.
TS146-148.
EI46-50 E.
EI101-127A.
C87-89B.
SP60-67B.
His-Lys-Val-Asp-Asn-Trp-Ile
C-terminus.
EI98-100B.
TS43-46D.
EI98-100A.
EI36-37C.

10.1 Peptide overlaps.

A summary of the peptide overlaps obtained for the staphylococcal type C enzyme are illustrated in figure 20. Although more than 90 percent of the 201 amino acid residues expected from the amino acid analysis of the protein can be found in the peptides shown, a number of overlaps are, as yet, tentative and others have yet to be elucidated. A study of some of the low yield peptides from several digests indicated heterogeneity of the protein, in spite of the 'clean' protein N-terminal sequence determinations and the electrophoretic evidence of purity. Perhaps the best example of heterogeneity is shown by the tryptic peptides TS 146 - 148 and TS 116 - 121 :

Peptide TS 146 - 148.

Sequence : His-Val-Gly-Arg.

Yield : 450 n moles.

Peptide TS 116 - 121.

Sequence : His-Val-Gly-Val-Arg.

Yield : 150 n moles.

Since peptide TS 116 - 121 differs from peptide TS 146 - 148 by the insertion of a single valine residue, such a small difference in protein structures would not be detectable by most, if not all, methods used in judging protein purity. Other possible examples of heterogeneity are peptides TS 76 - 77 and TS 18 - 22 :

Peptide TS 76 - 77.

Sequence : Phe-Tyr-Ser-Glx-Asx-Asx-Lys. (2 amides)

Yield : 700 n moles.

Peptide TS 18 - 22.

Sequence : Ser-Gln-(Asp,Ser)-Lys.

Yield : 300 n moles.

suggesting an Asx to Ser substitution, and the sequences :

-Pro-Ser-Leu-Ile-Arg-

(included in peptide SP 77 - 86 F) and the low yield tryptic peptide TS 127 - 129 : Leu-Arg.

suggesting deletion of a single isoleucine residue or a leu for ile substitution and cleavage between the two leu residues.

Since the C221 plasmid exists as multicopy replicons in S. aureus (52) mutation of one (or more) plasmids at one (or more) sites would produce a family of homologous proteins. Since the enzyme(s) were purified by affinity chromatography it is likely that the proteins are functionally equivalent and are all capable of catalysing the acetylation of chloramphenicol. Since the N-terminal determinations on protein samples showed no evidence of heterogeneity it is likely that such heterogeneity arises from substitutions at a limited number of sites where changes in structure can be accommodated without loss of activity.

10.2 Active site studies.

Since the R-factor type II CAT variants are very sensitive to inhibition by the thiol reagent DTNB (15), serious consideration has been given to the possibility that all of the CAT enzymes function via the formation of a covalent acyl-enzyme intermediate involving an essential cysteine residue. However, the results of the kinetic experiments on the staphylococcal variants (Chapter III, 3.2) suggest that binding of the two substrates is random. Double reciprocal plots showing the effects of concentrations of one substrate with the other held constant gave intersecting lines on the $1/[S]$ axis in each case; the parallel lines expected for a 'ping-pong' mechanism, such as that involving an acyl-enzyme intermediate, were not observed. These results are consistent with a kinetic analysis of the type I (R-factor JR66) enzyme (15).

Chemical modification experiments (Chapter III, 3.3) show no evidence for an essential cysteine residue since the thiol specific reagents DTNB and NTCB do not inactivate the staphylococcal enzymes.

Notwithstanding these results, an attempt was made to isolate an acyl-enzyme intermediate by an adaptation of the method dsecribed by Henkin and Abeles (68):

2mg (80 nanomoles) of the type C (plasmid C221) CAT in 1ml was dialysed extensively against 50 mM Tris HCl, pH 7.5, and $[^{14}C]$ AcCoA (25μC/μmole) added to give a final concentration of 1 mM. DTNB (1 mM final concentration) was added to act as a chemical 'scavenger' to remove CoASH in order to prevent deacylation of any acyl-enzyme intermediate formed. After 10 minutes at room temperature, the pH was lowered by the addition of HCl to pH 6.0 and the enzyme immediately applied to a Sephadex G-25 column equilibrated with 50 mM potassium phosphate buffer, pH 6.0, maintained at 4°C. Only 3.6 percent of the protein eluted from the column was found to contain $[^{14}C]$ label, this small amount of label incorporated being attributed to non specific adsorption or entrapment (as noted by Henkins and Abeles) of radioactive AcCoA by the protein (68).

Whereas the results of the kinetic experiments and chemical modification experiments show neither evidence for an acyl-enzyme intermediate nor the direct involvment of a cysteine residue in the mechanism of catalysis, the results do implicate the involvment of histidine residue(s). Thus, diethyl-pyrocarbonate is a potent inhibitor of the staphylococcal CAT variants, the pH dependence of iodoacetamide inactivation gives an apparent pK of 6.3, and the photo-oxidation pH profile suggests a pK of 6.9 for the labile group. The observation that iodoacetic acid does not inactivate the staphylococcal variants (or, indeed, all CAT variants studied) whereas iodoacetamide does

may indicate the presence of an ionized aspartate or glutamate residue at or near the active site.

In an attempt to identify the amino acid residue(s) involved in catalysis the reagent [^{14}C] iodoacetamide has been used extensively. When a type C staphylococcal variant (plasmid C221) was inactivated with [^{14}C] iodoacetamide, 0.99 moles of [^{14}C] iodoacetamide were incorporated per mole enzyme monomer, indicating that the reagent was reacting with a single residue in each enzyme monomer. Amino acid analysis of iodoacetamide inactivated enzyme showed the formation of 0.7 moles of 3-Cm histidine per mole enzyme monomer. No Cm-cysteine was identified in the analysis. A single radioactive peptide was isolated from an elastase digest of [^{14}C] iodoacetamide inactivated C221 CAT (see Chapter VII) and had the sequence :

His-3-Cm His-Ala-Val-Cm Cys-Asp-Gly-Tyr-His-Ala.

the [^{14}C] label being incorporated into residue 2 of the peptide.

In summary, cysteine residues appear to play no obvious part in the mechanism of catalysis of the type C staphylococcal enzyme, whereas a unique histidine residue has been implicated.

When the analogous experiments were performed on a type I (R-factor JR66) CAT variant it was found that 1.97 moles of [^{14}C] iodoacetamide were incorporated per mole enzyme monomer and amino acid analysis of the inactivated protein showed 0.7 moles of 3-Cm histidine per mole of enzyme monomer and 1.3 moles Cm cysteine per mole of enzyme monomer.

Two radioactive peptides have been obtained in high yield from chymotryptic digests of [^{14}C] iodoacetamide inactivated type I (R-factor) CAT (73):

	31	32	33	34	35	36	37	38
1.	Gln-	Ser-	Val-	Ala-	Gly-	Cys-	Thr-	Tyr.

178 179 180 181 182 183 184
2. His-Ala-Val-Cys-Asp-Gly-Phe

As yet the exact location of the [^{14}C] label has not been identified; however, from the finding that Cm-cysteine and 3-Cm histidine are found in amino acid analyses of hydrolysates of the inactivated protein and, by analogy with the results of the staphylococcal enzyme, it can be expected that the [^{14}C] label is localised in cysteine 36 and histidine 178. Indeed, a [^{14}C] labelled peptide from an elastase digest of [^{14}C] iodoacetamide inactivated type I CAT has been isolated which was sequenced as :

177 178 179 180
His-His-Ala-Val

Since dansyl 3-Cm histidine co-chromatographs with histidine, residue 178 was identified as histidine rather than 3-Cm histidine.

10.3 Primary sequence comparisons within the CAT system.

At present the N-terminal sequences of ten CAT variants is known (figure 21), the complete sequence of a type I (R-factor JR66) CAT has been determined (figure 22) and greater than 90 percent of the sequence of a type C staphylococcal variant has been determined in the course of this study (figure 20) enabling a comparison of the primary structures of a number of enzymes which catalyse the acetylation of chloramphenicol.

N-terminal comparisons (figure 21).

When the N-terminal sequences are aligned to give maximum obvious sequence homology only two amino acids are found to be invariant, residues 16 (tryptophan) and 18 (arginine), using the type I CAT (R-factor JR66) numbering scheme. The N-terminus of the Streptomyces acrimycini variant is the only exception. The N-terminal sequence of this enzyme is unusual in that four proline residues are found in the first fifteen residues and it should be remembered that this variant is an example of the 'larger' CAT

enzymes, having a monomer molecular weight of 25,000 daltons rather than the more common molecular weight of 24,000 daltons (20). The most obvious interpretation of these observations is that the S. acrimycini variant owes its greater molecular weight at least in part to the additional residues at the N-terminus of the protein. It should be noted, however, that the type III variant (R-factor R387) is also a high molecular weight variant but this enzyme shows homology with the other N-termini.

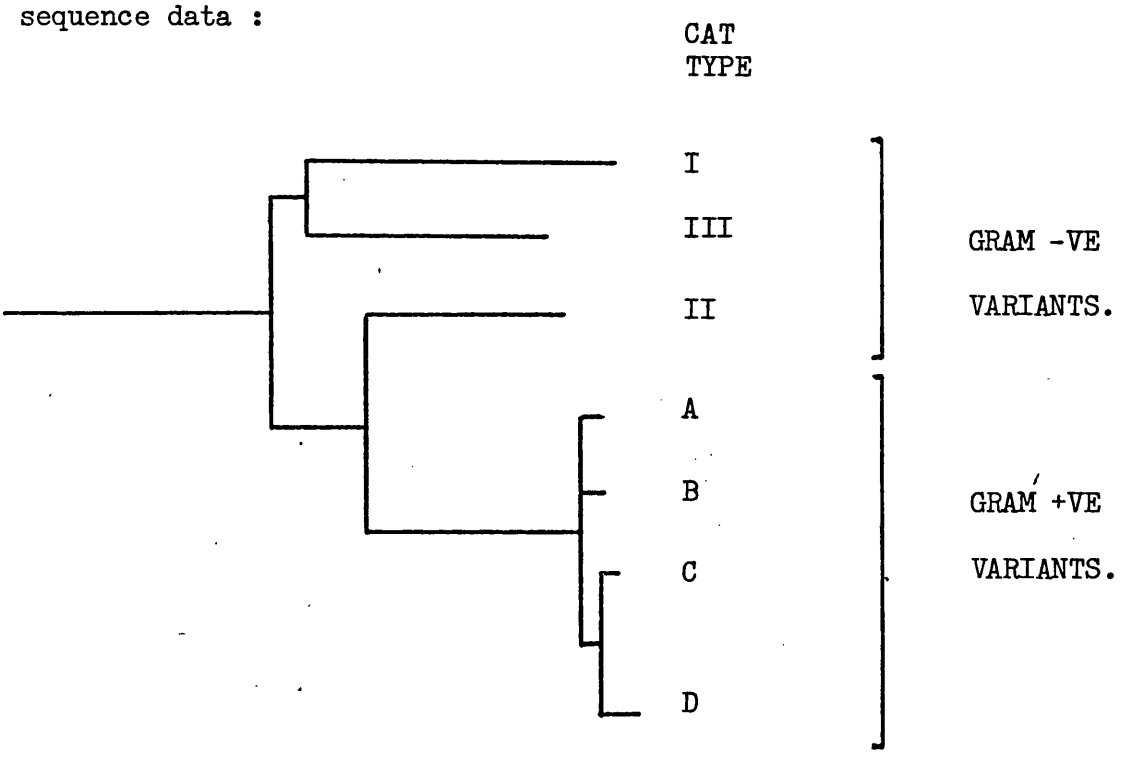
The two type I variants (R-factors JR66 and R429) are identical, as are the two type II variants (that specified by R-factor S-a and the enzyme from H. parainfluenzae). All the variants from the Enterobacteriaceae have an N-terminal methionine residue. When the N-terminal sequences are aligned to give maximal homology it can be seen that the type I variants possess an additional five residues at the N-terminus.

As had been anticipated from the results of immunological studies (15, 16, 17) the staphylococcal variants show a high degree of sequence homology, only two variable positions (residues 12 and 15 using the type I numbering scheme) being found.

Although the lengths of N-terminal sequences obtained to date are too short to allow a rigorous statistical analysis an attempt has been made to produce a dendrogram to show the possible relationships of the CAT variants using a computer programme based on the method of Moore, Goodman and Barnabas (69). A dissimilarity matrix was first constructed for the three R-factor variants and the four staphylococcal variants :

	I	II	III	A	B	C	D
I	0						
II	77	0					
III	68	68	0				
A	88	50	83	0			
B	88	50	83	6	0		
C	88	68	75	6	6	0	
D	88	68	75	13	13	6	0

The figures are the percentage non-homology when comparing any pair of CAT variants. The N-terminal six residues of the type I CAT and the N-terminal methionine residue of the types II and III CAT sequences were not included in the sequence comparison. The results of the computer analysis indicated that the following dendrogram best represented the N-terminal sequence data :



It must be stressed, however, that the data set used to produce this tree was very small, conservative substitutions were not incorporated in the analysis, and as with all analyses of this type, a constant evolutionary rate is assumed. This last assumption is particularly questionable when proteins coded for by extrachromosomal genes are compared, since transfer of a plasmid from one species to another can occur with consequent changes of both intra and extracellular environments so producing changes in evolutionary pressure.

C-terminal comparisons.

The C-terminal sequences of a type I variant (R-factor JR66) and a staphylococcal type C (plasmid C221) variant are now known :

	195	196	197	198	199	200	201	202	203	204
TYPE I	Gln	Tyr	Cys	Asp	Glu	Trp	Gly	Ala	Gly	Gln

TYPE C	His	Lys	Val	Asp	Asn	Trp	Ile

As with the N-terminal sequences definite homology seems to exist, although a number of amino acid differences are evident.

Initially it was hoped that a knowledge of the complete primary structure of the type I CAT variant would aid the determination of the primary sequence of the type C staphylococcal variant by finding analogous peptides and so determining their relative position in the polypeptide chain. However, very few homologous sequences seem to exist (compare figure 20 and figure 22). An exception is the region containing the iodoacetamide sensitive histidine residue :

	177	178	179	180	181	182	183	184	185	186
TYPE I	His	His	Ala	Val	Cys	Asp	Gly	Phe	His	Val

TYPE C	His	His	Ala	Val	Cys	Asp	Gly	Tyr	His	Ala

Only two conservative substitutions are found in this region, a Phe to Tyr and a Val to Ala substitution. Since this region is thought to be included in the 'active site' of the CAT enzymes mutational events in this area would be expected to affect enzymic activity and cause loss of resistance to the antibiotic. Because of the functional importance of this region the amino acid sequence would be highly conserved.

The only other obvious sequence homology between the two variants is the sequence :

His-Val-Gly-Arg

In the type I sequence this region is at the C-terminal end of the 'active site' region described above :

181 182 183 184 185 186 187 188
Cys-Asp-Gly-Phe-His-Val-Gly-Arg-

However, the homologous sequence from the type C staphylococcal variant is produced by tryptic digestion of the protein, implying that the residue adjacent to the histidine residue is lysine or arginine and the 'active site' region in this variant is known to be :

Cys-Asp-Gly-Tyr-His-Ala-Ser-(Leu,Phe).

(see figure 20)

The sequence His-Val-Gly-Arg must be in the vicinity of the 'active site' region since both peptides are included in the peptide SP60 - 67 B obtained from a staphylococcal protease digest of the type C variant.

The type I variant contains four cysteine residues in the sequence :

34 35 36 37 38
Ala-Gln-Cys-Thr-Tyr

71 72 73 74 75
Val-Ala-Cys-Tyr-Gly

179 180 181 182 183
Ala-Val-Cys-Asp-Gly

195 196 197 198 199
Gln-Tyr-Cys-Asp-Glu

The cysteine residue at position 36 of the sequence is sensitive to iodoacetamide as is a histidine residue at position 178 (see section 10.3). When the type C enzyme is inactivated with iodoacetamide however, only 3-Cm-histidine is observed on amino acid analysis. This observation is consistent with the sequence of the type C enzyme since the residue equivalent to the cysteine at position 36 in this protein is a lysine residue (figure 21). The cysteine residue at position 181 in the type I enzyme has an analogous cysteine residue in the type C sequence, included in the 'active site' region. The cysteine near the C-terminus of the type I variant (position 197) is replaced by a valine residue in the type C sequence. No peptide containing a cysteine residue equivalent to that at position 73 of the type I sequence has been found in the staphylococcal enzyme.

10.4 Secondary structure comparisons.

Many problems of protein structure, function and evolution can only be solved when the tertiary structure of the protein is known. At the present time only high resolution X-ray crystallographic analysis can yield such structural information. In view of the limited primary sequence homology between the type I and type C variants and the lack of X-ray data it was thought that a preliminary comparison of the predicted secondary structures might prove useful.

Many methods of secondary structure prediction are now available and there is considerable variation in the complexity of the prediction methods and in the results obtained. The predictive model of Chou and Fasman (70)

was chosen since it is capable of distinguishing helix, sheet, reverse turn and random coil conformations. The authors have proposed a mechanism for protein folding whereby regions which have high conformational parameters P_α or P_β (values for which have been derived from a study of the frequency of occurrence of each amino acid in α helices and β sheets) nucleate an α helix or β sheet which propagates in both directions until strong α helix or β sheet breaking amino acids terminate growth of the secondary structure. Using the parameters P_α and P_β they have formulated empirical rules for predicting the inititation and termination of α helical and β sheet regions in globular proteins. In addition Chou and Fasman have computed the frequency of occurrence of amino acids in reverse turns and devised rules for the prediction of reverse turns in proteins. Any region of sequence not predicted to be α helix, β sheet or reverse turn is assumed to be in the random coil conformation. Since the number of potential CAT sequences is large, the method of Chou and Fasman has been automated to a large extent through the use of a computer programme (see Appendix A.) Argos et al (71) concluded from their study of the predictive ability of a number of predictive methods that 'prediction algorithms can determine well the sequence and extent of gross secondary structural features in 'favourable' cases. This capability is especially useful when an unknown protein is suggested to have a similar structure to a known protein through chemical experiments or weak sequence homology.'

Secondary structure predictions of the N-termini of ten CAT variants are illustrated in figure 21. With the exception of Streptomyces acrimycini variant all variants show virtually identical N-terminal secondary structure, a short section of β sheet followed by α helix and (where the determined N-terminal sequences are long enough) another β sheet. The Streptomyces

acrimycini variant shows no predicted secondary structure (probably due to the high proportion of proline residues) and is assumed to be random coil. The type I CAT variants, which have an additional six residues at the N-terminus have a short section of random coil at the N-terminus.

The complete sequence and predicted secondary structure of a type I (R-factor JR66) variant is illustrated in figures 22 and 23. Figure 22 shows the uncorrected computer printout (see Appendix A for notation) and figure 23 shows the final predicted secondary structure after the necessary manual corrections have been made.

The predicted structure of the type I variant indicates a high proportion of both α helix and β sheet but only one reverse turn is predicted.

% α helix	= 37%
% β sheet	= 38%
% reverse turn	= 2%
% random coil	= 23%

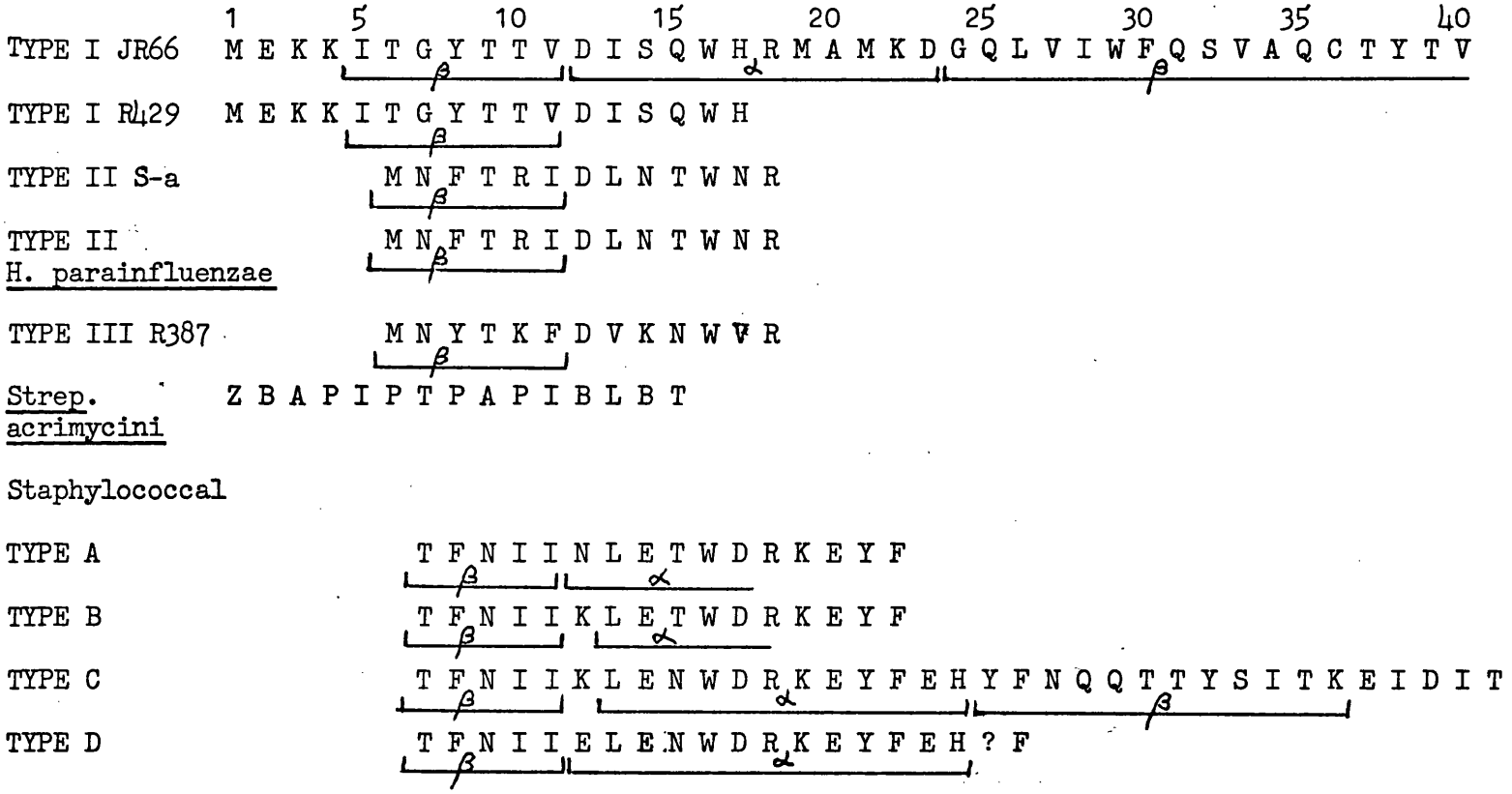
Circular dichroism studies (J. Carter and W.V. Shaw, unpublished results) indicate 30 percent α helix and 'considerable' β sheet secondary structure for this enzyme and the results are in good agreement with the predicted structure. The region which has been implicated to be included in the active site of the enzyme (residues 172 to 179) has the greatest α helix forming potential of the whole sequence ($P\alpha$ = 1.26) but also has a relatively high β sheet forming potential ($P\beta$ = 1.07). Chou and Fasman (70) suggest that the region of a polypeptide with the greatest α helix forming potential will act as the nucleation site for protein folding and, in the case of the type I CAT variant, folding of the polypeptide can be envisioned to occur around the active site of the protein.

The histidine residue implicated in the mechanism of catalysis (His 178) is found at the C-terminus of the strongly predicted α helix. Chou and Fasman (70) have noticed that residues involved in the mechanism of catalysis of enzymes are frequently found at helix-random coil boundaries and suggest that, since these regions are more flexible than the rigid inner helix, substrate binding and catalysis are facilitated.

figure : 21.

N-terminal primary structures and secondary structure predictions for all CAT variants studied at the primary sequence level.

Sequences are aligned to give both maximum primary and secondary structure homology.

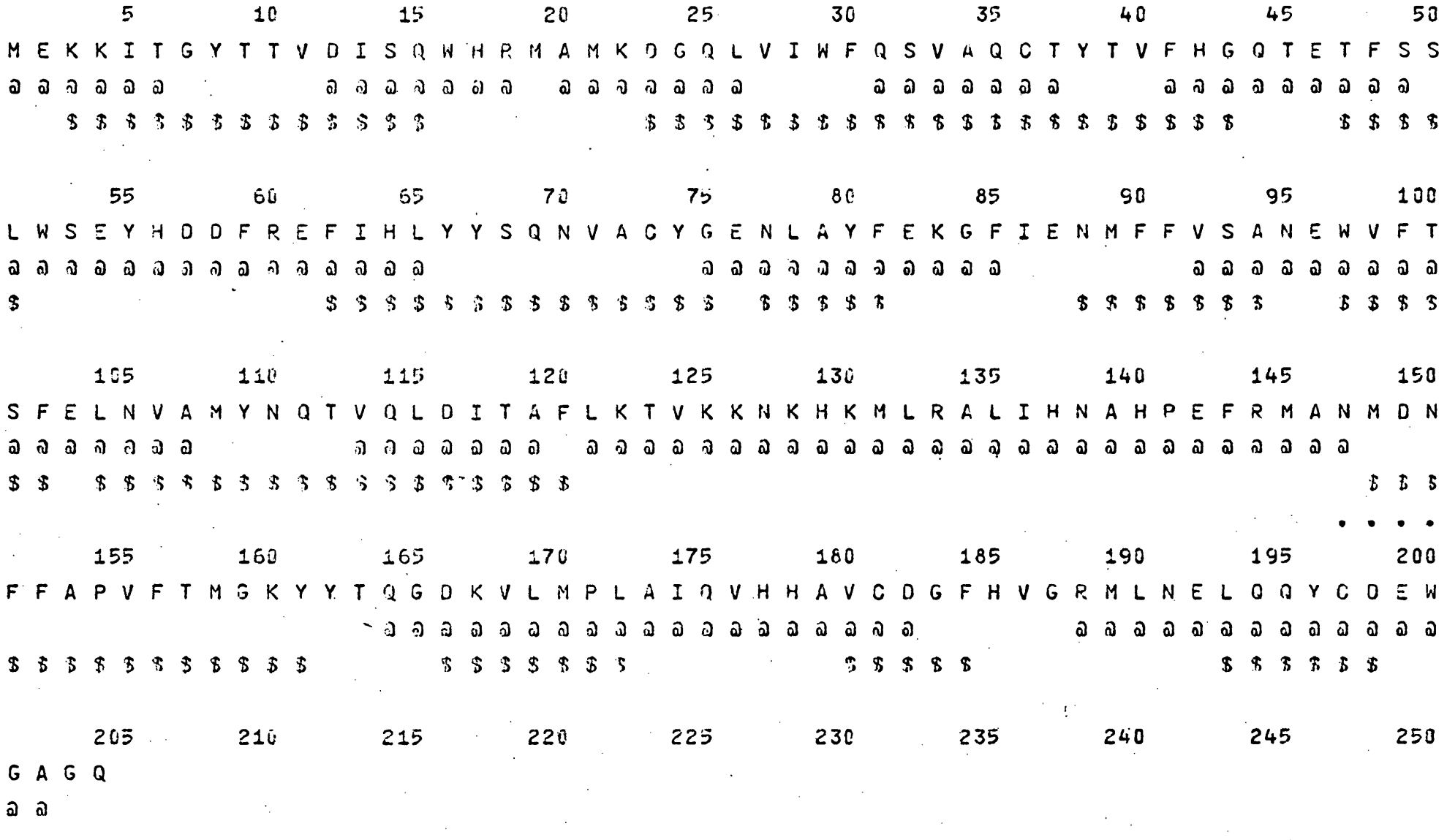

1 5 10 15 20 25 30 35 40
TYPE I JR66 M E K K I T G Y T T V D I S Q W H R M A M K D G Q L V I W F Q S V A Q C T Y T V
β
α
β
TYPE I RL29 M E K K I T G Y T T V D I S Q W H
β
TYPE II S-a M N F T R I D L N T W N R
β
TYPE II H. parainfluenzae M N F T R I D L N T W N R
β
TYPE III R387 M N Y T K F D V K N W V R
β
Strep. acrimycini Z B A P I P T P A P I B L B T
Staphylococcal
TYPE A T F N I I N L E T W D R K E Y F
β
α
TYPE B T F N I I K L E T W D R K E Y F
β
α
TYPE C T F N I I K L E N W D R K E Y F E H Y F N Q Q T T Y S I T K E I D I T
β
α
β
TYPE D T F N I I E L E N W D R K E Y F E H ? F
β
α

figures : 22 and 23.

Uncorrected secondary structure prediction of
R-factor type I CAT (JR66)

and

manually corrected secondary structure prediction
of R-factor type I CAT (JR66).


5 10 15 20 25 30 35 40 45 50
M E K K I T G Y T T V D I S Q W H R M A M K D G Q L V I W F Q S V A Q C T Y T V F H G Q T E T F S S
55 60 65 70 75 80 85 90 95 100
L W S E Y H D D F R E F I H L Y Y S Q N V A C Y G E N L A Y F E K G F I E N M F F V S A N E W V F T
105 110 115 120 125 130 135 140 145 150
S F E L N V A M Y N Q T V Q L D I T A F L K T V K K N K H K M L R A L I H N A H P E F R M A N M D N
155 160 165 170 175 180 185 190 195 200
F F A P V F T M G K Y Y T Q G D K V L M P L A I Q V H H A V C D G F H V G R M L N E L Q Q Y C D E W
205 210 215 220 225 230 235 240 245 250
G A G Q

```
        5         10        15        20        25        30        35        40        45        50
M E K K I T G Y T T V D I S Q W H R M A M K D G Q L V I W F Q S V A Q C T Y T V F H G Q T E T F S S
                  ə ə ə ə ə ə ə ə ə ə ə                                       ə ə ə ə ə ə ə
      $ $ $ $ $ $ $                     $ $ $ $ $ $ $ $ $ $ $ $ $ $ $ $ $ $

        55        60        65        70        75        80        85        90        95        100
L W S E Y H D D F R E F I H L Y Y S Q N V A C Y G E N L A Y F E K G F I E N M F F V S A N E W V F T
ə ə ə ə ə ə ə ə ə ə ə ə ə                           ə ə ə ə ə ə ə ə ə
                        $ $ $ $ $ $ $ $ $ $ $                     $ $ $ $ $ $ $     $ $ $ $

      105       110       115       120       125       130       135       140       145       150
S F E L N V A M Y N Q T V Q L D I T A F L K T V K K N K H K M L R A L I H N A H P E F R M A N M D N
                                                  ə ə ə ə ə ə ə ə ə ə
$ $   $ $ $ $ $ $ $ $ $ $ $ $ $ $ $ $

                                                                                    • • • •
      155       160       165       170       175       180       185       190       195       200
F F A P V F T M G K Y Y T Q G D K V L M P L A I Q V H H A V C D G F H V G R M L N E L Q Q Y C D E W
                        ə ə ə ə ə ə ə ə ə ə ə                     ə ə ə ə ə ə ə ə ə ə ə
$ $ $ $ $ $ $ $ $ $                                                       • • •

      205       210       215       220       225       230       235       240       245       250
G A G Q
ə ə
```

CHAPTER XI.

SUMMARY AND DISCUSSION.

Previous studies have shown that four electrophoretically distinct variants of CAT are found in staphylococci (16). Each is a tetrameric enzyme composed of four identical monomers of molecular weight 24,000 daltons. Immunological studies have revealed a high degree of cross-reactivity, indicating that all four staphylococcal variants are likely to show considerable structural homology (15, 16, 17)

All CAT variants from staphylococci have been purified by the method of affinity chromatography using a derivative of chloramphenicol, coupled through a six carbon atom 'spacer' to an agarose matrix. The chromatographic behaviour of the four variants on such affinity resins has been determined, and considerable differences have been observed in their binding and elution properties. Indeed, the use of affinity chromatography can discriminate between closely related CAT variants, both among those isolated from staphylococci and between the enteric (R-factor) variants (19). The use of affinity chromatography has simplified the purification of CAT to such a degree that a homogenous preparation of most CAT variants can be achieved in a single day from crude cell extracts and with an overall purification yield of greater than 90 percent in most instances. Although the four staphylococcal variants show different elution properties on affinity columns and differences in susceptibility to heat denaturation, their Michaelis constants (Km) are virtually identical for both chloramphenicol ($2.6 \pm 0.2 \mu M$) and AcCoA ($54 \pm 8 \mu M$). Their catalytic similarity is also obvious on the basis of chemical modification studies since all four CAT variants are inactivated by the same reagents and at virtually identical rates. Amino acid analyses of the four variants are very similar although significant differences are found in the numbers of aspartic acid, lysine and 'small' hydrophobic amino acid residues (valine, isoleucine and leucine). Tryptic peptide maps of the

four variants are similar although the type D enzyme appears to lack several basic peptides. Tryptic digests of types A and B also yield a tryptophan positive, acidic peptide which has a different electrophoretic mobility at pH 3.5 when compared with the analogous peptide in the type C and D variants, the peptide which is, in fact, known to be the C-terminal tryptic peptide in the type C enzyme.

Considerable difficulties were encountered in coupling CAT proteins to solid phase supports and, although the technology and chemistry of solid phase sequencing is relatively straightforward, there appear to be limitations to the available methods of coupling peptides and proteins to solid phase supports. Despite these difficulties, N-terminal sequences have been obtained and the staphylococcal variants show approximately 90 percent homology in the first 17 residues of their N-terminal sequences. Substitutions are observed at only two positions in the sequences (residues 6 and 9) suggesting that the tertiary structure can accommodate such substitutions without a measurable loss in the ability of the enzyme to catalyse the acetylation of chloramphenicol.

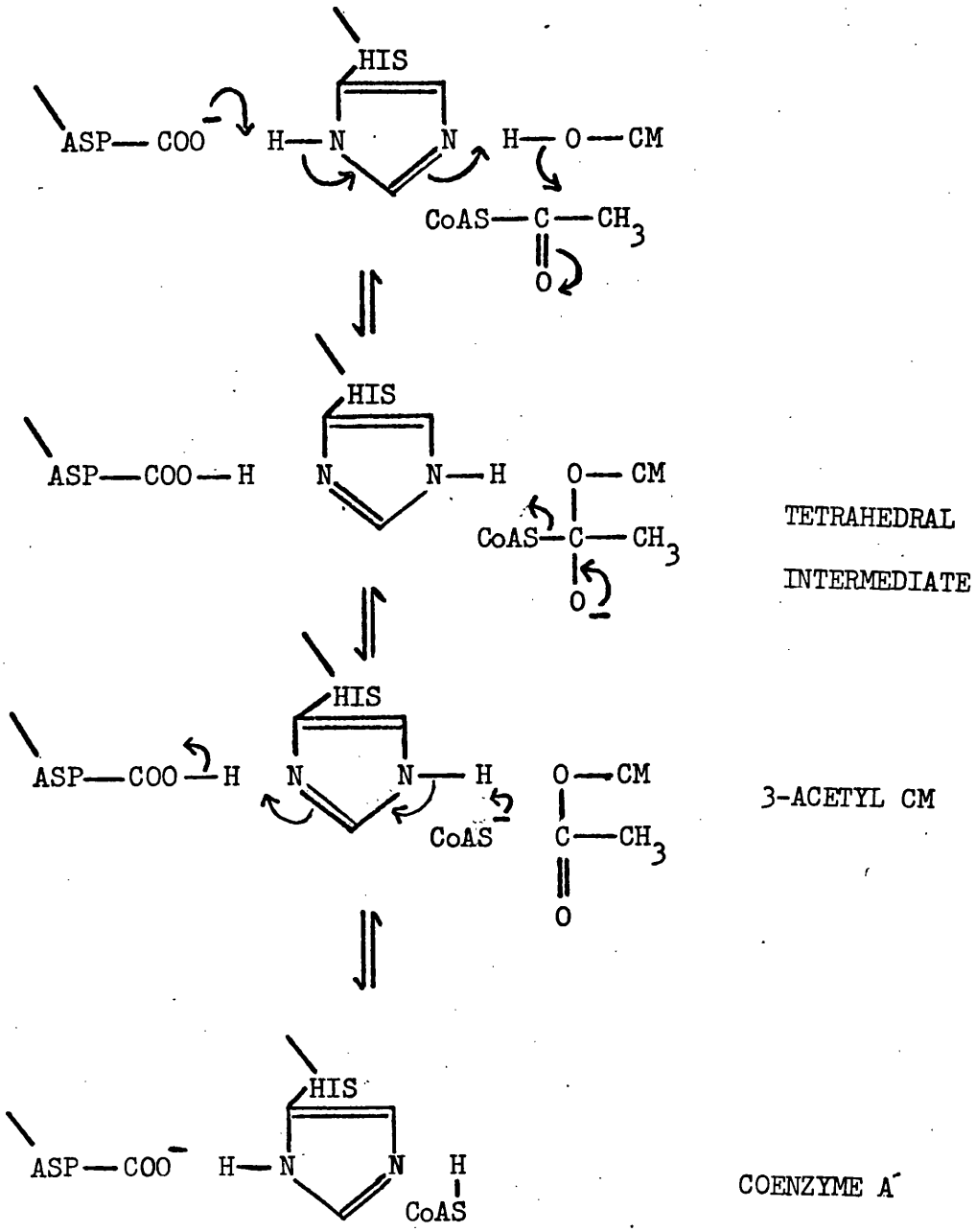
Although greater than 90 percent of the covalent structure of a type C staphylococcal variant has been determined in the course of these studies, a number of further peptide overlaps are still required and certain sequences need clarification before the complete primary structure is available. In view of the adaptability of the solid phase sequencer towards the sequence determination of large lysine containing peptides and the difficulties encountered in the purification of peptides resulting from cyanogen bromide cleavage of the type C protein, it is likely that future work will concentrate on the purification and sequencing of peptides generated by tryptic digestion of citraconylated Cm CAT. It is hoped that the known heterogeneity of certain sections of the protein sequence will not be a

major obstacle to achieving a complete primary structure of this protein.

A comparison of the amino acid sequence of the intensively studied type I R-factor CAT variant and the available primary structure of the type C staphylococcal enzyme shows very little obvious evidence of sequence homology. The exception to this is the potential active site region (residues 177 - 183) where the two sequences are identical. Chemical modification studies on the four staphylococcal variants and photo-oxidation studies on a type C variant implicate a histidine residue being involved in the catalytic mechanism. Specific sulphydryl reagents do not inactivate the staphylococcal variants and the reactive cysteine residue (position 36) found in a type I R-factor variant is replaced by lysine in the staphylococcal type C protein suggesting that this cysteine residue plays no part in catalysis. The finding that iodoacetamide inactivates all CAT enzymes studied whereas iodoacetate does not suggests that there may be an ionized aspartyl or glutamyl residue at, or near, the active site. The results of the active site experiments bear some resemblance to a proposed mechanism of catalysis of chymotrypsin and related serine proteases. Although such enzymes are best known for their proteolytic activity, they can also be regarded as acyl transfer enzymes. The active centre of chymotrypsin (76) consists of an aspartyl residue (residue 102) hydrogen bonded to a histidine residue (residue 57) to form a charge - relay system. A serine (residue 195) is acylated by the substrate as an intermediate in the hydrolysis of esters.

Although a number of reaction mechanisms are consistent with the active site studies on the CAT variants, possibly the simplest consists of a single encounter mechanism analogous to the reverse reaction of the deacylation step of chymotrypsin with the chloramphenicol hydroxyl serving as the acyl acceptor in place of the serine hydroxyl of chymotrypsin and acetyl -

Coenzyme A being the acyl donor:



Although the acidic residue is presented as an aspartyl residue, a glutamyl residue could perform the same function. The imidazole group is acting as a general base to abstract a proton from the hydroxyl of CM. The N3 of the histidine residue would be expected to react with iodoacetamide by an S_N2 displacement reaction since the N3 nitrogen atom would be a strong nucleophile due to proton abstraction from the N1

nitrogen atom by the acidic amino acid residue. The finding that the CAT enzymes catalyse the deacylation of AcCoA in the absence of the acyl acceptor chloramphenicol is also consistent with the proposed mechanism of catalysis since, in the model proposed, water could act as the acyl acceptor leading to the formation of acetate. Although kinetic analysis indicates that the mechanism is sequential and not a 'ping - pong' type the initial velocity patterns do not discriminate between a rapid-equilibrium random and an ordered mechanism. In addition, one cannot determine the order of substrate participation in an ordered mechanism from this information. Product inhibition studies are required to discriminate between these possibilities. The involvement of an acidic residue in the mechanism is inferred from the fact that iodoacetamide inactivates CAT enzymes whereas iodoacetate does not. The most commonly used chemical modification procedure for determining the importance of carboxyl groups in catalysis involves the use of a water soluble carbodiimide and a nucleophile (eg. glycinamide) at pH 4.5 to 5.0. This type of modification has not been attempted since the CAT enzymes are rapidly inactivated by pH values below 5.7.

The CAT variants as a whole show varying degrees of sequence homology. The type I (R-factor) variants have identical N-terminal sequences although they can be distinguished by their elution behaviour from affinity resins (L. Packman, unpublished results). The two type II enteric (R-factor) variants studied also have identical N-terminal sequences, although both enzymes are known to be specified by plasmids in different organisms; the plasmid S-a type II variant was first detected in Shigella isolates (26) whereas the variant associated with the putative plasmid 'CT' was isolated from E. coli following transfer of the chloramphenicol resistance determinant from Haemophilus parainfluenzae (27). Although the type B

staphylococcal variant studied was purified from Staphylococcus epidermidis, CAT enzymes with identical electrophoretic properties are also found in Staphylococcus aureus. These two examples suggest the possibility of interspecies and, in the case of the type II R-factor enzymes, intergeneric gene transfer. The rare type III R-factor variant specified by plasmid R387 shows a greater degree of sequence homology with the type I variants than the type II enzymes (Chapter X, 10.3). However, although the type III enzyme has a 'large' monomer molecular weight (25,000 daltons) in common with the enzyme from Streptomyces acrimycini, the latter two variants show no N-terminal sequence homology.

The type I (R-factor JR66) variant and the type C staphylococcal variant only show weak N-terminal primary sequence homology although predicted N-terminal secondary structures and active site regions are very homologous. The CAT enzymes thus form a 'family' which have similar gross properties, ie. all are tetrameric proteins of approximately the same molecular weight, catalysing the acetylation of chloramphenicol. The R-factor variants of any given electrophoretic 'type' are highly homologous, as are the four staphylococcal variants, but less homology is seen amongst the R-factor types I, II and III. The least sequence homology is found between the staphylococcal sub-group and the R-factor type I variants.

Several other examples of enzyme 'families' have been studied in considerable detail, notably the β-lactamases (42) and the serine proteinases (77, 78). The β-lactamase enzymes cleave the β-lactam ring of the penicillins and the closely related cephalosporins so as to destroy antibiotic activity. They are found in a diverse range of micro-organisms and are often specified by extrachromosomal genes. Ambler has determined the primary amino acid sequence of β-lactamases from S.aureus and Bacillus

licheniformis (42, 79) and a number of other β-lactamase enzymes are under study. The enzymes from S. aureus and B. licheniformis are 42 percent homologous at the primary sequence level and secondary structure predictions (see appendix A) indicate a greater degree of homology of the secondary and, by inference, tertiary structures. Studies on the primary structure of a frequently observed β-lactamase found in E. coli (plasmid TEM) are well advanced and reveal considerable homology (17%) with the β-lactamases of Gram positive organisms (R.P. Ambler, personal communication). Over 80 percent of the sequence of a β-lactamase from B. cereus has been reported and 60 percent of the proposed sequence is homologous with the enzyme from B. licheniformis and 40 percent homologous with the S. aureus β-lactamase (80).

Perhaps the most intensively studied enzyme family consists of the serine proteinases (78, 81). This group includes chymotrypsin, trypsin, elastase and thrombin in mammals and subtilisin, amongst others, in bacteria. The mammalian enzymes have primary sequence homologies of 30 to 50 percent but are not homologous in primary structure when compared with subtilisin. X-ray crystallographic analysis of the mammalian enzymes has shown that the tertiary structures of these enzymes have been highly conserved such that most of the amino acid differences occur on the exterior surface of the proteins and rather fewer substitutions in the closely packed interior of the molecules. The amino acid residues involved in the proposed charge-relay mechanism of catalysis (Asp-102, His-57 and Ser-195 in chymotrypsin) all occur in identical spacial conformations in other members of the 'family'. The specificity differences of chymotrypsin and trypsin can be accounted for by a single amino acid difference, Asp-189 in trypsin replaces Ser-189 in chymotrypsin allowing ionic interaction between lysine and arginine residues of the substrate with Asp-189 in trypsin. The large number of differences in

amino acids which are exposed to the enviroment appear to have no bearing on the specificity differences of these enzymes. Unlike the mammalian serine proteinases, subtilisin shows no primary sequence homology nor any obvious secondary or tertiary structure similarity, although an aspartic acid, a histidine and a serine residue are found in essentially the same spacial configuration. It has been proposed that this is an example of 'convergent' evolution wherein unrelated primordial proteins have evolved to give identical mechanisms of catalysis.

By comparison with the β-lactamases and the serine proteinases the CAT variants show a much greater degree of primary sequence homology within groups eg. the R-factor type I or the staphylococcal variants, although dissimilar groups eg. the type I enzymes compared with the staphylococcal enzymes, have much less primary sequence homology. Overall the CAT variants appear to be a very diverse family when primary sequences are considered. In view of the active site homologies between the distantly related type I R-factor and the type C staphylococcal variants, and the predicted secondary structure homologies, it seems likely that all the CAT enzymes have evolved from a common ancestral protein.

Studies of other o-acetyltransferase enzymes (eg. serine acetyltransferase and thiogalactoside acetyltransferase) might be considered in the hope that an acetyltransferase could be found which had a different substrate specificity but which had structural, kinetic or active site features in common with the CAT family, suggesting a 'prototype' acetyltransferase from which the CAT and other acetyltransferases might have evolved. To date, no such candidate has been found.

Studies on experimental enzyme evolution (78) have indicated that, for enzymes which may be regarded as catalysts of complex reactions ie. involving

more than one substrate or stereospecific reactions, the development of new substrate specificities is an exceedingly rare event and is likely to occur by multistep mutation. This is in contrast with the results of experimental enzyme evolution of a Pseudomonas amidase (84) where single amino acid substitutions can lead to large differences in substrate specificity. Presumably the amidase enzymes, catalysing a simple hydrolytic reaction involving chemically simple compounds, have less structural constraints and single amino acid substitutions can significantly affect substrate specificity without interfering with the catalytic process. It seems likely, however, that evolution toward a high degree of specificity may be accompanied by constraints on the efficiency of catalysis. Hard data on these interesting points is, however, sadly lacking.

Gene duplication followed by mutational divergence has been proposed as a mechanism of and prerequisite for the development of new enzyme specificities (78). Gene duplication allows extensive mutation to occur in a 'silent' copy of the duplicated gene without affecting the phenotype or viability of the organism. Should selection demand a new enzyme specificity, this model predicts that the organism can then draw upon a large 'silent gene pool' for the determinant of the required enzymic function.

Preliminary studies by Jacobson and Shaw (85) suggested that a strain of Proteus mirabilis mutated from sensitivity to chloramphenicol to resistance with high frequency, and it was proposed that this organism contained an enzyme which acetylated an unknown substrate and could be mutated to an enzyme possessing a high affinity for chloramphenicol. More recent work by S. Harford (unpublished) indicated that the majority of wild type strains of Proteus may contain plasmids specifying CAT and that the increase in observed resistance is due in fact to an increase in enzyme levels via a gene

amplification mechanism which may involve either plasmid copy numbers or tandem duplications. Although the precise mechanism of this effect is not yet clear, it now seems certain that the hypothesis of 'evolution' by mutation in Proteus is no longer tenable.

Ambler and Meadway(79) have discussed the possible origins of the β-lactamase enzymes in S. aureus and B. licheniformis and their discussion is applicable to the CAT system :

1. Penicillinase first evolved in an ancestor common to both staphylococci and bacilli. The penicillinase gene has been lost from most of the descendants, but still survives in recognisably homologous forms in S. aureus and B. licheniformis.
2. An enzyme with penicillinase activity has evolved on at least two independent occasions from an enzyme with a different activity.
3. The penicillinase gene evolved in one or other of the genera subsequent to their divergence, and has reached the other genus by a process such as intergeneric transduction or transformation.
4. The two penicillinases have evolved quite independently, and the similarity in amino acid sequence is a result of analogy rather than homology.

In view of the similarities between the CAT enzymes, particularly the active site homology between the type I R-factor and the type C staphylococcal variants, the fourth possibility would appear to be unlikely. Possibility 1, although possible, does not appear likely since the CAT gene is rarely chromosomal (see Chapter II, Materials : bacterial strains) and is found in many genera. Since CAT is frequently coded for by an extrachromosomal gene, possibility 3 would appear to be a likely explanation, however this would necessitate gene transfer between widely unrelated genera, although transfer could occur in a step-wise fashion via one or more intermediate species.

Possibility 2, where CAT activity had evolved in all CAT species known, would appear to be an unlikely explanation on its own since no authentic case is known wherein a mutation in a bone fide CAT negative organism has led to CAT synthesis.

In the final analysis, the most likely explanation of the origin and occurrence of the CAT enzymes is a combination of possibilities, 2 and 3 (or 1 and 3) where a CAT enzyme has evolved from an enzyme with a different specificity on one (or more) occasions and once CAT activity had evolved, intraspecies, interspecies (eg. type B CAT in S. epidermis and S. aureus) and intergeneric (eg. type II CAT in Shigella and H. parainfluenzae) transfer has occurred.

Appendix A.

Secondary structure prediction, computer programme, its limitations and its evaluation.

```
      PROGRAM PRODDD(INPUT,OUTPUT,TAPE5=INPUT,TAPE6=OUTPUT)
C                                  138.
C*****************************************************************
C
C     PROGRAM:  PRODDD
C      AUTHOR:  DEREK PEACOCK & JOHN FITTON
C        DEPT:  BIOCHEMISTRY
C        DATE:  JUNE 1976
C
C              THIS PROGRAM CALCULATES THE PREDICTED THREE
C              DIMENSIONAL STRUCTURE FO A PROTEIN FROM IST
C              AMINO AICD SEQUENCE USING THE MEHTOD OF
C              CHOU,P.Y. AND FASMAN,G.D. BIOCHEMISTRY 13,
C              PGS 222-245(1974).
C
C*****************************************************************
C
C         DATA INPUT IS IN THE FOLLOWING FORMAT
C
C           1.TITLE  OF RUN            80 CHARACTERS
C
C           2.NUMBER OF SEQUENCES      ANY WHERE
C
C           3.NUMBER OF RESIDUES
C             IN THE FIRST SEQUENCE      ANY WHERE
C
C           4.NAME OF FIRST SEQUENCE   80 CHARACTERS
C
C           5.FIRST SEQUENCE           1 SPACE,1LETTER IN GROUPS OF 36
C
C            LAST THREE ITEMS REPEATED
C
C
C     PROSEQ = PROTEIN AMINO ACID SEQUENCES
C     AHVAL  = ALPHA HELIX VALUES FOR ANY ONE PROTEIN SEQUENCE
C     BSVAL  = BETA SHEET VALUES FOR ANY ONE SEQUENCE
C     BTVAL  = BETA TURN VALUES FOR THE BETA TURNS IN ANY ONE SEQUENCE
C
C     ALPHA  = ALPHA HELIX PROBABILITIES FOR EACH AMINO ACID
C     BETAS  = BETA SHEET VALUES FOR EACH AMINO ACID
C     BETAT  = BETA TURN VALUES FOR EACH AMINO ACID
C     FIVAL  = FOUR FI VALUES FOR EACH AMINO ACID
C
C     HELIX  = STORES THE POSITIONS OF THE ALPHA HELIX REGIONS
C     SHEET  = STORES THE POSITIONS OF BETA SHEET REGIONS
C     TURN   = STORES THE POSITIONS OF BETA TURNS
C
C     NRES   = STORES THE NUMBER OF RESIDUES IN EACH SEQUENCE
C     NR     = NUMBER OF RESIDUES FOR ANY ONE SEQUENCE
C     NS      = NUMBER OF SEQUENCES
C
C     PA     = ALPHA HELIX PROBABILITY TOTAL FOR A GROUP OF RESIDUES
C     PB     = BETA SHEET PROBABILITY TOTAL FOR A GROUP OF RESIDUES
C     PT     = BETA TURN PROBABILITY TOTAL FOR A GROUP OF RESIDUES
C     FI     = FI TOTALS FOR A GROUP OF FOUR RESIDUES
C
C     AMINOS = A LIST OF THE AMINO ACIDS ACCEPTED BY THIS PROGRAM
C
C     TITLE  = A TITLE FOR THE COMPUTOR RUN
C     NAME   = THE NAME OF EACH PROTEIN
C
      IMPLICIT INTEGER(A-Z)
      REAL ALPHA,BETAS,BETAT,AHVAL,BSVAL,BTVAL,FIVAL,PA,PB,PT,FI,PE,P
      COMMON /S,PROSEQ,AHVAL,BSVAL,LAST
      DIMENSION HELIX(301),SHEET(301),TURN(301),NRES(10)
      DIMENSION PROSEQ(10,301),NAME(10,8),TITLE(8),RES(10)
      DIMENSION AMINOS(22),ALPHA(22),BETAS(22),BETAT(22),FIVAL(4,22)
      DIMENSION AHVAL(301),BSVAL(301),BTVAL(301),PE(301),P(301)
      DIMENSION PHRASE(6)
      DATA(AMINOS(I),I=1,22)/"A","C","D","E","F","G","H","I",
     + "K","L","M","N","P","Q","R","S","T","V","W","Y","-","."/
      DATA(ALPHA(I),I=1,22)/1.45,0.54,0.98,1.53,1.12,0.53,1.24,1.00,
     + 1.07,1.34,1.13,0.73,0.55,1.17,0.79,0.73,0.82,1.14,
     + 1.07,0.67,0.0,0.0/
      DATA(BETAS(I),I=1,22)/0.93,1.11,0.77,0.32,1.47,.78,0.68,1.53,
     + 0.71,1.17,1.00,0.72,0.43,1.05,0.86,0.77,1.15,1.58,1.14,1.24,
     + 0.00,0.00/
      DATA(BETAT(I),I=1,22)/0.75,0.96,1.36,0.71,0.41,1.55,0.80,
     + 0.59,0.93,0.53,0.71,1.55,1.35,0.92,0.95,1.47,1.13,0.45,
     + 1.03,1.05,0.0,0.0/
      DATA((FIVAL(I,J),I=1,4),J=1,22)/0.86,0.89,0.52,.60,
     + 1.22,.27,1.22,0.95,1.90,1.15,1.72,0.81,0.57,0.57,0.98,0.49,
     + 0.32,0.21,0.74,0.42,1.20,1.20,1.94,1.63,1.75,0.32,1.11,0.10,
     + 0.58,0.45,0.32,0.83,0.57,1.09,0.63,1.09,0.32,0.37,0.41,0.83,
```

```
     + 0.87,0.87,0.22,0.87,1.59,1.00,2.24,0.88,1.34,2.77,0.36,0.80,
     + 0.73,1.00,0.33,1.14,0.80,1.32,0.76,1.04,1.50,1.43,1.35,1.35,
     + 1.05,1.40,0.85,1.15,0.33,0.61,0.25,0.45,0.73,0.18,0.72,2.00,
     + 1.14,0.71,1.14,1.29,0.0,0.0,0.0,0.0,0.0,0.0,0.0,0.0/
       CALL PAPER(1)
       CALL GPSTOP (30)
       DO 10 I = 1,301
       HELIX(I) = " "
       SHEET(I) = " "
       TURN(I) = " "
    10 CONTINUE
       DO 20 I = 1,22
       WRITE(6,1001)(AMINOS(I),ALPHA(I),BETAS(I),BETAT(I),
      + (FIVAL(J,I),J=1,4))
    20 CONTINUE
       CALL READCS (PHRASE)
       WRITE(6,1040)(PHRASE(I),I=1,8)
       READ(5,*)(NS)
       DO 120 IS=1,NS
       READ(5,*)(NRES(IS))
       NR = NRES(IS)
       IR = 1
       IR36 = IR+35
       READ(5,1000)(NAME(IS,L),L=1,8)
   100 READ(5,1010)(PROSEQ(IS,KR),KR=IR,IR36)
       IR = IR+36
       IR36 = IR36+36
       IF(IR.GT.NR)GOTO 120
       IF(IR36.GT.NR)IR36 = NR
       GOTO 100
   120 CONTINUE
C
CC          THE AMINO ACID SEQUENCE IS TRANSLATED INTO
CC          A SEQUENCE OF ALPHA HELIX FORMING VALUES
CC          IN THE VARIABLE AHVAL.
C
       DO 500 IS = 1,NS
       NR = NRES(IS)
       WRITE(6,1050)(NAME(IS,L),L=1,8)
       WRITE(6,1125)
       ERROR = 0
       DO 140 IR = 1,NR
       DO 130 IA = 1,22
       IF(AMINOS(IA).NE.PROSEQ(IS,IR))GOTO 130
       AHVAL(IR) = ALPHA(IA)
       BSVAL(IR) = BETAS(IA)
       BTVAL(IR) = BETAT(IA)
       GOTO 140
   130 CONTINUE
       WRITE(6,1020)(PROSEQ(IS,IR),IR)
       ERROR = ERROR +1
   140 CONTINUE
C
CC          THE INPUT DATA IS WRITTEN OUT IN BLOCKS OF 25
C
       R1 = 1
       R2 = 25
   141 IF(R2.GT.NR)R2=NR
       DO 145 I=1,5
       RES(I) = (R1+5*I)-1
   145 CONTINUE
       WRITE(6,1100)(RES(I),I=1,5)
       WRITE(6,1060)(PROSEQ(IS,IR),IR=R1,R2)
       WRITE(6,1070)(AHVAL(IR),IR=R1,R2)
       WRITE(6,1070)(BSVAL(IR),IR=R1,R2)
       WRITE(6,1070)(BTVAL(IR),IR=R1,R2)
       R1 = R1+25
       R2 = R2+25
       IF(R1.LT.NR)GOTO 141
       IF(ERROR.GT.0)GOTO 500
C
CC          A SEARCH IS MADE FOR A HELIX NUCLEUS OF SIX RESIDUES
C
       IR = 0
   150 IR = IR+1
       LAST = NR -5
       IF(IR.GT.LAST)GOTO 400
       ABREAK=0
       AFORM=0
       PA =0
       PB = 0
       LYS = 0
       DO 160 NN =1,6
       KR = IR + NN-1
       PB = PB + BSVAL(KR)
```

```
      IF(NN.LE.3)GOTO 152
      IF(PROSEQ(IS,KR).NE."P")GOTO 151
      AFORM = 0
      ABREAK=2
      GOTO 160
  151 IF(NN.LE.5)GOTO 153
      IF(PROSEQ(IS,KR).NE."R")GOTO 153
      AFORM = AFORM +1
      PA = PA + 1.1
      GOTO 160
  152 IF(PROSEQ(IS,KR).NE."D")GOTO 153
      AFORM = AFORM +1
      PA = PA + 1.1
      GOTO 160
  153 PA = PA + AHVAL(KR)
      IF(PROSEQ(IS,KR).EQ."P")GOTO 150
      IF(AHVAL(KR).LE.0.75)ABREAK = ABREAK+1
      IF(PROSEQ(IS,KR).NE."K".OR.PROSEQ(IS,KR).NE."I")GOTO 154
      LYS = LYS +1
      GOTO 160
  154 IF(AHVAL(KR).GT.1.0)AFORM = AFORM +1
  160 CONTINUE
      IF(LYS.GT.5)AFORM = AFORM +3
      IF(LYS.GT.3)AFORM = AFORM +2
      IF(LYS.GT.1)AFORM = AFORM +1
      WRITE(6,1080)(IR,KR,PA,PB,ABREAK,AFORM)
      IF(PA.LT.5.18)GOTO 150
      IF(PB.GT.PA)GOTO 150
      IF(ABREAK.GT.1)GOTO 150
      IF(AFORM.LT.3)GOTO 150
C           ANY HELIX NUCLEUS IS EXTENDED BY FOUR RESIDUES
C            IN BOTH DIRECTIONS UNTIL A REGION OF LOW HELIX
C            PROBABILITY IS FOUND
C
      EX = "A"
      R1 = IR
      R2 = KR
      KS = IS
      LAST = NR
      CALL EXTEND(R1,R2,EX)
      EX = IR - R1
      IF(EX.EQ.0)GOTO 180
      DO 170 XR = 1,EX
      IF(XR.GT.3)GOTO 180
      YR = IR + 3 - XR
      IF(PROSEQ(IS,YR).NE."P")GOTO 170
      R1 = R1 + XR
      GOTO 180
  170 CONTINUE
  180 IF(PROSEQ(IS,R2).EQ."R")AFORM = AFORM +1
      DO 190 I = R1,R2
      HELIX(I) = "D"
  190 CONTINUE
      IR = R2
      GOTO 150
C
CC            SEARCH FOR BETA SHEET NUCLEATION SITES
C
  400 IR= 0
  550 IR = IR +1
      LAST = NR -5
      IF(IR.GT.LAST)GOTO 580
      ABREAK = 0
      AFORM = 0
      PA = 0
      PB = 0
      DO 560 NN = 1,5
      KR = IR + NN-1
      PA = PA + AHVAL(KR)
      PB = PB + BSVAL(KR)
      IF(PROSEQ(IS,KR).EQ."E")GOTO 550
      IF(PROSEQ(IS,KR).EQ."P")GOTO 550
      IF(NN.GT.3)GOTO 555
      IF(PROSEQ(IS,KR).EQ."R")GOTO 550
      IF(PROSEQ(IS,KR).EQ."H")GOTO 550
  555 IF(BSVAL(KR).LE.0.75)ABREAK = ABREAK +1
      IF(BSVAL(KR).GT.1.0)AFORM = AFORM +1
  560 CONTINUE
      WRITE(6,1060)(IR,KR,PA,PB,ABREAK,AFORM)
      IF(PB.LT.5.25)GOTO 550
      IF(ABREAK.GE.2)GOTO 550
      IF(AFORM.LE.2)GOTO 550
      IF(PA.GT.PB)GOTO 550
C
C           EXTEND BETA SHEET NUCLEUS IN BOTH DIRECTIONS
```

```
C           BY FOUR RESIDUES UNTIL A REGION OF LOW PROBABILITY
C           IS FOUND
            EX = "B"
            R1 = IR
            R2 = KR
            KS = IS
            LAST = NR
            CALL EXTEND(R1,R2,EX)
            IF(PROSEQ(IS,R1).EQ."E")R1 = R1 + 1
            IF(PROSEQ(IS,R1).EQ."R")R1 = R1 + 1
            IF(PROSEQ(IS,R1).EQ."H")R1 = R1 + 1
            IF(PROSEQ(IS,R2).EQ."P")R2 = R2 - 1
            IF(PROSEQ(IS,R2).EQ."E")R2 = R2 - 1
            IF(PROSEQ(IS,R2).EQ."W")R2 = R2 - 1
            DO 570 JR = R1,R2
            SHEET(JR) = "$"
        570 CONTINUE
            IR = R2
            GOTO 550
C
CC                  SEARCH FOR BETA TURNS
C
        580 IR = 0
            IX = 0
            LAST = NR -4
        595 IR = IR +1
            IF(IR.GT.LAST)GOTO 700
            PA = 0
            PB = 0
            PT = 0
            DO 600 NN = 1,4
            KR = IR + NN-1
            PT = PT + BTVAL(KR)
            PB = PB + BSVAL(KR)
            PA = PA + AHVAL(KR)
        600 CONTINUE
            WRITE(6,1140)(PA,PB,PT,IR,KR)
            IX=IX+1
            PE(IX)=1.
            DO 1620 NN=1,4
            KR=NN+IR-1
            DO 1610 IA=1,22
            IF (AMINOS(IA).NE.PROSEQ(IS,KR)) GO TO 1610
            PE(IX)=PE(IX)*FIVAL (NN,IA)
            GO TO 1620
       1610 CONTINUE
       1620 CONTINUE
            P(IX) = 0.
            IF(PB.GT.PT)GOTO 595
            IF(PA.GT.3.5)GOTO 595
            PT = 1
            DO 620 NN = 1,4
            KR = NN + IR -1
            DO 610 IA = 1,22
            IF(AMINOS(IA).NE.PROSEQ(IS,KR))GOTO 610
            PT = PT * FIVAL(NN,IA)
            GOTO 620
        610 CONTINUE
        620 CONTINUE
            WRITE(6,1150)(PT)
            P(IX) = PT
            IF(PT.LT.2.0)GOTO 595
            DO 630 JR = IR,KR
            TURN(JR) = "."
        630 CONTINUE
            GOTO 595
C
CC
C
        700 R1 = 1
            R2 = 50
            WRITE(6,1120)
        410 IF(R2.GT.NR)R2=NR
            DO 590 I=1,10
            RES(I) = (R1+5*I)-1
        590 CONTINUE
            WRITE(6,1101)(RES(I),I=1,10)
            WRITE(6,1130)(PROSEQ(IS,I),I=R1,R2)
            WRITE(6,1130)(HELIX(I),I=R1,R2)
            WRITE(6,1130)(SHEET(I),I=R1,R2)
            WRITE(6,1130)(TURN(I),I=R1,R2)
            R1 = R1 + 50
            R2 = R2 + 50
            IF(R1.LT.NR)GOTO 410
        500 CONTINUE
```

```
      WRITE(6,1120)
      WRITE(6,1120)
      WRITE(6,1170)
      WRITE(6,1180)
      WRITE(6,1190)
      WRITE(6,1200)
      WRITE(6,1210)
      WRITE(6,1220)
      WRITE(6,1230)
      WRITE(6,1240)
 1000 FORMAT(8A10)
 1001 FORMAT(/5X,A1,2X,F4.2,2X,F4.2,2X,F4.2,5X,4(F5.3,2X))
 1010 FORMAT(36(1X,A1))
 1020 FORMAT(/5X,"INVALID CHARACTER '",A1,"' FOUND AT POSITION",I3)
 1030 FORMAT(//20X,"ALPHA HELIX REGION STARTS AT",I3," AND ENDS AT",
 1040 FORMAT("1"/////20X,8A10)
 1050 FORMAT("1"/////5X,8A10)
 1060 FORMAT(/,5X,25(2X,A1,2X))
 1070 FORMAT(/,5X,25(1X,F4.2))
 1080 FORMAT(" ","IR = ",I3,2X,"KR = ",I3,2X,"PA = ",F6.2,
     + 2X,"PB = ",F6.2,2X,"ABREAK = ",I2,2X,"FORM = ",I2)
 1100 FORMAT(//5X,5(20X,I3,2X))
 1101 FORMAT(/10(7X,I3))
 1120 FORMAT("1",/////)
 1125 FORMAT(" "////)
 1130 FORMAT(/50(1X,A1))
 1140 FORMAT(/"PA = ",F6.2,2X,"PB = ",F6.2,2X,"PT = ",F6.2,2X,
     + "IR = ",I3,2X,"KR = ",I3,2X)
 1150 FORMAT(/"BETA TURN VALUE = ",F6.6)
 1170 FORMAT(" ","MANUAL RULES FOR SEC. STRUCT. TERMINATION")
 1180 FORMAT(" ","P CANNOT OCCURE IN THE INNER HELIX OR C TERMINUS")
 1190 FORMAT(" ","P D E PREFER HELIX N TERM. H K R PREFER C TERM")
 1200 FORMAT(" ","E IS RARE IN BETA SHEETS")
 1210 FORMAT(" ","CHARGED RESIDUES RARE IN BETA SHEET TERMINII")
 1220 FORMAT(" ","W OCCURES MUST AT BETA SHEET N TERMINUS")
 1230 FORMAT(" ","IF B TURN NEAR 2.0 & NEXTO B SHEET MAYBE A B SHEET
 1240 FORMAT(" ","BSHEET MAYBE PRED WRONGLY AFTER STRONG BTURN")
      FI=1./(12.5*2.54)
      CALL PSPACE (1.*FI,30.*FI,1.*FI,30.*FI)
      CALL MAP (0.,29.,0.,29.)
      CALL BORDER
      CALL CHRMAG (8)
      CALL PLOTCS (10.,27.8,PHRASE,80)
      CALL PLOTCS (15.,1.2,15H RESIDUE NUMBER,15)
      CALL CHRORI (1.)
      CALL PLOTCS (1.5,8.,16H BETA TURN VALUE,16)
      CALL CHRORI (0.)
      CALL PSPACE (5.*FI,28.*FI,4.*FI,28.*FI)
      IX=IX+1
      CALL MAP  (0.,FLOAT(IX),0.,10.1)
      CALL AXESSI (20.,1.)
      IX=IX-1
      CALL POINT (1.,PE(1))
      DO 1 IP = 2,IX
      CALL JOIN (FLOAT(IP),PE(FLOAT(IP)))
    1 CONTINUE
      CALL POINT (0.,2.)
      CALL JOIN (FLOAT(IX),2.)
      CALL THICK (2)
      CALL DENSTY (3)
      CALL POINT (1.,P(1))
      DO 6 IP=2,IX
      IN=IP-1
      IF (P(IP).EQ.0..AND.P(IN).EQ.0.) GO TO 7
      CALL JOIN (FLOAT(IP),P(FLOAT(IP)))
      GO TO 6
    7 CALL POSITN (FLOAT(IP),P(FLOAT(IP)))
    6 CONTINUE
      CALL FRAME
      CALL GREND
      STOP
      END
      SUBROUTINE EXTEND(R1,R2,EX)
      IMPLICIT INTEGER(A-Z)
      REAL AHVAL,BSVAL,PA,PB,PX
      COMMON KS,PROSEQ,AHVAL,BSVAL,LAST
      DIMENSION PROSEQ(10,301),AHVAL(301),BSVAL(301)
C
C          EXTEND BACKWARDS BY FOUR RESIDUES
C
      IR = R1
  100 IR = IR - 1
      IF(IR.LT.1)GOTO 130
      PA =0
      PB = 0
```

```
      PRO = 0
      DO 110 NN = 1, 4
      KR = IR + NN - 1
      IF(EX.EQ."B")GOTO 105
      IF(PROSEQ(KS,KR).NE."P")GOTO 105
      PA = PA + 1.1
      GOTO 106
  105 PA = PA + AHVAL(KR)
  106 PB = PB + BSVAL(KR)
  110 CONTINUE
      WRITE(6,1000)(IR,KR,PA,PB,EX)
      IF(EX.EQ."A")GOTO 120
      IF(PROSEQ(KS,KR).EQ."E")PRO = 1
      IF(PROSEQ(KS,KR).EQ."R")PRO = 1
      IF(PROSEQ(KS,KR).EQ."H")PRO = 1
      PX = PA
      PA = PB
      PB = PX
  120 IF(PA.LT.4.0)GOTO 130
      IF(PA.LT.PB)GOTO 130
      IF(PROSEQ(KS,KR).EQ."P")GOTO 130
      IF(PRO.EQ.0)GOTO 160
      GOTO 130
C
CC          EXTEND FORWARDS BY ONE RESIDUE IN GROUPS OF FOUR
C
  130 R1 = IR + 1
      IR = R2
  140 IR = IR + 1
      IF(IR.GT.LAST)GOTO 180
      PA =0
      PB =0
      PRO =0
      DO 150 NN = 1, 4
      KR = IR - NN + 1
      PA = PA + AHVAL(KR)
      PB = PB + BSVAL(KR)
      IF(PROSEQ(KS,KR).EQ."P")PRO=1
  150 CONTINUE
      WRITE(6,1000)(IR,KR,PA,PB,EX)
      IF(EX.EQ."A")GOTO 160
      IF(PROSEQ(KS,KR).EQ."E")PRO = 1
      PX = PA
      PA = PB
      PB = PX
  160 IF(PA.LT.4.0)GOTO 180
      IF(PA.LT.PB)GOTO 130
      IF(EX.EQ."B")GOTO 140
      IF(PRO.EQ.0)GOTO 140
  180 R2 = IR - 1
 1000 FORMAT(" ","R1 = ",I3,2X,"R2 = ",I3,2X,"PA = ",F6.2,
     + 2X,"PB = ",F6.2,2X,A1)
      RETURN
      END
```

The computer programme was designed to incorporate as many of the constraints of the Chou and Fasman predictive model (70) as possible. Manual corrections to the computer prediction are necessary since several constraints were unquantitated or difficult to programme. Thus the computer printout tends to overpredict secondary structures and the following points should be noted when an uncorrected printout is under study (see figures 23 & 22 for JR66 corrected and uncorrected printouts).

1. P, D and E prefer helix N-termini.
2. H, K and R prefer helix C-termini.
3. Charged residues are rare in beta sheet termini.
4. If a reverse turn has a value 2.0 to 2.5 and next to a strongly predicted beta sheet, the turn may be an extension of the sheet.
5. Short beta sheets are often overpredicted after a strongly predicted (above 3.0) reverse turn and are eliminated if containing charged residues.

Rules 4 and 5 are additional to Chou and Fasman (although noted by them) and by the application of these rules a better prediction of sheets and turns was obtained.

Format of computer printout.

line 1 = residue number.

line 2 = sequence.

line 3 = predicted helix (H).

line 4 = predicted sheet ($).

line 5 = predicted turn (•).

Predicted turns take precedence over sheets which take precedence over helices provided all the manual checks are taken into account.

Many methods have been proposed to measure the agreement between prediction and observation to assess the power of predictive schemes and the

percentage agreement factors %k and Qk (70) have been used in this thesis. The percentage of residues predicted correctly in conformational state k is given by :

$$\%k = \frac{100\ (nk - \text{number incorrect})}{nk}$$

where n = number of residues in conformational state k

k = helix, sheet or turn

The percentage of residues correctly predicted which are not in the conformational state k is given by :

$$\%Nk = \frac{100\ (Nnk - \text{number incorrect})}{Nnk}$$

where Nk = number of residues not in the conformational state k.

The mean of %k and %Nk (designated Qk) gives an overall estimate of agreement between prediction and observation.

$$Qk = \frac{\%k + \%Nk}{2}$$

A comparison of $\%\alpha$, $\%\beta$, $Q\alpha$ and $Q\beta$ generated by the programme for 5 proteins are compared with Chou and Fasman's predictions (table 12) and also the combined computer method of Argos et al (71).

Randomised sequences of proteins used in assessing the computer programme give values of :

$\%\alpha$	10 - 30%	
$\%\beta$	5 - 20%	
%t	10 - 25%	
$Q\alpha$	45 - 55%	
$Q\beta$	40 - 55%	
Qt	40 - 55%	(Q statistical average 50%)

Since the computer programme is being used to predict the secondary structures of CAT proteins of unknown secondary structure the programme was also used to predict the secondary structures of two analogous proteins which show more primary sequence homology than the type I R-factor CAT and the type C staphylococcal CAT's, these proteins being the penicillinases (β-lactamases) sequenced by Ambler (42). Penicillinases from S.aureus PC1 and B. licheniformis show 42 percent homology in their primary sequences. The results of the secondary structure predictions are shown in figures 24 and 25. Discounting the 5 residue 'additions' to the N and C termini of the B. licheniformis enzyme the proteins have the following secondary structure correlations :

$\%\alpha$	= 73	$Q\alpha$	= 78
$\%\beta$	= 56	$Q\beta$	= 76
%t	= 66	Qt	= 79

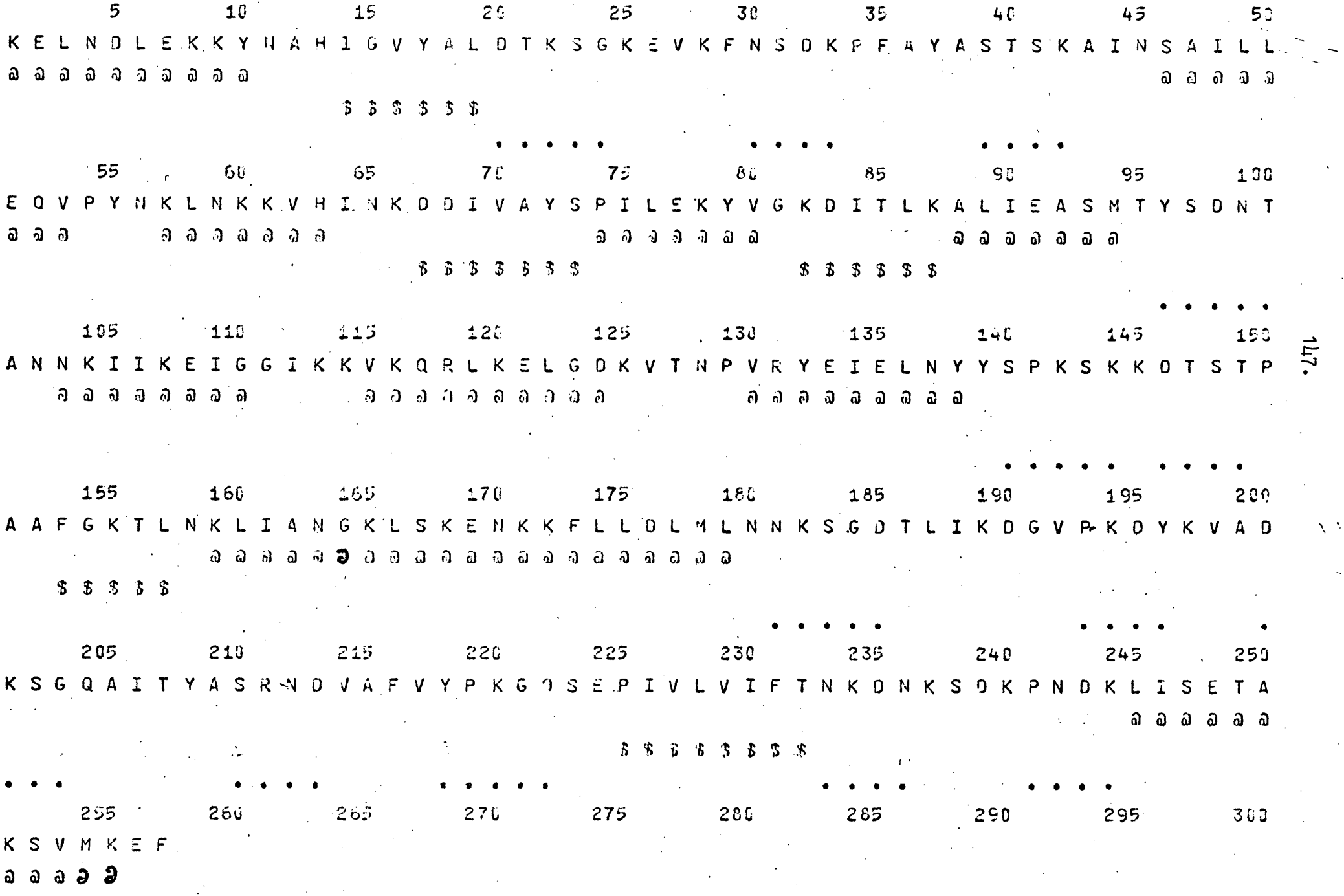
indicating a greater degree of secondary structure homology than primary sequence homology.

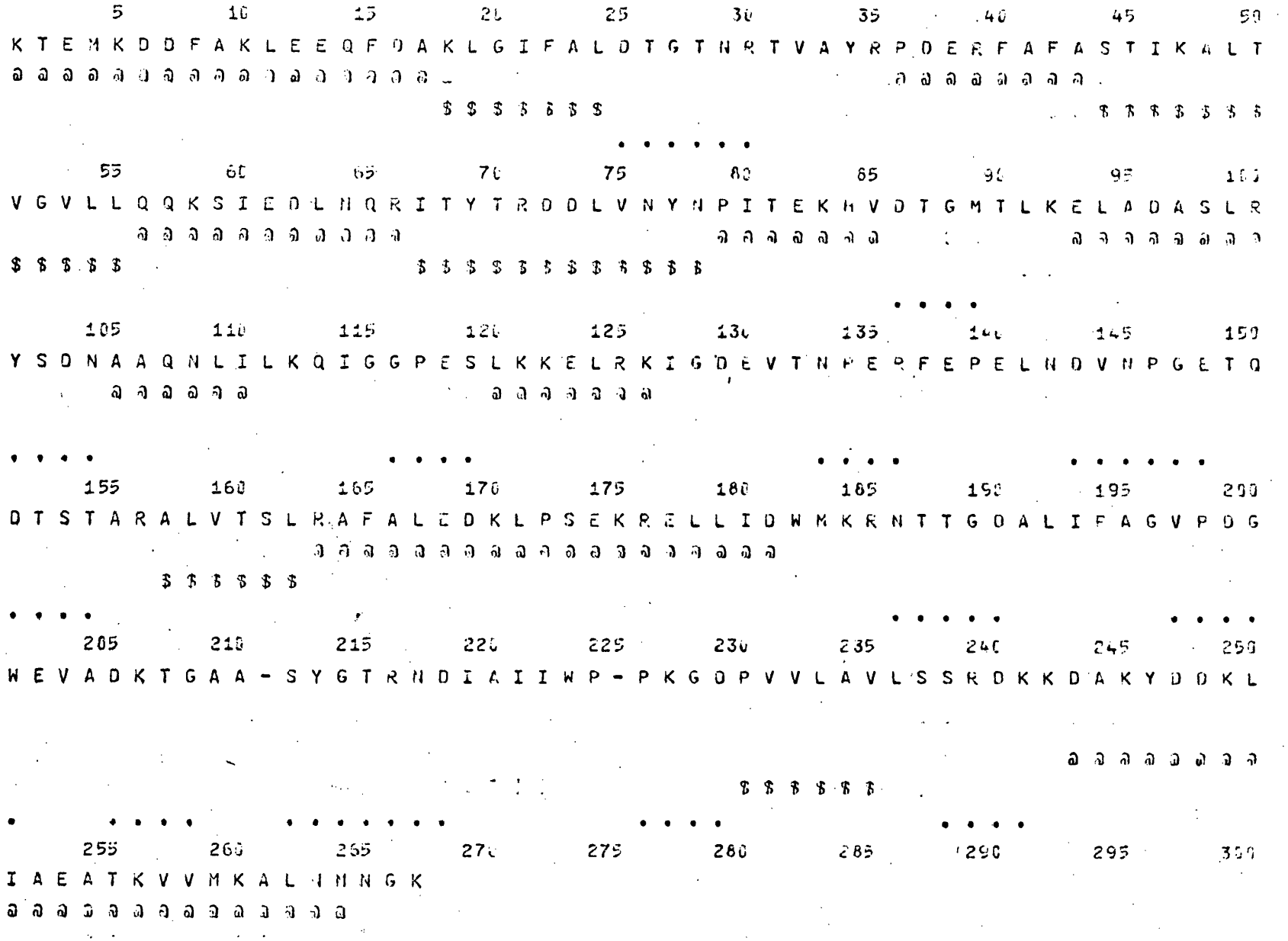
Table 12. Comparison of predictive ability of computer programmes.

Protein.	Fitton and Peacock.				Chou and Fasman.				Argos et al.			
	$\%\alpha$	$\%\beta$	$Q\alpha$	$Q\beta$	$\%\alpha$	$\%\beta$	$Q\alpha$	$Q\beta$	$\%\alpha$	$\%\beta$	$Q\alpha$	$Q\beta$
Adenylate kinase	81	57	80	76	-	-	81	84	-	-	82	84
Lysozyme	69	81	81	89	91	71	94	83	72	35	79	61
Subtilisin BPN	70	74	80	78	67	100	80	91	63	48	76	63
Trypsin inhibitor	80	74	88	86	87	95	90	95	47	64	71	79
Cytochrome C	90	100	80	98	71	100	80	89	66	-	81	-

figures : 24 and 25.

Secondary structure predictions of S. aureus penicillinase PC1 and B. licheniformis penicillinase.


5 10 15 20 25 30 35 40 45 50
K E L N D L E K K Y N A H I G V Y A L D T K S G K E V K F N S D K P F A Y A S T S K A I N S A I L L
55 60 65 70 75 80 85 90 95 100
E Q V P Y N K L N K K V H I N K D D I V A Y S P I L E K Y V G K D I T L K A L I E A S M T Y S D N T
105 110 115 120 125 130 135 140 145 150
A N N K I I K E I G G I K K V K Q R L K E L G D K V T N P V R Y E I E L N Y Y S P K S K K D T S T P
155 160 165 170 175 180 185 190 195 200
A A F G K T L N K L I A N G K L S K E N K K F L L D L M L N N K S G D T L I K D G V P K D Y K V A D
205 210 215 220 225 230 235 240 245 250
K S G Q A I T Y A S R N D V A F V Y P K G D S E P I V L V I F T N K D N K S D K P N D K L I S E T A
255 260 265 270 275 280 285 290 295 300
K S V M K E F


5 10 15 20 25 30 35 40 45 50
K T E M K D D F A K L E E Q F D A K L G I F A L D T G T N R T V A Y R P D E R F A F A S T I K A L T
55 60 65 70 75 80 85 90 95 100
V G V L L Q Q K S I E D L N Q R I T Y T R D D L V N Y N P I T E K H V D T G M T L K E L A D A S L R
105 110 115 120 125 130 135 140 145 150
Y S D N A A Q N L I L K Q I G G P E S L K K E L R K I G D E V T N P E R F E P E L N D V N P G E T Q
155 160 165 170 175 180 185 190 195 200
D T S T A R A L V T S L R A F A L E D K L P S E K R E L L I D W M K R N T T G D A L I F A G V P D G
205 210 215 220 225 230 235 240 245 250
W E V A D K T G A A - S Y G T R N D I A I I W P - P K G D P V V L A V L S S R D K K D A K Y D D K L
255 260 265 270 275 280 285 290 295 300
I A E A T K V V M K A L I N N G K

1

Overall it can be seen that the programme of Fitton and Peacock gives somewhat better predictions than that of Argos et al but is inferior to the predictions of Chou and Fasman. As noted by Argos et al the predictors often modify the results obtained from a strict adherence to their scheme as there are overlapping regions of predicted secondary structure and this may account for the superior prediction, in some cases, of Chou and Fasman.

Although the method of Argos et al is based on a combination of five predictive methods, the method of Chou and Fasman, included in the combined programme, is, perhaps, one of the best single methods in the combined programme.

As a final check of the predictive scheme the prediction of Triose phosphate isomerase from Bacillus stearothermophilus was compared with a prediction made manually using the Chou and Fasman rules (S. Artavarnis Ph. D. thesis, Cambridge, 1974.) performed by Artavarnis.

The correlation found was :

$$\%\alpha = 91 \qquad Q\alpha = 93$$

$$\%\beta = 97 \qquad Q\beta = 95$$

indicating very good agreement between the results of the present author and an independent predictor.

ADDENDUM.

All predictions presented in this thesis are based on the P_α, P_β and Pt values derived from a data base of 19 proteins ([70]). Chou and Fasman have recently published P_α and P_β values based on a study of 29 proteins ([72]) and the programme of Fitton and Peacock has now been modified to incorporate the new data. Preliminary results indicate only minor differences in predictions using a 19 protein and a 29 protein data base.

Revised data set. Based on a study of 29 proteins.

(P_α & P_β values only, reverse turn data not yet available).

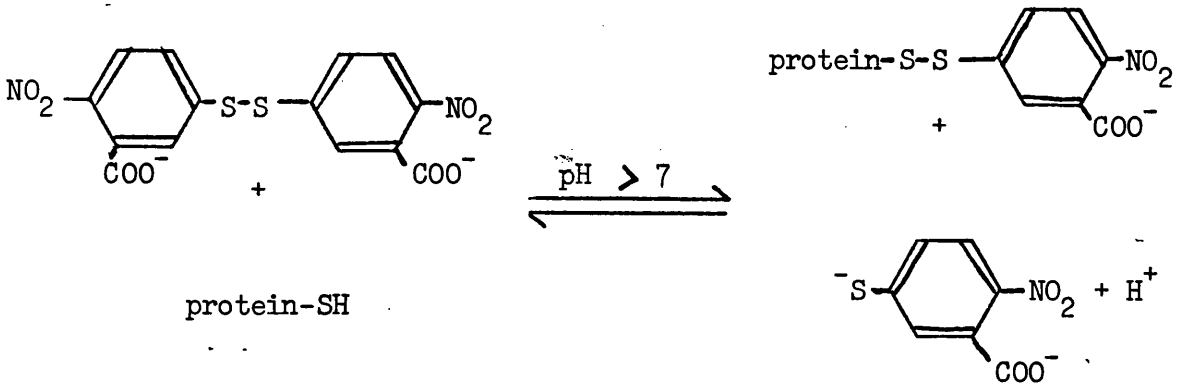
```
C
      IMPLICIT INTEGER(A-Z)
      REAL ALPHA,BETAS,BETAT,AHVAL,BSVAL,BTVAL,FIVAL,PA,PB,PT,FI,PE,P
      COMMON KS,PROSEQ,AHVAL,BSVAL,LAST
      DIMENSION HELIX(301),SHEET(301),TURN(301),NRES(10)
      DIMENSION PROSEQ(10,301),NAME(10,8),TITLE(8),RES(10)
      DIMENSION AMINOS(22),ALPHA(22),BETAS(22),BETAT(22),FIVAL(4,22)
      DIMENSION AHVAL(301),BSVAL(301),BTVAL(301),PE(301),P(301)
      DIMENSION PHRASE(8)
      DATA(AMINOS(I),I=1,22)/"A","C","D","E","F","G","H","I",
     + "K","L","M","N","P","Q","R","S","T","V","W","Y","-",".","/
      DATA(ALPHA(I),I=1,22)/1.42,0.7,1.01,1.51,1.13,0.57,1.0,1.08,
     + 1.16,1.21,1.45,0.67,0.57,1.11,0.98,0.77,0.83,1.06,
     + 1.08,0.69,0.0,0.0/
      DATA(BETAS(I),I=1,22)/0.83,1.19,0.54,0.37,1.38,0.75,0.87,1.6,
     + 0.74,1.3,1.05,0.89,0.55,1.1,0.93,0.75,1.19,1.7,1.37,1.47,
     + 0.0,0.0/
      DATA(BETAT(I),I=1,22)/0.76,0.96,1.38,0.71,0.41,1.55,0.80,
     + 0.59,0.93,0.53,0.71,1.55,1.35,0.92,0.95,1.47,1.13,0.45,
     + 1.03,1.06,0.0,0.0/
      DATA((FIVAL(I,J),I=1,4),J=1,22)/0.86,0.89,0.52,.60,
     + 1.22,.27,1.22,0.95,1.90,1.15,1.72,0.81,0.57,0.57,0.98,0.49,
     + 0.32,0.21,0.74,0.42,1.20,1.28,1.94,1.63,1.75,0.32,1.11,0.16,
     + 0.58,0.45,0.32,0.83,0.57,1.09,0.83,1.09,0.32,0.37,0.41,0.83,
```

Appendix B.

Reactions of chemical modification reagents.

5,5'-dithiobis-(2-nitrobenzoic acid) (DTNB).

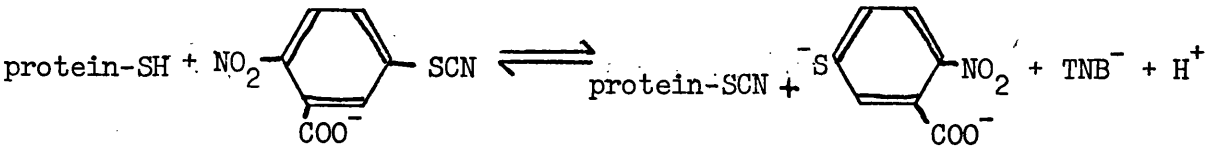
DTNB specifically reacts with free sulphydryl groups of proteins :



A mixed disulphide is formed and one molecule of free thionitrobenzoate anion (TNB^-) is released for each sulphydryl group that reacts in the protein. TNB^- has a strong yellow colour with a maximum absorbance at 412nm (A_{412nm} = 1.36 x $10^4 M^{-1}$ cm^{-1} at pH 8.0). The reaction of DTNB with sulphydryl groups can be readily reversed by adding an excess of thiol groups which competitively displace the TNB^- group (82).

Nitrothiocyanobenzoic acid (NTCB).

This reagent reacts with proteins in an analogous way to DTNB, with the S-cyano derivative being formed instead of the mixed disulphide derivative (83).



NTCB should be preferable to DTNB since the cyano group is smaller and does not carry a negative charge. Consequently it is less likely to impose steric hindrance within the protein.

Iodoacetamide (INH_2) and Iodoacetic acid (IAA).

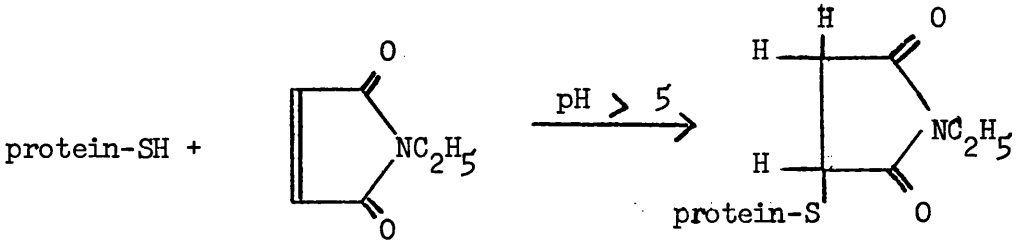
Iodoacetamide and iodoacetic acid can react with sulphydryl, imidazole (74), thioester and amino groups depending on the pH of the reaction (82). Of these the sulphydryl group is the most reactive and reactivity increases with increase in pH values since the reactive species is the ionic form of this group :

$$\text{protein-S}^- + ICH_2COO^- \xrightarrow{pH>7} \text{protein-S-}CH_2COO^- + I^-$$

However, to avoid unnecessary reaction with amino groups pH values should not be too high; the optimum for modification of sulphydryl groups is usually around pH 7-8.

N-ethylmaleimide (NEM).

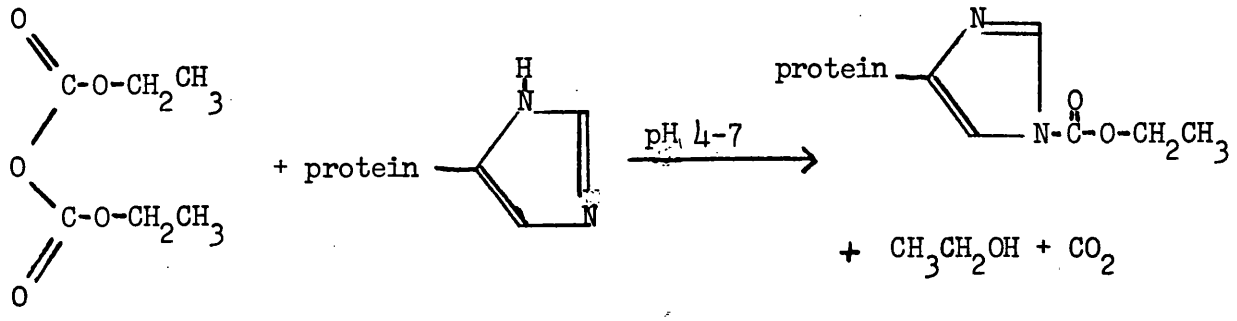
NEM has been widely used as a sulphydryl-specific reagent both in the determination of the number of sulphydryl groups in a protein and also in establishing the effect of modification on enzyme activity.



It also reacts with amino and imidazole groups.

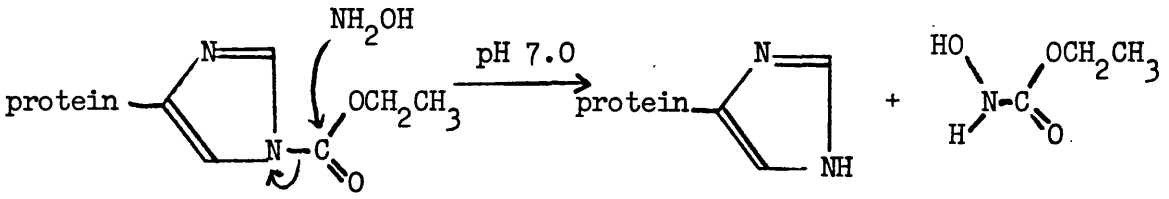
Diethylpyrocarbonate (DEPC).

DEPC is both an ester and an anhydride, but because of resonance involving the ester group it is less reactive than many anhydrides. It is slowly hydrolysed by water with a half life at 25°C, pH 7.0, of 25 minutes (56) to give two equivalents of ethanol and carbon dioxide. At low pH values DEPC reacts principally with imidazole groups of proteins and is consequently a useful reagent for the chemical modification of histidine residues :



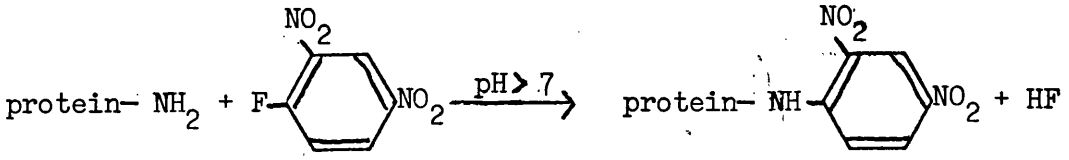
The reaction of DEPC with the imidazole groups of proteins can also cause polypeptide chain cleavage (75).

The modification of histidine residues can be readily reversed at neutral pH values by hydroxylamine :



However, reversal of modified lysine and arginine groups cannot be obtained in this way (56).

1-fluoro-2,4-dinitrobenzene(FDNB).



Originally used by Sanger for the N-terminal determination of insulin FDNB also reacts with thiol and imidazole groups.

REFERENCES.

(1) Rebstock, M.C., Crooks, H.M., Controulis, J. & Bartz, Q.R. (1949)
J. Amer. Chem. Soc. 71, 2458-2462

(2) Ehrlich, J.E., Bartz, Q.R., Smith, R.M., Joslyn, D.A. & Burkholder, P.R. (1947)
Science 106, 417

(3) Brock, T.D. (1964)
Exper. Chemother. 3, 119-169

(4) Controulis, J., Rebstock, M.C. & Crooks, H.M. (1949)
J. Amer. Chem. Soc. 71, 2463-2468

(5) Hurwitz, C. & Braun, C. (1968)
Biochim. Biophys. Acta 157, 392-403

(6) Nierhaus, D. & Nierhaus, K.H. (1973)
Proc. Natl. Acad. Sci. (USA) 70, 2224-2228

(7) Watanabe, T. (1963)
Bacteriol. Rev. 27, 87-115

(8) Shaw, W.V. (1974)
Biochem. Soc. Trans. 2, 834-836

(9) Suzuki, Y. & Okamoto. S. (1976)
J. Biol. Chem. 242, 4722-4730

(10) Shaw, W.V. (1967)
J. Biol. Chem. 242, 687-693

(11) Shaw, W.V. & Unowsky, J. (1968)
J. Bacteriol. 95, 1976-1978

(12) Shaw, W.V. (1975) (Ed. J.H. Hash).
In: 'Methods in Enzymology', 43, 737-755, Academic Press, New York.

(13) Fitton, J.E., Packman, L.C., Harford. S., Zaidenzaig, Y. & Shaw, W.V. (1977)
In: 'Microbiology - 1977' (Proc. 2 nd. A.S.M. Conf. Extrachromsomal Elements, Jackson, Wyoming); In Press.

(14) Shaw, W.V. (1971)
Ann. N. Y. Acad. Sci. 182, 234-242

(15) Foster, T.J. & Shaw, W.V. (1973)
Antimicrob. Ag. & Chemother. 3, 99-104

(16) Sands, L.C. & Shaw, W.V. (1973)
Antimicrob. Ag. & Chemother. 3, 299-305

(17) Shaw, W.V. & Brodsky, R.F. (1968)
J. Bacteriol. 95, 28-36

(18) Shaw, W.V., Bentley, D.W. & Sands, L.C. (1970)
J. Bacteriol. 104, 1095-1105

(19) Zaidenzaig, Y. & Shaw, W.V. (1976)
FEBS Letters 62, 266-271

(20) Zaidenzaig, Y., Fitton, J.E., Packman, L.C., Harford, S., & Shaw, W.V. (1977)
Eurø. J. Biochem. (submitted).

(21) Shaw, W.V., Sands, L.C. & Datta, N. (1972)
Proc. Natl. Acad. Sci. 69, 3049-3053

(22) Garber, N. & Zipser, J. (1970)
Biochim. Biophys. Acta 220, 341-343

(23) Shaw, W. & Brodsky, R.F. (1968)
Antimicrob. Ag. & Chemother. - 1967:, 257-263

(24) Benveniste, R. & Davies, J. (1971)
FEBS Letters 14, 293-296

(25) Shaw, W.V., Sands, L.C. & Datta, N. (1972)
Proc. Natl. Acad. Sci. 69, 3049-3053

(26) Hedges, R. & Datta, N. (1971)
Nature 234, 220-222

(27) Shaw, W.V., Bouanchaud, D. & Goldstein, F. (1977)
Antimicrob. Ag. & Chemother. In Press.

(28) Shaw, W.V. & Hopwood, D.A. (1976)
J. Gen. Microbiol. 94, 159-166

(29) Haag, R., Süssmuth, R. & Lingens, F. (1976)
FEBS Letters 63, 62-64

(30) Sabath, L.D., Gerstein, D.A., Loder, P.B. & Finland, M. (1967)
Antimicrob. Ag. & Chemother. (1967) 264-270

(31) Lowry, O.H., Rosebrough, N.J., Farr, A.L., & Randall, R.J. (1951)
J. Biol. Chem. 193, 265-275

(32) Schroeder, W.A. (1967) (Ed.C.H.W. Hirs).
In: 'Methods in Enzymology' 11, 351-361 ,Academic Press,New York.

(33) Brown, J.R. & Hartley, B.S. (1966)
Biochem. J. 101, 214-228

(34) Milstein, C. (1966)
Biochem.J. 101, 338-351

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

(36) Laemmli, U.K. (1970)
Nature 227, 680-685

(37) Harley-Mason, J. & Archer, A.A.P.G. (1958)
Biochem. J. 69, 60p.

(38) Yamada, S. & Itano, H.A. (1966)
Biochim. Biophys. Acta 130, 538-540

(39) Liu, T.Y. & Chang, Y.H. (1971)
J.Biol.Chem. 246, 2842-2843

(40) Grey, W.R. & Hartley, B.S. (1963)
Biochem. J. 89, 379-380

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

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

(43) Solid Phase Methods in Protein Sequence Analysis (1975)
Proc. 1ST International Conf. (Ed. R.A. Lausen).
Pierce Chemical Co.

(44) Wachter, E., Machleidt, W., Hofner, H. and Otto, J. (1973)
FEBS Letters 35, 97 - 102.

(45) Bridgen, J. (1975)
FEBS Letters 50, 159 - 162.

(46) Horn, M.J., and Lausen, R.A. (1973)
FEBS Letters 36, 285 - 288.

(47) Lausen, R.A., Horn, M.J. and Bonner, A.G. (1972)
FEBS Letters 21, 67 - 70.

(48) Previero, A., Derancourt, J., Colletti - Previero, M.A. and Lausen, R.A. (1973)
FEBS Letters 33, 135 - 138.

(49) Solid Phase Sequencing Handbook (1976)
L.K.B. Biochrom Ltd. Milton Rd.,Cambridge.

(50) Bridgen, J., Graffeo, A., Karger, B.L. & Waterfield, M. (1975)
'The identification of PTH amino acids.'
In : 'Instrumentation in amino acid sequence analysis.' pp 111-146 (1975) Academic Press, London.

(51) Chase, J.F.A. & Tubbs, P.K. (1969)
Biochem. J. 111, 225 - 235.

(52) Novick, R. (1976)
J. Bacteriol. 127, 1177 - 1187.

(53) Colman, R.F. (1969)
Biochemistry 8, 888 - 898.

(54) In:'Methods in Enzymology' 11, (1967), RIORDAN, J.F. & VALLEE, B.L.
Section VIII, (Ed. C.H.W. Hirs.)Academic Press,New York. P.541

(55) Mühlrad, A., Hegyi, G., & Horanyi, M. (1969)
Biochim. Biophys. Acta 181, 184-190

(56) Melchior, W.B. & Fahney, D.E. (1970)
Biochemistry 9, 251-257

(57) Weil, L. & Buchert. A.R. (1951)
Arch.Biochem. Biophys. 34, 1-15

(58) Weil, L. (1965)
Arch.Biochem. Biophys. 110, 57-68

(59) IUPAC-IUB Commission (1968)
J. Biol. Chem. 243, 3557-3559

(60) IUPAC-IUB Commission (1972)
J. Biol. Chem. 247, 977-983

(61) Konigsberg, W. (1967) (Ed. C.H.W. Hirs).
In: 'Methods in Enzymology' 11, 461-469 Academic Press,New York.

(62) Needleman, S.B. (Ed.) (1977)
In: 'Protein Sequence Determination' 2nd. Edition.
Springer-Verlag, New York

(63) Bridgen, J. (1976)
Biochemistry 15, 3600-3604

(64) Rebstock, M.C., Moersch, G.W., Moore, A.C. & Vandenbelt, J.M. (1951)
J. Amer. Chem. Soc. 73, 3666-3670

(65) De Moss, J.A. (1977)
J. Biol. Chem. 252, 1696-1701

(66) Singhal, R.P. & Atassi, M.Z. (1971)
Biochemistry 10, 1756-1762

(67) Houmard, J. & Drapeau, G.R. (1972)
Proc. Natl. Acad. Sci. (USA) 69, 3506-3509

(68) Henkin, J. & Abeles, R.H. (1976)
Biochemistry 15, 3472-3479

(69) Moore, G.W., Goodman, M. & Barnabas, J. (1973)
J. Theor. Biol. 38, 423-457

(70) Fasman, G.D. & Chou, P.Y. (1974)
Biochemistry 13, 211-244

(71) Argos, P., Swarz, James & Swarz, John (1976)
Biochim. Biophys. Acta 439, 261-273

(72) Fasman, G.D., Chou, P.Y. & Adler, A.J. (1976)
Biophys. J. 16, 1201-1238

(73) Zaidenzaig, Y. & Shaw, W.V. (1977)
Euro. J. Biochem. In Press.

(74) Crestfield, A.M., Stein, W.H. & Moore, S. (1963)
J. Biol. Chem. 238, 2413-2419

(75) Loosemore, M.J. & Pratt, R.F. (1976)
FEBS Letters 72, 155-158

(76) Hartley, B.S. (1974)
Ann. N.Y. Acad. Sci. 227, 438-445

(77) Shotton, D.M. & Hartley, B.S. (1970)
Nature 225, 802-806

(78) Hartley, B.S. (1974)
In: 'Evolution in the Microbial World' 151-182
Cambridge University Press. (EDS.M.J. Carlile & J.J. Skehel).

(79) Ambler, R.P. & Meadway, R.J. (1969)
Nature 222, 24-26

(80) Thatcher, D.R. (1975)
Biochem. J. 147, 313-326

(81) Hartley, B.S. (1970)
Miami Winter Symposium 1, 191-209 North-Holland Pub.,Amsterdam.

(82) Means, G.E. & Feeney, R.E. (1971)
In: 'The Chemical Modification of Proteins'
Holden-Day Publishers.

(83) Dagani, Y. & Patchornik, A. (1974)
Biochemistry(USA) 13, 1-11

(84) Clarke, P.H. (1974)
In: 'Evolution in the Microbial World' 183-218
Cambridge University Press. (EDS.M.J. Carlile & J.J. Skehel)

(85) Shaw, W.V. (1971)
Trans. Assoc. Amer. Phys. 84, 190-199

(86) Brownlee, G.G. (1972)
In: Laboratory techniques in biochemistry & molecular biology
Determination of sequences in RNA. p. 65.
North-Holland Pubs. 1972